Poznań University of Technology

Four perspectives of confirmation

Salvatore Greco, Roman Słowiński, Izabela Szczęch

Research Report RA–01/2014

Institute of Computing Science

Piotrowo 2, 60–965 Poznań, Poland

Four perspectives of confirmation

Salvatore Greco[1,2], Roman Słowiński[3,4], Izabela Szczęch[3]

[1] Department of Economics & Business, University of Catania, Italy
salgreco@unicit.it
[2] Portsmouth Business School, Operations & Systems Management
University of Portsmouth, United Kingdom
[3] Institute of Computing Science, Poznań University of Technology, Poland
{Roman.Slowinski, Izabela.Szczech}
@cs.put.poznan.pl
[4] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Abstract. In the context of evaluation of if-then rules induced from data and properties that evaluation measures should possess, we discuss the fundamental property of confirmation. We distinguish four perspectives of such confirmation and compare them. We show logical equivalence of those four perspectives and propose a general definition of confirmation.

Keywords: Bayesian confirmation, strong Bayesian confirmation, confirmation measures, rule evaluation

1. Introduction

The evaluation of rule patterns induced from datasets is a valid and necessary step in the data mining process allowing to filter out rules that are useless or irrelevant. It is commonly done using measures of interest such as *confidence, support*, *rule interest function, lift* etc. (see [1], [7], [10], [12], [15] for a survey). Among the variety of measures, users are commonly interested in using those which reward the rules in which the premise confirms the conclusion (such measures are often referred to as measures of confirmation). But along with such a demand, there arises a question what does it mean that a premise *confirms* the conclusion? What does the *confirmation* stand for and how can we quantify it?
In this article we aim at answering the above questions by studying the concept of confirmation from four perspectives. These four perspectives derive from the Bayesian confirmation measure and interpret in an

exhaustive way probabilistic relationship of confirmation type between premise and conclusion. These perspectives lead, moreover, to four mathematical formulations of confirmation. We show the logical equivalence of all four formulations and propose a general formulation that has the advantage of being always defined (as opposed to the other four formulations that become undefined in certain conditions).
The rest of the paper is organized as follows. In the next Section we present preliminaries. Section 3 is the main part of this work- it distinguishes four perspectives of confirmation, shows their logical equivalence and proposes a general formulation of confirmation. Next, in Section 4 popular measures of confirmation are defined and their relation with the four perspectives is discussed. The paper ends with conclusions.

Identification of different perspectives of confirmation and their comparison should be regarded as the first step towards identification of properties that valuable confirmation measures should possess. Thus, our future work shall concentrate on finding desirable properties in the context of different confirmation perspectives. As a result, we hope to determine the most useful and meaningful measures (i.e., enjoying the desirable properties) that should be recommended as a tool for evaluation of rules induced from data.

2. Preliminaries about data, rules and supporting observations

A dataset is composed of a number of observations, called objects, described by a number of variables. The objects constitute a universe U from which rules are induced. Each rule is a consequence relation denoted by $E \to H$, read as "*if E, then H*". It consists of a premise (evidence) E and a conclusion (hypothesis) H. In the context of a particular dataset, the relation between E and H may be quantified by four non-negative numbers a, b, c and d, corresponding to all possible cases of truth and falsity of E and H, presented in a 2 x 2 contingency table (see Table 1).

Table 1. Contingency table of E and H for rule $E \rightarrow H$

	H	$\neg H$	$\sum$
E	a	c	$a+c$
$\neg E$	b	d	$b+d$
$\sum$	$a+b$	$c+d$	$\lvert U \rvert$

Precisely, the number of all objects in U supporting both the premise and the conclusion of a rule is quantified by a; b reflects the number of objects for which the premise in not satisfied, but the conclusion is, etc. Using a, b, c and d is common and intuitive for data mining techniques since all observations are collected in an information table describing each object by a set of variables. However, a, b, c and d can also be used to estimate probabilities: e.g., $\Pr(E)=(a+c)/|U|$ or $\Pr(H)=(a+b)/|U|$, $\Pr(H|E)=a/(a+c)$ (which, however, is only defined when $a + c > 0$).
The notation based on a, b, c and d can be effectively used for defining such interestingness measures as confidence, support, rule interest function, lift or measures of confirmation (see Section 4).

3. Four perspectives of confirmation

A common expectation with respect to the behavior of interestingness measures used for evaluation of rules is that they obtain:

- values > 0 when the premise of a rule confirms its conclusion,
- values = 0 when the rule's premise and conclusion are neutral to each other,
- values < 0 when the premise disconfirms the conclusion.

Such requirements are referred to (informally) as the property of confirmation and thus, measures acting according to them are called confirmation measures.

Moving to a formal definition of the concept of confirmation, we aim at reflecting the statement "E confirms H" in a more quantitative fashion. A commonly used definition of confirmation, called *Bayesian confirmation*, requires that an interestingness measure $c(H,E)$ satisfies the following conditions:

$$\mathrm{c}(H,E)\begin{cases} >0 \text{ if } \Pr(H \mid E) > \Pr(H), \\ =0 \text{ if } \Pr(H \mid E) = \Pr(H), \\ <0 \text{ if } \Pr(H \mid E) < \Pr(H). \end{cases} \qquad (1)$$

The definition of Bayesian confirmation identifies confirmation with an increase in the probability of the conclusion provided by the premise, neutrality with the lack of influence of the premise on the probability of conclusion, and finally disconfirmation with a decrease of probability of the conclusion imposed by the premise [2]. In the literature [5], [11], such understanding of confirmation is sometimes also called *incremental Bayesian confirmation*, as opposed to the *absolute confirmation* which assumes that E confirms H, if some kind of a threshold $k\in(0, 1)$ is surpassed by the conditional probability of H given E. This article however shall not cover the absolute confirmation.

It is important to stress that the Bayesian confirmation is not the only definition of confirmation. In the literature (see also [5], [11]), there are three other ways of expressing that E confirms H:

- $\Pr(H|E) > \Pr(H|\neg E)$,
- $\Pr(E|H) > \Pr(E)$,
- $\Pr(E|H) > \Pr(E|\neg H)$.

This gives four perspectives in which confirmation can be considered. Below, we propose a way of systematizing them, pointing out the differences between them and showing their logical equivalence.

To better distinguish the perspectives of confirmation let us call them the following way (for clarity of the presentation, in brackets we put only the conditions under which a measure should obtain positive values, as the conditions for neutrality and negative values are analogously formed):

(i) Bayesian confirmation ($\Pr(H|E) > \Pr(H)$),
(ii) strong Bayesian confirmation ($\Pr(H|E) > \Pr(H|\neg E)$),
(iii) likelihoodist confirmation ($\Pr(E|H) > \Pr(E)$),
(iv) strong likelihoodist confirmation ($\Pr(E|H) > \Pr(E|\neg H)$).

Those perspectives can be naturally grouped into pairs reaching to the debate between Bayesians and likelihoodists about confirmation's probabilistic interpretation [6]. Let us note that rule $E\rightarrow H$ in the Bayesian viewpoint, corresponds to rule $H\rightarrow E$ in the likelihoodist approach.

All those four perspectives have different philosophical background and motivations. They emphasize different faces of confirmation:

- the Bayesian confirmation states that E confirms H if H is more probable with E rather than without E, where "without E" means without knowing if E or $\neg E$ is true,
- the strong Bayesian confirmation stresses that E confirms H if H is more probable with E rather than with $\neg E$,
- the likelihoodist confirmation says that E confirms H if E is more probable with H rather than without H, where "without H" means without knowing if H or $\neg H$ is true,
- finally, the strong likelihoodist confirmation states that E confirms H if E is more probable with H rather than with $\neg H$.

To provide an interpretation of the four perspectives of confirmation let us use an illustrative example, in which the premise E is the evidence that a patient suffered from a fever and the conclusion H reflects that the patient had a flu. Then:

- in case of Bayesian confirmation (i) if flu is more probable with fever rather than without knowing whether the fever occurred or not, then fever confirms flu,
- in case of strong Bayesian confirmation (ii) if flu is more probable with fever rather than with no fever, than fever confirms flu,
- in case of likelihoodist confirmation (iii) if fever is more probable with flu rather than without knowing whether the flu occurred of not, then fever confirms flu,
- in case of strong likelihoodist confirmation (iv) if fever is more probable with flu rather than with no flu, then fever confirms flu.

Let us stress that the difference between those four perspectives of confirmation does not only come from different philosophical backgrounds, motivations or interpretations. The particular formulations in terms of probabilities or frequencies involving a, b, c and d also result in differences with respect to undefined situations they may lead to.
In particular, the perspective of Bayesian confirmation in terms of probabilities is formulated as $\Pr(H|E) > \Pr(H)$, which can be estimated by the non-negative frequencies as $a/(a+c)>(a+b)/|U|$. Clearly, such formulation requires that $\Pr(E)\neq 0$ (or more precisely $\Pr(E)>0$) or equivalently $a+c \neq 0$ (or more precisely $a+c > 0$).

The perspective of strong Bayesian confirmation has even stronger requirements, since the formulation $\Pr(H|E) > \Pr(H|\neg E)$ (or equivalently $a/(a+c)>b/(b+d)$) in order to be defined wants that $\Pr(E)\neq 0$ and $\Pr(\neg E)\neq 0$ (or equivalently $a+c \neq 0$ and $b+d \neq 0$).
Analogous considerations for the perspectives of likelihoodist confirmation and strong likelihoodist confirmation lead to requiring that $\Pr(H)\neq 0$ (or equivalently $a+b \neq 0$) or that $\Pr(H)\neq 0$ and $\Pr(\neg H)\neq 0$ (or equivalently $a+b \neq 0$ and $c+d \neq 0$), respectively.

3.1. Logical equivalence of four perspectives of confirmation

As the above considerations show, the four perspectives of confirmation should be regarded as alternative ways of formalizing this concept. Nevertheless, it is important to notice that the four formulations are logically equivalent, provided they do not lead to undefined values. By logical equivalence we understand that the conditions which need to be satisfied to switch between positive, zero and negative values are the same for all the formulations. Thus, they are not the same, but they "switch" in the same situations, which we will demonstrate below.

Let us observe that the situation of confirmation with respect to Bayesian confirmation is represented by the following inequality: $\Pr(H|E)>\Pr(H)$. Using the non-negative frequencies a, b, c and d, it can be expressed as $a/(a+c) > (a+b)/|U|$ (of course, we require that $a+c \neq 0$). Simple mathematical transformations show that $a/(a+c) > (a+b)/|U|$ iff $a|U| > (a+b)(a+c)$, which can be further simplified to $ad\text{-}bc > 0$. Thus, provided that Bayesian confirmation is defined (i.e., $a+c \neq 0$), the $ad-bc\{>,=,<\}0$ are the conditions for switching between situation of confirmation, neutrality and disconfirmation, respectively.

Regarding the strong Bayesian confirmation, the situation of confirmation is represented as $\Pr(H|E)>\Pr(H|\neg E)$, which can be also expressed as $a/(a+c) > b/(b+d)$ (of course, we require that $a+c \neq 0$ and $b+d \neq 0$). Simple mathematical transformations show that $a/(a+c) > b/(b+d)$ iff $a(b+d) > b(a+c)$, which can be further simplified to $ad-bc > 0$. Thus, provided that the definition of the strong Bayesian confirmation is defined (i.e., $a+c \neq 0$ and $b+d \neq 0$), $ad-bc\{>,=,<\}0$ are the conditions for switching between situation of confirmation, neutrality and disconfirmation, respectively.

Analogous transformations can be performed for likelihoodist and strong likelihoodist confirmations, showing that again the $ad-bc\{>,=,<\}0$ are the conditions for switching between situation of confirmation, neutrality and disconfirmation, respectively.

Summing up, we can formulate a general conclusion, that there are four (i) - (iv) alternative, different formulations of confirmation, but, provided they are defined, they all boil down to the following *general definition of confirmation* expressed in terms of the non-negative *a*, *b*, *c* and *d*, as:

$$\mathrm{c}(H,E)\begin{cases} >0 & \text{if } ad-bc>0, \\ =0 & \text{if } ad-bc=0, \\ <0 & \text{if } ad-bc<0. \end{cases} \tag{2}$$

The logical equivalence of Bayesian confirmation, strong Bayesian confirmation, likelihoodist confirmation and strong likelihoodist confirmation with the "$ad-bc$" formulation is true provided that none of the (i)-(iv) formulations is undefined, which means that all the following sums: $a+c$, $b+d$, $a+b$ and $c+d$ are non-zero.

The above general definition of confirmation has the advantage over the (i)-(iv) formulations of never being undefined. The fact that there are no denominators in it guarantees that for any dataset, and thus any particular contingency table with *a*, *b*, *c* and *d*, definition (2) determines whether we are in the situation of confirmation, neutrality of disconfirmation. On the other hand, working with Bayesian confirmation, strong Bayesian confirmation, likelihoodist confirmation or strong likelihoodist confirmation we can also obtain the undesirable undefined situations (when $a+c=0$, or $b+d=0$, or $a+b=0$, or $c+d=0$).

4. Popular measures of confirmation

Measures that possess the property of confirmation defined as Bayesian confirmation strong Bayesian confirmation, likelihoodist confirmation or strong likelihoodist confirmation are referred to as *confirmation measures* or *measures of confirmation*. Sometimes an adjective comes into the denotation e.g. Bayesian confirmation measures.

Due to the logical equivalence of all the four analyzed formulations of confirmation (i)-(iv) we can conclude that a measure satisfying the Bayesian confirmation must also satisfy strong Bayesian confirmation, likelihoodist confirmation or strong likelihoodist confirmation, as long as we exclude undefined values. It is thus legitimate to call such measures simply measures of confirmation (or confirmation measures).
Such measures quantify the degree to which the premise E provides "support for or against" the conclusion H [5], the degree to which E confirms/disconfirms H. By using confirmation measures in the rule evaluation process, we aim at limiting the set of rules proposed to the user [16].
Let us observe, that the constraints put on a measure by any of the four (i)-(iv) formulations of confirmation are that a measure assigns positive values in the situation when confirmation occurs, negative values in case of disconfirmation, and zero otherwise. In consequence of that many alternative, non-equivalent measures of confirmation have been defined. Among the most commonly used ones, there are (see also [4]):

Table 2. Popular measures of confirmation

$D(H,E) = \Pr(H\|E) - \Pr(H) = \frac{a}{a+c} - \frac{a+b}{\|U\|}$	[2]
$M(H,E) = \Pr(E\|H) - \Pr(E) = \frac{a}{a+b} - \frac{a+c}{\|U\|}$	[13]
$S(H,E) = \Pr(H\|E) - \Pr(H\|\neg E) = \frac{a}{a+c} - \frac{b}{b+d}$	[3]
$N(H,E) = \Pr(E\|H) - \Pr(E\|\neg H) = \frac{a}{a+b} - \frac{c}{c+d}$	[14]
$C(H,E) = \Pr(E \wedge H) - \Pr(E)\Pr(H) = \frac{a}{\|U\|} - \frac{(a+c)(a+b)}{\|U\|^2}$	[2]
$F(H,E) = \frac{\Pr(E\|H) - \Pr(E\|\neg H)}{\Pr(E\|H) + \Pr(E\|\neg H)} = \frac{ad-bc}{ad+bc+2ac}$	[9]
$Z(H,E) = \begin{cases} 1 - \frac{\Pr(\neg H\|E)}{\Pr(\neg H)} = \frac{ad-bc}{(a+c)(c+d)} & \text{in case of confirmation} \\ \frac{\Pr(H\|E)}{\Pr(H)} - 1 = \frac{ad-bc}{(a+c)(a+b)} & \text{in case of disconfirmation} \end{cases}$	[4]

$A(H,E) = \begin{cases} \dfrac{\Pr(E\|H) - \Pr(E)}{1 - \Pr(E)} = \dfrac{ad - bc}{(a+b)(b+d)} & \text{in case of confirmation} \\ \dfrac{\Pr(H) - \Pr(H\|\neg E)}{1 - \Pr(H)} = \dfrac{ad - bc}{(b+d)(c+d)} & \text{in case of disconfirmation} \end{cases}$	[8]
$c_1(H,E) = \begin{cases} \alpha + \beta A(H,E) & \text{in case of confirmation when } c = 0 \\ \alpha Z(H,E) & \text{in case of confirmation when } c > 0 \\ \alpha Z(H,E) & \text{in case of disconfirmation when } a > 0 \\ -\alpha + \beta A(H,E) & \text{in case of disconfirmation when } a = 0 \end{cases}$	[8]
$c_2(H,E) = \begin{cases} \alpha + \beta Z(H,E) & \text{in case of confirmation when } b = 0 \\ \alpha A(H,E) & \text{in case of confirmation when } b > 0 \\ \alpha A(H,E) & \text{in case of disconfirmation when } d > 0 \\ -\alpha + \beta Z(H,E) & \text{in case of disconfirmation when } d = 0 \end{cases}$	[8]
$c_3(H,E) = \begin{cases} A(H,E)Z(H,E) & \text{in case of confirmation} \\ -A(H,E)Z(H,E) & \text{in case of disconfirmation} \end{cases}$	[8]
$c_4(H,E) = \begin{cases} \min(A(H,E), Z(H,E)) & \text{in case of confirmation} \\ \max(A(H,E), Z(H,E)) & \text{in case of disconfirmation} \end{cases}$	[8]

Application of popular confirmation measures (e.g. all measures from Table 2 except for measure $C(H,E)$) for evaluation of rules induced from datasets may lead to obtaining undefined values in particular cases, e.g. for any rule characterized by a contingency table with $a+c = 0$ an undefined value of measure $D(H,E)$ is obtained. Such situations are inconvenient and troublesome for data analysts and are generally undesirable. In such situations a measure simply cannot evaluate a rule. In our opinion, the more often undefined values can occur for a measure, the less useful the measure is.

Let us observe, however, that taking advantage of the general definition of the confirmation (2), we can avoid situations when for a particular measure, the answer id undefined. Since the condition $ad-bc = 0$ requires that a measure obtains value 0, we can assume that any measure (despite its actual definition) defaults to 0 whenever $ad-bc = 0$. This way, we would not even calculate the value of a measure but basing on the fact that $ad-bc = 0$ assign it to 0. Such an approach would allow us to

eliminate some undefined values, e.g. for a contingency table where $a = c = 0$, we would obtain an undefined value of measure $D(H,E)$ (the denominator would be equal to 0), however using the general definition of confirmation, we see that $a = c = 0$ results in the situation of neutrality because $ad-bc = 0$, and thus instead of an undefined value we could say that $D(H,E)$ states neutrality.
Let us stress that the above proposition is inspired by practical experiments, in which dealing with undefined values perturbs the rule evaluation procedure.

5. Conclusions

It is commonly expected that in rules presented as a result of data mining tasks, the premise should confirm the conclusion. The paper aimed at answering what such a confirmation means. In particular we have distinguished four perspectives of confirmation and compared them. We have arrived at the conclusion that all of those perspectives are logically equivalent and can be boiled down to a general formulation of confirmation.
The results presented in this paper are a proper starting point towards determination which properties of measures are desirable in the light of different confirmation perspectives. Thus, our future work will concentrate on the identifying which properties are valuable for measures of confirmation. Possession of such properties would then be regarded as responding to the user's expectations, and particular measures would be highly useful for practical applications.

6. References

[1] Bramer, M., 2007. *Principles of Data Mining*, Springer-Verlag, New York Inc.

[2] Carnap, R., 1962. *Logical Foundations of Probability*, 2nd ed., University of Chicago Press, Chicago.

[3] Christensen, D., 1999. Measuring confirmation. *Journal of Philosophy*, vol. 96, 437-461.

[4] Crupi V., Tentori, K., Gonzalez, M., 2007. On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*, vol. 74, 229-252.

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

[6] Fitelson B., 2007. Likelihoodism, Bayesianism, and Relational Confirmation. *Synthese*, vol. 156, no. 3, 473-489.

[7] Geng, L., Hamilton, H.J., 2006. Interestingness Measures for Data Mining: A Survey. *ACM Computing Surveys*, vol. 38, no. 3, artic 9.

[8] Greco, S., Słowiński, R., Szczęch, I., 2012. Properties of rule interestingness measures and alternative approaches to normalization of measures. *Information Sciences*, vol.216, 1-16.

[9] Kemeny, J., Oppenheim, P., 1952. Degrees of factual support. *Philosophy of Science*, 19, 307-324.

[10] Lenca, Ph., Meyer, P., Vaillant, B., Lallich, S., 2008. On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. *European J. of Operational Research*, vol. 184, no. 2, 610-626.

[11] Maher, P., 2005. *Confirmation Theory*. The Encyclopedia of Philosophy (2nd ed.). Macmillan Reference, USA.

[12] McGarry, K., 2005. A survey of interestingness measures for knowledge discovery. *The Knowledge Engineering Review*, vol. 20, no.1, 39-61.

[13] Mortimer, H., 1988. *The Logic of Induction*, Paramus, Prentice Hall.

[14] Nozick, R., 1981. *Philosophical Explanations*, Clarendon Press, Oxford (UK).

[15] Omiecinski, E., 2003. Alternative interest measures for mining associations in databases. *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 1, 57-69.

[16] Szczęch I., 2009. Multicriteria Attractiveness Evaluation of Decision and Association Rules. *Transactions on Rough Sets X*, LNCS series, vol. 5656, Springer, Berlin, pp.197-274.