BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 63, No. 1, 2015
DOI: 10.1515/bpasts-2015-0036

Visualization support for the analysis of properties of interestingness measures

R. SUSMAGA* and I. SZCZĘCH

Institute of Computing Science, Poznań Univesity of Technology, 2 Piotrowo St., 60-965 Poznań, Poland

Abstract. The paper considers a particular group of rule interestingness measures, called Bayesian confirmation measures, which have become the subject of numerous, but often exclusively theoretical studies. To assist and enhance their analysis in real-life situations, where time constraints may impede conducting such time consuming procedures, a visual technique has been introduced and described in this paper. It starts with an exhaustive and non-redundant set of contingency tables, which consists of all possible tables having the same number of observations. These data, originally 4-dimensional, may, owing to an inherent constraint, be effectively represented as a 3-dimensional tetrahedron, while an additional, scalar function of the data (e.g. a confirmation measure) may be rendered using colour.

Dedicated analyses of particular colour patterns on this tetrahedron allow to promptly perceive particular properties of the visualized measures. To illustrate the introduced technique, a set of 12 popular confirmation measures has been selected and visualized. Additionally, a set of 9 popular properties has been chosen and the visual interpretations of the measures in terms of the properties have been presented.

Key words: visualization, interestingness measures, confirmation measures, properties of confirmation measures.

1. Introduction

Development of various visualization techniques and tools becomes increasingly important and common. In "Information Visualization in Data Mining and Knowledge Discovery" [1], the following purposes of visualization are named: "Human beings look for structure, features, patterns, trends, anomalies, and relationships in data. Visualization supports this by presenting the data in various forms with differing interactions. A visualization can provide a qualitative overview of large and complex data sets, can summarize data, and can assist in identifying regions of interest and appropriate parameters for more focused quantitative analysis. In an ideal system, visualization harnesses the perceptual capabilities of the human visual system".

Following that trend, we propose to exploit visualization of interestingness measures in order to facilitate and support the analysis of their properties. Interestingness measures are commonly used to evaluate patterns (e.g. expressed in form of decision rules) induced from data sets. The plurality of measures proposed and discussed in the literature created the need for and the rapid growth of analysis of properties possessed by the interestingness measures. The determination of their properties groups the measures according to their behaviour in particular situations (a feature especially useful in cases of idiosyncratic data sets, e.g. sets with heavily unbalanced classes, as in [2]) and makes it easier to choose the right measure.

The visualization that we propose has been conceived to aid the process of analysis of measures with respect to their properties. Our technique, as exemplified through Matlab®-based implementation, allows to conduct a preliminary determination of properties possessed by each visualized measure. It also facilitates finding counter-examples that rule out particular properties. Our visualization significantly eases the property analysis for the measures already known in the literature, but also for newly developed measures (e.g. automatically generated).

In the paper, the visualization results are presented for a particular group of interestingness measures called *confirmation measures*, designed for evaluation of decision rules, in the form of "*if* premise, *then* conclusion". The confirmation measures are characterised by the fact that they obtain:

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

Our analysis using measure visualization focuses on the following properties:

- property M, of monotonic dependency of the measure on the number of objects satisfying (supporting) or not the premise or the conclusion of the rule [3–5],
- property Ex_1, and its generalization to weak Ex_1, assuring that any conclusively confirmatory rule is assigned a higher value of a 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 [6, 7],
- property L (*Logicality*), and its generalization to weak L, indicating the conditions under which measures should obtain their maximal or minimal values [6–8],

*e-mail: susmaga@cs.put.poznan.pl

Unauthenticated
Download Date | 4/9/15 9:29 AM

- properties of symmetry, being a whole set of properties, characterising how the value of a confirmation measure relates to its value obtained after the rule's premise and conclusion are interchanged and/or negated [3, 6, 8–11].

Let us also note that another property closely related to L, Ex_1 and their generalizations is *maximality/minimality*, proposed in [12].

This paper builds on the technique proposed in [13, 14], in which the idea of visualizing the confirmation measures using barycentric coordinates was introduced, and on the preliminary results, concerning the visual analysis of properties of confirmation measures, presented in [15].

The technique presented in this paper focuses on visualizing the whole domains of different measures used in the data analysis process. In particular, it may concern measures applied at early stages, e.g. interestingness measures used to evaluate and filter patterns (e.g. decision rules). This is what differentiates our technique from classic data visualization approaches, commonly applied in KDD and ML, which are basically concerned with representing selected evaluations of applied tools, e.g. the performance of classifiers. In most typical applications these are usually two-dimensional characteristics, e.g. ROC curves.

The rest of the paper is organized as follows. Section 2 defines popular confirmation measures and presents an overview of their common properties. Section 3 demonstrates the used visualization technique. Visualization-based analysis of confirmation measures with respect to their properties is conducted in Sec. 4. Final remarks and conclusions are contained in Sec. 5.

2. Confirmation measures and their properties

This paper concentrates on a particular group of interestingness measures called confirmation measures. They evaluate rule patterns induced from a set of objects U described by a non-empty finite set of attributes A with respect to their relevance and utility [16]. In particular, confirmation measures quantify the degree to which the evidence in the rule's premise E provides support *for* or *against* the hypothesised piece of evidence in the rule's conclusion H [8].

Formally, for a rule $E \rightarrow H$, an interestingness measure $c(H, E)$ has the property of Bayesian confirmation when it satisfies the following conditions:

$$c(H,E)\begin{cases} > 0 & \text{when} \quad P(H|E) > P(H), \\ = 0 & \text{when} \quad P(H|E) = P(H), \\ < 0 & \text{when} \quad P(H|E) < P(H). \end{cases} \tag{1}$$

Thus, the confirmation is interpreted as an increase in the probability of the conclusion H provided by the premise E (similarly for the neutrality and the disconfirmation).

Let us observe that in the context of a particular set of objects U, the relation between E and H may be quantified by four non-negative numbers a, b, c and d, briefly represented in a 2×2 table (Table 1). Notice that every confirmation measure $c(H, E)$ is a scalar function f of a, b, c and d.

Table 1
An exemplary contingency table of the rule's premise and conclusion

	H	$\neg H$	Σ
E	a	c	$a+c$
$\neg E$	b	d	$b+d$
Σ	$a+b$	$c+d$	n

At the same time a, b, c and d can be used to estimate different probabilities: e.g. the probability of the premise is expressed as $P(E) = (a + c)/n$, the conditional probability of the conclusion given the premise is $P(H|E) = P(H \cap E)/P(E) = a/(a + c)$. This allows to reformulate the (1) conditions in the following manner:

$$c(H,E)\begin{cases} > 0 & \text{when} \quad \dfrac{a}{a+c} > \dfrac{a+b}{n}, \\ = 0 & \text{when} \quad \dfrac{a}{a+c} = \dfrac{a+b}{n}, \\ < 0 & \text{when} \quad \dfrac{a}{a+c} < \dfrac{a+b}{n}. \end{cases} \tag{2}$$

Notice that ([17]) given $a + c \neq 0$ and $n \neq 0$, the formulation (2) is logically equivalent to

$$c(H,E)\begin{cases} > 0 & \text{when} \quad ad - bc > 0, \\ = 0 & \text{when} \quad ad - bc = 0, \\ < 0 & \text{when} \quad ad - bc < 0. \end{cases} \tag{3}$$

The main difference between formulations (2) and (3) concerns situations when $a + c = 0$ and $n = 0$ (although the second condition will never occur, as n is always assumed to be positive), because in those cases the measure is undefined according to formulation (2).

Let us additionally stress that the list of alternative, non-equivalent measures of confirmation is quite large [6, 18]. It is due to the fact that the (1) conditions do not impose any constraints on the measures except for requiring when the measures should obtain positive, negative or zero values. The set of 12 selected, popular confirmation measures is presented in Table 2. For further information concerning those measures refer to: $D(H, E)$ – [19], $M(H, E)$ – [20], $S(H, E)$ – [21], $N(H, E)$ – [22], $C(H, E)$ – [23], $F(H, E)$ – [24], $Z(H, E)$ – [6], $A(H, E)$–$c_4(H, E)$ – [7].

The definitions of measures $Z(H, E)$, $A(H, E)$, $c_1(H, E)$, $c_2(H, E)$, $c_3(H, E)$ and $c_4(H, E)$ in Table 2 are formulated for the following (defined) situations: that of confirmation ($ad - bc > 0$) and that of disconfirmation ($ad - bc < 0$). In the third possible (defined) situation, i.e. when $ad - bc = 0$, the measures default to 0. For reasons of brevity the last situation is omitted from the definitions.

Let us point out that some of the confirmation measures (in particular c_1, c_2, c_3, c_4) can be regarded as *derived*, because they were formed so that their definition would benefit from valuable properties of their constituents [7].

Unauthenticated
Download Date | 4/9/15 9:29 AM

Table 2
Popular confirmation measures

$D(H,E) = P(H\|E) - P(H) = \dfrac{a}{a+c} - \dfrac{a+b}{n}$
$M(H,E) = P(E\|H) - P(E) = \dfrac{a}{a+b} - \dfrac{a+c}{n}$
$S(H,E) = P(H\|E) - P(H\|\neg E) = \dfrac{a}{a+c} - \dfrac{b}{b+d}$
$N(H,E) = P(E\|H) - P(E\|\neg H) = \dfrac{a}{a+b} - \dfrac{c}{c+d}$
$C(H,E) = P(E \wedge H) - P(E)P(H) = \dfrac{a}{n} - \dfrac{(a+c)(a+b)}{n^2}$
$F(H,E) = \dfrac{P(E\|H) - P(E\|\neg H)}{P(E\|H) + P(E\|\neg H)} = \dfrac{ad-bc}{ad+bc+2ac}$
$Z(H,E) = \begin{cases} 1 - \dfrac{P(\neg H\|E)}{P(\neg H)} = \dfrac{ad-bc}{(a+c)(c+d)} & \text{in case of confirmation} \\ \dfrac{P(H\|E)}{P(H)} - 1 = \dfrac{ad-bc}{(a+c)(a+b)} & \text{in case of disconfirmation} \end{cases}$
$A(H,E) = \begin{cases} \dfrac{P(E\|H) - P(E)}{1 - P(E)} = \dfrac{ad-bc}{(a+b)(b+d)} & \text{in case of confirmation} \\ \dfrac{P(H) - P(H\|\neg E)}{1 - P(H)} = \dfrac{ad-bc}{(b+d)(c+d)} & \text{in case of disconfirmation} \end{cases}$
$c_1(H,E) = \begin{cases} \alpha + \beta A(H,E) & \text{in case of confirmation when } c = 0 \\ \alpha Z(H,E) & \text{in case of confirmation when } c > 0 \\ \alpha Z(H,E) & \text{in case of disconfirmation when } a > 0 \\ -\alpha + \beta A(H,E) & \text{in case of disconfirmation when } a = 0 \end{cases}$
$c_2(H,E) = \begin{cases} \alpha + \beta Z(H,E) & \text{in case of confirmation when } b = 0 \\ \alpha A(H,E) & \text{in case of confirmation when } b > 0 \\ \alpha A(H,E) & \text{in case of disconfirmation when } d > 0 \\ -\alpha + \beta Z(H,E) & \text{in case of disconfirmation when } d = 0 \end{cases}$
$c_3(H,E) = \begin{cases} A(H,E)Z(H,E) & \text{in case of confirmation} \\ -A(H,E)Z(H,E) & \text{in case of disconfirmation} \end{cases}$
$c_4(H,E) = \begin{cases} \min(A(H,E), Z(H,E)) & \text{in case of confirmation} \\ \max(A(H,E), Z(H,E)) & \text{in case of disconfirmation} \end{cases}$

The 12 selected confirmation measures obtain values ranging from -1 to $+1$, except for measures $D(H,E)$ and $M(H,E)$ whose values approach -1 or $+1$ only for n approaching $+\infty$. Moreover, measure $C(H,E)$ originally obtains values from $-1/4$ to $+1/4$ (regardless of n), so a simple linear transformation (a multiplication by 4) has been introduced and all further results concern the transformed $C(H,E)$ (for a detailed discussion of these aspects of the measures see [25]). Finally, the parametrized measures ($c_1(H,E)$ and $c_2(H,E)$, both depending on parameters α and β) have been computed with the values of $\alpha = \beta = 1/2$.

The ongoing argument between Bayesians and Likelihoodists about the probabilistic clarification of what confirmation should mean resulted in defining measures that come in pairs, some of them bayesian-inspired and the others likelihoodist-inspired (e.g. $D(H,E)$ and $M(H,E)$, $S(H,E)$ and $N(H,E)$, $Z(H,E)$ and $A(H,E)$, $c_1(H,E)$ and $c_2(H,E)$; measure $F(H,E)$ has no popular counterpart, but it can easily be defined). Pairs of such measures have exchangeable, bayesian-likelihoodist, properties, and may be treated as adjoint. Other measures do not form pairs, often being self-adjoint in this respect (e.g. $C(H,E)$, $c_3(H,E)$ and $c_4(H,E)$).

To bring even further characterization of measures and to choose a suitable one for a particular application and for particular user's expectations, many properties have been proposed and compared in the literature [9, 12, 16, 26]. In the context of confirmation measures the following properties are often analysed: property M, Ex_1 and weak Ex_1, L and weak L, and a group of symmetry properties. Each of those properties groups the measures according to similarities in their behaviour.

2.1. Property of monotonicity M. In [4] the application of confirmation measures to the evaluation of usefulness of decision rules has been considered. The discussion led to formulation of a desirable property of monotonicity M for confirmation measures.

Property M requires that a confirmation measure $c(H,E)$ is a function:

- non-decreasing with respect to a and d, and
- non-increasing with respect to b and c.

Unauthenticated
Download Date | 4/9/15 9:29 AM

It means that any evidence in which the premise E and the conclusion H (or, analogously, $\neg E$ and $\neg H$) hold together is required to increase (or at least not decrease) the confirmation of the rule $E \rightarrow H$. Thus, e.g. arrival of new objects supporting the rule cannot lower the value of the measure. On the other hand, with respect to b (or, analogously, with respect to c) property M means that any evidence in which $\neg E$ and H hold (or, analogously, E and $\neg H$ hold) is required to decrease (or at least not increase) the confirmation of the rule $E \rightarrow H$.

The property M has also proved to be very useful regarding multicriteria evaluation of patterns in form of rules [3,5].

2.2. Property Ex_1 and its generalization to weak Ex_1. To handle the plurality of alternative confirmation measures, property Ex_1 considering inductive logic as an extrapolation from classical deductive logic, has been proposed in [6].

Given any $V > 0$, let us introduce a function v as in [6]:

$$v(H,E) = \begin{cases} V, & \text{when} \quad E \models H, \\ -V, & \text{when} \quad E \models \neg H, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

For any argument (H, E) function v assigns it the same positive value V (e.g. $+1$) when the premise E of the rule entails the conclusion H (i.e. $E \models H$). The value of the opposite sign, $-V$ (e.g. -1), is assigned when 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.

According to [6], the relationship between the logical implication or refutation of H by E, and the conditional probability of H subject to E should fulfil the following principle Ex_1 based on function v:

$$\begin{aligned} &\text{if} \quad v(H_1,E_1) > v(H_2,E_2), \\ &\text{then} \quad c(H_1,E_1) > c(H_2,E_2). \end{aligned} \tag{5}$$

Measures satisfying property Ex_1 rank the conclusively confirmatory rules (e.g. the rule: "*if x is a jack, then x is a face-card*") higher than non-conclusively confirmatory rules. On the very bottom of the ranking are conclusively disconfirmatory rules (e.g. "*if x is seven of spades, then x is a face-card*"). Property Ex_1 concentrates on cases of entailment or refutation, which are equivalent to situations when $P(H|E) = 1$ (entailment) or $P(H|E) = 0$ (refutation). Thus, measures possessing property Ex_1 should place rules for which there are no counterexamples (i.e. $c = 0$) on top ranking places (high, positive value of the confirmation measure) and rules for which there are no positive examples ($a = 0$) on low ranking places (low, negative value of the confirmation measure). Measure $Z(H,E)$ was introduced in [6] as enjoying Ex_1.

As observed in [7], property Ex_1 implies that if a confirmation measure $c(H,E)$ reaches its maximum, then $c = 0$. However, it is also possible that $c = 0$ and the measure does not reach its maximal value. The consideration for minimal values are analogous and focus on $a = 0$.

The ranking of rules depending on entailment and refutation done by measures possessing Ex_1 property seems naturally desirable. Such an approach, however, can sometimes result in an inconvenient ranking of rules, as the considerations are boiled down to two situations: when there are no counterexamples to the rule, and when there are no positive examples to the rule [7].

To eliminate such undesirable rule rankings, a generalization of Ex_1 into weak Ex_1 has been proposed in [7]. Weak Ex_1 is based on the fact 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 $\neg E$ is present. In this context weak Ex_1 guarantees that a confirmation measure $c(H,E)$ cannot attain its maximal value unless $P(H|E) = 1$ (or equivalently, $c = 0$, i.e. there are no counterexamples to the rule) and $P(H|\neg E) = 0$ (or equivalently, $b = 0$, i.e. there are no objects satisfying the rule's conclusion but not its premise). Analogously, for the case of disconfirmation weak Ex_1 guarantees that a confirmation measure $c(H,E)$ cannot attain its minimal value unless $a = 0$ and $d = 0$. Thus, in case of confirmation, weak Ex_1 implies that if a confirmation measure reaches its maximal value, then $c = b = 0$. However, it is possible that $c = b = 0$ and still the measure does not reach its maximum.

2.3. Property L and its generalization to weak L. The logicality property L can be regarded as closely related to Ex_1 as it also concerns cases of entailment or refutation of the rule's conclusion by its premise [6,23].

Formally, a confirmation measure $c(H,E)$ enjoys property L under the following conditions:

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

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 the rule, i.e. when $a = 0$ [7]. Let us observe that property L implies that if $c = 0$, then a confirmation measure $c(H,E)$ reaches its maximum, however, $c(H,E)$ can also reach its maximum when $c > 0$.

Thus, if both properties L and Ex_1 are satisfied, then $c(H,E)$ reaches its maximum if and only if $c = 0$.

Similarly as with Ex_1, for measures possessing L there occurs a danger of obtaining unwanted rule rankings. To solve this problem a generalization of logicality L into weak L has been proposed [7], stating that a confirmation measure $c(H,E)$ possesses the weak L property when it satisfies the following conditions:

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

Now, weak L implies that if $c = b = 0$, then a confirmation measure $c(H,E)$ reaches its maximum, and when $a = d = 0$, then $c(H,E)$ obtains its minimal value.

Unauthenticated
Download Date | 4/9/15 9:29 AM

2.4. Properties of symmetry. A set of popular confirmation measures has been analysed in [9] from the viewpoint of the following four properties of symmetry, earlier also discussed in [23]:

- evidence symmetry (ES): $c(H,E) = -c(H,\neg E)$,
- commutativity (also referred to as inversion) symmetry (IS): $c(H,E) = c(E,H)$,
- hypothesis symmetry (HS): $c(H,E) = -c(\neg H,E)$,
- total (also referred to as evidence-hypothesis) symmetry (EHS): $c(H,E) = c(\neg H,\neg E)$,

The above symmetry properties consider how the value of a confirmation measure $c(H,E)$ relates to its value obtained for the situation when the rule's premise is negated (evidence symmetry), when the premise and conclusion switch positions (inversion symmetry), when the conclusion is negated (hypothesis symmetry) or, finally, when both premise and conclusion are negated (evidence-hypothesis symmetry).

More recently, in [6] the authors have argued for a more systematic analysis of symmetry properties. They extended the group of symmetries to include evidence-inversion symmetry, hypothesis-inversion symmetry and evidence-hypothesis-inversion symmetry. The idea behind these symmetries is however analogous to that in [9], thus, in this paper we shall only focus on ES, IS, HS and EHS. A detailed discussion on whether particular symmetries are desirable or not for interestingness measures can be found in [6, 9–11].

Finding the analysis of the above properties valuable and useful for the determination of truly meaningful confirmation measures, in the next sections we propose a particular visualization technique that supports and speeds up the process of property analysis of measures, especially useful with rapidly (i.e. in an automated way) generated measures (when there may be no time for detailed analysis of their properties).

3. The visualization technique

Some characteristics of the considered confirmation measures can best be demonstrated with regard to a particular data set (this includes the visualization technique to be presented in this chapter). In our analysis, the data set consist of an exhaustive and non-redundant set of contingency tables. Given a constant $n > 0$ (the total number of observations), it is generated as the set of all possible $\left[\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right]$ tables satisfying $a + b + c + d = n$. The set thus contains exactly one copy of each such contingency table. The set reveals all areas that can be possibly occupied by the four-dimensional domain of a measure, even those that would be omitted when using a real-life data set. Thus, our approach gives an insight into all possible behaviours and features of the visualized measures. Such a construction of the data set gives the results the necessary generality, rather than leaving them application-specific.

The total number of contingency tables t in the set is given by $t = (n+1)(n+2)(n+3)/6$. That makes it a polynomial (not exponential), but nevertheless, rapid growth, with the number t becoming considerable: the number of all tables for $n = 128$ equals $t = 366145$, while for $n = 1000$ (a typical number of objects in a non-trivial, benchmark classification data set) t exceeds hundreds of millions.

The data set comprises thus t rows and 4 columns, with the columns representing a, b, c and d. Because the four columns correspond to four degrees of freedom, visualization of such data in the form of a scatter-plot would require four dimensions. Owing to the condition $a+b+c+d = n$, the number of degrees of freedom is reduced by one, which means that is it possible to represent such data in three dimensions.

A simple progression of geometrical representations of constrained data sets is as follows:

- any one value, $a \geq 0$, satisfying $a = n > 0$ may be represented as a single point (zero dimensions),
- any two values, $a \geq 0$ and $b \geq 0$, satisfying $a+b = n > 0$ may be represented as a point of a segment (one dimension),
- any three values, $a \geq 0$, $b \geq 0$ and $c \geq 0$, satisfying $a + b + c = n > 0$ may be represented as point of a triangle (two dimensions),
- any four values, $a \geq 0$, $b \geq 0$, $c \geq 0$ and $d \geq 0$, satisfying $a+b+c+d = n > 0$ may be represented as a point of a tetrahedron (three dimensions).

Each of the above shapes (point, segment, triangle, tetrahedron) constitutes a simplex in the corresponding number of dimensions (0, 1, 2 and 3, respectively). Notice that the simplex is always a convex hull of its vertices (so any vector representing the coordinates of a point of the simplex may be thus represented as a convex combination of vectors representing the coordinates of the vertices). Because, as stated above, one needs three dimensions to represent the constrained a, b, c and d variables, the three-dimensional simplex, e.g. the tetrahedron, will be used in all further visualizations.

The tetrahedron consists of four vertices, six edges and four faces. A skeleton view (only edges visible) of an exemplary tetrahedron is depicted in Fig. 1, representing what will be referred to as the standard view, which is constructed as follows.

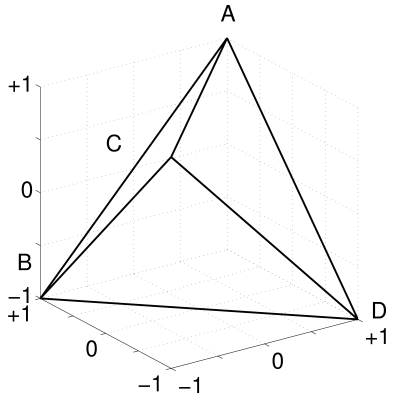

Fig. 1. A skeleton visualization of the tetrahedron

Imagine the 8 possible vectors of the form $\left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]$, where $x, y, z \in \{-1, +1\}$, constituting the vertices of a $2 \times 2 \times 2$ cube. The tetrahedron in question is inscribed into this cube

Unauthenticated
Download Date | 4/9/15 9:29 AM

with its four vertices A, B, C and D coinciding with the following cube's vertices: $A : \begin{bmatrix}+1\\+1\\+1\end{bmatrix}$, $B : \begin{bmatrix}-1\\+1\\-1\end{bmatrix}$, $C : \begin{bmatrix}-1\\-1\\+1\end{bmatrix}$ and $D : \begin{bmatrix}+1\\-1\\-1\end{bmatrix}$. Apart from those, the tetrahedron features six edges (segments): AB, AC, AD, BC, BD and CD, and four faces (triangles): ABC, BCD, CDA and DAB.

In the standard view (see Fig. 1), the tetrahedron vertices can be seen as follows:

- vertex C: in the centre,
- vertex A: above C,
- vertices B and D: below C, on the left and right, respectively.

The viewing angles (azimuth, elevation) in this view is $(-35°, 22°)$, which approximately corresponds to viewing the tetrahedron along (normalized) vector $\begin{bmatrix}0.53\\0.76\\-0.38\end{bmatrix}$, the viewing distance being ≈ 17.32 (i.e. from about point $\begin{bmatrix}-9.18\\-13.16\\6.58\end{bmatrix}$ towards point $\begin{bmatrix}0\\0\\0\end{bmatrix}$).

Assuming $a \geq 0$, $b \geq 0$, $c \geq 0$, $d \geq 0$ and $a+b+c+d = n > 0$, the original a, b, c and d values (i.e. the barycentric coordinates of objects in some universe of consideration, in this case: of contingency tables from the exhaustive and non-redundant set of such tables) may be converted to their corresponding x, y and z values (i.e. the Cartesian coordinates of elements of a simplex, in this case: of points in the tetrahedron defined above) using the formula:

$$\begin{bmatrix}x\\y\\z\end{bmatrix} = \frac{a}{n}\begin{bmatrix}+1\\+1\\+1\end{bmatrix} + \frac{b}{n}\begin{bmatrix}-1\\+1\\-1\end{bmatrix} + \frac{c}{n}\begin{bmatrix}-1\\-1\\+1\end{bmatrix} + \frac{d}{n}\begin{bmatrix}+1\\-1\\-1\end{bmatrix} = \frac{1}{n}\begin{bmatrix}a-b-c+d\\a+b-c-d\\a-b+c-d\end{bmatrix}.$$

Thanks to the assumed conditions fractions $\frac{a}{n}$, $\frac{b}{n}$, $\frac{c}{n}$ and $\frac{d}{n}$ satisfy $\frac{a}{n} \geq 0$, $\frac{b}{n} \geq 0$, $\frac{c}{n} \geq 0$, $\frac{d}{n} \geq 0$ as well as $\frac{a}{n} + \frac{b}{n} + \frac{c}{n} + \frac{d}{n} = 1$, which makes them proper coefficients of a convex combination of four vectors (in this case: the four vertices of the tetrahedron).

Simultaneously, because $a+b+c+d = n$ and $n > 0$, it is impossible for a, b, c and d to satisfy $a+b+c+d = 0$. But it is generally possible for them to satisfy the more specific, marginal conditions, with the following implications:

- Condition $a+b+c = 0$. In this case $d = n$. The corresponding data matrix is then of the form $\left[\begin{smallmatrix}0&0\\0&n\end{smallmatrix}\right]$ and, as such, corresponds to a point that is situated in the vertex D of the tetrahedron (notice that of all points of the tetrahedron, this vertex is maximally 'removed' from vertices A, B and C). Additional remarks:
  - The actual position of the point is strictly determined.
  - Analogous reasoning holds for all the other vertices (and corresponding conditions).
- Condition $a+b = 0$. In this case $c+d = n$. The corresponding data matrix is then of the form $\left[\begin{smallmatrix}0&n_1\\0&n_2\end{smallmatrix}\right]$, where $n_1+n_2 = n$ and, as such, corresponds to a point that is situated in the edge CD of the tetrahedron (notice that of all points of the tetrahedron, this edge is maximally 'removed' from vertices A and B). Additional remarks:
  - The actual position of the point within the edge is determined by n_1 and n_2: if $n_1 = 0$, then the point is situated in vertex D, if $n_2 = 0$, then the point is situated in vertex C, otherwise the point is situated strictly inside the convex hull of C and D; in particular, if $n_1 = n_2$, which implies $n_1 > 0$ and $\frac{n_1}{n} = \frac{n_1}{n_1+n_2} = \frac{n_1}{n_1+n_1} = \frac{n_1}{2 \cdot n_1} = \frac{1}{2}$ (similarly: $\frac{n_2}{n} = \frac{1}{2}$), it is equidistant from C and D.
  - Analogous reasoning holds for all the other edges (and corresponding conditions).
- Condition $a = 0$. In this case $b+c+d = n$. The corresponding data matrix is then of the form $\left[\begin{smallmatrix}0&n_2\\n_1&n_3\end{smallmatrix}\right]$, where $n_1+n_2+n_3 = n$ and, as such, corresponds to a point that is situated in the face BCD of the tetrahedron (notice that of all points of the tetrahedron, this face is maximally 'removed' from vertex A). Additional remarks:
  - The actual position of the point within the face is determined by n_1, n_2 and n_3: if $n_1 = 0$, then the point is situated in edge CD, if $n_2 = 0$, then the point is situated in edge BD, if $n_3 = 0$, then the point is situated in edge BC, otherwise the point is situated strictly inside the convex hull of B, C and D; in particular, if $n_1 = n_2 = n_3$, which implies $n_1 > 0$ and $\frac{n_1}{n} = \frac{n_1}{n_1+n_2+n_3} = \frac{n_1}{n_1+n_1+n_1} = \frac{n_1}{3 \cdot n_1} = \frac{1}{3}$ (similarly: $\frac{n_2}{n} = \frac{n_3}{n} = \frac{1}{3}$), it is equidistant from B, C and D.
  - Analogous reasoning holds for all the other faces (and corresponding conditions).

If $n_1 > 0$, $n_2 > 0$, $n_3 > 0$ and $n_4 > 0$, then the point corresponding to table $\left[\begin{smallmatrix}n_1&n_3\\n_2&n_4\end{smallmatrix}\right]$, is situated strictly inside the tetrahedron. In particular, if $n_1 = n_2 = n_3 = n_4$, the point occupies its centre (and is thus equidistant from vertices A, B, C and D). Otherwise at least one inequality between n_i and n_j (with $i \neq j$) holds, and the point is not in the centre. The principle is that whenever $n_i > n_j$, then the point is closer to the vertex corresponding to n_i than to that corresponding to n_j (e.g. if $n_1 > n_2$, then the point is closer to A than to B).

Summarizing, the condition $a+b+c+d = n$ makes it possible to visualize the originally four-dimensional data in three dimensions, and three dimensions may be naturally visualized in three-dimensional plots. Now, using colour makes it possible to include a fourth dimension into the visualization, which leaves room for an additional variable. The visualization method may thus visualize any such variable, in particular, any (also non-linear) function $f(a, b, c, d)$ of a, b, c and d (further referred to as the operational function).

Unauthenticated
Download Date | 4/9/15 9:29 AM

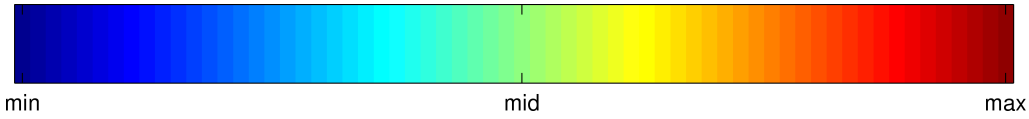

Fig. 2. The colour maps for the defined values of the operational function

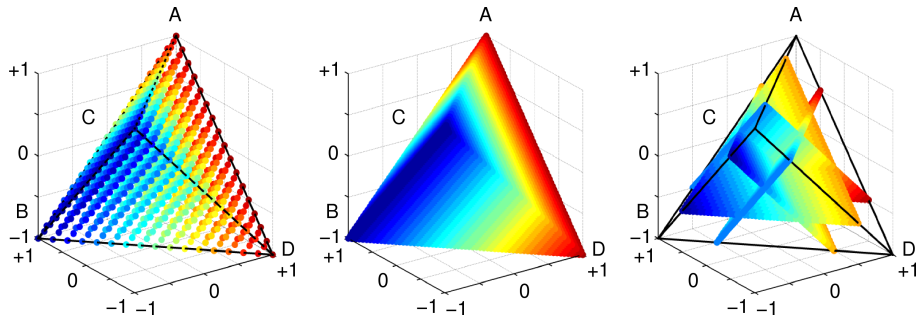

Fig. 3. Three tetrahedron visualizations of $f(a,b,c,d) = (a+d) - (b+c)$

The actual colour map (i.e. the correspondence ⟨colour range⟩ ↔ ⟨operational function range⟩) used in the following visualizations (see Fig. 2) is as follows:

- dark blue ↔ minimal values of the function,
- pale green ↔ middle values of the function,
- dark brown ↔ maximal values of the function.

The number of colour shades in the map is limited to 64 only in order to emphasize the value changes. Non-numeric values of the operational function, i.e. $+\infty$, NaN and $-\infty$, if produced, will be depicted in special colours (i.e. colours not occurring in the map). Since in this paper we focus on measures of confirmation and the only occurring undefined values are NaNs, they will be depicted in magenta.

Exemplary tetrahedron visualizations with the operational function $f(a,b,c,d) = (a+d) - (b+c)$ are shown in Fig. 3. The values of this function range from $-n$ (edge BC of the tetrahedron) to $+n$ (edge AD of the tetrahedron), so the colour map in Fig. 2 should be interpreted accordingly. The tetrahedron depicted in Fig. 3(left) consists of relatively few points ($t = 969$, created for $n = 16$), which renders it semi-transparent and thus provides some insight into its internal structure. All further visualizations show solid tetrahedra, made of opaque faces and thus showing no details of their interiors (see Fig. 3(middle), which shows a tetrahedron created for $n = 64$, with only the faces visible). If interior views are required, tetrahedron cuts with planes running parallel to the faces may be shown (see Fig. 3(right), $n = 64$, with selected internals visible).

The externals of the tetrahedron (as in Fig. 3(middle)) actually represent the extreme values of the arguments of the visualized function (as exemplified by the marginal conditions, e.g. the point in vertex D is reached when $a + b + c = 0$, points on the edge CD are reached when $a + b = 0$, etc.), which may turn out to be particularly useful in some applications. For brevity, all the further visualizations will be of the solid type (although other forms of visualization, including animation, are also available). However, because static figures show only one particular side of the tetrahedron, two views are usually presented:

- Standard view (left-hand side). The viewing angles (azimuth, elevation): $(-35\,^\circ, 22\,^\circ)$, viewing vector: $\begin{bmatrix} 0.53 \\ 0.76 \\ -0.38 \end{bmatrix}$, viewing distance: ≈ 17.32.
- Rotated view (right-hand side). The viewing angles (azimuth, elevation): $(145\,^\circ, 22\,^\circ)$, viewing vector: $\begin{bmatrix} -0.53 \\ -0.76 \\ -0.38 \end{bmatrix}$, viewing distance: ≈ 17.32.

An exemplary 2-view tetrahedron visualization of the function $f(a,b,c,d) = (a+d) - (b+c)$ is shown in Fig. 4.

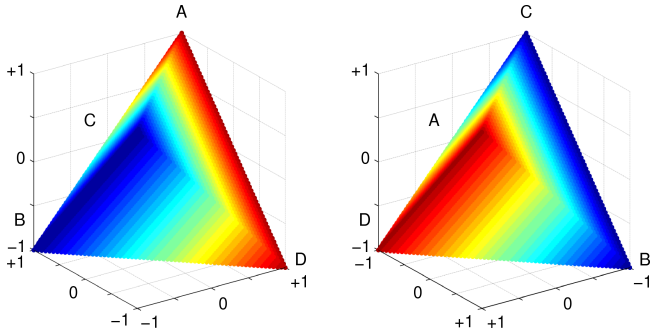

Fig. 4. A 2-view visualization of $f(a,b,c,d) = (a+d) - (b+c)$

An alternative to the 2-view visualization of the tetrahedron (i.e. the standard view and the rotated view), which is 3-dimensional, is a 2-dimensional view of the net (i.e. a set of planar triangles, which, when folded along selected edges, become the faces) of the tetrahedron. An exemplary 2-dimensional visualization of this net, in which the four faces of the tetrahedron form a parallelogram consisting of two distinct rhombi, is presented in Fig. 5.

Unauthenticated
Download Date | 4/9/15 9:29 AM

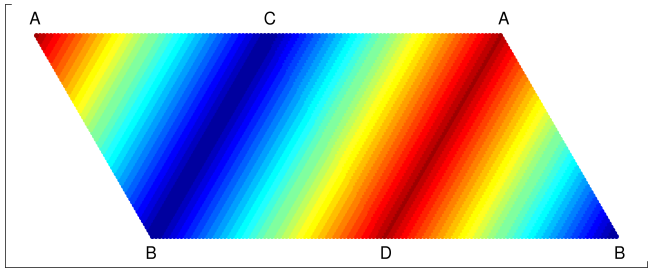

Fig. 5. A 2D 'parallelogram' visualization of $f(a, b, c, d) = (a + d) - (b + c)$

Notice that each of the three edges: AB, AC and BD is actually shown twice in the 'parallelogram' visualization. In other respects, in particular as far as the faces are concerned, the 2-dimensional visualization is more economical than the 3-dimensional visualization, as it shows each face only once, while 3D 2-view shows each of the two faces: ABC and ACD twice (similary for all the edges).

Described conventional visualizations of operational functions constitute only some of the numerous capabilities of the presented visualization technique, which formally include static (figures) and dynamic (animations) ones. A short summary of static capabilities [13, 14] (including those not demonstrated in this paper) is as follows:

- generalized views of single operational functions (e.g. Figs. 3(middle), 4, 5) capable of demonstrating the general properties of the functions, including:
  - gradient profiles of the faces and edges,
  - locations of extreme (minimal/maximal) values,
  - locations of (potentially existing) undefined values,
- specialized views of single/multiple operational functions:
  - regions of interest of single functions, i.e. only points satisfying pre-defined conditions (e.g. Fig. 3(right)),
  - differences between pairs of operational functions [13, 14],
  - variances of sets of operational functions [13],
  - (weighted) means of sets of operational functions,
  - directional derivatives of single functions,
  - variability levels of single functions.

4. Finding properties using visualization

Having established the visualization technique based on presenting a coloured tetrahedron for a particular confirmation measure (the colour reflecting the values of the measure), let us now consider how it facilitates the analysis of such measures. In case of the 12 selected confirmation measures (which include the modified $C(H, E)$ measure) the following holds:

- $\min\ f(a, b, c, d) = -1$ (dark blue),
- $\text{mid}\ f(a, b, c, d) = 0$ (pale green),
- $\max\ f(a, b, c, d) = +1$ (dark brown),

Additionally, NaN values of $f(a, b, c, d)$ are rendered in magenta.

The measures are depicted, in the form of the 'parallelogram' visualizations, in Figs. 6–16. Notice that out of every two rhombi that constitute the parallelogram, the left-hand side ones correspond to non-positive values of the measures, while the right-hand side ones to non-negative values.

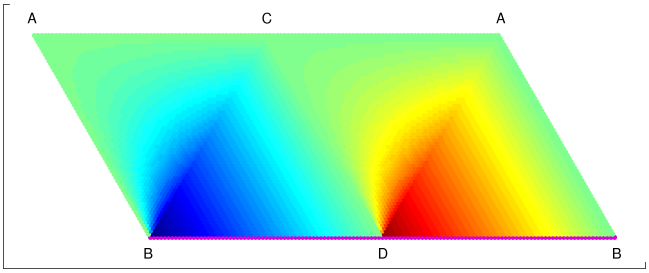

Fig. 6. A 2D 'parallelogram' visualization of $D(H, E)$

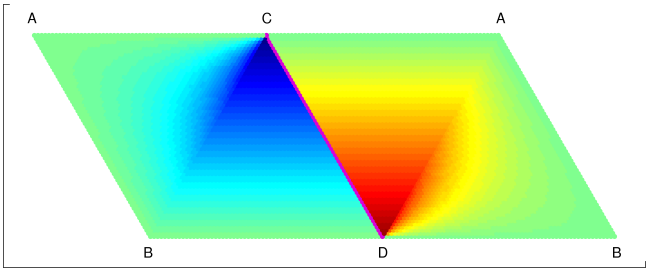

Fig. 7. A 2D 'parallelogram' visualization of $M(H, E)$

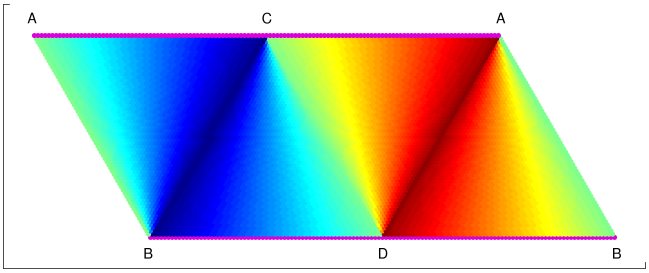

Fig. 8. A 2D 'parallelogram' visualization of $S(H, E)$

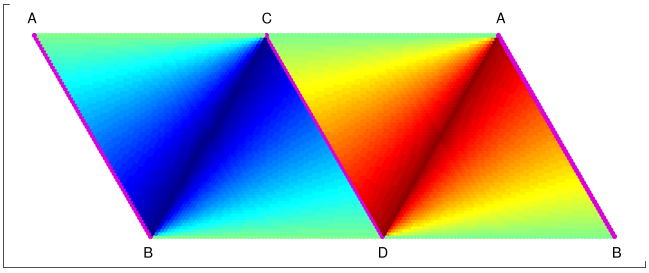

Fig. 9. A 2D 'parallelogram' visualization of $N(H, E)$

Unauthenticated
Download Date | 4/9/15 9:29 AM

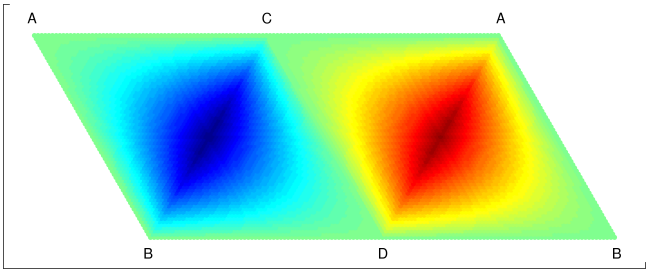

Fig. 10. A 2D 'parallelogram' visualization of $C(H, E)$

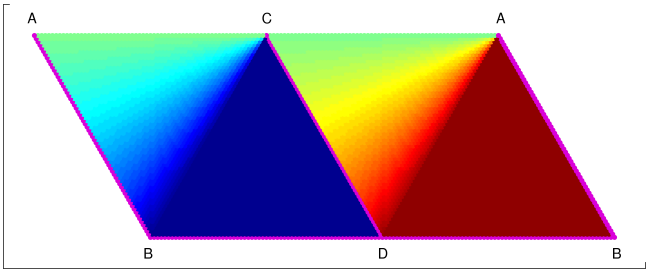

Fig. 11. A 2D 'parallelogram' visualization of $F(H, E)$

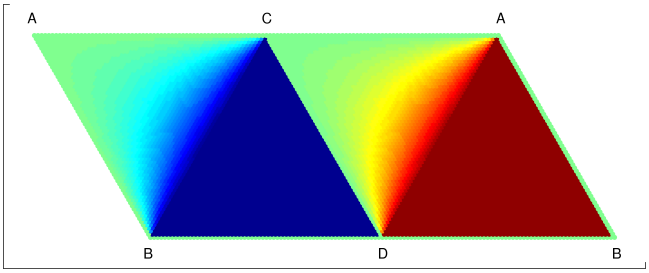

Fig. 12. A 2D 'parallelogram' visualization of $Z(H, E)$

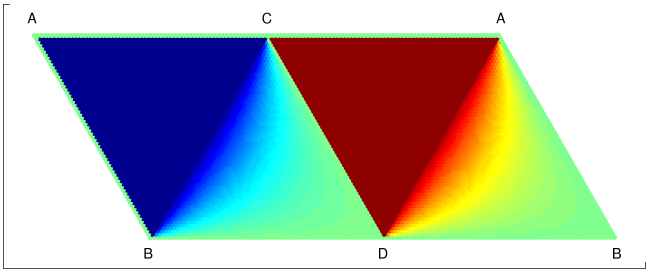

Fig. 13. A 2D 'parallelogram' visualization of $A(H, E)$

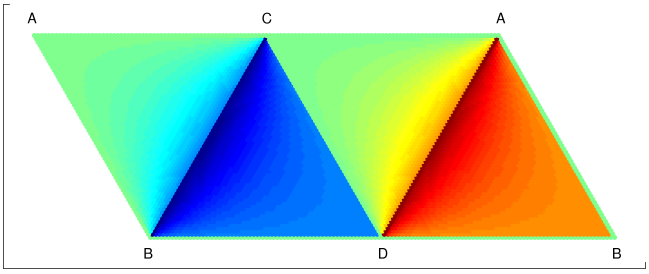

Fig. 14. A 2D 'parallelogram' visualization of $c_1(H, E)$

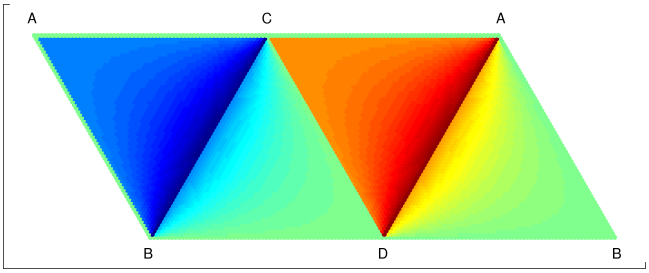

Fig. 15. A 2D 'parallelogram' visualization of $c_2(H, E)$

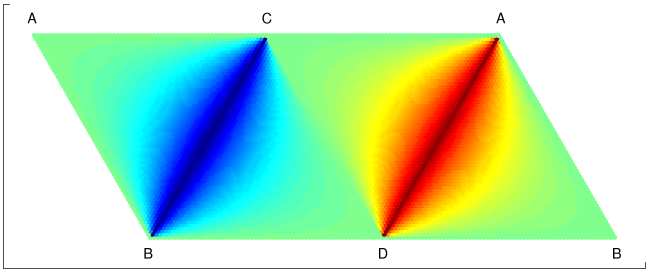

Fig. 16. A 2D 'parallelogram' visualization of $c_3(H, E)$ and $c_4(H, E)$ (these two measures, although different in general, obtain identical values in the external parts of the tetrahedron)

Further basic properties of the measures are easily discernible, as they are implied by the location of extreme (maximal and minimal) and intermediate values in the tetrahedron, and the general forms of transitions between these. Additional information is provided by the location of undefined values (if present).

The following subsections provide analyses of the measures with respect to property M, Ex_1, weak Ex_1, L and weak L as well as to four selected symmetries.

4.1. Visual analysis of property M. Property of monotonicity M demands that a measure $c(H, E)$ (by definition being a scalar function $f(a, b, c, d)$ of a, b, c and d) is a function that is non-decreasing with respect to a and d, and non-increasing with respect to b and c. Out of the 12 selected confirmation measures the following possess this property: $S(H, E)$, $N(H, E)$, $F(H, E)$, $Z(H, E)$, $A(H, E)$, $c_1(H, E)$, $c_2(H, E)$, $c_3(H, E)$ and $c_4(H, E)$.

Assuming that the values of the analysed measures range from -1 to $+1$ (as it is for the 12 selected confirmation measures), we expect dark blue or dark brown values to express their extremes. In this context, the "non-decreasing with a and d" condition should be reflected in the visualization as colours changing towards dark brown (increase of confirmation) around vertices A and D. On the other hand, the colours are expected to change towards dark blue (towards stronger disconfirmation) around vertices B and C, in order to fulfil the "non-increasing with b and c" condition.

The visualization allows thus to quickly identify situations (counterexamples) being contrary to the demands of proper-

Unauthenticated
Download Date | 4/9/15 9:29 AM

ty M. Any case of colour change different than towards dark brown (dark blue) around vertices A and D (B and C) shows that the visualized measure does not satisfy property M. Let us stress, however, that thorough analysis of property M requires insight into the tetrahedron, as potential counterexamples can also lie inside the shape.

The most conspicuous (but not sole) counterexamples in Figs. 6, 7 and 10 are located in the AD edge, where the colour changes from dark brown at vertex D to pale green at vertex A (Figs. 6 and 7) and from dark brown in the middle of AD to pale green at vertex A (Fig. 10). This indicates the decrease of the measure values with a, contrary to the demands of the property.

On the other hand, there are no observable counterexamples to this property in Figs. 8, 9 and 11–16 which, together with additional analysis of the inside of the tetrahedrons, determines the possession of the property by the corresponding measures.

4.2. Visual analysis of properties Ex_1 and weak Ex_1. Property Ex_1 implies that if a confirmation measure $c(H, E)$ reaches its maximum (minimum) then $c = 0$ ($a = 0$). Out of the 12 selected confirmation measures the following possess this property: $F(H, E)$, $Z(H, E)$, $c_1(H, E)$.

Assuming that the dark brown (dark blue) colour expresses the maximal (minimal) value of the visualized measure, this translates into expectation that the dark brown (dark blue) colour cannot be found anywhere else than on the ABD (BCD) face of the tetrahedron. The ABD (BCD) face contains all points for which $c = 0$ ($a = 0$). Let us stress that property Ex_1 does not require the whole ABD (BCD) face to be dark brown (dark blue), as it is possible that $c = 0$ ($a = 0$) and still the visualized measure $c(H, E)$ does not reach its maximal (minimal) value.

It is also important to notice that verification of possession of Ex_1 using our visualization technique requires insight into the tetrahedron, as any dark brown (dark blue) point inside the shape, i.e. not on the ABD (BCD) face, should be interpreted as a counterexample, ruling out Ex_1 from the set of properties of the visualized measure.

Clear counterexamples to the demands of this property can be seen in Fig. 13 as the whole ACD face consists of dark brown points, violating the demand that they occur only in ABD face.

On the other hand, there are no observable counterexamples to this property in Figs. 11, 12 and 14 which, together with additional analysis of the inside of the tetrahedrons, determines the possession of the property by the corresponding measures.

As to weak Ex_1 property, it implies that if a confirmation measure $c(H, E)$ reaches its maximum (minimum) then $c = b = 0$ ($a = d = 0$). Out of the 12 selected confirmation measures the following possess this property: $S(H, E)$, $N(H, E)$, $c_1(H, E)$, $c_2(H, E)$, $c_3(H, E)$ and $c_4(H, E)$.

In the context of visualization this translates into expectation that the dark brown (dark blue) colour can only be found on the AD (BC) edge of the tetrahedron, as it contains all points for which $c = b = 0$ ($a = d = 0$). Let us observe that the weak Ex_1 does not require the whole AD (BC) edge to be dark brown, as it is possible that $c = b = 0$ ($a = d = 0$) and still the visualized measure $c(H, E)$ does not reach its maximal value.

Similarly, as with Ex_1 property, determination of weak Ex_1 requires insight into the tetrahedron, as any dark brown (dark blue) point inside the shape, i.e. not on the AD (BC) edge should be interpreted as a counterexample, ruling out weak Ex_1 from the set of properties of the visualized measure.

Counterexamples to the demands of weak Ex_1 property can be seen in Fig. 13 as the whole ACD face consists of dark brown points, violating the demand that they occur only in AD edge.

On the other hand, there are no observable counterexamples to this property in Figs. 8, 9 and 14–16 which, together with additional analysis of the inside of the tetrahedrons, determines the possession of the property by the corresponding measures.

4.3. Visual analysis of properties L and weak L. Property L implies that if $c = 0$ ($a = 0$) then a confirmation measure $c(H, E)$ obtains its maximal (minimal) value. Out of the 12 selected confirmation measures the following possess this property: $F(H, E)$ and $Z(H, E)$.

Should the dark brown (dark blue) colour stand for the maximum (minimum) of the visualized measure, we would expect the whole ABD (BCD) face to be dark brown (dark blue). This face contains all points for which $c = 0$ ($a = 0$). It is important to notice that the whole ABD (BCD) face must be dark brown (dark blue) if the visualized measure is to possess the L property. We do not demand, however, that the dark brown (dark blue) points lie only on the respective faces, as the visualized measure can reach its maximum (minimum) even when $c > 0$ ($a > 0$). This demand distinguishes determination of possession of property Ex_1 from property L.

Moreover, contrary to property Ex_1, when considering property L, we do not need any insight into the tetrahedron. We are only interested in the points for which $c = 0$ ($a = 0$), thus the external ABD (BCD) face is enough for such analyses. In this light, any non-dark brown (non-dark blue) point on the ABD (BCD) face should be considered as a counterexample, ruling out L from the set of properties of the visualized measure.

All counterexamples in Figs. 6–10 and 13–16 are located in the ABD (BCD) faces, which do contain points other than dark brown (dark blue). This indicates non-maximal (non-minimal) values for $c = 0$ ($a = 0$), contrary to the demands of the property.

On the other hand, there are no counterexamples to this property in Figs. 11 and 12, which determines the possession of the property by the corresponding measures.

As to weak L property, it implies that if $c = b = 0$ ($a = d = 0$) then a confirmation measure $c(H, E)$ obtains its maximal (minimal) value. Out of the 12 selected confirmation measures the following possess this property: $S(H, E)$,

Unauthenticated
Download Date | 4/9/15 9:29 AM

$N(H,E)$, $F(H,E)$, $Z(H,E)$, $A(H,E)$, $c_1(H,E)$, $c_2(H,E)$, $c_3(H,E)$ and $c_4(H,E)$.

Again, assuming that the dark brown (dark blue) colour stands for the maximum (minimum) of the visualized measure, we expect the whole AD (BC) edge to be dark brown (dark blue). Those edges contain all points for which ($c = b = 0$). It is important to notice that the whole AD (BC) edge must be dark brown (dark blue) if the visualized measure possesses the weak L property. Similarly, as with property L, we do not demand that the dark brown (dark blue) points lie only on the respective edges. Again, this differentiates the weak Ex_1 property from weak L.

Moreover, for the analysis of a measure with respect to weak L, we do not need any insight into the tetrahedron as we are only interested in the points for which $c = b = 0$ ($a = d = 0$). Nevertheless, any non-dark brown (non-dark blue) point on the AD (BC) edge should be considered as a counterexample, ruling out weak L from the set of properties of the visualized measure.

All counterexamples in Figs. 6, 7, and 10 are located in the AD (BC) edges, which do contain points other than dark brown (dark blue). This indicates non-maximal (non-minimal) values for $c = b = 0$ ($a = d = 0$), contrary to the demands of the property.

On the other hand, there are no counterexamples to this property in Figs. 8, 9 and 11–16, which determines the possession of the property by the corresponding measures.

4.4. Visual analysis of symmetry properties ES, IS, HS and EHS.

Evidence symmetry. The property of evidence symmetry demands that a measure obtains the same values, but of the opposite sign, for rules $E \rightarrow H$ and $\neg E \rightarrow H$, i.e. $c(H,E) = -c(H,\neg E)$. Let us assign to the $E \rightarrow H$ rule a contingency table with a, b, c and d frequencies, and to the $\neg E \rightarrow H$ rule a table with a', b', c' and d' frequencies. Then, the requirement of the evidence symmetry can be formulated as $c(H,E) = f(a,b,c,d) = -c(H,\neg E) = -f(a',b',c',d') = -f(b,a,d,c)$. A row exchange in the contingency table can be observed here, as $a = b'$, $b = a'$, $c = d'$ and $d = c'$.

In this context, a verification of possession of evidence symmetry in the 'parallelogram' visualization can be done by rotating the parallelogram by 180° about its middle (which leads to the exchange of the left-hand side and the right-hand side rhombi, with their orientation changed to upside-down) and reversing the colour map. If the rotated and 'recoloured' parallelogram is not the same as the original one, then the visualized measure does not possess the evidence symmetry. The visualization allows thus to quickly identify situations (counterexamples) being contrary to demands of the evidence symmetry. Let us stress, however, that thorough analysis of ES requires insight into the tetrahedron, as potential counterexamples can also lie inside the shape.

In Figs. 6, 7 and 11–15 the counterexamples to evidence symmetry are easily seen, as the rotated and 'recoloured' parallelograms are not the same as the original ones.

On the other hand, there are no observable counterexamples to this property in Figs. 8–10 and 16 which, together with additional analysis of the inside of the tetrahedrons, determines the possession of the evidence symmetry by the corresponding measures.

Inversion symmetry. Moving on to the inversion symmetry, it requires that a measure obtains the same values for rules $E \rightarrow H$ and $H \rightarrow E$, i.e. $c(H,E) = c(E,H)$. A measure possessing the inversion symmetry can be actually regarded as an interestingness measure of itemsets instead of rules, as the measure does not distinguish the position of the premise and the conclusion (they can interchange without any influence on the value of the measure). Associating a, b, c and d with $E \rightarrow H$ and a', b', c' and d' with $H \rightarrow E$, the requirement of the inversion symmetry can be formulated as $c(H,E) = f(a,b,c,d) = c(E,H) = f(a',b',c',d') = f(a,c,b,d)$. A transpose operation in the contingency table can be observed here, as $a = a'$, $b = c'$, $c = b'$ and $d = d'$.

A measure depicted using the 'parallelogram' visualization possesses the inversion symmetry provided that the left-hand side rhombus is symmetric with respect to its longer diagonal and the right-hand side rhombus is symmetric with respect to its shorter diagonal. Naturally, a thorough analysis of IS requires insight into the tetrahedron, as potential counterexamples can also lie inside the shape. Let us observe that since the inversion symmetry does not concern opposite signs (as e.g. evidence symmetry does) there is no need to reverse the colour map.

In Figs. 6–9 and 11 the counterexamples to inversion symmetry are easily seen in either rhombi, as neither the left-hand side nor the right-hand side one possesses the required symmetry. Moreover, Figs. 12–14 and 15 show counterexamples to this property, but only in the area of non-negative values, i.e. in the right-hand side rhombus.

On the other hand, there are no observable counterexamples to this property in Figs. 10 and 16 which, together with additional analysis of the inside of the tetrahedrons, determines the possession of the inversion symmetry by the corresponding measures.

Hypothesis symmetry. As to the hypothesis symmetry, it demands that a measure obtains the same values, but of the opposite sign, for rules $E \rightarrow H$ and $E \rightarrow \neg H$, i.e. $c(H,E) = -c(\neg H,E)$. Let us assign to the first rule a contingency table with a, b, c and d frequencies, and to the latter a table with a', b', c' and d' frequencies. Then, the requirement of the hypothesis symmetry can be formulated as $c(H,E) = f(a,b,c,d) = -c(\neg H,E) = -f(a',b',c',d') = -f(c,d,a,b)$. A column exchange in the contingency table can be observed here, as $a = c'$, $b = d'$, $c = a'$ and $d = b'$.

Using the 'parallelogram' visualization technique, the verification of possession of hypothesis symmetry can be easily done by shifting the left-hand side rhombus to the right, the right-hand side rhombus to the left (which leads to the exchange of the rhombi, with their orientation unchanged) and reversing the colour map. If the re-shifted and 'recoloured'

Unauthenticated
Download Date | 4/9/15 9:29 AM

parallelogram is not the same as the original one, then the visualized measure does not possess the hypothesis symmetry. Again, a thorough analysis of HS requires insight into the tetrahedron, as potential counterexamples can also lie inside the shape.

Figure 7 depicts the counterexamples to hypothesis symmetry of measure $M(H, E)$, as the left-hand side and the right-hand side rhombi differ even after reversing the colour map.

On the other hand, there are no observable counterexamples to this property in Figs. 6 and 8–16 which, together with additional analysis of the inside of the tetrahedrons, determines the possession of the hypothesis symmetry by the corresponding measures.

Evidence-hypothesis symmetry. Finally, the evidence-hypothesis symmetry is defined as a composition of ES and HS. It requires that a confirmation measure obtains the same values for rules $E \rightarrow H$ and $\neg E \rightarrow \neg H$, i.e. $c(H, E) = c(\neg H, \neg E)$. Associating a, b, c and d with $E \rightarrow H$ and a', b', c' and d' with $\neg E \rightarrow \neg H$, the requirement of the evidence-hypothesis symmetry can be formulated as $c(H, E) = f(a, b, c, d) = c(\neg H, \neg E) = f(a', b', c', d') = f(d, c, b, a)$. A 180° rotation of the contingency table can be observed here, as $a = d'$, $b = c'$, $c = b'$ and $d = a'$.

A measure in the 'parallelogram' visualization may be verified for this type of symmetry by rotating both the left-hand side and the right-hand side rhombi by 180° about their respective middle points. If any of the rotated rhombi is not the same as the original one, then the visualized measure does not possess the evidence-hypothesis symmetry. Naturally, a thorough analysis of EHS requires insight into the tetrahedron, as potential counterexamples can also lie inside the shape. Again, just as with inversion symmetry, there is no need to reverse the colour map.

In Figs. 6, 7 and 11–15 the counterexamples to evidence-hypothesis symmetry are easily seen, as the rotated rhombi are not the same as the original ones.

On the other hand, there are no observable counterexamples to this property in Figs. 8–10 and 16 which, together with additional analysis of the inside of the tetrahedrons, determines the possession of the evidence-hypothesis symmetry by the corresponding measures.

5. Conclusions

The paper presents a visual technique designed to assist and enhance the analysis of interestingness measures, in particular confirmation measures. The technique may be advantageous in real-life situations, especially when time constraints impede conducting in-depth, theoretical analyses of large numbers of such measures (e.g. generated in an automatic way). Using a data set that consists of an exhaustive and non-redundant set of contingency tables (which constitute arguments to these measures), a 3-dimensional tetrahedron is created and visualized. This shape is constructed so that the positions of its individual points represent the contingency tables, while the colours of the points represent values of the visualized measure.

Dedicated analyses of particular colour patterns on this tetrahedron allow to promptly perceive distinct properties of the visualized measures. Basic properties are determined by the location of undefined, extreme and intermediate values in the tetrahedron, and the general forms of transitions between these. Additionally, counterexamples to requirements and conditions imposed by some specialized properties may easily be found, ruling out certain properties from the set of properties of the visualized measure. If such a condition requires an insight into the interior of the tetrahedron, this may be carried out using specialized views (including animation). Of course it must be always kept in mind that the visual information is only one way of analysing the interestingness measures, and, as such, however useful, it may not convey answers to all possible questions regarding measure properties.

To illustrate the introduced technique with real-life confirmation measures and real-life properties, a set of 12 popular confirmation measures has been selected. They have been visually analysed in terms of 9 commonly used properties. The performed analyses reveal that out of the considered properties the symmetries are most "visually-friendly", in that they generate colour patterns that are easiest to determine. Further well supported properties are logicality and weak logicality.

It should be stressed that the visual analyses introduced and described in this paper are universal enough to be applied to basically all newly defined interestingness measures. Thus, they greatly improve the general comprehension of the measures and, as such, may positively influence the way in which they will be defined in the future.

Acknowledgements. The authors' research was supported by the Polish National Science Center, the grant no. DEC-2013/11/B/ST6/00963.

REFERENCES

[1] U. Fayyad, G. Grinstein, and A. Wierse, *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann, Berlin, 2002.

[2] K. Napierała and J. Stefanowski, "Bracid: a comprehensive approach to learning rules from imbalanced data", *J. Intelligent Information Systems* 39 (2), 335–373 (2012).

[3] I. Brzezińska, S. Greco, and R. Słowiński, "Mining pareto-optimal rules with respect to support and anti-support", *Engineering Applications of Artificial Intelligence* 20 (5), 587–600 (2007).

[4] S. Greco, Z. Pawlak, and R. Słowiński, "Can bayesian confirmation measures be useful for rough set decision rules?", *Engineering Applications of Artifficial Intelligence* 17, 345–361 (2004).

[5] I. Szczęch, "Multicriteria attractiveness evaluation of decision and association rules", *Transactions on Rough Sets X, LNCS Series* 5656, 197–274 (2009).

[6] V. Crupi, K. Tentori, and M. Gonzalez, "On bayesian measures of evidential support: Theoretical and empirical issues", *Philosophy of Science* 74, 229–252 (2007).

Unauthenticated
Download Date | 4/9/15 9:29 AM

[7] S. Greco, R. Słowiński, and I. Szczęch, "Properties of rule interestingness measures and alternative approaches to normalization of measures", *Information Sciences* 216, 1–16 (2012).
[8] B. Fitelson, "Studies in Bayesian confirmation theory", *PhD Thesis*, University of Wisconsin, Madison, 2001.
[9] E. Eells and B. Fitelson, "Symmetries and asymmetries in evidential support", *Philosophical Studies* 107 (2), 129–142 (2002).
[10] S. Greco, R. Słowiński, and I. Szczęch, "Analysis of symmetry properties for bayesian confirmation measures", *Rough Sets and Knowledge Technology – 7th Int. Conf.* (RSKT 2012), 207–214 (2012).
[11] S. Greco, R. Słowiński, and I. Szczęch, "Finding meaningful bayesian confirmation measures", *Fundamenta Informaticae* 127 (1–4), 161–176 (2013).
[12] D.H. Glass, "Confirmation measures of association rule interestingness", *Knowlegde Based Systems* 44, 65–77 (2013).
[13] R. Susmaga and I. Szczęch, "Can interestingness measures be usefully visualized?", *Int. J. Applied Mathematics and Computer Science* (2015), (to be published).
[14] R. Susmaga and I. Szczęch, "Visualization of interestingness measures", *Proc. 6th Language & Technology Conf.: Human Language Technologies as a Challenge for Computer Science and Linguistics* 1, 95–99 (2013).
[15] R. Susmaga and I. Szczęch, "Visual-based detection of properties of confirmation measures", *Lecture Notes in Computer Science, ISMIS* 8502, 133–143 (2014).
[16] L. Geng and H. Hamilton, "Interestingness measures for data mining: A survey", *ACM Computing Surveys* 38, 3 (2006).
[17] S. Greco, R. Słowiński, and I. Szczęch, "Property of confirmation", in *Tech. Rep.* RA-01/14, Poznań University of Technology, Poznań, 2014.
[18] B. Fitelson, "The plurality of bayesian measures of confirmation and the problem of measure sensitivity", *Philosophy of Science* 66, 362–378 (1999).
[19] E. Eells, *Rational Decision and Causality*, Cambridge University Press, Cambridge, 1982.
[20] H. Mortimer, *The Logic of Induction*, Paramus, Prentice Hall, 1988.
[21] D. Christensen, "Measuring confirmation", *J. Philosophy* 96, 437–461 (1999).
[22] R. Nozick, *Philosophical Explanations*, Clarendon Press, Oxford, 1981.
[23] R. Carnap, *Logical Foundations of Probability*, University of Chicago, Chicago, 1962.
[24] J. Kemeny and P. Oppenheim, "Degrees of factual support", *Philosophy of Science* 19, 307–324 (1952).
[25] R. Susmaga and I. Szczęch, "Statistical significance of bayesian confirmation measures", in *Tech. Rep.*, RA-010/12, Poznań University of Technology, Poznań, 2012.
[26] P. Lenca, P. Meyer, B. Vaillant, and S. Lallich, "On selecting interestingness measures for association rules: user oriented description and multiple criteria decision aid", *Eur. J. Operational Research* 184 (2), 610–626 (2008).

Unauthenticated
Download Date | 4/9/15 9:29 AM