Int. J. Appl. Math. Comput. Sci., *2015, Vol. 25, No. 2, 323–336*
DOI: 10.1515/amcs-2015-0025

CAN INTERESTINGNESS MEASURES BE USEFULLY VISUALIZED?

ROBERT SUSMAGA [a,*], IZABELA SZCZĘCH [a]

[a] Institute of Computing Science
Poznań University of Technology, Piotrowo 2, 60-965 Poznań, Poland
e-mail: {Robert.Susmaga,Izabela.Szczech}@cs.put.poznan.pl

The paper presents visualization techniques for interestingness measures. The process of measure visualization provides useful insights into different domain areas of the visualized measures and thus effectively assists their comprehension and selection for different knowledge discovery tasks. Assuming a common domain form of the visualized measures, a set of contingency tables, which consists of all possible tables having the same total number of observations, is constructed. These originally four-dimensional data may be effectively represented in three dimensions using a tetrahedron-based barycentric coordinate system. At the same time, an additional, scalar function of the data (referred to as the operational function, e.g., any interestingness measure) may be rendered using colour. Throughout the paper a particular group of interestingness measures, known as confirmation measures, is used to demonstrate the capabilities of the visualization techniques. They cover a wide spectrum of possibilities, ranging from the determination of specific values (extremes, zeros, etc.) of a single measure, to the localization of pre-defined regions of interest, e.g., such domain areas for which two or more measures do not differ at all or differ the most.

Keywords: visualization, interestingness measures, confirmation measures, barycentric coordinates.

1. Introduction

Rapid progress in data mining and knowledge discovery techniques has increased, over the recent years, our ability to extract answers from data. Presenting these answers in meaningful ways, a task difficult in itself, may employ means like data visualization techniques. Data visualization can provide graphical metaphors for data manipulation and comprehension; it is thus natural that the development of different tools within knowledge discovery in databases (KDD) and machine learning (ML), like association/decision rule inductors, regression model generators, classifiers, etc., is accompanied by the development of various visualization approaches (Ware, 2004).

Following that trend, we propose some visualization techniques to facilitate and support the analyses of interestingness measures, commonly used to evaluate rule patterns mined from data (Geng and Hamilton, 2006; Tan *et al*., 2002; Shaikh *et al*., 2013). The induction of *if-then* rules from data sets usually requires an evaluation step to limit the number of rules presented to the user, and quantitative measures of interest are often used for such a filtration process (Agrawal *et al*., 1993; Morzy and Zakrzewicz, 2003). It is not easy, though, to choose an appropriate measure for a particular application. The visualization techniques that we propose aim at revealing the recesses of interestingness measures, and thus at directing the users toward the measures that act according to their expectations. It is done by visualizing the values obtained by a measure for an exhaustive and non-redundant set of contingency tables. This way we gain an insight into all areas of the domain that the visualized measure can possibly occupy, and which could otherwise be omitted and thus remain undiscovered while working on real-life data.

The analyses facilitated by our visualization techniques cover a wide spectrum of possibilities, ranging from determination of a measure's extremes or the areas for which its value is undefined, to visualization of the areas of the data set for which two or more measures differ the most. One could then, e.g., decide to work with a couple of measures that react to different (types of) objects in the data set, or could choose to use measures that are not ordinally equivalent. Thus, the visualization enriches our knowledge on the features and the behaviour

[*] Corresponding author

of the visualized measures. Moreover, it eases defining new measures and facilitates the analyses of the newly developed ones (e.g., automatically generated).

In this paper, the illustrative application of our techniques, as exemplified through a MATLAB-based implementation, is presented for a particular group of interestingness measures, called *confirmation measures*. These particular measures are designed for the evaluation of decision rules in the form of "*if* premise, *then* conclusion". The confirmation measures are characterised by the fact that they obtain

- positive values, when the premise of a rule confirms its conclusion,
- zero values, when the rule's premise and conclusion are neutral to each other,
- negative values, when the premise of a rule disconfirms its conclusion.

This paper builds on the main ideas proposed by Susmaga and Szczęch (2013), regarded as our preliminary results, which are now extended and refined. In particular, the description of the proposed visualization techniques, including the idea of barycentric coordinates, both in two and three dimensions, has been presented in more detail. Moreover, the group of the described and analysed confirmation measures is extended. The range of analyses applied to these measures now also includes specialized views of coefficients specific to groups of measures (e.g., their standard deviations). The proposed approach is also the basis for investigating the measures with regard to their selected properties (monotonicity, symmetry, etc.), further described in Susmaga and Szczęch (2014).

Let us observe that the techniques presented in this paper are in many aspects different from data visualization approaches commonly applied in KDD and ML, which are basically concerned with representing graphically selected evaluations of employed tools, e.g., the performance of classifiers. In most typical applications these are usually two-dimensional characteristics, e.g., ROC curves (Alaíz-Rodríguez *et al.*, 2008; Drummond and Holte, 2006; Hernández-Orallo *et al.*, 2011; Zhou *et al.*, 2014), although more dimensional approaches are also attempted (Everson and Fieldsend, 2006). These characteristics are constructed for particular data sets and particular data analysis tools, with the main purpose of describing and controlling the data analysis process (e.g., the convergence of classification results).

The proposed approach, on the other hand, focuses on visualizing the whole domains of different measures that are used in the data analysis process. This may concern measures applied at early stages, e.g., interestingness measures used to evaluate and filter patterns (e.g., decision rules) that contribute to the classifiers under construction, but also performance measures used to evaluate the classifiers that are already constructed. The required feature of the measures to be visualized is a four-dimensional, real-valued domain. Incidentally, this makes them also actually difficult to represent visually. Our approach solves this particular difficulty by rendering the originally four-dimensional domains in three dimensions using a tetrahedron-based barycentric coordinate system. The techniques applied are comprehensively illustrated by their sample application to a set of selected confirmation measures.

The rest of the paper is organized as follows. Section 2 demonstrates the proposed visualization techniques. Section 3 defines popular confirmation measures and presents the application of the visualization techniques to those measures. It also recounts some conclusions drawn from the visualization-based analyses. A summary of the approach and final remarks are contained in Section 4.

2. Visualization techniques

Our visualization techniques aim at facilitating and supporting the analyses of various characteristics of interestingness measures. Such measures are commonly used to evaluate rules induced from a sample of a larger reality, represented in the form of a set of objects. A rule induced from such a set consists of a *premise* "*if* E" (referring to an existing piece of evidence, E), and a *conclusion* "*then* H" (referring to a hypothesised piece of evidence, H). Below, we shall use the common, shortened denotation $E \to H$ (read as "*if* E*, then* H").

In the context of a particular set of objects, the relation between E and H may be quantified by four non-negative integers a, b, c and d, briefly represented in a 2×2 table (see Table 1). The number of all objects in the set satisfying both the premise and the conclusion of a rule is expressed by a, b stands for the number of objects for which the premise in not satisfied, but the conclusion is, etc. Let us observe that a, b, c and d can also be used to estimate 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 $P(H|E) = P(H \cap E)/P(E) = a/(a + c)$ (which, however, is only defined when $a + c > 0$). The notation based on a, b, c and d can be effectively used for defining interestingness measures, e.g., support and confidence (Agrawal *et al.*, 1993), lift (IBM, 1996), gain (Fukuda *et al.*, 1996) or measures of confirmation (see Section 3).

There exist a great deal of different features/parameters of interestingness measures, e.g., characterization of their extreme values, zero values, non-numeric values (e.g., ∞), gradients, etc., which, when well comprehended, allow the users to choose a measure for a particular application more competently.

In this article we focus on the characteristics of

Table 1. Contingency table of the rule's premise E and conclusion H.

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

interestingness measures that can best be demonstrated with regard to particular data. For the purpose of our visualization, an exhaustive and non-redundant set of contingency tables shall be used. 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 table. The total number of contingency tables t in the set is given by $t = (n+1)(n+2)(n+3)/6$. We use n reaching up to 256 (for which $t = 2862209$) in further analyses.

The resulting data set comprises thus t rows and 4 columns, with the columns representing a, b, c and d. Because, in general, four independent columns correspond to four degrees of freedom, visualization of such data in the form of a scatter-plot would formally require four dimensions. Owing to the condition $a + b + c + d = n$, however, the number of degrees of freedom is reduced to three, which means that it is possible to visualize such data in three dimensions using barycentric (Floater *et al.*, 2006; Warren, 2003) coordinates.

The general idea of barycentric coordinates in two dimensions is as follows. Let $\delta(P, XY)$ denote the (Euclidean) distance from a point P to a segment XY. Given an equilateral triangle XYZ with the length of each side equal to $s = 2/\sqrt{3}$, the equality

$$\delta(P, XY) + \delta(P, YZ) + \delta(P, ZX) = \frac{\sqrt{3}}{2}s = \frac{\sqrt{3}}{2} \cdot \frac{2}{\sqrt{3}} = 1$$

holds for all points P inside the triangle. At the same time, for each such a point, the combination of values $\delta(P, XY)$, $\delta(P, YZ)$ and $\delta(P, ZX)$ differs (with each distance ranging from 0 to 1). This means that these distances can in general represent three variables, e.g., x, y and z, provided these variables range from 0 to r, where $r > 0$ is a constant, and they satisfy $x + y + z = r$. This is because, given values x_0, y_0 and z_0, satisfying $x_0 + y_0 + z_0 = r$, it is always possible to find a point P inside XYZ, such that the distances $\delta(P, XY)$, $\delta(P, YZ)$ and $\delta(P, ZX)$ are equal to x_0/r, y_0/r and z_0/r, respectively, and thus proportional to the values x_0, y_0 and z_0.

Conversely, the position of each point inside the triangle represents three values, known as the barycentric coordinates of this point. In particular,

- $x_0 + y_0 = 0$. In this case $z_0 = r$. This corresponds to a point that is situated in the vertex Z of the triangle.
- $x_0 = 0$. In this case $y_0 + z_0 = r$. This corresponds to a point that is situated at the edge YZ of the tetrahedron.

Otherwise, i.e., when $x_0 > 0$, $y_0 > 0$ and $z_0 > 0$, the corresponding point is situated strictly inside the triangle. In particular, if $x_0 = y_0 = z_0$, the point occupies the centre of the shape, otherwise it is not at the centre. For example, if $x_0 > y_0$, then the point is closer to the vertex X than to the vertex Y.

In particular, for any real $r > 0$,

- the point X has coordinates $[r, 0, 0]$,
- the point Y has coordinates $[0, r, 0]$,
- the point Z has coordinates $[0, 0, r]$.

Now, consider points P_0, P_1 and P_2, visualized in barycentric coordinates in Fig. 1:

- point P_0, located in the middle of the triangle, has coordinates $[r/3, r/3, r/3]$,
- point P_1, located in the middle of the YZ edge, has coordinates $[0, r/2, r/2]$,
- point P_2, located towards the vertex X, has coordinates $[r/2, r/4, r/4]$,

Notice the reduction of the dimensionality (degrees of freedom) in the underlying data: while each of the depicted points has three original (barycentric) coordinates, it has only two planar ones, which allows representing it on a two-dimensional plane. The reduction is implied by the fact that each set $[x_0, y_0, z_0]$ of the original coordinates satisfies $x_0 + y_0 + z_0 = r$ (every such a constraint reduces the number of degrees of freedom by one).

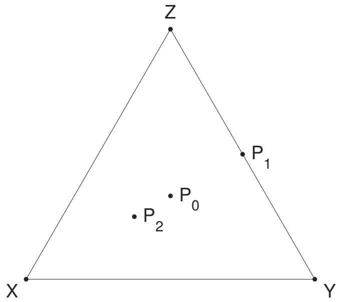

Fig. 1. P_0, P_1 and P_2 in a barycentric coordinate system.

The barycentric coordinates may also be applied to four variables, resulting in a three-dimensional representation, in which case the equilateral triangle must be replaced with a regular tetrahedron (further shortly referred to as a 'tetrahedron'). The tetrahedron $ABCD$ consists of four vertices A, B, C and D, six edges AB, AC, AD, BC, BD and CD, and four faces ABC, BCD, CDA and DAB. A sample tetrahedron is depicted in Fig. 2.

The corresponding barycentric interpretations of exemplary cases are as follows:

- $a+b+c=0$. 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 at the vertex D of the tetrahedron.
- $a+b=0$. 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 at the edge CD of the tetrahedron.
- $a=0$. 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.

Otherwise, i.e., when $n_1 > 0$, $n_2 > 0$, $n_3 > 0$ and $n_4 > 0$, 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 the centre of the shape, otherwise it is not in the centre. In general, if $n_i > n_j$ (with $i \neq j$), 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 vertex A than to vertex B).

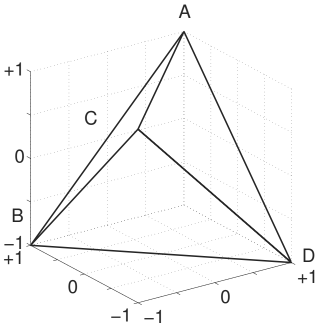

Fig. 2. Skeleton visualizations of the tetrahedron.

The particular, three-dimensional (3D) view of the tetrahedron, as used throughout the paper (and referred to as the standard view), is constructed as follows. Let $\begin{bmatrix}x\\y\\z\end{bmatrix}$ be a vector representing a point in 3D space. In this space, imagine a $2 \times 2 \times 2$ cube of the following coordinates:

$$\begin{bmatrix}1\\1\\1\end{bmatrix},\quad \begin{bmatrix}1\\-1\\1\end{bmatrix},\quad \begin{bmatrix}-1\\-1\\1\end{bmatrix},\quad \begin{bmatrix}-1\\1\\1\end{bmatrix},$$
$$\begin{bmatrix}1\\1\\-1\end{bmatrix},\quad \begin{bmatrix}1\\-1\\-1\end{bmatrix},\quad \begin{bmatrix}-1\\-1\\-1\end{bmatrix},\quad \begin{bmatrix}-1\\1\\-1\end{bmatrix}.$$

The tetrahedron in question, with its four vertices A, B, C and D, is inscribed into this cube so that

$$A: \begin{bmatrix}1\\1\\1\end{bmatrix},\quad B: \begin{bmatrix}-1\\1\\-1\end{bmatrix},\quad C: \begin{bmatrix}-1\\-1\\1\end{bmatrix},\quad D: \begin{bmatrix}1\\-1\\-1\end{bmatrix}.$$

The combination of the viewing angles (azimuth, elevation) in the standard view is $(-35\,^\circ, 22\,^\circ)$. The standard view is accompanied by the so-called rotated view, angles $(145\,^\circ, 22\,^\circ)$, designed to depict the DAB face of the tetrahedron (not visible in the standard view). The combination of these two views will be collectively referred to as the 2-view 3D visualization of the tetrahedron.

The described procedure makes it possible to visualize the four-dimensional data set in three dimensions. This leaves room for an additional variable, which may be represented as colour. It is thus possible to visualize a function $f(a,b,c,d)$, further referred to as the operational function (e.g., an interestingness measure). In the following, it is additionally assumed that the value set of this function is a real interval $[r,s]$, with $r < s$, so that its values may be rendered using a pre-defined colour map (see, e.g., Healey (1996) and Ware (2004) for a discussion of colour maps and data visualization in general).

The actual colour map[1] (see Fig. 3) used in the following visualizations ranges from black (corresponding to r), through grey, up to white (corresponding to s). The number of shades in the map is limited to 16 only in order to emphasize the value changes. Non-numeric values, i.e., $+\infty$, NaN and $-\infty$, if generated by a particular function, may be depicted as special characters. In this paper the only occurring undefined values are NaN's. They are consistently depicted as '*'; however, to avoid massive and thus incomprehensible overlapping, these characters are interspaced with blanks.

A sample 2-view 3D visualization for the function $f(a,b,c,d) = ad - bc$, with $a+b+c+d = n = 64$, is shown in Fig. 4. In this particular case $r = -1024$, $s = 1024$ and there are no undefined values.

An alternative to the 3D visualization of the tetrahedron is a two-dimensional (2D) view of the net (i.e., a set of planar triangles, which, when folded along selected edges, become the faces) of the tetrahedron. A sample 'parallelogram' visualization of the net of the tetrahedron is presented in Fig. 5.

Notice that both the 3D visualization (Fig. 4) as well as the 2D 'parallelogram' visualization (Fig. 5) of a 'solid' tetrahedron show only extreme values of the arguments of the visualized function. If areas located strictly inside

[1] Grey scale is used throughout this paper. For colour renderings of the figures, see `http://www.cs.put.poznan.pl/iszczech/publications/amcs-2014.pdf`.

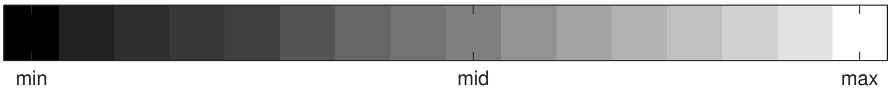

Fig. 3. Colour map for the defined values of the operational function.

the tetrahedron have to be additionally visualized, various variants of the visualization may be generated, see Fig. 6.

Summarizing, the capabilities of the visualization techniques include

- regular views of any operational function,
- specialized views of a region of interest, i.e., only points satisfying some pre-defined conditions, e.g., $f(a,b,c,d)=0$, of any operational function,
- specialized views of multiple operational functions:
  - differences between two functions,
  - standard deviations/variances/means of a number of functions.

3. Application of the visualization techniques

In this paper the application of the visualization techniques is performed on a particular group of interestingness measures called Bayesian confirmation measures. They 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 (Fitelson, 2001). More formally, for a rule $E \to H$, an interestingness measure $c(H,E)$ has the property of Bayesian confirmation (i.e., it is a confirmation measure) when it satisfies the following conditions (1):

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

The same conditions may be equivalently formulated in terms of the a, b, c and d:

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

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).

The research on using confirmation measures for evaluation of rules shows that those measures play an important part in identification of the most interesting rules (Greco *et al.*, 2004; Pawlak, 2002; 2004). Let us stress that the list of alternative, ordinally non-equivalent measures of confirmation is quite large (Crupi *et al.*, 2007; Fitelson, 1999) due to the fact that the property of Bayesian confirmation does not favour any single measure as the most adequate. Definitions of 12 commonly used confirmation measures are listed in Table 2.

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, the measure $C(H,E)$ originally gets 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 the brevity and clarity of presentation, 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 omit the situation of neutrality, in which the measures default to 0. Finally, the parametrized measures $c_1(H,E)$ and $c_2(H,E)$ are computed for the values of $\alpha=\beta=1/2$.

3.1. Regular views of confirmation measures.

Taking particular confirmation measures as operational functions, the regular views of the measures may be used to practically compare their general configurations of values and gradient profiles. Consider 2D 'parallelogram' visualizations for measures $C(H,E)$ and $F(H,E)$ in Figs. 7 and 8, which depict the external areas of the tetrahedron[2]. Such visualizations allow us to instantly notice fundamental differences in the measures, e.g., between their gradient profiles. Observe that while $C(H,E)$ manifests a 'concentric' gradient in pairs of faces, the measure $F(H,E)$ is characterized by constant values (and thus no gradient) in two faces and a 'radial' gradient in the other two. Such a regular view based analysis allows us to tentatively conclude about the ordinal equivalence of the visualized measures, an especially important issue for multi-criteria evaluation of the rules. In the case of $C(H,E)$ and $F(H,E)$, the different gradient profiles in the external areas of the corresponding tetrahedrons constitute conclusive counterexamples to the ordinal equivalence of those measures. In general, however, this kind of equivalence analysis may require an insight into the tetrahedra.

[2] Notice that the '*' characters are interspaced with blanks only to increase the figure's comprehensibility, and in fact the edges AB, BD and CD in $F(H,E)$ consist entirely of NaN's.

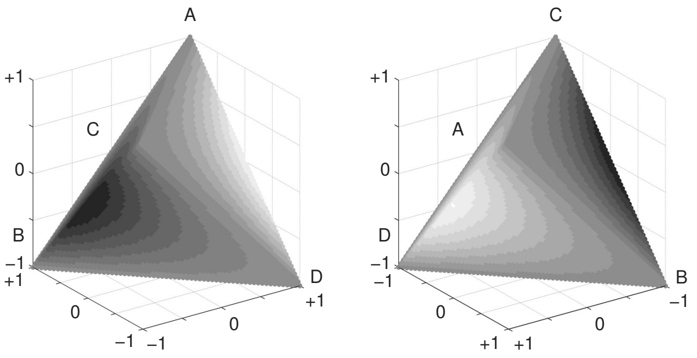

Fig. 4. 2-view 3D visualization of $f(a,b,c,d) = ad - bc$.

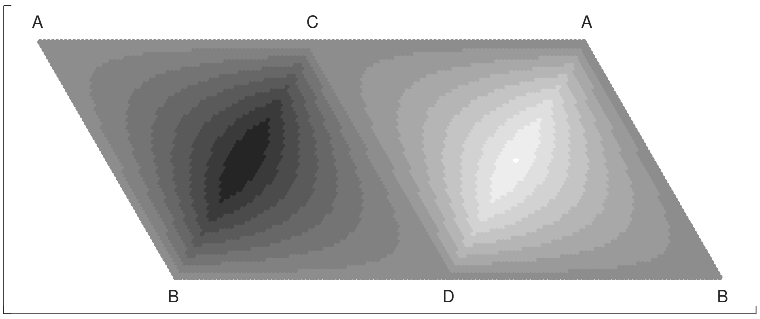

Fig. 5. 2D 'parallelogram' visualization of $f(a,b,c,d) = ad - bc$.

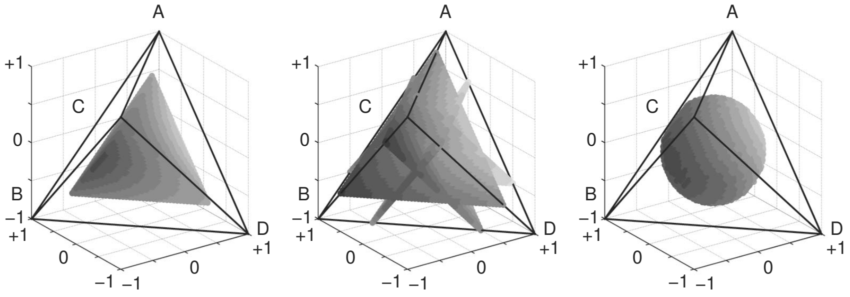

Fig. 6. Three specialized visualizations of the inside of the tetrahedron for $f(a,b,c,d) = ad - bc$.

Table 2. Popular confirmation measures.

$D(H,E) = P(H\mid E) - P(H) = \frac{a}{a+c} - \frac{a+b}{n} = \frac{ad-bc}{n(a+c)}$	Eells, 1982
$M(H,E) = P(E\mid H) - P(E) = \frac{a}{a+b} - \frac{a+c}{n} = \frac{ad-bc}{n(a+b)}$	Mortimer, 1988
$S(H,E) = P(H\mid E) - P(H\mid \neg E) = \frac{a}{a+c} - \frac{b}{b+d} = \frac{ad-bc}{(a+c)(b+d)}$	Christensen, 1999
$N(H,E) = P(E\mid H) - P(E\mid \neg H) = \frac{a}{a+b} - \frac{c}{c+d} = \frac{ad-bc}{(a+b)(c+d)}$	Nozick, 1981
$C(H,E) = P(E \wedge H) - P(E)P(H) = \frac{a}{n} - \frac{(a+c)(a+b)}{n^2} = \frac{ad-bc}{n^2}$	Carnap, 1962
$F(H,E) = \frac{P(E\mid H) - P(E\mid \neg H)}{P(E\mid H) + P(E\mid \neg H)} = \frac{\frac{a}{a+b} - \frac{c}{c+d}}{\frac{a}{a+b} + \frac{c}{c+d}} = \frac{ad-bc}{ad+bc+2ac}$	Kemeny and Oppenheim, 1952
$Z(H,E) = \begin{cases} 1 - \frac{P(\neg H\mid E)}{P(\neg H)} = \frac{ad-bc}{(a+c)(c+d)} & \text{in the case of confirmation} \\ \frac{P(H\mid E)}{P(H)} - 1 = \frac{ad-bc}{(a+c)(a+b)} & \text{in the case of disconfirmation} \end{cases}$	Crupi *et al.*, 2007
$A(H,E) = \begin{cases} \frac{P(E\mid H) - P(E)}{1 - P(E)} = \frac{ad-bc}{(a+b)(b+d)} & \text{in the case of confirmation} \\ \frac{P(H) - P(H\mid \neg E)}{1 - P(H)} = \frac{ad-bc}{(b+d)(c+d)} & \text{in the case of disconfirmation} \end{cases}$	Greco *et al.*, 2012
$c_1(H,E) = \begin{cases} \alpha + \beta A(H,E) & \text{in the case of confirmation when } c = 0 \\ \alpha Z(H,E) & \text{in the case of confirmation when } c > 0 \\ \alpha Z(H,E) & \text{in the case of disconfirmation when } a > 0 \\ -\alpha + \beta A(H,E) & \text{in the case of disconfirmation when } a = 0 \end{cases}$	Greco *et al.*, 2012
$c_2(H,E) = \begin{cases} \alpha + \beta Z(H,E) & \text{in the case of confirmation when } b = 0 \\ \alpha A(H,E) & \text{in the case of confirmation when } b > 0 \\ \alpha A(H,E) & \text{in the case of disconfirmation when } d > 0 \\ -\alpha + \beta Z(H,E) & \text{in the case of disconfirmation when } d = 0 \end{cases}$	Greco *et al.*, 2012
$c_3(H,E) = \begin{cases} A(H,E)Z(H,E) & \text{in the case of confirmation} \\ -A(H,E)Z(H,E) & \text{in the case of disconfirmation} \end{cases}$	Greco *et al.*, 2012
$c_4(H,E) = \begin{cases} \min(A(H,E), Z(H,E)) & \text{in the case of confirmation} \\ \max(A(H,E), Z(H,E)) & \text{in the case of disconfirmation} \end{cases}$	Greco *et al.*, 2012

3.2. Specialized views of regions of interest. In their analyses of the confirmation measures, users may be interested in discovering regions for which the considered measures satisfy some pre-defined conditions, e.g., $c(H,E) = 0$ (the neutral value) or $c(H,E) = +1$ (the maximal value). Supporting the user with such specialized views is important for at least two reasons: it allows testing for the existence of such regions and identifying the localizations of these regions within the tetrahedron, translating them uniquely to particular domain values of a, b, c and d.

An important characteristic region for confirmation measures is the neutrality zone. By the definition of the property of confirmation, all such measures assume zero values for the same arguments. Thus, the neutrality region is common for all confirmation measures. Figure 9 depicts this region, rendered using a non-standard colour map, which lends the view the necessary perspective, but whose colours do not translate to the constant values of the measure. Its characteristic saddle-like shape divides the tetrahedron into two subregions of positive, i.e., confirmatory values (around edge AD), and negative, i.e., disconfirmatory values (around edge BC).

Figures 10 and 11 depict other sample regions of interest, for which $|C(H,E)| = 0.5$ and $|c_1(H,E)| = 0.5$. Again, the grey colour map is used only to provide the necessary perspective and the colours do not translate to values of the measures (constant also in this case). Notice

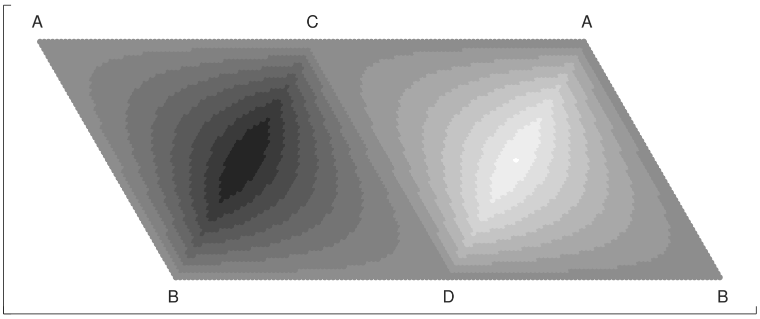

Fig. 7. 2D ‘parallelogram’ visualization of $C(H, E)$.

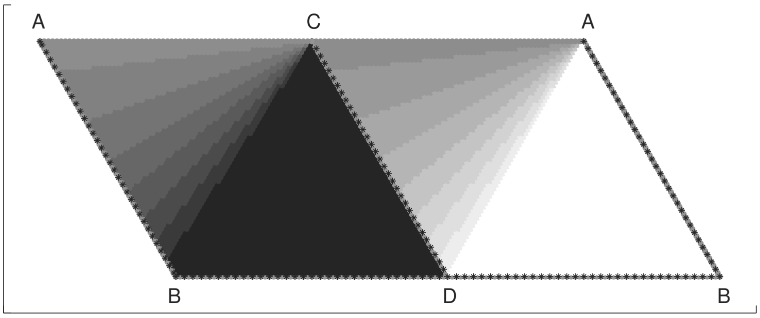

Fig. 8. 2D ‘parallelogram’ visualization of $F(H, E)$.

the full symmetry of the regions manifested by $C(H, E)$ (coinciding with edges BC and AD, while avoiding the other ones). Simultaneously, the regions for $c_1(H, E)$ approach the edges BC, BD and CD, while avoiding the other edges.

Other sample regions of interest are presented in Fig. 12. It depicts both extreme (-1 and $+1$) and non-numeric (NaN) values of the measure $S(H, E)$. Its analysis reveals that the non-numeric values occur in two disjoint localizations (i.e., at edges BD and AC) in the tetrahedron. A similar remark concerns the extreme values (-1 at edge BC, $+1$ at edge AD).

3.3. Specialized views of differences between measures.

Since the set of available measures may be considerable, the practitioners are forced to choose only subsets of measures for their particular applications. To guide them towards the most suitable solutions, our visualization techniques provide specialized views, allowing, among other things, identification of those arguments (i.e., values of a, b, c and d) for which two given measures differ only insignificantly (similarity of the measures) or differ considerably (dissimilarity of the measures). Notice that in the case of confirmation measures the difference of any two such measures belongs to $(-1, +1)$, thus the colour map in Fig. 3 is to be interpreted as for regular views of the measures.

Consider $c_3(H, E)$ and $c_4(H, E)$, commonly defined using $Z(H, E)$ and $A(H, E)$. Figures 13 (exterior) and 14 (interior) show the difference $c_3(H, E) - c_4(H, E)$. A visual analysis instantly reveals that $c_3(H, E)$ and $c_4(H, E)$ manifest both similarities and dissimilarities. While being identical in all the faces of the tetrahedron (Fig. 13), they differ inside the shape (Fig. 14), with $c_4(H, E)$ exceeding $c_3(H, E)$ around the edge AD and $c_3(H, E)$ exceeding $c_4(H, E)$ around the edge BC.

A slightly different situation concerns measures $c_1(H, E)$ and $c_2(H, E)$. Figures 15 (exterior) and 16 (interior) show the difference $c_1(H, E) - c_2(H, E)$. Observe that $c_1(H, E)$ exceeds $c_2(H, E)$ by a constant (the difference being $+1/2$) in faces ABC and ABD, while $c_1(H, E)$ is exceeded by $c_2(H, E)$ by a constant (the difference being $-1/2$) in faces BCD and ACD. Inside the shape this dependency is to some degree preserved, although the differences are no longer constant.

3.4. Specialized views of standard deviations among measures.

When a group of more than two measures is involved in the comparison, calculation of the difference could be substituted with, e.g., the variance or standard deviation (i.e., square root of the variance). This identifies the regions of the tetrahedron (and thus values of a, b, c and d), where the measures of the group vary the least (low standard deviation or variance) or the most

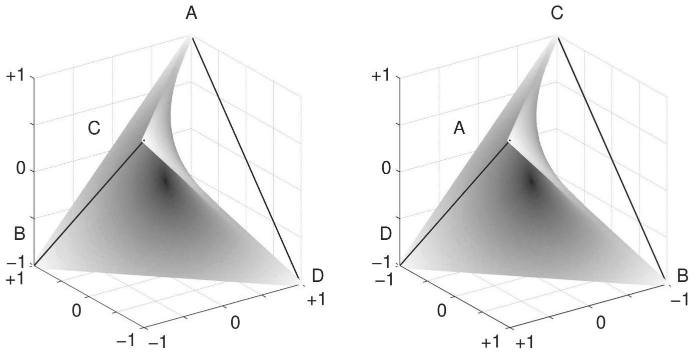

Fig. 9. 2-view 3D visualization of the neutral regions (common for all confirmation measures).

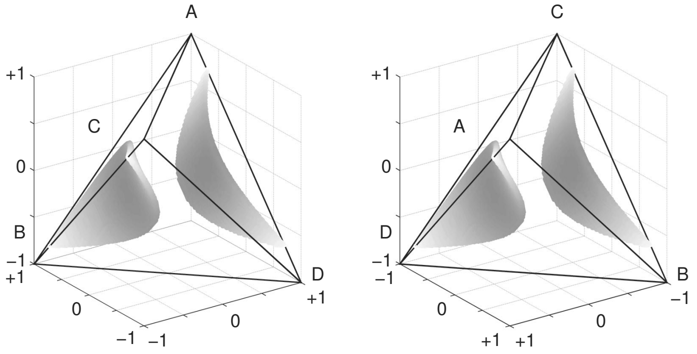

Fig. 10. 2-view 3D visualization of $|C(H,E)| = 0.5$ regions.

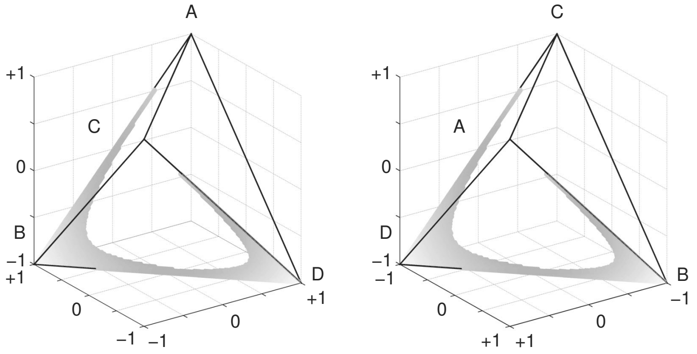

Fig. 11. 2-view 3D visualization of $|c_1(H,E)| = 0.5$ regions.

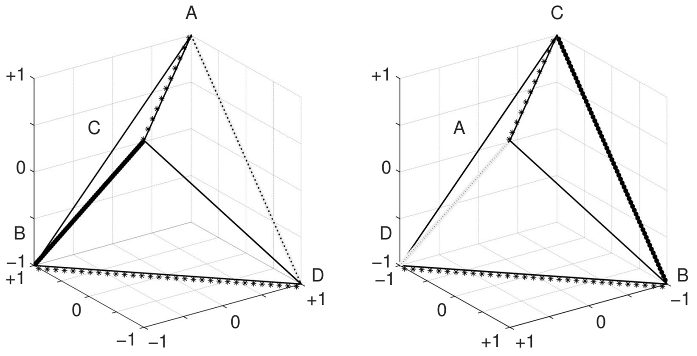

Fig. 12. 2-view 3D visualization showing regions with extreme or non-numeric values of $S(H,E)$.

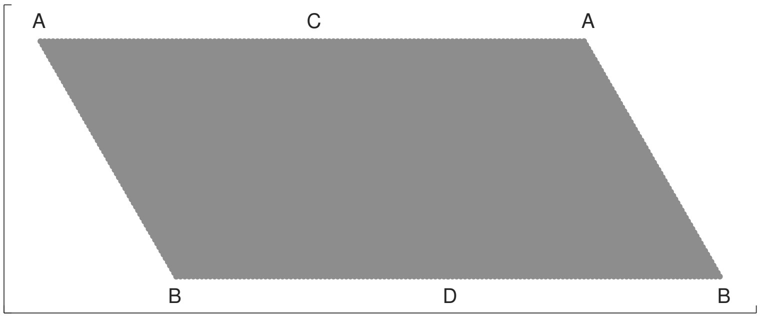

Fig. 13. 2D ‘parallelogram’ visualization of $c_3(H,E) - c_4(H,E)$.

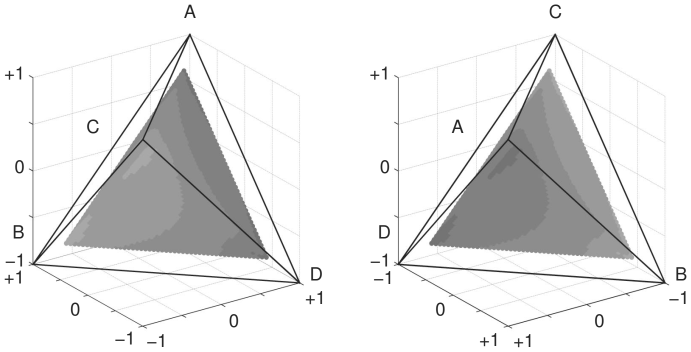

Fig. 14. 2-view 3D interior visualization of $c_3(H,E) - c_4(H,E)$.

(high standard deviation or variance). The practitioners could then decide on using only one representative of a low-variant group (as such measures tend to produce fairly consistent evaluations), whereas highly variant groups of measures may require more representatives.

A popular division of confirmation measures establishes two groups of measures: those inspired by the Bayesian and those inspired by the likelihoodist point of view. Among the 12 selected confirmation measures considered in this paper, $D(H,E)$, $S(H,E)$, $Z(H,E)$ and $c_1(H,E)$ belong to the first group, while $M(H,E)$, $N(H,E)$, $A(H,E)$ and $c_2(H,E)$ to the second one. Submitting any of these groups to such a type of analysis reveals how consistent their evaluations are and for which arguments the greatest differences among those evaluations occur.

Figures 17 (exterior) and 18 (interior) show the standard deviation of the Bayesian, while Figs. 19 (exterior) and 20 (interior) show the standard deviation of the likelihoodist measures. The values within the analysed groups of measures belong to $[0, max]$ (where $max \approx 0.5$), and thus the colour map in Fig. 3 should be interpreted accordingly. Notice that the standard deviations are generally higher in the faces of the tetrahedron, with different faces manifesting different profiles of the standard deviation. As far as the Bayesian measures are concerned, the deviation is maximal in faces BCD and ABD, with the marked increase towards the edge CD (face BCD) and towards the edge AB (face ABD). In the case of the likelihoodist measures, the deviation is maximal in faces ABC and ACD, with the marked increase towards edges AB and AC (face ABC) and towards the edge AC (face ACD).

4. Conclusions

The paper presents visualization techniques for interestingness measures, which provide practical insights into different details of the analysed measures. The originally four-dimensional arguments of the measures are effectively represented in three dimensions using a tetrahedron-based barycentric coordinate system, with values of any operational function, e.g., an interestingness measure, rendered as colour.

The visual analyses allow us to instantly detect and locate interesting characteristics of the measures, which would otherwise have to be laboriously derived from the analytic definitions. The presented techniques are principally capable of visualizing single interestingness measures, regions of interest, i.e., only arguments satisfying some pre-defined conditions, differences between pairs of measures or standard

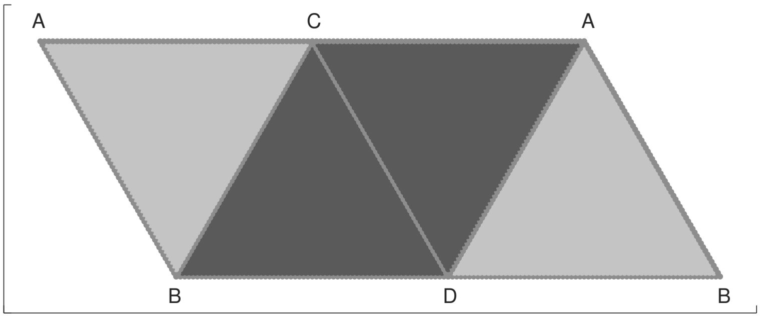

Fig. 15. 2D ‘parallelogram’ visualization of $c_1(H, E) - c_2(H, E)$.

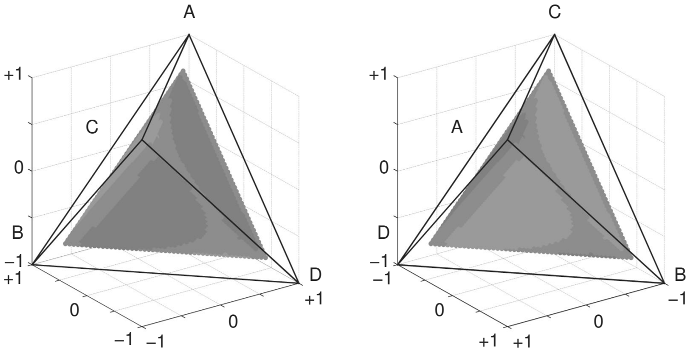

Fig. 16. 2-view 3D interior visualization of $c_1(H, E) - c_2(H, E)$.

deviations/variances/means of sets of measures. They might be thus in particular used to demonstrate how newly designed measures differ from the existing ones. Examples of applications of the visualization techniques are presented and discussed in detail for the group of 12 popular confirmation measures.

Further research includes applications of the introduced techniques to other types of measures, e.g., the so-called classifier performance measures, like sensitivity, specificity, F_1-score, etc. (in fact, the approach is potentially applicable to any interval-valued coefficient defined for a 2×2 contingency table). Moreover, it is possible to devise series of experiments designed to compare classifiers that incorporate different (sets of) measures. These experiments could verify if measures with desirable visual features are the most beneficial for classifier performance. In particular, they could illustrate the benefits of using single representatives of groups of measures that produce fairly similar evaluations of rules (such similarities can be easily detected by the visual analysis of the group’s standard deviation).

To sum up, enriching our knowledge on the features and the behaviour of the visualized measures, the approach helps to determine, e.g., if the visualized measures are identical or similar in particular domain regions, or if they are ordinally equivalent. The gained insights can swiftly guide the practitioners towards interestingness measures that best reflect their expectations.

Acknowledgment

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

References

Agrawal, R., Imielinski, T. and Swami, A. (1993). Mining associations between sets of items in massive databases, *Proceedings of the 1993 ACM-SIGMOD International Conference on Management of Data, Washington, DC, USA*, pp. 207–216.

Alaíz-Rodríguez, R., Japkowicz, N. and Tischer, P.E. (2008). Visualizing classifier performance on different domains, *ICTAI 2008, Dayton, OH, USA*, pp. 3–10.

Carnap, R. (1962). *Logical Foundations of Probability*, 2nd Edn., University of Chicago Press, Chicago, IL.

Christensen, D. (1999). Measuring confirmation, *Journal of Philosophy* **96**(9): 437–461.

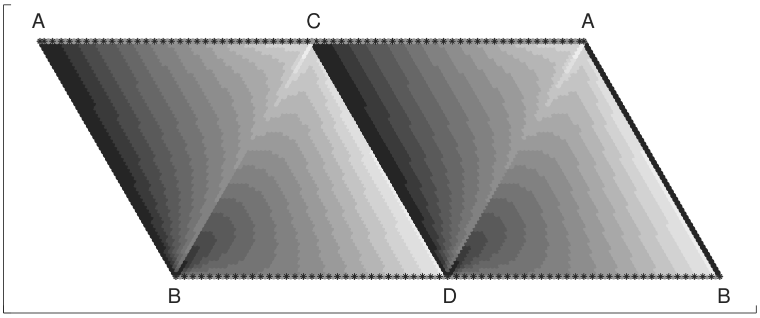

Fig. 17. 2D ‘parallelogram’ visualization of the standard deviation of the Bayesian measures.

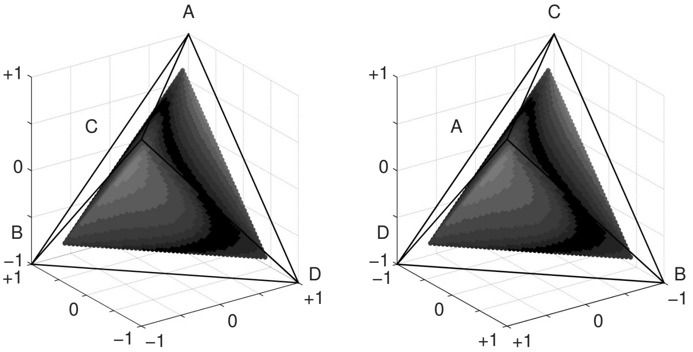

Fig. 18. 2-view 3D interior visualization of the standard deviation of the Bayesian measures.

Crupi, V., Tentori, K. and Gonzalez, M. (2007). On Bayesian measures of evidential support: Theoretical and empirical issues, *Philosophy of Science* **74**(2): 229–252.

Drummond, C. and Holte, R.C. (2006). Cost curves: An improved method for visualizing classifier performance, *Machine Learning* **65**(1): 95–130.

Eells, E. (1982). *Rational Decision and Causality*, Cambridge University Press, Cambridge.

Everson, R.M. and Fieldsend, J.E. (2006). Multi-class ROC analysis from a multi-objective optimisation perspective, *Pattern Recognition Letters* **27**(8): 918–927.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity, *Philosophy of Science* **66**: 362–378.

Fitelson, B. (2001). *Studies in Bayesian Confirmation Theory*, Ph.D. thesis, University of Wisconsin, Madison, WI.

Floater, M.S., Hormann, K. and Kos, G. (2006). A general construction of barycentric coordinates over convex polygons, *Advances in Computational Mathematics* **24**(1–4): 311–331.

Fukuda, T., Morimoto, Y., Morishita, S. and Tokuyama, T. (1996). Data mining using two-dimensional optimized association rules: Scheme, algorithms, and visualization, *Proceedings of the 1996 ACM-SIGMOD International Conference on the Management of Data, Montreal, Quebec, Canada,* pp. 13–23.

Geng, L. and Hamilton, H. (2006). Interestingness measures for data mining: A survey, *ACM Computing Surveys* **38**(3), Article no. 9.

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

Greco, S., Słowiński, R. and Szczęch, I. (2012). Properties of rule interestingness measures and alternative approaches to normalization of measures, *Information Sciences* **216**: 1–16.

Healey, C. (1996). Choosing effective colors for data visualization, *Proceedings of the 7th Conference on Visualization, VIS’96, San Francisco, CA, USA*, pp. 263–270.

Hernández-Orallo, J., Flach, P.A. and Ramirez, C.F. (2011). Brier curves: A new cost-based visualisation of classifier performance, *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, WA, USA*, pp. 585–592.

IBM (1996). Intelligent miner user guide, Version 1, Release 1, *Technical report*, International Business Machines, San Jose, CA.

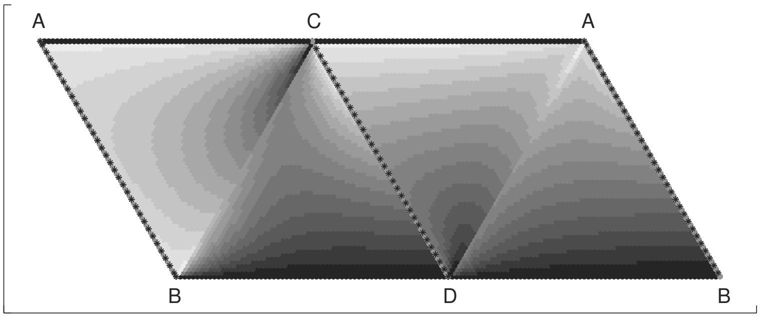

Fig. 19. 2D ‘parallelogram’ visualization of the standard deviation of the likelihoodist measures.

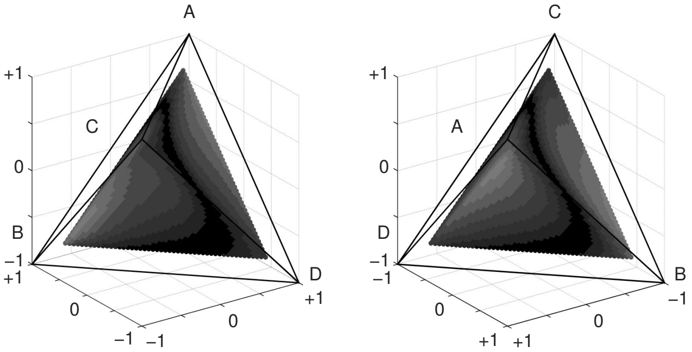

Fig. 20. 2-view 3D interior visualization of the standard deviation of the likelihoodist measures.

Kemeny, J. and Oppenheim, P. (1952). Degrees of factual support, *Philosophy of Science* **19**(4): 307–324.

Mortimer, H. (1988). *The Logic of Induction*, Prentice Hall, Paramus, NJ.

Morzy, T. and Zakrzewicz, M. (2003). Data mining, *in* J. Blazewicz, W. Kubiak, T. Morzy and M.E. Rusinkiewicz (Eds.), *Handbook on Data Management Information Systems*, Springer, Heidelberg, pp. 487–565.

Nozick, R. (1981). *Philosophical Explanations*, Clarendon Press, Oxford.

Pawlak, Z. (2002). Rough sets, decision algorithms and Bayes’ theorem, *European Journal of Operational Research* **136**(1): 181–189.

Pawlak, Z. (2004). Some issues on rough sets, *Transactions on Rough Sets I*, Elsevier Science Publishers, New York, NY, pp. 1–58.

Shaikh, M., McNicholas, P.D., Antonie, M.L. and Murphy, T.B. (2013). Standardizing interestingness measures for association rules, *Computing Research Repository,* `http://arxiv.org/abs/1308.3740`.

Susmaga, R. and Szczęch, I. (2013). Visualization of interestingness measures, *Proceedings of the 6th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, Poznań, Poland*, pp. 95–99.

Susmaga, R. and Szczęch, I. (2014). Visual-based detection of properties of confirmation measures, *in* T. Andreasen, H. Christiansen, J.C.C. Talavera and Z.W. Ras (Eds.), *Proceedings of the 21st International Symposium on Methodologies for Intelligent Systems, ISMIS 2014*, Lecture Notes in Computer Science, Vol. 8502, Springer, Heidelberg, pp. 133–143.

Tan, P., Kumar, V. and Srivastava, J. (2002). Selecting the right interestingness measure for association patterns, *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada*, pp. 32–41.

Ware, C. (2004). *Information Visualization: Perception for Design, 2nd Edition*, Morgan Kaufmann, Waltham, MA.

Warren, J. (2003). On the uniqueness of barycentric coordinates, *in* R. Goldman and R. Krasauskas (Eds.), *Topics in Algebraic Geometry and Geometric Modeling*, Contemporary Mathematics, Vol. 334, American Mathematical Society, Providence, RI, USA, pp. 93–99.

Zhou, Y., Wischgoll, T., Blaha, L.M., Smith, R. and Vickery, R.J. (2014). Visualizing confusion matrices for multidimensional signal detection correlational methods, *Proceedings of the SPIE Conference on Visualization and Data Analysis, San Francisco, CA*.

Robert Susmaga received his M.Sc. and Ph.D. degrees in computing science from the Poznań University of Technology, Poland, in 1994 and 2001, respectively, where he has been working at the Laboratory of Intelligent Decision Support Systems, Institute of Computing Science, ever since. His research interests focus on various elements of machine learning, knowledge discovery and data mining, and include the analysis and visualization of multidimensional data.

Izabela Szczęch received her B.Eng., M.Sc. and Ph.D. degrees in computing science from the Poznań University of Technology, Poland, in 2002, 2004 and 2008, respectively. Currently she is an assistant professor at the same university, at the Laboratory of Intelligent Decision Support Systems, Institute of Computing Science. She works mainly on topics related to data mining, and in particular to measures of rule interestingness and properties of interestingness measures.

Received: 11 March 2014
Revised: 19 July 2014