Monotonicity of a Bayesian confirmation measure in rule support and confidence

Izabela Brzezińska
Institute of Computing Science, Poznań University of Technology,
Piotrowo 3a, 60-965 Poznań, Poland
e-mail: izabela.brzezinska@cs.put.poznan.pl

Roman Słowiński
Institute of Computing Science, Poznań University of Technology,
and Institute for Systems Research, Polish Academy of Sciences, Warsaw, Poland
e-mail: roman.slowinski@cs.put.poznan.pl

Abstract:
In knowledge discovery and data mining many measures of interestingness have been proposed in order to reveal different characteristics of the discovered knowledge patterns. Among these measures, an important role is played by Bayesian confirmation measures, which express in what degree a piece of evidence (premise) confirms a hypothesis (conclusion). In this paper, we are considering knowledge patterns in form of "*if*... *then*..." decision rules with a fixed conclusion. We are investigating the question of monotonic relationship between a particular Bayesian confirmation measure on one side, and rule support and confidence, on the other side. We prove that rules which are optimal according to the confirmation measure, are included in the set of non-dominated rules with respect to both rule support and confidence.

Keywords: Knowledge discovery, Decision rules, Bayesian confirmation measures, Monotonicity in rule support and confidence

1. Introduction

In data mining and knowledge discovery, the discovered knowledge patterns are often expressed in a form of "*if*..., *then*..." rules. They are consequence relations representing correlation, association, causation etc. between independent and dependent attributes. If the division into independent and dependent 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 very large, but only a few of them are likely to be useful for the domain expert analysing the data. In order to increase the relevance and utility of selected rules, quantitative measures, also known as attractiveness or interestingness measures (metrics), have been proposed and studied. Among them there are: confidence and support [1], gain [7], conviction [3], and many others. The proposed measures have been introduced to capture different characteristics of rules. Bayardo and Agrawal [2] have proved, however, 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 [7], laplace [4] [17], lift [11], conviction [3], an unnamed measure proposed by Piatetsky-Shapiro [15]. This is a practically useful result that allows to identify, for a given conclusion, the most interesting rules according to several interestingness measures by solving an optimized rule mining problem with respect to rule support and confidence only.

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 [6]. Among the most well-known Bayesian confirmation measures proposed in the literature, an important role is played by a confirmation measure denoted in [6] and other studies by f, which has the property of hypothesis symmetry, found desirable by Eells and Fitelson [5], as well the property of monotonicity proposed by Greco et al. [9]. However, according to our knowledge, the confirmation measure f has not yet been the subject of Bayardo and Agrawal's analysis or any other work of this sort. Therefore, the objective of this paper is to verify whether rules, with a fixed hypothesis, that are best according to the confirmation measure f are included in the set of non-dominated rules with respect to rule support and confidence.

The paper is organized as follows. In the next section there are preliminaries on decision rules and their quantitative description. In section 3, we investigate the question of monotonic relationship between confirmation measure f on one side, and rule support and confidence, on the other side. 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 $\phi \to \psi$ (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], at once we only consider rules with the same conclusion, not the whole set of all possible rules induced from a dataset.

2.1. Monotonicity

For x belonging to a set ordered by the relation $>$ and for the values of g belonging to a set ordered by the relation $\leq$, a function $g(x)$ is understood to be monotone in x, if $x_1 < x_2$ implies that $g(x_1) \le g(x_2)$.

2.2. Support and confidence measures of rules

With every rule induced from information table S measures called *support* and *confidence* are often 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:

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

Note, that it can be regarded as a conditional probability (frequency) with which conclusion ψ evaluates to true, given that premise ϕ evaluates to true.

2.3. Bayesian confirmation measure f

In general, measures of confirmation quantify the strength of confirmation that premise ϕ gives to conclusion ψ. There are many confirmation measures proposed in literature, but the confirmation measure f bares a very special place among them for having the desirable properties of hypothesis symmetry of Eells and Fitelson [5], as well the property of monotonicity of Greco et al. [9]. The 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 other authors advocating for this measure are Good [8], Heckerman [10], Pearl [14], and Fitelson [6].

The confirmation measure f is defined as:

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

where $\Pr(\phi \mid \psi)$ is the conditional probability with which premise ϕ evaluates to true given that conclusion ψ evaluates to true. From the Bayes' theorem we have:

$$\Pr(\phi \mid \psi) = \frac{\Pr(\phi \wedge \psi)}{\Pr(\psi)}.$$

If the probability $\Pr(\mathbf{o})$ is understood as the ratio of the number of objects in U having property $\mathbf{o}$ to the cardinality of U, then it is easy to observe that, $\Pr(\phi \mid \psi) = conf(\psi \rightarrow \phi)$. Thus, the confirmation measure f can be expressed as

$$f(\phi \rightarrow \psi) = \frac{conf(\psi \rightarrow \phi) - conf(\neg\psi \rightarrow \phi)}{conf(\psi \rightarrow \phi) + conf(\neg\psi \rightarrow \phi)} \qquad (1)$$

2.4. Partial order on rules in terms of rule support and confidence

As in [2], let us denote the partial order on rules in terms of rule support and confidence as $\preceq_{sc}$, and define it in the following manner:

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

$sup(r_1) \le sup(r_2) \wedge conf(r_1) < conf(r_2)$, or

$sup(r_1) < sup(r_2) \wedge conf(r_1) \le conf(r_2)$.

Moreover, $r_1 =_{sc} r_2$ if and only if

$sup(r_1) = sup(r_2) \wedge conf(r_1) = conf(r_2)$.

2.5. Implication of a total order by $\preceq_{sc}$

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 measure f 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*** $\preceq_{sc}$ if

$r_1 \prec_{sc} r_2 \Rightarrow r_1 \preceq_t r_2$, and

$r_1 =_{sc} r_2 \Rightarrow r_1 =_t r_2$.

It has been proved by Bayardo and Agrawal [2] that if a total order $\preceq_t$ is implied by 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 support and confidence, i.e. among the most interesting rules according to support-confidence partial order. 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 and be sure that the rules optimal with respect to that total order are in the discovered set.

Moreover, Bayardo and Agrawal [2] have showed 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 $\preceq_{sc}$:

- g(r) is monotone in rule support over rules with the same confidence, and
- g(r) is monotone in confidence over rules with the same rule support.

Hence, to fulfill the objective of this paper, we shall verify whether these conditions hold when the confirmation measure f is the g(r) rule value function. Thus, we shall check whether the confirmation measure f is monotone in rule support when the confidence is held constant, as well as in confidence, when the rule support remains unchanged.

3. Analysis of the monotonicity of *f* in rule support and confidence

Let us consider the confirmation measure f given as in (1). Below, we will try to transform confirmation measure f such that, for given U and ψ, it only depends on confidence and support of rule $\phi\rightarrow\psi$.

Remark that:

$$f(\varphi \rightarrow \psi) = \frac{\dfrac{\sup(\psi \rightarrow \varphi)}{\sup(\psi)} - \dfrac{\sup(\neg\psi \rightarrow \varphi)}{\sup(\neg\psi)}}{\dfrac{\sup(\psi \rightarrow \varphi)}{\sup(\psi)} + \dfrac{\sup(\neg\psi \rightarrow \varphi)}{\sup(\neg\psi)}} =$$

$$=\frac{sup(\psi\to\varphi)sup(\neg\psi)-sup(\neg\psi\to\varphi)sup(\psi)}{sup(\psi\to\varphi)sup(\neg\psi)+sup(\neg\psi\to\varphi)sup(\psi)}.$$

One can make the following observations:

- $sup(\neg\psi)=|U|-sup(\psi)$,
- $sup(\phi)=\dfrac{sup(\phi\to\psi)}{conf(\phi\to\psi)}$,
- $sup(\phi\to\psi)=sup(\psi\to\phi)$,
- $sup(\neg\psi\to\phi)+sup(\psi\to\phi)=sup(\phi)$, so
- $sup(\neg\psi\to\phi)=sup(\phi)-sup(\psi\to\phi)$.

Hence,

$$f(\phi\to\psi)=[sup(\phi\to\psi)(|U|-sup(\psi))-$$
$$-(sup(\phi)-sup(\phi\to\psi))sup(\psi)]\,/\,[sup(\phi\to\psi)$$
$$(|U|-sup(\psi))+(sup(\varphi)-sup(\varphi\to\psi))sup(\psi)]=$$

$$=[|U|sup(\phi\to\psi)-sup(\psi)sup(\phi\to\psi)-sup(\phi)sup(\psi)+$$
$$+sup(\phi\to\psi)sup(\psi)]/[|U|sup(\phi\to\psi)-sup(\psi)sup(\phi\to\psi)+$$
$$+sup(\phi)sup(\psi)-sup(\phi\to\psi)sup(\psi)]=$$

$$=[|U|sup(\phi\to\psi)-\frac{sup(\phi\to\psi)}{conf(\phi\to\psi)}sup(\psi)]/[|U|sup(\phi\to\psi)-$$
$$-2sup(\psi)sup(\phi\to\psi)+\frac{sup(\phi\to\psi)}{conf(\phi\to\psi)}sup(\psi)]=$$

$$=\frac{sup(\varphi\to\psi)\left[|U|-\dfrac{sup(\psi)}{conf(\varphi\to\psi)}\right]}{sup(\varphi\to\psi)\left[|U|-2\,sup(\psi)+\dfrac{sup(\psi)}{conf(\varphi\to\psi)}\right]}=$$

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

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

Finally, we have

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

For the confirmation measure *f* we need to verify whether:

- confirmation measure *f* is monotone in rule support when the confidence is held constant, and
- confirmation measure *f* is monotone in confidence when the rule support is held constant.

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 confidence, we have a constant value of the confirmation measure *f*, no matter what the rule support is. Hence, according to the definition of monotonicity given in paragraph 2.1, confirmation measure *f* is monotone in rule support when the confidence is held constant. Therefore, the first of the sufficient conditions for proving that the total order $\leq_t$ defined over the confirmation measure *f* is implied by partial order $\leq_{sc}$ is held.

Now, let us verify whether the confirmation measure *f* is monotone in confidence for a constant value of rule support.

For the clarity of the presentation, let us express the confirmation measure *f* as a function of confidence, still regarding $|U|$ and $sup(\psi)$ as constants greater than 0:

$$f(x)=\frac{ax-b}{cx+b},$$

where $x=conf(\phi\to\psi)$, $a=|U|$, $b=sup(\psi)$, $c=|U|-2\,sup(\psi)$.

It is easy to observe that:

- $a=|U|>0$, and
- $0<b\leq|U|$.

In order to verify the monotonicity of *f* in confidence, let us calculate the derivative of the above function.

$$f'(x)=\frac{(ax-b)'(cx+b)-(cx+b)'(ax-b)}{(cx+b)^2}.$$

Since there is a square in the denominator, it must always be a positive number. Hence, the sign of the derivative depends on the sign of the nominator. The nominator is equal to:

$$(ax-b)'(cx+b)-(cx+b)'(ax-b)=$$
$$=a(cx+b)-c(ax-b)=$$
$$=acx+ab-acx+bc=$$
$$=ab+bc=b(a+c)$$

As $b>0$, and $a+c=|U|+|U|-2sup(\psi)=2|U|-2sup(\psi)>0$ for $|U|\geq sup(\psi)$, the whole derivative is always not smaller than 0. Therefore, confirmation measure *f* is monotone in confidence.

Thus, both of Bayardo and Agrawal's sufficient conditions for proving that a total order $\leq_t$ defined over a confirmation measure *f* is implied by partial order $\leq_{sc}$ are held. This means that, for a class of rules with a fixed conclusion, rules optimal according to the confirmation measure *f* will be found in the set of rules that are best with respect to both rule support and confidence.

4. Conclusions

Bayardo and Agrawal have proved in [2] that total orders of many interestingness measures such as gain, laplace, lift, conviction, the one proposed by Piatetsky-Shapiro, etc. are implied by the support-confidence partial order $\leq_{sc}$. This result showed that the most-interesting rules according to any of the above measures, are included in the set of Pareto-optimal (i.e. non-dominated) rules with respect to both rule support and confidence.

In this paper, for a class of rules with the same conclusion, we have analyzed the monotonicity of the confirmation measure *f* in order to verify whether this measure is also implied by the partial order $\leq_{sc}$. We have shown that the confirmation measure

f is constant and independent of the rule support when confidence is held fixed, and that it is monotone in confidence. Hence, it is implied by the partial order $\leq_{sc}$. Therefore, rules optimal according to this confirmation measure lie on the upper support-confidence Pareto-border, i.e. they are included in the set of rules that are non-dominated with respect to both rule support and confidence, for a fixed conclusion. In other words, for a given conclusion, finding the set of support-confidence Pareto-optimal rules also guarantees finding rules optimal according to the confirmation measure f.

Let us stress that the above result does not deny the interest of the confirmation measure f in expressing the attractiveness of decision rules. It just states the monotonicity of f in confidence of decision rules for a fixed conclusion. This result does not refer, however, to utility of scales in which confirmation $f(\phi\rightarrow\psi)$ and confidence $conf(\phi\rightarrow\psi)$ are expressed. While the confidence $conf(\phi\rightarrow\psi)$ is the truth value of the knowledge pattern "*if* ϕ, *then* ψ", the confirmation measure $f(\phi\rightarrow\psi)$ says to what extend ψ is satisfied more frequently when ϕ is satisfied rather than when ϕ is not satisfied. In other words, f says what is a "value of information" that ϕ adds to the credibility of ψ. Remark that the semantics of these two values are different (see [16]).

The difference of semantics and utility of the two values, $conf(\phi\rightarrow\psi)$ and $f(\phi\rightarrow\psi)$, 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 6". Given three different premises: ϕ_1="the result is divisible by 3", ϕ_2="the result is divisible by 2" and ϕ_3="the result is divisible by 1", we get, respectively: $conf(\phi_1\rightarrow\psi)=1/2$, $f(\phi_1\rightarrow\psi)=2/3$, $conf(\phi_2\rightarrow\psi)=1/3$, $f(\phi_2\rightarrow\psi)=3/7$, and $conf(\phi_3\rightarrow\psi)=1/6$, $f(\phi_3\rightarrow\psi)=0$. While this example acknowledges the monotonicity of confirmation in confidence, it clearly shows that the value of f has a more useful interpretation than $conf$, in particular, in case of rule $\phi_3\rightarrow\psi$, which can also be read as "in any case, the result is 6"; indeed, the "any case" does not add any information which could confirm that the result is 6, and this fact is expressed by $f(\phi_3\rightarrow\psi)=0$.

The difference of semantics and utility of the two measures can be seen even better when considering two different hypotheses and the same 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, $conf(\phi\rightarrow\psi_1)=1/3$, $f(\phi\rightarrow\psi_1)=3/7$ and $conf(\phi\rightarrow\psi_2)=2/3$, $f(\phi\rightarrow\psi_2)=-3/7$. 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 it confirms conclusion ψ_1, which is expressed by a positive value of $f(\phi\rightarrow\psi_1)=3/7$.

Finally, as semantics of $f(\phi\rightarrow\psi)$ is more useful than that of $conf(\phi\rightarrow\psi)$, and as both these measures are monotonically linked while being independent of the support, it would be reasonable to search for the most interesting rules taking into account just confirmation $f(\phi\rightarrow\psi)$ and support $sup(\phi\rightarrow\psi)$.

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] Clark, P. and Boswell, P.: Rule Induction with CN2: Some Recent Improvements. [In]: *Machine Learning: Proc. of the Fifth European Conference*, 151-163 (1991).

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

[6] Fitelson, B.: *Studies in Bayesian Confirmation Theory*. Ph.D. Thesis, University of Wisconsin, Madison (2001).

[7] 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).

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

[9] Greco, S., Pawlak, Z. and Słowiński, R.: Can Bayesian Confirmation measures be useful for rough set decision rules? *Engineering Applications of Artificial Intelligence*, 17: 345-361 (2004).

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

[11] International Business Machines: *IBM Intelligent Miner User's Guide*, Version 1, Release 1 (1996).

[12] 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 24th Very Large Data Bases Conf.*, 380-391 (1998).

[13] 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).

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

[15] Piatetsky-Shapiro, G.: Discovery, Analysis, and Presentation of Strong Rules. Chapter 12 [In]: *Knowledge Discovery in Databases*, AAAI/MIT Press (1991).

[16] Słowiński, R. and Greco, S.: 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 (2005).

[17] Webb, G.I.: OPUS: An Efficient Admissible Algorithm for Unordered Search. *Journal of Artificial Intelligence Research*, 3: 431-465 (1995).