Properties of rule interestingness measures and alternative approaches to normalization of measures

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

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

[2] Institute of Computing Science, Poznań University of Technology, 60–965 Poznań, Poland
{Roman.Slowinski, Izabela.Szczech }@cs.put.poznan.pl

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

Abstract. We are considering properties of interestingness measures of rules induced from data. These are: Bayesian confirmation property, two properties related to the case of entailment or refutation, called Ex_1 and logicality L, and a group of symmetry properties. We propose a modification of properties Ex_1 and L, called weak Ex_1, and weak L, that deploy the concept of confirmation in its larger sense. We demonstrate that properties Ex_1 and L do not fully reflect such understanding of the confirmation concept, and thus, we propose to substitute Ex_1 by weak Ex_1 and L by weak L. Moreover, we introduce four new approaches to normalization of confirmation measures in order to transform measures so that they would obtain desired properties. The analysis of the results of the normalizations of the confirmation measures takes into account all considered properties. We advocate for two normalized confirmation measures: measure *Z* considered in the literature, and newly proposed measure *A*. Finally, we provide some ideas for combining them in a single measure keeping all desirable properties.

Keywords: Rule interestingness measures; Properties of measures; Confirmation; Normalization

1. Introduction

One of the main objectives of data mining process is to identify "*valid*, *novel*, *potentially useful*, and *ultimately comprehensible* knowledge from databases" [9], [33]. The discovered knowledge (patterns) is often expressed in a form of "*if*..., *then*..." rules, which are consequence relations reflecting relationship, association, causation, etc., between independent (i.e. those in the premise of the rule) and dependent (i.e. those in the conclusion of the rule) attributes. The number of rules discovered in databases is often overwhelmingly large rising an urgent need to identify the most useful ones and filter out those that are irrelevant. In order to help to deal with this problem, various quantitative measures of rule interestingness (attractiveness) have been proposed and studied. Among the most commonly used interestingness measures there are *support*, *confidence*, *lift*, *rule interest function* (for a survey on interestingness measures see [3], [15], [25], [27], [31]).

Each of the measures proposed in the literature has been introduced to reflect different characteristics of rules. For example, measures like support [1] value generality (also referred to as coverage) of the rule, i.e. favor rules that cover a relatively large subset of a dataset. In opposition, there are measures that bring forth peculiarity, believing that patterns far away from other discovered knowledge, according to some distance measure, may be unknown to the user and therefore interesting. The list of characteristics that are emphasized by different measures is long and covers conciseness, reliability, novelty, surprisingness, utility, actionability, among others [15].

Generally, interestingness measures can be categorized as *objective* and *subjective* measures. The first group can be established through statistical arguments derived from data to determine whether a rule is interesting or not. No knowledge about the user or application is needed. For example, rules that cover only very few objects from the dataset, and can therefore capture spurious relationships in data, are discarded by objective measures [21].

On the other hand, the group of subjective measures takes into account both the data and the user, thus, those measures require interaction with the user to obtain information about the user's background knowledge and expectations. Subjective measures regard a rule as uninteresting unless it reveals unexpected information about the data or provides knowledge that can lead to profitable actions [39], [40]. Thus, for subjective evaluation criteria rare cases in the data are often interesting and rules that cover them are of high value.

All in all, objective measures depend on the structure of the rules and the underlying data used in the discovery process, whereas the subjective measures also rely on the class of users who examine the rule [35].

Moreover, measuring the interestingness of discovered patterns receives recently much attention from researchers developing the paradigm of granular computing (see, e.g., the rough-set-based granular computing in [2], [17], [18], [32], [37], [38]).

A common conclusion stemming from this broad interest in measuring attractiveness of discovered rules is that there is no single way that would work the best on any real-life problem. The literature is a rich resource of ordinally non-equivalent measures that reflect different characteristics of rules and rank them in different ways. As there is no agreement which measure is the best, the choice of an interestingness measure for a particular application is a non-trivial task that should closely relate to the domain of application and should take advantage of available domain knowledge.

To help to analyze objective measures and to choose one for a certain application, some properties have been proposed. They express the user's expectations towards the behavior of measures in particular situations. Those expectations can be of various types, e.g., one could desire to use only such measures that reward the rules having a greater number of objects supporting the pattern. In general, properties group the measures according to similarities in their characteristics, thus using the measures which satisfy the desirable properties one can avoid considering unimportant rules. Different properties have been proposed and surveyed in [5], [8], [15], [16], [19], [25], [36], [39].

Among the commonly used properties of rule interestingness measures there are:

- *property of confirmation* related to quantification of the degree to which the premise of the rule provides evidence for or against the conclusion [5], [12];
- *property* Ex_1 assuring that any conclusively confirmatory rule is assigned a higher value of interestingness measure than any rule which is not conclusively confirmatory, and any conclusively disconfirmatory rule is assigned a lower value than any rule which is not conclusively disconfirmatory [7], [20];
- *property* L, called *logicality*, for which any conclusively confirmatory rule is assigned the maximum value, and any conclusively disconfirmatory rule is assigned the minimum value [7], [12]; properties Ex_1 and L can be regarded as strongly related, as both of them deal with the behavior of confirmation measures in cases of conclusive confirmation or conclusive disconfirmation;
- *properties of symmetry* being a whole set of properties that describe desirable and undesirable behavior of measures in cases when the premise or conclusion in not satisfied, or when the premise and conclusion switch positions in a rule [4], [7], [8], [12].

This paper concentrates on the abovementioned properties of objective interestingness measures. We propose a modification of properties Ex_1 and L, called *weak* Ex_1, and *weak* L, that deploy the concept of confirmation in its larger sense. In fact, according to the deep meaning of the confirmation concept, a confirmation measure should give an account of the credibility that it is more probable to have the conclusion when the premise is present, rather than when the premise is absent. We demonstrate that properties Ex_1 and L do not fully reflect such understanding of the confirmation concept, and thus, we propose to substitute Ex_1 by weak Ex_1 and L by weak L.

Moreover, since Crupi et al. [7] represent Bayesian approach to defining Ex_1 and L, we enrich their point of view by considering also likelihoodist counterparts of those properties, denoted as L-Ex_1 and L-L, respectively.

Next, we introduce four new approaches to normalization of confirmation measures in order to transform measures so that they would obtain desired properties. The analysis of the results of the normalizations of the confirmation measures considers property Ex_1, L-Ex_1, weak Ex_1, L, L-L, weak L, and the properties of symmetry.

As the final contribution, we propose a new measure *A* that fulfils all the desirable symmetry properties. Its strength lies in the fact that it does not possess the property Ex_1, but its likelihoodist counterpart L-Ex_1. On the basis of these remarks, we argument that measure *A* and measure *Z* proposed by Crupi et al. [7], should be considered as complementary tools for assessing the quality of rules. At the end, we provide some ideas for combining them in a single measure keeping all desirable properties.

2. Preliminaries

A *rule* induced from a dataset on a universe U shall be denoted by $E\to H$ (read as "*if E*, *then H"*). It consists of a premise (evidence) E and a conclusion (hypothesis) H.

In general, by $sup(\gamma)$ we denote the number of objects in the dataset for which γ is true, e.g., $sup(E)$ is the number of objects in the dataset satisfying the premise, and $sup(H, E)$ is the number of objects satisfying both the premise and the conclusion of a $E\to H$ rule.

Moreover, the following notation shall be used throughout the paper: $a=sup(H, E)$, $b=sup(H, \neg E)$, $c=sup(\neg H, E)$, $d=sup(\neg H, \neg E)$. Observe that b can be interpreted as the number of objects that do not satisfy the premise but satisfy the conclusion of the $E\to H$ rule. Analogously, $c=sup(\neg H, E)$ can be construed as the number of objects in the dataset that satisfy the premise but do not satisfy the conclusion of the $E\to H$ rule, and $d=sup(\neg H, \neg E)$ can be interpreted as the number of objects in the dataset that do not satisfy neither the premise nor the conclusion of the $E\to H$ rule. Moreover, the following relations occur: $a+c=sup(E)$, $a+b=sup(H)$, $b+d=sup(\neg E)$, $c+d=sup(\neg H)$, and the cardinality of the dataset U, denoted by $|U|$, is the sum of a, b, c and d.

Reasoning in terms of a, b, c and d is natural and intuitive for data mining techniques since all observations are gathered in some kind of an information table describing each object by a set of attributes. However, a, b, c and d can also be regarded as frequencies that can be used to estimate probabilities: e.g., $\Pr(E)=(a+c)/|U|$ or $\Pr(H)=(a+b)/|U|$.

3. Desirable properties of objective measures

The problem of choosing an appropriate interestingness measure for a certain application is non-trivial because the number and variety of measures proposed in the literature is overwhelming. Therefore, there naturally arises a need to analyze theoretical properties of measures, which allow for grouping the measures and unveiling relationships between them.

3.1. Property of Bayesian confirmation

Formally, an interestingness measure $c(H, E)$ has the property of Bayesian confirmation if and only if it satisfies the following conditions:

$$c(H,E)\begin{cases} >0 \;\; if \;\Pr(H\mid E)>\Pr(H), \\ =0 \;\; if \;\Pr(H\mid E)=\Pr(H), \\ <0 \;\; if \;\Pr(H\mid E)<\Pr(H). \end{cases} \tag{BC}$$

In the contemporary literature [12], [26], this conception of confirmation is also known as *incremental Bayesian confirmation*, in order to distinguish it from the *absolute confirmation* which assumes that the premise E confirms the conclusion H, if some kind of a threshold $k\in(0, 1)$ is exceeded by the probability of H given E. This article concentrates only on the incremental Bayesian confirmation, which shall be denoted as confirmation for simplicity.

The (BC) definition 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 [5].

Estimating probabilities in terms of frequencies, the (BC) conditions can be expressed in terms of *a*, *b*, *c* and *d*:

$$c(H,E)\begin{cases} >0 \ \ if \ \ \dfrac{a}{a+c} > \dfrac{a+b}{|U|}, \\ =0 \ \ if \ \ \dfrac{a}{a+c} = \dfrac{a+b}{|U|}, \\ <0 \ \ if \ \ \dfrac{a}{a+c} < \dfrac{a+b}{|U|}. \end{cases}$$

The above conditions are used to check if for a given rule $E \rightarrow H$ we have the case of confirmation or the case of disconfirmation, respectively.

If one adopts Kolmogorov theory of probability [24] (i.e., assumes that Pr is a Kolmogorov probability function) there are many different, but logically equivalent, ways of expressing that *E* confirms *H*:

$\Pr(H|E) > \Pr(H)$

$\Pr(H|E) > \Pr(H|\neg E)$

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

Since they are equivalent (see also [12], [26]), one can also express the (BC) conditions as:

$$c(H,E)\begin{cases} >0 \ \ if \ Pr(H \mid E) > Pr(H \mid \neg E), \\ =0 \ \ if \ Pr(H \mid E) = Pr(H \mid \neg E), \\ <0 \ \ if \ Pr(H \mid E) < Pr(H \mid \neg E). \end{cases} \tag{BC'}$$

To avoid ambiguity, we shall denote the above formulation as (BC'). Taking this formulation into account, it can be concluded that *E* confirms *H* when *E* raises the probability of *H*, and *E* raises the probability of *H* if the probability of *H* given *E* is higher than the probability of *H* given non *E*.

Measures that possess the property of confirmation are referred to as *confirmation measures* or *measures of confirmation*. They quantify the degree to which the premise *E* provides "support for or against" the conclusion *H* [12]. Thus, for a given rule $E \rightarrow H$, interestingness measures with the property of confirmation express the credibility of the following proposition: *H is satisfied more frequently when E is satisfied, rather than when E is not satisfied*. This interpretation stresses the very valuable semantics of the property of confirmation. By using the interestingness measures that possess this property one can filter out rules which are misleading and disconfirm the user, and this way, limit the set of induced rules only to those that are meaningful [36].

The only constraints that the conditions (BC) put on a measure is that it assigns positive values in the situation when confirmation occurs, negative values in case of disconfirmation, and zero otherwise. As a result many alternative, non-equivalent measures of confirmation have been proposed. Now, the catalogue of confirmation measures available in the literature is quite large and the condition (BC) itself does not favor one single measure as the most adequate [11]. The most commonly used ones are gathered in Table 1 (selection provided in [7]):

Table 1. Popular confirmation measures

$D(H,E) = Pr(H \mid E) - Pr(H) = \frac{a}{a+c} - \frac{a+b}{\lvert U \rvert}$	Carnap [5]
$S(H,E) = Pr(H \mid E) - Pr(H \mid \neg E) = \frac{a}{a+c} - \frac{b}{b+d}$	Christensen [6]
$M(H,E) = Pr(E \mid H) - Pr(E) = \frac{a}{a+b} - \frac{(a+c)}{\lvert U \rvert}$	Mortimer [28]
$N(H,E) = Pr(E \mid H) - Pr(E \mid \neg H) = \frac{a}{a+b} - \frac{c}{c+d}$	Nozick [30]
$C(H,E) = Pr(E \wedge H) - Pr(E)Pr(H) = \frac{a}{\lvert U \rvert} - \frac{(a+c)(a+b)}{\lvert U \rvert^2}$	Carnap [5]
$R(H,E) = \frac{Pr(H \mid E)}{Pr(H)} - 1 = \frac{a \lvert U \rvert}{(a+c)(a+b)} - 1$	Finch [10]
$G(H,E) = 1 - \frac{Pr(\neg H \mid E)}{Pr(\neg H)} = 1 - \frac{c \lvert U \rvert}{(a+c)(c+d)}$	Rips [34]

3.2. Properties Ex_1 and L

To handle the plurality of alternative confirmation measures, Crupi, Tentori and Gonzalez [7] have proposed a property (principle) Ex_1 resorting to considering inductive logic as an extrapolation from classical deductive logic. On the basis of classical deductive logic they construct a function *v*:

$$v(H,E) = \begin{cases} \textit{the same positive value, denoted as } V, \textit{ if } E \models H; \\ \textit{the same negative value, denoted as } -V, \textit{ if } E \models \neg H; \\ 0, \textit{ otherwise.} \end{cases}$$

For any argument (*H*, *E*) function *v* assigns it the same positive value *V* (e.g., +1) if and only if the premise *E* of the rule entails the conclusion *H* (i.e. $E \models H$). The same value but of opposite sign $-V$ (e.g., -1) is assigned if and only if the premise *E* refutes the conclusion *H* (i.e. $E \models \neg H$). In all other cases (i.e. when the premise is neither conclusively confirmatory nor conclusively disconfirmatory for the conclusion) function *v* obtains value 0.

From definition, any confirmation measure *c*(*H, E*) agrees with function *v*(*H, E*) in the way that if *v*(*H, E*) is positive, then the same is true of *c*(*H, E*), and when *v*(*H, E*) is negative, so is *c*(*H, E*). According to Crupi et al., the relationship between the logical implication or refutation of *H* by *E*, and the conditional probability of *H* subject to *E* should go further and demand fulfillment of the following principle (Ex_1):

$$\textit{if } v(H_1,E_1) > v(H_2,E_2) \textit{ then } c(H_1,E_1) > c(H_2,E_2) \qquad (Ex_1)$$

Property Ex_1 guarantees that the measure will assign a greater value for any conclusively confirmatory rule (i.e. such that $E \models H$) than for any rule which is not conclusively confirmatory. Moreover, rules that are conclusively disconfirmatory (i.e. such that $E \models \neg H$) will obtain smaller values of interestingness measures than any rule which is not conclusively disconfirmatory.

Let us now explain the consequences of property Ex_1 by considering three rules:

- $E \rightarrow H$, such that there are only positive examples to the rule and no counterexamples, i.e. $a>0$ and $c=0$, which implies that $v(H, E) = 1$;
- $E' \rightarrow H'$, such that there are some positive examples and some counterexamples to the rule, i.e. $a'>0$ and $c'>0$, which implies that $v(H',E') = 0$;
- $E'' \rightarrow H''$, such that there are no positive examples and some counterexamples to the rule, i.e. $a''=0$ and $c''>0$, which implies that $v(H'',E'') = -1$.

If a confirmation measure satisfies Ex_1, then the ordering based on function v: $v(H, E) > v(H', E') > v(H'', E'')$, implies the following relations between confirmation measures: $c(H,E) > c(H',E') > c(H'',E'')$.

Another property which is closely related to Ex_1 is *logicality* property L, discussed among others by Carnap [5] and Fitelson [13], that can be expressed as follows:

- $c(H, E)$ is maximal when $E \models H$
- and $c(H, E)$ is minimal when $E \models \neg H$. (L)

In terms of conditional probability, property L can be expressed as follows:

- $c(H, E)$ attains its maximum if $\Pr(H|E)=1$
- and $c(H, E)$ attains its minimum if $\Pr(H|E)=0$.

Equivalently, a confirmation measure possessing property L obtains its maximum when there are no counterexamples to a rule, i.e. $c=0$, and obtains its minimum when there are no positive examples to a rule, i.e. when $a=0$.

Crupi et al. [7] note that L and Ex_1 are independent in the sense that there are confirmation measures satisfying L but not Ex_1, as well as there are confirmation measures satisfying Ex_1 but not L. As an example of the first situation, consider a measure assigning a fixed value, e.g. 1, in any case of confirmation, another fixed value, e.g. -1, in any case of disconfirmation, and 0 in case of neutrality. The case of confirmation corresponds to $\Pr(H|E)>\Pr(H)$, which means that $ad>bc$. The case of disconfirmation corresponds to $\Pr(H|E)<\Pr(H)$, which means that $ad<bc$. The case of neutrality corresponds to $\Pr(H|E)=\Pr(H)$, which means that $ad=bc$.

As an example of the second situation, where a confirmation measure satisfies Ex_1 but not L, let us propose measure $c_1(H, E)$ defined as follows:

$$c_1(H,E)=\begin{cases} \alpha+\beta\dfrac{ad\text{-}bc}{(a+b)(b+d)} & \textit{in case of confirmation if } c=0, \\ \alpha\dfrac{ad\text{-}bc}{(a+c)(c+d)} & \textit{in case of confirmation if } c>0, \\ \alpha\dfrac{ad\text{-}bc}{(a+b)(a+c)} & \textit{in case of disconfirmation if } a>0, \\ -\alpha+\beta\dfrac{ad\text{-}bc}{(b+d)(c+d)} & \textit{in case of disconfirmation if } a=0, \end{cases}$$

where $\alpha+\beta=1$, and $\alpha>0$, $\beta>0$. First, observe that $c_1(H, E)$ is a confirmation measure because it is positive when $\Pr(H|E)>\Pr(H)$ (i.e. when $ad>bc$), it is equal to 0 when $\Pr(H|E)=\Pr(H)$ (i.e. when $ad=bc$), and it is negative when $\Pr(H|E)<\Pr(H)$ (i.e. when $ad>bc$). Moreover, measure $c_1(H, E)$ satisfies Ex_1 because all the cases for which $E|=H$ (i.e., conclusively confirmatory cases) get a greater value of $c_1(H, E)$ than cases for which $E|=H$ is not true (i.e., cases which are not conclusively confirmatory). In fact, if $E|=H$ then $c_1(H, E)=\alpha+\beta(ad-bc)/[(a+b)(b+d)]$, and if $E|=H$ is not true then $c_1(H, E)=\alpha(ad-bc)/[(a+c)(c+d)]$.

One can observe that $\alpha+\beta(ad-bc)/[(a+b)(b+d)] > \alpha \geq \alpha(ad-bc)/[(a+c)(c+d)]$, and, therefore, the ordering of conclusively confirmatory cases and non-conclusively confirmatory cases is the same in terms of function v and measure $c_1(H, E)$. Similarly for the disconfirmation, all the cases for which $E \models \neg H$ (i.e., conclusively disconfirmatory cases) receive smaller values of $c_1(H, E)$ than cases for which $E \models \neg H$ is not true (i.e., cases which are not conclusively disconfirmatory). Therefore, the ordering of conclusively disconfirmatory cases and non-conclusively disconfirmatory cases is also the same in terms of function v and measure $c_1(H, E)$. Thus, we can conclude that measure $c_1(H, E)$ satisfies property Ex_1.

However, complete absence of counterexamples (c=0), i.e., the case when $E \models H$, is not enough to result in the maximal value of measure $c_1(H, E)$. The maximum (i.e., $c_1(H, E) = 1$) is obtained when c=0 and b=0. Analogous observation holds in case of the minimal value of $c_1(H, E)$. The absence of positive examples (a=0), i.e., the case when $E \models \neg H$, is not a sufficient condition for the measure to reach the minimum. Measure $c_1(H, E)$ obtains its minimal value when both a=0 and d=0. Thus, $c_1(H, E)$ satisfies property Ex_1 but not L.

Nevertheless, even if properties Ex_1 and L are independent, they are strongly related, because both of them deal with the behavior of the confirmation measure in case when $E \models H$ and when $E \models \neg H$.

To better explain the relationship between properties Ex_1 and L let us use an example of drawing cards from a standard deck. We review three possible situations: conclusively confirmatory, non-conclusively confirmatory (or disconfirmatory), and conclusively disconfirmatory rules.

A rule: r_1≡"if x is seven of spades then x is black"

is conclusively confirmatory as the premise (drawing seven of spades) entails the conclusion that the drawn card is black. Since $E \models H$, it is clear that there are no counterexamples to the rule, i.e. c=0. Such conclusively confirmatory rule should also be assigned maximal value V of a function $v(H, E)$. For such a rule an interestingness measure possessing property L should assign a maximal value (e.g., +1).

As an example of a rule which is neither conclusively confirmatory nor conclusively discomfirmatory let us consider a rule:

r_2≡"if x is black then x is seven of spades".

Drawing a black card, we can be lucky to get a seven of spades, but it is not a 100% sure situation, therefore the premise does not entail the conclusion and the rule is not conclusively confirmatory. The rule is not conclusively disconfirmatory either as the premise does not refute the conclusion as there is a chance of drawing that seven. For such rules, neither conclusively confirmatory nor conclusively discomfirmatory, function $v(H, E)$ obtains value 0, which implies that confirmation measures with property Ex_1 should assign to such rules values smaller than to rules conclusively confirmatory (for which $v(H, E)=V$, where $V>0$).

Finally, an example of conclusively disconfirmatory rule could be the following:

r_3≡"if x is seven of spades then x is red".

One can never get a red card having drawn a seven of spades, thus the premise in the above rule completely disconfirms the conclusion. For such a rule $E \models \neg H$ and there are no positive examples to the rule (i.e. a=0). Thus, such rule will be assigned a minimal value (e.g., -1) by a measure possessing property L. Moreover, function $v(H, E)$ shall assign the value $-V$ (where $-V < 0$) to a conclusively disconfirmatory rule. Therefore, a confirmation measure satisfying Ex_1 assigns to such rule a smaller value than to rules which are neither conclusively confirmatory nor conclusively disconfirmatory. Consequently, confirmation measure satisfying Ex_1 assigns to conclusively disconfirmatory rules a smaller value than to conclusively confirmatory.

In terms of above rules r_1, r_2 and r_3, for properties L and Ex_1 we have:

- c(black, seven spades) = $c(r_1) = 1$ and c(red, seven spades) = $c(r_3) = -1$ if the measure possesses property L,
- c(black, seven spades) = $c(r_1) >$ c(seven spades, black) = $c(r_2) >$ c(red, seven spades) = $c(r_3)$ if the measure satisfies property Ex_1.

Concluding, interestingness measures satisfying property L and/or Ex_1 have the ability to rank the rules in such a way that those in which the premise entails the conclusion (e.g., the rule: *if x is seven of spades then x is black*) are on top of the ranking, those in which the premise refutes the conclusion (e.g., *if x is seven of spades then x is red*) are on the very bottom, and rules which are neither 100% sure nor 100% false are in between.

Remark that in case of confirmation, both Ex_1 and L concern situations of entailment, which is equivalent to $\Pr(H|E)=1$. However, confirmation should express how much it is more probable to have H when E is present rather than when E is absent. Thus, the requirement $\Pr(H|E)=1$ is not sufficient, and properties Ex_1 and L should be modified to take into account also the value of $\Pr(H|\neg E)$. In particular, $\Pr(H|\neg E)$ should be equal to zero for maximal confirmation in case of entailment. Analogical requirements concern the case of disconfirmation. These considerations lead to new properties Ex_1 and L, called weak Ex_1 and weak L, which are described in the next point.

3.2.1. Desirable modification of properties Ex_1 and L into properties weak Ex_1 and weak L

Properties Ex_1 and L can be regarded as one-sided because they focus on situations when $E \models H$ (i.e. there are no counterexamples to a rule and c=0), and situations when $E \models \neg H$ (i.e. there are no positive examples to a rule and a=0).

In our opinion, the concept of confirmation is too complex and rich to be boiled down simply to verification whether there are no counterexamples or no positive examples. We claim that it is also important how the measure behaves in intermediate cases – between the absence of counterexamples and absence of positive examples for a rule.

Let us explain our opinion by taking into account the formulation of (BC') conditions which state that: when E confirms H, this means that E raises the probability of H, and E raises the probability of H if the probability of H given E is higher than the probability of H given $\neg E$. We believe that it is reasonable to conclude that, *in case of confirmation*, a confirmation measure $c(H, E)$ should express *how much it is more probable to have H when E is present rather than when E is absent*.

Analogously, let us interpret (BC') conditions as: E disconfirms H, which means that E decreases the probability of H, and E decreases the probability of H if the probability of H given E is smaller than the probability of H given $\neg E$. We believe that it is reasonable to conclude that *in case of disconfirmation* a confirmation measure $c(H, E)$ should express *how much it is less probable to have H when E is present rather than when E is absent*.

Taking into account such interpretations, we can formulate a property called *weak* Ex_1, which generalizes the original Ex_1 property:

$$\text{if } v(H_1,E_1) > v(H_2,E_2) \text{ and } v(H_1,\neg E_1) < v(H_2,\neg E_2) \text{ then } c(H_1,E_1) > c(H_2,E_2) \qquad \text{(weak Ex}_1\text{)}$$

Property weak Ex_1 guarantees that a confirmation measure $c(H, E)$ cannot attain its maximal value unless the two following conditions are satisfied:

1) $E \models H$, or equivalently, $\Pr(H|E)=1$, or equivalently, $c=sup(\neg H, E)=0$.

2) $\neg E \models \neg H$, or equivalently, $\Pr(H|\neg E)=0$, or equivalently, $b=sup(H, \neg E)=0$.

Let us supplement that for a given dataset, $E \models H$

$$\Leftrightarrow \Pr(H \mid E) = \frac{sup(H,E)}{sup(H,E)+sup(\neg H,E)} = \frac{a}{a+c} = 1 \Leftrightarrow c = 0$$

Moreover, $\neg E \models \neg H$

$$\Leftrightarrow \Pr(H \mid \neg E) - \frac{sup(H,\neg E)}{sup(H,\neg E)+sup(\neg H,\neg E)} = \frac{b}{b+d} = 0 \Leftrightarrow b = 0$$

By the following example let us explain the advantage of considering weak Ex_1 property instead of property Ex_1 in case of confirmation.

Let us consider the following two cases:

- Case 1: a=100, b=99, c=0, d=1;
- Case 2: a=99, b=0, c=1, d=100.

In case 1 the value of a confirmation measure should be greater than in case 2 if Ex_1 holds. However, if we use the idea that a confirmation measure $c(H, E)$ should express how much it is more probable to have H when E is present rather than when E is absent, one can see that $\Pr(H|E) = 1$ and $\Pr(H|\neg E) = 0.99$ in case 1, while in case 2 $\Pr(H|E) = 0.99$ and $\Pr(H|\neg E) = 0$. In other words, if Ex_1 holds, passing from $\neg E$ to E, we assign a greater value of a confirmation measure when we have a 1% increment of the probability of H (case 1) rather than when the same increment is of 99% (case 2). If we consider a confirmation measure that satisfies weak Ex_1, we do not demand that $c(H, E)$ should have a greater value in case 1 rather than in case 2, nor vice versa. Thus, the paradox disappears under conditions of weak Ex_1 property.

Analogously, property weak Ex_1 guarantees that the confirmation measure $c(H, E)$ cannot attain its minimal value unless the two following conditions are satisfied:

3) $E \not\models H$, or equivalently, $\Pr(H|E)=0$, or equivalently, $a=sup(H, E)=0$.

4) $\neg E \models H$, or equivalently, $\Pr(H|\neg E)=1$, or equivalently, $d=sup(\neg H, \neg E)=0$.

Let us supplement that for a given dataset, $E \not\models H$

$$\Leftrightarrow \Pr(H \mid E) = \frac{sup(H,E)}{sup(H,E)+sup(\neg H,E)} = \frac{a}{a+c} = 0 \Leftrightarrow a = 0$$

Moreover, $\neg E \models H$

$$\Leftrightarrow \Pr(H \mid \neg E) = \frac{sup(H,\neg E)}{sup(H,\neg E)+sup(\neg H,\neg E)} = \frac{b}{b+d} = 1 \Leftrightarrow d = 0$$

The following example explains the advantage of considering weak Ex_1 property instead of Ex_1 property in case of disconfirmation.

Let us consider the following two cases

- Case 3: a=0, b=1, c=100, d=99;
- Case 4: a=1, b=100, c=99, d=0.

In case 3 the disconfirmation should be greater than in case 4 if Ex_1 holds i.e. the value of a confirmation measure should be smaller in case 3 than in case 4. Moreover, one can see that $\Pr(H|E) = 0$ and $\Pr(H|\neg E) = 0.01$ in case 3, while in case 4 $\Pr(H|E) = 0.01$ and $\Pr(H|\neg E) = 1$. According to our interpretation of the (BC') conditions, in case of disconfirmation, a confirmation measure $c(H, E)$ should express how much it is less probable to have H when E is present rather than when E is absent. However, it is clear that if Ex_1 holds, passing from $\neg E$ to E, we should have a smaller value of confirmation measure (greater disconfirmation) when we have a 1% decrement of probability of H (case 3) rather than when the same decrement is of 99% (case 4). If we consider a confirmation measure that satisfies weak Ex_1, we do not demand that $c(H, E)$ should have a smaller value in case 3 rather than in case 4, nor vice versa. Thus the paradox disappears under conditions of weak Ex_1.

Analogously as with property Ex_1, we can generalize property L into property *weak* L. In fact, in case of confirmation conditions 1) and 2) are the only to ensure the maximal degree of confirmation, because the increment of the probability of H when passing from $\neg E$ to E is maximal when $\Pr(H|E)=1$ and $\Pr(H|\neg E)=0$.

In our opinion, only rules with such full confirmation should be assigned the maximal value of confirmation measures, while in some cases, as in the above case 1, it is reasonable not to assign that maximal value. Rules for which $\neg E \models \neg H$ is not verified should not obtain maximal values of confirmation measures despite the fact that $E \models H$ occurs, and vice versa. Thus, we claim that the requirement from property L of complete absence of counterexamples (c=0) is not a sufficient condition to assign the maximal value of a confirmation measure. It needs to be supplemented by the condition: b=0.

Analogously, in case of disconfirmation, conditions 3) and 4) are the only to ensure the maximal degree of disconfirmation, because the decrement of the probability of H when passing from $\neg E$ to E is maximal when $\Pr(H|E)$=0 and $\Pr(H|\neg E)$=1.

In our opinion, only rules with such full disconfirmation should be assigned the minimal value of confirmation measures, while in some cases, as in the above case 3, it is reasonable not to assign that minimal value. Rules for which $\neg E \models H$ is not verified should not obtain minimal values of confirmation measures despite the fact that $E \models \neg H$ occurs, and vice versa. Thus, we claim that the requirement from property L of complete absence of positive examples (a=0) is not a sufficient condition to assign the minimal value of a confirmation measure. It needs to be supplemented by the condition: d=0.

On the basis of the above considerations we can generalize property L into property weak L as follows:

- $c(H, E)$ is maximal when $E \models H$ and $\neg E \models \neg H$,
- and $c(H, E)$ is minimal when $E \models \neg H$ and $\neg E \models H$. (weak L)

3.3. Properties of symmetry

Bayesian confirmation measures are also often considered with respect to their symmetry properties. Eells and Fitelson analysed in [8] a set of best-known confirmation measures from the viewpoint of four properties of symmetry introduced by Carnap in [5]. Later, Crupi et al. [7] presented an extended and systematic treatment of symmetry properties gathered up as principle Ex_2. By a symmetry they mean a function $\sigma(H, E)$ which is obtained from (H, E) by applying the negation operator ($\neg$) to either H or E (or both), and/or by inverting them. On the whole they propose to analyze a confirmation measure $c(H, E)$ with respect to seven such symmetry functions (Table 2).

Table 2. Basic symmetry functions

$E(H, E)$:	$c(H, E) = -c(H, \neg E)$
$H(H, E)$:	$c(H, E) = -c(\neg H, E)$
$EI(H, E)$:	$c(H, E) = -c(\neg E, H)$
$HI(H, E)$:	$c(H, E) = -c(E, \neg H)$
$I(H, E)$:	$c(H, E) = c(E, H)$
$EH(H, E)$:	$c(H, E) = c(\neg H, \neg E)$
$EHI(H, E)$:	$c(H, E) = c(\neg E, \neg H)$

However, the analysis is said to be done separately for the case of confirmation and for the case of disconfirmation. Using examples of drawing cards from a standard deck, Crupi et al. point out which of the symmetries are desired and which are definitely unwanted. For instance, the symmetry I is undesired in case of confirmation as for a rule: *if Jack was drawn, then the card is a face*, *Jack* does not confirm *face* with the same strength as the *face* confirms *Jack*, i.e. $c(H, E) \neq c(E, H)$. On the other hand, symmetry I is desirable in case of disconfirmation, as for an exemplary rule: *if the drawn card is an Ace, then it is a face*, the strength with which an *Ace* disconfirms *face* is the same as the strength with which the *face* disconfirms an *Ace*, i.e. $c(H, E) = c(E, H)$.

Summing up, Ex_2 can be stated as follows:

Given a confirmation measure $c(H, E)$, in case of confirmation the following conditions have to be satisfied:

(E^+) $c(H, E) \neq -c(H, \neg E)$,
(H^+) $c(H, E) = -c(\neg H, E)$,
($E\Gamma^+$) $c(H, E) \neq -c(\neg E, H)$,
($H\Gamma^+$) $c(H, E) = -c(E, \neg H)$,
(Γ^+) $c(H, E) \neq c(H, E)$,
(EH^+) $c(H, E) \neq c(\neg H, \neg E)$,
($EH\Gamma^+$) $c(H, E) = c(\neg E, \neg H)$,

while, in case of disconfirmation:

(E^-) $c(H, E) \neq -c(H, \neg E)$,
(H^-) $c(H, E) = -c(\neg H, E)$,
($E\Gamma^-$) $c(H, E) = -c(\neg E, H)$,
($H\Gamma^-$) $c(H, E) \neq -c(E, \neg H)$,
(Γ^-) $c(H, E) = c(E, H)$,
(EH^-) $c(H, E) \neq c(\neg H, \neg E)$,
($EH\Gamma^-$) $c(H, E) \neq c(\neg E, \neg H)$.

Considering the formulation of a confirmation measure in terms of quantities a, b, c, d, and denoting by $f^+(a,b,c,d)$ the formulation of the confirmation measure in case of confirmation, and by $f^-(a,b,c,d)$ the formulation of the confirmation measure in case of disconfirmation, we can formulate the above conditions as in Table 3.

Table 3. Ex_2 symmetry properties

(E^+) $f^+(a,b,c,d) \neq -f^-(b,a,d,c)$	(E^-) $f^-(a,b,c,d) \neq -f^+(b,a,d,c)$
(H^+) $f^+(a,b,c,d) = -f^-(c,d,a,b)$	(H^-) $f^-(a,b,c,d) = -f^+(c,d,a,b)$
($E\Gamma^+$) $f^+(a,b,c,d) \neq -f^-(b,d,a,c)$	($E\Gamma^-$) $f^-(a,b,c,d) = -f^+(b,d,a,c)$
($H\Gamma^+$) $f^+(a,b,c,d) = -f^-(c,a,d,b)$	($H\Gamma^-$) $f^-(a,b,c,d) \neq -f^+(c,a,d,b)$
(Γ^+) $f^+(a,b,c,d) \neq f^+(a,c,b,d)$	(Γ^-) $f^-(a,b,c,d) = f^-(a,c,b,d)$
(EH^+) $f^+(a,b,c,d) \neq f^+(d,c,b,a)$	(EH^-) $f^-(a,b,c,d) \neq f^-(d,c,b,a)$
($EH\Gamma^+$) $f^+(a,b,c,d) = f^+(d,b,c,a)$	($EH\Gamma^-$) $f^-(a,b,c,d) \neq f^-(d,b,c,a)$

4. Different approaches to understanding the concept of extreme values of confirmation measures

Having observed that confirmation measures D, S, M, N, C, R, G (defined earlier on) are contrary to Ex_1, Crupi et al. [7] proposed to normalize them by dividing each of them by the maximum (minimum, respectively) that the measure obtains when $E \models H$, i.e. when the rule's premise entails its conclusion ($E \models \neg H$, respectively).

The meaning of *what a maximum or minimum that a confirmation measure obtains in case of confirmation or disconfirmation is*, constitutes however a very broad subject and the approach applied by Crupi et al. is only one of many ways to answer that question. Exploiting different understandings of the concept of extreme confirmation or disconfirmation, we propose and analyze four, alternative to Crupi et al., schemas allowing to determine the maximum (or minimum) of any confirmation measure in situation when the premise entails (or refutes) the conclusion. Each of those schemas eventually leads to a different normalization. The normalization proposed by Crupi et al. is also explained by us in terms of a, b, c and d in this section.

4.1. Approach inspired by Nicod

The Nicod's criterion presented in [29] says that an evidence confirms a rule $E \rightarrow H$ if and only if it satisfies both the premise and the conclusion of the rule, and disconfirms it if and only if it satisfies the premise but not the conclusion of the rule. Thus, objects for which the premise and the conclusion is supported are considered as positive examples for the rule and objects satisfying the premise but not the conclusion are counter-examples. Moreover, according to Nicod's criterion an evidence that does not satisfy the premise is neutral with respect to the rule. It means that objects for which the premise is not satisfied are irrelevant to the rule, no matter whether they support the conclusion or not.

Let us look at Hempel's [23] rule: *if x is a raven then x is black* from the point of view of Nicod's criterion. In this situation, a *black raven* supports the rule and is a positive example of that rule, *non-black ravens* are against it and are considered as counter-examples, and everything which is *not a raven* can be ignored.

Now, let us propose a schema, based on Nicod's criterion, for determination of maximum (or minimum) of a confirmation measure. Following Nicod's directives, the only objects that are relevant to a rule are positive and counter-examples. It brings us to an observation that a measure will obtain its maximum when all counter-examples change into positive examples. That would take place when all *non-black ravens* change into *black ravens*. It means that the number of positive examples should take over all counter-examples (i.e. $a' \rightarrow a+c$), and the number of counter-examples should drop to 0 (i.e. $c' \rightarrow 0$). The number of *black non-ravens* and *non-black non-ravens* should remain unchanged (i.e. $b' \rightarrow b$ and $d' \rightarrow d$).

Moving on to determination of the minimal value of a measure, let us observe that it will be obtained when all positive examples change into counter-examples (i.e. $c' \rightarrow a+c$ and $a' \rightarrow 0$). Again the number of *black non-ravens* and *non-black non-ravens* remains unchanged (i.e. $b' \rightarrow b$ and $d' \rightarrow d$) as these are objects irrelevant to the rule.

Putting the above considerations together we obtain the following schema inspired by Nicod's approach (Table 4).

Table 4. Normalization schema inspired by Nicod

Maximum	Minimum
$a' \rightarrow a+c$	$a' \rightarrow 0$
$b' \rightarrow b$	$b' \rightarrow b$
$c' \rightarrow 0$	$c' \rightarrow a+c$
$d' \rightarrow d$	$d' \rightarrow d$

For measure D, its maximal value according to Nicod's criterion is: $1-[(a+b+c)/|U|]=d/|U|$ and the minimal: $-b/|U|$.

4.2. Bayesian approach

The ongoing argument between Bayesians and Likelihoodist about the proper probabilistic explication of confirmation [14] inspired us to distinguish the two following approaches to determination of extremes of a measures: Bayesian and likelihoodist. Bayesian approach is related to the idea that the evidence confirms the hypothesis, if the hypothesis is more frequent with the evidence rather than with ¬evidence. Analogously, the evidence disconfirms the hypothesis, if ¬hypothesis is more frequent with the evidence rather than with ¬evidence. Thus, determination of measure's extremes based on this approach should consider a rule from the perspective of its conclusion.

Following Bayesian approach, let us observe that a measure will obtain its maximum if all *black non-ravens* change into *black ravens* (i.e. $a'\to a+b$ and $b'\to 0$), and all *non-black ravens* change into *non-black non-ravens* (i.e. $d'\to c+d$ and $c'\to 0$). It is due to the fact that when there are no *black non-ravens* (i.e. $b'=0$), the hypothesis of being *black* is more frequent with the premise of being a *raven* rather than with ¬premise of being a *non raven*, which means that the premise confirms the rule's conclusion. Moreover, when there are no *non-black ravens* (i.e. $c'=0$), the ¬hypothesis of being *non-black* is disconfirmed as it is more frequent with the ¬premise of being a *non-raven* rather than with the premise of being a *raven*. Disconfirmation of ¬hypothesis is desirable as it results in confirmation of the hypothesis.

To get the minimum of a measure we need to reverse the above situation: all *black ravens* should change into *black non-ravens* (i.e. $b'\to a+b$ and $a'\to 0$), and all *non-black non-ravens* change into *non-black ravens* (i.e. $c'\to c+d$ and $d'\to 0$). Here, in case of the minimal value of a measure we want the situation to be as disconfirmning as possible. It will occur when there are no *black ravens* (i.e. $a'=0$), because then the hypothesis of being *black* is more frequent with the ¬premise of being a *non raven* rather than with premise of being a *raven*. Moreover, when there are no *non-black non-ravens* (i.e. $d'=0$), the ¬hypothesis of being *non-black* is confirmed as it is more frequent with the premise of being a *raven* rather than with the ¬premise of being a *non-raven*. From the fact that the ¬hypothesis is confirmed we can conclude that the hypothesis is disconfirmed.

Table 5 sums up Bayesian approach to determination of a measure's extremes.

Table 5. Bayesian normalization schema

Maximum	Minimum
$a'\to a+b$	$a'\to 0$
$b'\to 0$	$b'\to a+b$
$c'\to 0$	$c'\to c+d$
$d'\to c+d$	$d'\to 0$

The maximal value of measure D obtained using Bayesian approach is: $(c+d)/|U|$ and the minimal: $-(a+b)/|U|$.

4.3. Likelihoodist approach

The likelihoodist approach is based on the idea that the evidence confirms the hypothesis, if the evidence is more frequent with the hypothesis rather than with ¬hypothesis, and in this context, analogously, the evidence disconfirms the hypothesis, if the evidence is more frequent with ¬hypothesis rather than with the hypothesis. Thus, one can informally say that likelihoodists look at the rule from the perspective of its premise.

According to likelihoodist approach, a measure will obtain its maximum if all *non-black ravens* change into *black ravens* (i.e. $a' \rightarrow a+c$ and $c' \rightarrow 0$), and all *black non-ravens* change into *non-black non-ravens* (i.e. $d' \rightarrow b+d$ and $b' \rightarrow 0$). It results from the fact that when there are no *non-black ravens* (i.e. $c'=0$), the evidence of being *a raven* is more frequent with the hypothesis of being *black* rather than with ¬ hypothesis of being *non black*, which means that the premise confirms the rule's conclusion. Moreover, when there are no *black non-ravens* (i.e. $b'=0$), the ¬evidence of being a *non-raven* is more frequent with the ¬hypothesis of being *non-black* rather than with the hypothesis of being *black*. Thus we can conclude that hypothesis is disconfirmed by the ¬premise and as a result of that the hypothesis is confirmed by the premise.

To determine the minimum of a measure all *black ravens* should change into *non-black ravens* (i.e. $c' \rightarrow a+c$ and $a' \rightarrow 0$), and all *non-black non-ravens* change into *black non-ravens* (i.e. $b' \rightarrow b+d$ and $d' \rightarrow 0$). It means that a measure will obtain its most disconfirming values when there are no *black ravens* (i.e. $a'=0$), because then the evidence of being a *raven* is more frequent with the ¬hypothesis of being *non black* rather than with the hypothesis of being *black*. Moreover, when there are no *non-black non-ravens* (i.e. $d'=0$), the ¬evidence of being a *non-raven* is more frequent with the hypothesis of being *black* rather than with the ¬hypothesis of being *non-black*. Since the ¬hypothesis is confirmed by the premise we can conclude that the hypothesis is disconfirmed.

The likelihoodist approach to determination of a measure's maximum or minimum is presented in Table 6.

Table 6. Likelihoodist normalization schema

Maximum	Minimum
$a' \rightarrow a+c$	$a' \rightarrow 0$
$b' \rightarrow 0$	$b' \rightarrow b+d$
$c' \rightarrow 0$	$c' \rightarrow a+c$
$d' \rightarrow b+d$	$d' \rightarrow 0$

The maximal value of D obtained using likelihoodist approach is: $(b+d)/|U|$ and the minimal: $-(b+d)/|U|$.

4.4. Crupi et al. approach

Having proved that none of the measures: D, S, M, N, C, R or G satisfies the desirable property Ex_1, Crupi et al. [7] showed an easy way to transform them into measures that do fulfil Ex_1. They have presented formulas to which the considered measures reduce when $E \models H$ and when $E \models \neg H$ and proposed to normalize the measures by dividing them by the obtained formulas. The result of the division by the formula obtained when $E \models H$ is the normalized measure in case of confirmation (i.e. when $\Pr(H|E) \geq \Pr(H)$), and the division by the absolute value of the formula obtained when $E \models \neg H$ gives the normalized measure in case of disconfirmation (i.e. when $\Pr(H|E) < \Pr(H)$).

Interestingly, it turned out that the considered measures all gave the same result after that transformation, i.e.

$D_{\text{norm}} = S_{\text{norm}} = M_{\text{norm}} = N_{\text{norm}} = C_{\text{norm}} = R_{\text{norm}} = G_{\text{norm}}$.

Crupi et al. labelled the newly obtained measure of confirmation Z:

$$Z(H,E) = \begin{cases} 1 - \dfrac{Pr(\neg H \mid E)}{Pr(\neg H)} = \dfrac{ad - bc}{(a+c)(c+d)} & \textit{in case of confirmation,} \\[2ex] \dfrac{Pr(H \mid E)}{Pr(H)} - 1 = \dfrac{ad - bc}{(a+c)(a+b)} & \textit{in case of disconfirmation.} \end{cases}$$

Let us observe that in case of confirmation $Z=G$ (measure G was discussed by Rips [34]) and in case of disconfirmation $Z=R$ (measure R was considered by Finch [10]).

Crupi et al. [7] have showed that measure Z satisfies the symmetries gathered up as property Ex_2. It also possesses the valuable property M [20] making it a meaningful tool for assessing the interestingness of rules.

Let us now come back to the way Crupi et al. obtained formulas to which the considered measures D, S, M, N, C, R, G reduce when $E \models H$ and when $E \models \neg H$, as it is the essence of their normalization. Their article [7] provides the calculated formulas expressing the maximal (minimal) confirmation (disconfirmation) for each of the analyzed measures. Since, the approach of Crupi et al. brings such interesting results, we propose to explain it in terms of a, b, c and d. The requirements that must be used to obtain the meaning of the concept of maximal confirmation, exploited by Crupi et al., are the following:

- there must be no counterexample and therefore $c'=0$,
- the number of objects in the universe must be maintained and therefore $a'+b'+c'+d'=a+b+c+d$,
- the number of objects for which the premise is satisfied should remain unchanged and therefore $a'+c'=a+c$,
- the number of objects for which the conclusion is true should remain the same and therefore $a'+b'=a+b$.

The considerations about the minimal confirmation are analogous. In Table 7, we summarize the schema to determine the extremes of measures based on the concept of Crupi et al.

Table 7. Normalization schema of Crupi et al.

Maximum	Minimum
$a' \to a+c$	$a' \to 0$
$b' \to b-c$	$b' \to a+b$
$c' \to 0$	$c' \to a+c$
$d' \to d+c$	$d' \to d-a$

The maximal value of D obtained using the approach of Crupi et al. is:$1-[(a+b)/|U|] = (c+d)/|U|$ and the minimal: $-(a+b)/|U|$.

4.5. The likelihoodist counterpart of the approach of Crupi et al.

From our analysis with respect to symmetry properties in Ex_2, there also emerged an observation that there exists a confirmation measure other than Z that satisfies Ex_2 in exactly the same manner that measure Z. In terms of a, b, c and d, it is defined as:

$$A(H,E)=\begin{cases}\dfrac{Pr(E\mid H)-Pr(E)}{1\text{ - }Pr(E)}=\dfrac{ad-bc}{(a+b)(b+d)}=M_{L^+} & \textit{in case of confirmation}\,,\\[2ex] \dfrac{Pr(H)-Pr(H\mid\neg E)}{1\text{ - }Pr(H)}=\dfrac{ad-bc}{(b+d)(c+d)}=G_{L^-} & \textit{in case of disconfirmation}\,.\end{cases}$$

Observe that $A(H, E)$ is a confirmation measure because it is positive when $\Pr(H|E)>\Pr(H)$ (i.e. when $ad>bc$), it is equal to 0 when $\Pr(H|E)=\Pr(H)$ (i.e. when $ad=bc$), and it is negative when $\Pr(H|E)<\Pr(H)$ (i.e. when $ad>bc$). Notice also that $A(H, E)=1$ if $H|=E$ (i.e. $b=0$) and $A(H, E)=-1$ if $\neg H|=E$ (i.e. $d=0$). This means that A satisfies the property weak L, but it does not satisfy the original property Ex_1 and L.

Furthermore, these observations suggest other adequacy requirements that we can consider as

- likelihoodist counterpart of Ex_1, denoted as L-Ex_1: *if* $v(E_1, H_1) > v(E_2, H_2)$, *then* $c(H_1, E_1) > c(H_2, E_2)$,
- and as likelihoodist counterparts of L, denoted as L-L: $c(H, E)$ is maximal when $H \models E$ and $c(H, E)$ is minimal when $H \models \neg E$.

Moreover, let us observe that measure A can be obtained from measures D, S, M, N, C, R, G if we consider a likelihoodist counterpart of the normalization proposed by Crupi et al. (Table 8).

Table 8. Likelihoodist counterpart of normalization schema of Crupi et al.

Maximum	Minimum
$a' \to a+b$	$a' \to a-d$
$b' \to 0$	$b' \to b+d$
$c' \to c-b$	$c' \to c+d$
$d' \to d+b$	$d' \to 0$

Let us remind that the Bayesians look at a rule from the viewpoint of its conclusion, where as Likelihoodists consider a rule from the view point of its premise. Thus, a rule $E \to H$ in the Bayesian perspective, corresponds to rule $H \to E$ in the likelihoodist perspective [14].

5. Results of different normalizations of measures

Each of the schemas presented by us to determine the extremes of measures eventually leads to a different normalization. Table 9 presents the results of normalizations using Nicod's, Bayesian, likelihoodist and Crupi's et al. approaches as well as the likelihoodist counterpart of the approach of Crupi et al.

For the sake of the presentation, the definitions of the analyzed measures were simplified by basic mathematical transformations (column 1). The next five columns contain results for different normalization schemas, for each measure there are two rows containing the normalized measure in case of confirmation (the first row) and disconfirmation (the second row). The notation we used assumes that lower indexes signify the applied normalization (N stands for Nicod and L for likelihoodist), and that the case of confirmation is marked by a "+" and the case of disconfirmation by a "– " (e.g D_{N^+} stands for measure D normalized in case of confirmation, using the approach inspired by Nicod).

Table 9. Results of alternative normalizations

Definition of a measure	Nicod's	Bayesian	Likelihoodist	Crupi et al.	Likelihoodist counterpart of Crupi et al.
$D(H,E)=\dfrac{ad-bc}{\lvert U\rvert(a+c)}$	$D_{N^+}=\dfrac{ad-bc}{d(a+c)}$	G	S	G	M_{L^+}
	$D_{N^-}=\dfrac{ad-bc}{b(a+c)}$	R	S	R	G_{L^-}
$S(H,E)=\dfrac{ad-bc}{(a+c)(b+d)}$	D_{N+}	S	S	G	M_{L^+}
	D_{N^-}	S	S	R	G_{L^-}
$M(H,E)=\dfrac{ad-bc}{\lvert U\rvert(a+b)}$	$M_{N^+}=\dfrac{(ad-bc)(a+b+c)}{d(a+b)(a+c)}$	N	$M_{L^+}=\dfrac{ad-bc}{(a+b)(b+d)}$	G	M_{L^+}
	R	N	R	R	G_{L^-}
$N(H,E)=\dfrac{ad-bc}{(a+b)(c+d)}$	$N_{N^+}=\dfrac{(ad-bc)(a+b+c)}{(a+b)(c+d)(a+c)}$	N	N	G	M_{L^+}
	$N_{N^-}=\dfrac{(ad-bc)(a+c+d)}{(a+b)(c+d)(a+c)}$	N	N	R	G_{L^-}
$C(H,E)=\dfrac{ad-bc}{\lvert U\rvert^2}$	D_{N+}	N	S	G	M_{L^+}
	D_{N^-}	N	S	R	G_{L^-}
$R(H,E)=\dfrac{ad-bc}{(a+b)(a+c)}$	M_{N^+}	G	M_{L^+}	G	M_{L^+}
	R	R	R	R	G_{L^-}
$G(H,E)=\dfrac{ad-bc}{(a+c)(c+d)}$	G	G	G	G	M_{L^+}
	$G_{N^-}=\dfrac{(ad-bc)(a+c+d)}{(a+c)(c+d)b}$	R	$G_{L^-}=\dfrac{ad-bc}{(c+d)(b+d)}$	R	G_{L^-}

Table 9 shows that in many situations the normalization transforms one measure from our list into another, e.g. measure D normalized using the likelihoodist approach reduces to measure S. Some measures are also invariant to certain normalizations, like in case of measure N normalized using Bayesian or likelihoodist approach. As it was mentioned before, the normalization proposed by Crupi et al. transforms all of the analyzed measures into the same formula called Z measure, which is equal to measure G in case of confirmation and to measure R in case of disconfirmation. On the other hand, the likelihoodist counterpart of the normalization proposed by Crupi et al. transforms all of the analyzed measures into measure that we called A, which is equal to M_{L+} in case of confirmation and to G_{L-} in case of disconfirmation.

Since the normalization of Crupi et al. was introduced as a tool for transforming the measures so they would satisfy the property Ex_1, we have analysed the results of different normalizations of measures D, S, M, N, C, R, G from the view point of this property. Our analysis was also extended by the weak Ex_1 property, L-Ex_1, logicality property L, weak L, and L-L.

Theorem 1. Property analysis shows that:

- All of the analyzed confirmation measures after undergoing the normalization inspired by Nicod or normalization of Crupi et al. satisfy the logicality L property.
- All of the analyzed confirmation measures after undergoing the normalization of Crupi et al. satisfy the Ex_1 condition. All of the analyzed confirmation measures after undergoing the normalization inspired by the likelihoodist counterpart of Crupi et al. satisfy the L-L and L-Ex_1 properties.
- The measure N undergoing Nicod normalization is the only one satisfying both the logicality L and the L-L properties.
- All five normalization approaches guarantee transformations that satisfy the weak L property.
- The weak Ex_1 property is satisfied by measures S and N (before applying any transformations) and by all the measures that transformed results in S and N (for example the measure D undergoing the likelihoodist normalization).

Proof. The possession of property L can be verified by putting c=0 and a=0 and checking whether $c(H, E)$=1 in case when c=0, and whether $c(H, E)= -1$ in case when a=0. Analogously, possession of Ex_1 can be verified observing that the analyzed confirmation measures after undergoing the normalization of Crupi et al. reach their maximum only if c=0 and reach their minimum only if a=0. The possession of properties L-L and L-Ex_1 can be checked analogously to verification of L and Ex_1 condition taking into account b=0 and d=0 instead of c=0 and a=0. Weak L can be verified by putting b=c=0 and a=d=0 and checking whether $c(H, E)$=1 in case b=c=0 and $c(H, E)= -1$ in case a=d=0. Weak Ex_1 can be verified by checking that S and N reach their maximum only if b=c=0 and reach their minimum only if a=d=0.

Let us remark that properties Ex_1 and L are directional in the sense that:

- L implies that if c=0 then $c(H, E)$ reaches its maximum, however, $c(H, E)$ can also reach its maximum when c>0;
- Ex_1 implies that if $c(H, E)$ reaches its maximum then c=0, however, it is possible that c=0 and $c(H, E)$ does not reach its maximum;
- if both L and Ex_1 are satisfied then $c(H, E)$ reaches its maximum if and only if c=0.

Theorem 2. Measures D_N, S_N, M_N, N_N, C_N, R_N, and G_N (resulting from application of normalization inspired by Nicod) are ordinally non-equivalent to measure Z.

Proof. [22] Measure f is ordinally equivalent to measure g iff for any two rules r_1, r_2,:

$$f(\mathrm{r}_1)\begin{Bmatrix} > \\ = \\ < \end{Bmatrix} f(\mathrm{r}_2) \text{ iff } g(\mathrm{r}_1)\begin{Bmatrix} > \\ = \\ < \end{Bmatrix} g(\mathrm{r}_2)$$

The above condition needs to be fulfilled both in case of confirmation and disconfirmation. For Table 9 it is enough to consider measures D_N^+, M_N^+, N_N^+ and G_N^-.

The situation in which the number of objects in U is distributed over a, b, c and d is called scenario α. In scenario α, rule r: $E\rightarrow H$ is supported by a objects from U.

Let us prove, by counterexample, that in two exemplary scenarios α_1 and α_2 measures D_N^+, and M_N^+ produce rankings different than measure G:

α_1	a=90	b=8	c=1	d=1	U=100	$\boldsymbol{D_{N+}(\mathrm{r}_1)=0.90}$	$\boldsymbol{M_{N+}(\mathrm{r}_1)=0.91}$	$G(\mathrm{r}_1)$=0.45
α_2	a=70	b=16	c=4	d=10	U=100	$D_{N+}(\mathrm{r}_2)$=0.86	$M_{N+}(\mathrm{r}_2)$=0.90	$\boldsymbol{G(\mathrm{r}_2)=0.61}$

Measure G assigns r_2 greater value than to r_1, whereas measures D_N^+, and M_N^+ rank those rules the other way round.

Again, by counterexample, let us show that in two exemplary scenarios α_3 and α_4 measure N_N^+ produces different ranking than measure G:

α_3	a=70	b=1	c=19	d=10	U=100	$N_{N+}(\mathrm{r}_1)$=0.33	$\boldsymbol{G(\mathrm{r}_1)=0.26}$
α_4	a=55	b=2	c=26	d=17	U=100	$\boldsymbol{N_{N+}(\mathrm{r}_2)=0.37}$	$G(\mathrm{r}_2)$=0.25

Here, measure G assigns r_1 greater value, whereas measures N_N^+ favors r_2.

Finally, let us use scenarios α_1 and α_2 to prove that measure G_N^- produces different ranking than measure R:

α_1	a=90	b=8	c=1	d=1	U=100	$\boldsymbol{G_{N-}(\mathrm{r}_1)=5.18}$	$R(\mathrm{r}_1)$=0.009
α_2	a=70	b=16	c=4	d=10	U=100	$G_{N-}(\mathrm{r}_2)$=3.22	$\boldsymbol{R(\mathrm{r}_2)=0.099}$

Here, measure G_{N-} assigns r_1 greater value, whereas measures R favors r_2. □

The measures in Table 9, obtained by different normalization schemas, have also been considered with respect to the properties of symmetries (for detailed results see table A-D in the Appendix published as electronic supplementary material).

Theorem 3. The only normalizations that transform all of the analyzed measures into measures Z and A satisfying the symmetries of Ex_2 are the normalization of Crupi et al. and its likelihoodist counterpart. Moreover, measures *D*, *R* and *G* normalized using the Bayesian approach satisfy all symmetries gathered up in Ex_2.

Proof. The proof is in the Appendix published as electronic supplementary material.

Summing up, five different normalization schemas have been considered for a set of seven popular confirmation measures. Two normalizations turned out to be invariant for the analyzed set of measures, i.e. transformed all of them into one measure, being measure *Z* for the normalization of Crupi et al., and measure *A* for the likelihoodist counterpart of that normalization. Analysis of the results of normalizations with respect to symmetry properties (Ex_2) showed that only in case of normalization of Crupi et al., and its likelihoodist counterpart, the transformation for all of the considered seven measures lead to measures that fulfill Ex_2. We believe that there is no specific advantage of *Z* over *A* (and vice versa), and that measures *A* and *Z* should be considered as complementary confirmation measures. The complementarity of *A* and *Z* is based on at least three arguments:

1. *Complementarity with respect to measuring how much* Pr(*H*|*E*) *is greater or smaller than* Pr(*H*|¬*E*).

In fact we can write measure *Z*(*H*, *E*) in terms of Pr(*H*|*E*) and of Pr(*H*) as:

$$Z(H,E)=\begin{cases}\dfrac{Pr(H\mid E)-Pr(H)}{1-Pr(H)} & \textit{in case of confirmation}\\[2ex] \dfrac{Pr(H\mid E)-Pr(H)}{Pr(H)} & \textit{in case of disconfirmation}\end{cases}$$

while we can write measure *A*(*H*, *E*) in terms of Pr(*H*|¬*E*) and of Pr(*H*) as:

$$A(H,E)=\begin{cases}\dfrac{Pr(H)-Pr(H\mid\neg E)}{Pr(H)} & \textit{in case of confirmation}\\[2ex] \dfrac{Pr(H)-Pr(H\mid\neg E)}{1-Pr(H)} & \textit{in case of disconfirmation}\end{cases}$$

2. *Bayesian and likelihoodist inspirations.*

As shown above, measure *Z*(*H*, *E*) has been obtained using a normalization of Crupi et al. inspired by the Bayesian view corresponding to property Ex_1 for which the maximum of confirmation is reached when Pr(*H*|*E*)=1 and the minimum when Pr(*H*|*E*)=0, while measure *A*(*H*, *E*) has been obtained using a normalization being a likelihoodist counterpart of the approach of Crupi et al. corresponding to property L-Ex_1 for which the maximum of confirmation is reached when Pr(*E*|*H*)=1 and the minimum when Pr(*E*|*H*)=0.

3. *Applications to some exemplary cases.*

If we apply confirmation measures *Z*(*H*, *E*) and *A*(*H*, *E*) to the cases 1 and 3 presented in 3.2.1, we obtain:

Case 1: *a*=100, *b*=99, *c*=0, *d*=1; Pr(*H*|*E*)=1, Pr(*H*|¬*E*)=99/100; *Z*(*H*, *E*)=1, *A*(*H*, *E*)=1/199;

Case 3: *a*=0, *b*=1, *c*=100, *d*=99; Pr(*H*|*E*)=0, Pr(*H*|¬*E*)=1/100; *Z*(*H*, *E*)= −1, *A*(*H*, *E*)= −1/199.

In all above cases $Z(H, E)$ gives a bad representation while $A(H, E)$ gives a good representation of the confirmation and disconfirmation. However, one can build similar cases in which $Z(H, E)$ gives a good representation while $A(H, E)$ gives a bad representation of the confirmation and disconfirmation:
Case 1': a=100, b=0, c=99, d=1; $\Pr(H|E)=1/199$, $\Pr(H|\neg E)=0$; $Z(H, E)=1/199$, $A(H, E)=1$;
Case 3': a=99, b=1, c=100, d=0; $\Pr(H|E)=99/199$, $\Pr(H|\neg E)=1$; $Z(H, E)=-1/199$, $A(H, E)=-1$.

From these arguments there clearly arises a guideline to consider some average between the value of $Z(H, E)$ and $A(H, E)$. Effectively this is an interesting way to deal with the problem of measuring confirmation that we are developing in a companion paper.

6. Conclusions

Evaluation of knowledge represented by patterns induced from a dataset is an important and active research area. The literature is a rich source of considerations on this subject. In this paper, we focused on evaluation of patterns in form of "*if... then...*" rules by means of interestingness measures. In order to assess the usefulness and meaningfulness of such measures, we have investigated their properties. We have described and analysed property of Bayesian confirmation, property Ex_1 and L property, our proposal of modifications of Ex_1 and L called L-Ex_1, L-L, weak Ex_1 and weak L, and a group of symmetry properties Ex_2.
Next, normalization of confirmation measures as a way to transform the measures so that they would obtain property Ex_1, L, L-L, L-Ex_1,weak L or weak Ex_1 has been considered. A crucial step of such normalization is determination of the extremes of the measures in case of confirmation and disconfirmation. In this article, we have introduced four alternative approaches to this problem. Each of those approaches leads to different results and normalizations, as they consider the concept of maximal and minimal confirmation from different perspectives.
A set of seven confirmation measures, earlier analyzed by Crupi et al. [6], has been normalized using four new schemas and the schema resulting from the approach of Crupi et al. We have analyzed the results of the normalizations with respect to properties Ex_1, L, L-L, L- Ex_1,weak L and weak Ex_1. The conclusions that we have obtained show that Nicod and Crupi et al. approaches give ordinally non-equivalent normalized measure satisfying property L, while all of the considered approaches give normalized measure satisfying weak L. Moreover, no non-normalized confirmation measure satisfies property Ex_1, while only measures S and N (and all the measures that after normalization result in S or in N) fulfil the weak Ex_1 property.
Considering the results of normalizations with respect to properties of symmetries (Ex_2) we can conclude that only measures D, R and G normalized using Bayesian approach satisfied Ex_2 just like measure Z proposed by Crupi et al. and measure A proposed in this paper as a result of normalization based on a likelihoodist counterpart of the approach of Crupi et al.

7. Further investigations

Let us observe that another confirmation measure satisfying property weak Ex_1 is measure $c_1(H, E)$ presented in section 3. In fact, $c_1(H, E)$ is a specific combination of measures $Z(H, E)$ and $A(H, E)$, because it can be written as follows:

$$c_1(H,E)=\begin{cases}\alpha+\beta A(H,E) \textit{ in case of confirmation if } c=0,\\ \alpha Z(H,E) \textit{ in case of confirmation if } c>0,\\ \alpha Z(H,E) \textit{ in case of disconfirmation if } a>0,\\ -\alpha+\beta A(H,E) \textit{ in case of disconfirmation if } a=0.\end{cases}$$

Another confirmation measure that satisfies property weak Ex_1 is the following measure $c_2(H, E)$ that can be built using measures $Z(H, E)$ and $A(H, E)$ in an inverse way:

$$c_2(H,E)=\begin{cases}\alpha+\beta Z(H,E) \textit{ in case of confirmation if } b=0,\\ \alpha A(H,E) \textit{ in case of confirmation if } b>0,\\ \alpha A(H,E) \textit{ in case of disconfirmation if } d>0,\\ -\alpha+\beta Z(H,E) \textit{ in case of disconfirmation if } d=0.\end{cases}$$

It is interesting to note that $c_1(H, E)$ satisfies also property Ex_1, and $c_2(H, E)$ satisfies also property L-Ex_1.
One can imagine other confirmation measures that can be obtained from $A(H, E)$ and $Z(H, E)$, which satisfy the property weak Ex_1, for example:

$$c_3(H,E)=\begin{cases}A(H,E)Z(H,E) & \textit{in case of confirmation}\\ -A(H,E)Z(H,E) & \textit{in case of disconfirmation}\end{cases}$$

and

$$c_4(H,E)\begin{cases}min(A(H,E),Z(H,E)) & \textit{in case of confirmation}\\ max(A(H,E),Z(H,E)) & \textit{in case of disconfirmation.}\end{cases}$$

Observe that also measures $c_1(H,E)$ and $c_2(H,E)$ satisfy the symmetries (Ex_2) while this is not the case of measures $c_3(H,E)$ and $c_4(H,E)$ (that do not satisfy E^+,EI^+,I^+, EH^+, E^-, HI^-, EH^- EHI^-).
With respect to the special cases of section 5, we observe that all measures obtained by combining $A(H,E)$ and $Z(H,E)$ work well: in fact

a) using $c_1(H,E)$, in case 1 we would have $c_1(H,E)=\alpha+\beta A(H,E)$, in case 3 we would have $c_1(H,E)=-\alpha+\beta A(H,E)$, in case 1' and 3' we would have $c_1(H,E)=\alpha Z(H,E)$; in all above cases, for average values of α and β (let us say both of them around 0.5), the result can be considered reasonable;
b) using $c_2(H,E)$, in case 1 and 3 we would have $c_2(H,E)=\alpha A(H,E)$, in case 1' we would have $c_2(H,E)=\alpha+\beta Z(H,E)$, in case 3' we would have $c_2(H,E)=-\alpha+\beta Z(H,E)$; in all above cases, for average values of α and β (let us say both of them around 0.5), the result can be considered reasonable;
c) using $c_3(H,E)$ and $c_4(H,E)$, in case 1 and 3 we would have $A(H,E)$, and in case 1' and 3' we would have $Z(H,E)$, which is exactly what one would expect in such situations.

The above conducted property analysis brings us to the conclusion that measures *A* and *Z* should be regarded as complementary and that the future work ought to concentrate on measures combining *A* and *Z* in a proper way.

Acknowledgement.

The research of the second and third author has been supported by the Polish Ministry of Science and Higher Education.

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] Błaszczyński, J., Słowiński, R., Szeląg, M., 2011. Sequential covering rule induction algorithm for variable consistency rough set approaches, *Information Sciences*, vol 181, Issue 5, 987-1002.

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

[4] Brzezińska, I., Greco, S., Słowiński, R., 2007. Mining Pareto–optimal rules with respect to support and anti–support. *Engineering Applications of Artificial Intelligence*, vol. 20, no. 5, 587-600.

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

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

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

[8] Eells, E., Fitelson, B., 2002. Symmetries and asymmetries in evidential support. *Philosophical Studies*, vol. 107, no. 2, 129-142.

[9] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery: an overview. In: Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., Uthursamy, R. (eds.) *Advances in Knowledge Discovery and Data Mining*, AAAI Press, pp. 1-34.

[10] Finch, H.A., 1999. Confirming Power of Observations Metricized for Decisions among Hypotheses. *Philosophy of Science*, vol. 27, 293-307 and 391-404.

[11] Fitelson, B., 1999. The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity. *Philosophy of Science*, vol. 66, 362-378.

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

[13] Fitelson, B., 2006. Logical Foundations of Evidential Support. *Philosophy of Science*, vol. 73, 500-512.

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

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

[16] Ghosh, A., Nath, B.,2004. Multi-objective rule mining using genetic algorithms, *Information Sciences*, vol. 163, issues 1-3, 123-133.

[17] Greco, S., Matarazzo, B., Słowiński, R., 2005: Rough Membership and Bayesian Confirmation Measures for Parameterized Rough Sets. In: Ślęzak, D., Wang, G., Szczuka, M. Düntsch, I., Yao, Y. (eds.) *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, LNAI, vol. 3641, pp. 314-324. Springer, Heidelberg.

[18] Greco, S., Matarazzo, B., Słowiński, R., 2008. Parameterized rough set model using rough membership and Bayesian confirmation measures. *Int. J. Approx. Reasoning*, vol. 49, no. 2, 285-300.

[19] Greco, S., Pawlak, Z., Słowiński, R., 2004. Can Bayesian confirmation measures be useful for rough set decision rules? *Engineering Applications of Artificial Intelligence*, vol. 17, no.4, 345-361.

[20] Greco, S., Słowiński, R., Szczęch, I., 2008. Assessing the quality of rules with a new monotonic interestingness measure Z. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.), *Artificial Intelligence and Soft Computing (ICAISC 2008)*, LNAI, vol. 5097, pp. 556-565. Springer, Heidelberg.

[21] Greco, S., Słowiński, R., Szczęch, I., 2009. Analysis of monotonicity properties of some rule interestingness measures. *Control and Cybernetics*, vol.38,(1), 9-25.

[22] Greco, S., Słowiński, R., Szczęch, I., 2010. Alternative normalization schemas for Bayesian confirmation measures. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.), *Computational Intelligence for Knowledge-Based Systems Design (IPMU 2010)*, LNAI, vol. 6178, pp. 230-239. Springer-Verlag, Heidelberg.

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

[24] Kolmogorov, A., 1956. *Foundations of Probability* (2nd English ed.). Providence, Rhode Island, AMS Chelsea publishing.

[25] 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.

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

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

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

[29] Nicod, J., 1923. *Le problème de la logique de l'induction*. Alcan, Paris.

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

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

[32] Pedrycz, W., 2012. Allocation of Information Granularity in Optimization and Decision-Making Models: Towards Building the Foundations of Granular Computing. *European Journal of Operational Research*, to appear.

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

[34] Rips, L.J., 2001. Two Kinds of Reasoning. *Psychological Science*, vol. 12, 129-134.

[35] Silberschatz, A., Tuzhilin, A., 1996. What Makes Patterns Interesting in Knowledge Discovery Systems. *IEEE Transactions on Knowledge and Data Engineering*, vol. 8, no. 6, 970-974.

[36] 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.

[37] Ślęzak, D., 2005. Rough Sets and Bayes Factor. *Transactions on Rough Sets III*, LNCS series, vol. 3400, Springer-Verlag Berlin Heidelberg, pp.202-229.

[38] Ślęzak, D., Ziarko, W., 2005. The investigation of the Bayesian rough set model. *Int. J. Approx. Reasoning*, vol. 40, 81-91.

[39] Tan, P.-N., Kumar, V., Srivastava, J., 2002. Selecting the right interestingness measure for association patterns. In: *Proc. of the 8th Int. Conf. on Knowledge Discovery and Data Mining (KDD 2002)*. Edmonton, Canada, pp.32-41.

[40] Tan, P.-N., Kumar, V., Srivastava, J., 2004. Selecting the right objective measure for association analysis. *Information Systems* , vol. 29, no. 4, 293-313.