Boolean factors as a means of clustering of interestingness measures of association rules

Radim Belohlavek · Dhouha Grissa ·
Sylvie Guillaume · Engelbert Mephu Nguifo · Jan Outrata

Abstract Measures of interestingness play a crucial role in association rule mining. An important methodological problem, on which several papers appeared in the literature, is to provide a reasonable classification of the measures. In this paper, we explore Boolean factor analysis, which uses formal concepts corresponding to classes of measures as factors, for the purpose of clustering of the measures. Unlike the existing studies, our method reveals overlapping clusters of interestingness measures. We argue that the overlap between clusters is a desired feature of natural groupings

R. Belohlavek · J. Outrata (✉)
Data Analysis and Modeling Lab, Dept. Computer Science, Palacky University Olomouc,
17. listopadu 12, 771 46 Olomouc, Czech Republic
e-mail: jan.outrata@upol.cz

R. Belohlavek
e-mail: radim.belohlavek@acm.org

D. Grissa · E. Mephu Nguifo
LIMOS, Clermont Université, Université Blaise Pascal, BP 10448,
63000 Clermont-Ferrand, France

D. Grissa
e-mail: dgrissa@isima.fr

E. Mephu Nguifo
e-mail: mephu@isima.fr

S. Guillaume
LIMOS, Clermont Université, Université d'Auvergne, BP 10448,
63000 Clermont-Ferrand, France
e-mail: guillaum@isima.fr

D. Grissa · S. Guillaume · E. Mephu Nguifo
LIMOS, CNRS, UMR 6158, 63173 Aubiére, France

D. Grissa
LIPAH, Département d'Informatique, Faculté des Sciences de Tunis,
Campus Universitaire, 1060 Tunis, Tunisie

of measures and that because formal concepts are used as factors in Boolean factor analysis, the resulting clusters have a clear meaning and are easy to interpret. We conduct three case studies on clustering of measures, provide interpretations of the resulting clusters and compare the results to those of the previous approaches reported in the literature.

1 Introduction

An important problem in extracting association rules, well known since the early stage of association rule mining [1, 29], is the possibly huge number of rules extracted from data. A general way of dealing with this problem is to define the concept of rule interestingness: only association rules that are considered interesting according to some measure are presented to the user. The most widely used measures of interestingness are based on the well-known concept of support and confidence [1]. However, the suitability of these measures to extract interesting rules was challenged by several studies, see [12, 30] for an overview. Consequently, several other interest-ingness measures of association rules were proposed, see e.g. [4, 7, 13, 20, 22, 33, 35].

To understand better the behavior of various measures, several studies of the properties of interestingness measures appeared, see e.g. [13, 14, 22, 23, 28]. Those studies explore various properties of the measures that are considered important and attempt to provide reasonable classifications of the measures. For example, Vaillant et al. [36], as a part of their work on facilitating the choice of user-adapted interestingness measures [24–26], evaluated 20 interestingness measures according to 8 properties. The authors applied hierarchical clustering on the 20×8 measure–property matrix and obtained four clusters of measures. Tan et al. [33] studied 21 interestingness measures using 8 properties. They argued that each measure is useful for some applications but not for others and study when two measures are consistent with each other. By applying a new clustering approach, Huynh et al. [21] classified 34 interestingness measures with a correlation analysis. Geng and Hamilton [13] made a survey of 38 interestingness measures for rules and summaries with 11 properties. They also gave strategies to select the appropriate measures. Feno [10] evaluated 15 interestingness measures with 13 properties to describe their behaviour. Delgado et al. [9] provided a study of interestingness measures by means of so-called logical model. Furthermore, the authors proposed and justified the addition of two new principles for interestingness measures to the three proposed by Piatetsky-Shapiro [29]. Heravi and Zaiane [19] studied 53 objective measures for associative classification rules according to 16 properties and concluded that no single measure can be introduced as an obvious winner for associative classification. Surana et al. [31] addressed the problem of selecting an interestingness measure for mining rare association rules. For this purpose, they analyzed various properties of measures

and suggested a set of properties to be considered for selecting a measure. Through experimental results, they showed that the measures satisfying the suggested properties can efficiently discover rare association rules. Guillaume et al. [14] attempted to significantly extend the number of measures and properties under scrutiny. The resulting measure-property matrix was subject to two studies. Namely, Grissa et al. [15] describes how formal concept analysis may help highlighting interestingness measures with similar behavior to help a user during his choice. More importantly, Guillaume et al. [14] attempted to find natural clusters of measures by widely used clustering methods, the agglomerative hierarchical method and the K-means method. The proposed approach in [14] avoids the four pitfalls mentioned by Suzuki [32] concerning categorization of interestingness measures. For example, the fourth one (search bias) was avoided by testing various values of parameters, and aggregating the results of two different clustering algorithms.

The numerous previous works indicate the importance of categorization of interestingness measures into clusters. For one, the clusters provide us with a structured view on the many existing interestingness measures. Such a view helps to highlight the similarity and possibly other relationships between the measures. For second, the revealed clusters or groups of measures may be useful to help the user select interestingness measures. In particular, a group of measures may be thought of a representing a particular kind of interesting measures. In this view, a collection of clusters represents a collection of particular kinds of measures, i.e. particular points of view regarding the importance of association rules. If a user intends to take into account several such views represented by the clusters, he may opt to select one measure per cluster. This is the case in a recent paper by Bouker et al. [6] who propose to generate association rules using several different interestingness measures by means of aggregating the measure values.

A common feature of the above mentioned approaches to cluster interestingness measures is that the methods used only produce disjoint clusters of measures. However, since the clusters are supposed to correspond to significant classes of measures, one would naturally expect overlapping clusters rather than disjoint ones. Namely, a single measure may belong to a particular cluster because it shares some properties with other measures in that cluster (those characteristic for the cluster) and, at the same time, may belong to a different cluster because it shares some other properties with the measures in this different cluster. The aim of this paper is to explore the possibility of obtaining reasonable overlapping clusters of measures using Boolean factor analysis (BFA) and to compare the results with those of the other studies. In particular, we use the method recently developed in [3], apply it to the various measure-property Boolean matrices describing the measures reported in the literature, and take the discovered factors for (overlapping) clusters. We argue that the clusters obtained this way provide a reasonable classification of measures and that since the method uses formal concepts as factors, the clusters are easy to interpret.

The paper is organized as follows. Section 2 provides preliminaries on Boolean (binary) data, association rules, and Boolean factor analysis. Section 3 provides three case studies on clustering using by means of Boolean factors and a discussion regarding the interpretation of the clusters obtained and a comparison of the results to those reported in the literature. Section 4 concludes the paper and outlines some directions for future research.

2 Preliminaries

2.1 Boolean (binary) data

Let X be a set of objects (such as a set of customers, a set of functions or the like) and Y be a set of attributes (such as a set of products that customers may buy, a set of properties of functions). The information about which objects have which attributes may formally be represented by a binary relation I between X and Y, i.e. $I \subseteq X \times Y$, and may be visualized by a table (matrix) that contains 1s and 0s, according to whether the object corresponding to a row has the attribute corresponding to a column. We denote the entries of such matrix by I_{xy}. A data of this type is called *Boolean data* (or binary data). The triplet $\langle X, Y, I \rangle$ is called a *formal context* in formal concept analysis (FCA) [8, 11, 37] but other terms are used in other areas.

Such type of data appears in two roles in our paper. First, association rules, whose interestingness measures we analyze, are certain dependencies over Boolean data. Second, the information we have about the interestingness measures of association rules is in the form of Boolean data: the objects are interestingness measures and the attributes are their properties. An example is given by Table 2.

2.2 Association rules

An *association rule* [34] over a set Y of attributes is a formula

$$A \Rightarrow B \tag{1}$$

where A and B are sets of attributes from Y, i.e. $A, B \subseteq Y$. Let $\langle X, Y, I \rangle$ be a formal context. A natural measure of interestingness of association rules is based on the notions of confidence and support. The *confidence* and *support* of an association rule $A \Rightarrow B$ in $\langle X, Y, I \rangle$ are defined by

$$\text{conf}(A \Rightarrow B) = \frac{|A^\downarrow \cap B^\downarrow|}{|A^\downarrow|} \quad \text{and} \quad \text{supp}(A \Rightarrow B) = \frac{|A^\downarrow \cap B^\downarrow|}{|X|},$$

where $C^\downarrow$ for $C \subseteq Y$ is defined by $C^\downarrow = \{x \in X \mid \text{for each } y \in C : \langle x, y \rangle \in I\}$, i.e. $C^\downarrow$ is the set of all objects sharing all attributes from C, and $|D|$ denotes the size of a set D. Hence, the confidence of an association rule $A \Rightarrow B$ is the ratio of the number of objects sharing all attributes from both A and B to the number of objects sharing all attributes from A, while the support of the association rule is the ratio of the number of objects sharing all attributes from both A and B to the number of all objects from X.

An association rule is considered interesting if its confidence and support exceed some user-specified thresholds. However, the support-confidence approach suffers from certain weaknesses. Often, this approach as well as algorithms based on it leads to the extraction of an exponential number of rules. In addition, the limitation of the support measure is that sometimes many rules that are potentially interesting have a support lower than the specified threshold and get therefore lost. To address this problem, many other measures of interestingness have been proposed in the literature [12], mainly because they are effective for mining potentially interesting rules and capture various aspects of user interest. Because of lack of space, we do not describe all the measures to which we refer in this paper. Rather, we provide

the reader with references (mainly in Table 1) where the definitions may be found. Note that association rules are attributed to [2]. However, the concept of association rule itself as well as various measures of interestingness are particular cases of what is investigated in depth in [17], a book that develops logico-statistical foundations of the GUHA method [18].

2.3 Boolean factor analysis (BFA)

Let I be an $n \times m$ Boolean (binary) matrix. The aim in Boolean factor analysis (BFA), also refered to as factor analysis of (Boolean) binary data, is to find a decomposition

$$I = A \circ B \tag{2}$$

of I into an $n \times k$ Boolean matrix A and a $k \times m$ Boolean matrix B with $\circ$ denoting the Boolean product of matrices, i.e.

$$(A \circ B)_{ij} = \max_{l=1}^{k} \min(A_{il}, B_{lj}).$$

Table 1 Interestingness measures properties

No.	Property (basic description)	Ref.
P_1	Intelligibility of measure.	[26]
P_2	Easiness to fix a threshold to the rule.	[23]
P_3	Asymmetric role of A and B.	[23, 33]
P_4	Asymmetric measure in the sense of the negation of the consequent.	[23, 33]
P_5	Evaluates $A \Rightarrow B$ and $\bar{B} \Rightarrow \bar{A}$ similarly in the logical implication case.	[23]
P_6	Increasing in the number of examples or decreasing in the number of counter-examples.	[23, 29]
P_7	Increasing in data size.	[13, 33]
P_8	Decreasing in the size of antecedent and consequent.	[23, 29]
P_9	Fixed value in the independence case.	[23, 29]
P_{10}	Fixed value in the logical implication case.	[23]
P_{11}	Fixed value in the equilibrium case.	[5]
P_{12}	Particular values in the attraction case between A and B.	[29]
P_{13}	Particular values in the repulsion case between A and B.	[29]
P_{14}	Tolerance to a few counter-examples.	[16, 27]
P_{15}	Invariance in case of expansion of certain quantities.	[33]
P_{16}	Desired relationship between $A \Rightarrow B$ and $\bar{A} \Rightarrow B$ rules, namely $m(A \Rightarrow B) = -m(\bar{A} \Rightarrow B)$ for measure m.	[33]
P_{17}	Desired relationship between $A \Rightarrow B$ and $A \Rightarrow \bar{B}$ rules, namely $m(A \Rightarrow B) = -m(A \Rightarrow \bar{B})$ for measure m.	[33]
P_{18}	Desired relationship between $A \Rightarrow B$ and $\bar{A} \Rightarrow \bar{B}$ rules, namely $m(A \Rightarrow B) = m(\bar{A} \Rightarrow \bar{B})$ for measure m.	[33]
P_{19}	Antecedent size is fixed (if measure m is not based on a certain probabilistic model) or random (if m is based on a certain probabilistic model).	[23]
P_{20}	Descriptive (invariant with n) or statistical (increases with n) measure.	[23]
P_{21}	Discriminant measure (as n grows, measure is less able to distinguish the interesting rules).	[23]

The inner dimension, k, in the decomposition may be interpreted as the number of factors that may be used to describe the original data. Namely, $A_{il} = 1$ if and only if the lth factor applies to the ith object and $B_{lj} = 1$ if and only if the jth attribute is one of the manifestations of the lth factor. The factor model behind (2) has therefore the following meaning: the object i has the attribute j if and only if there exists a factor l that applies to i and for which j is one of the particular manifestations of factor l. Note that while the general aim is the same as in classical factor analysis, the nature of data and the algebra behind make BFA substantially different from classical factor analysis. We refer to [3] for further information and references to papers that deal with the problem of factor analysis and decompositions of Boolean matrices.

In [3], the following method for finding decompositions (2) with the number k of factors as small as possible has been presented. The method utilizes formal concepts of the formal context $\langle X, Y, I \rangle$ as factors, where $X = \{1, \ldots, n\}$, $Y = \{1, \ldots, m\}$ (objects and attributes correspond to the rows and columns of I). Recall that a formal concept of $\langle X, Y, I \rangle$ is any pair $\langle C, D \rangle$ of sets $C \subseteq X$ and $D \subseteq Y$ such that D is just the set of all attributes shared by all objects in C and, conversely, C is just the set of all objects that have all the attributes in D. Note also that the set of all formal concepts of $\langle X, Y, I \rangle$ equipped with a natural subconcept-superconcept hierarchy is called the concept lattice of $\langle X, Y, I \rangle$ and is denoted by $\mathcal{B}(X, Y, I)$ [11]. Let

$$\mathcal{F} = \{\langle C_1, D_1 \rangle, \ldots, \langle C_k, D_k \rangle\} \tag{3}$$

be a set of formal concepts of $\langle X, Y, I \rangle$, i.e. $\langle C_l, D_l \rangle$ are elements of the concept lattice $\mathcal{B}(X, Y, I)$. Consider the $n \times k$ Boolean matrix $A_{\mathcal{F}}$ and a $k \times m$ Boolean matrix $B_{\mathcal{F}}$ defined by

$$(A_{\mathcal{F}})_{il} = 1 \text{ iff } i \in C_l \quad \text{and} \quad (B_{\mathcal{F}})_{lj} = 1 \text{ iff } j \in D_l. \tag{4}$$

Denote by $\rho(I)$ the smallest number k, so-called Schein rank of I, such that a decomposition of I exists with k factors. The following theorem [3] shows that using formal concepts as factors as in (4) is optimal in that it enables us to reach the Schein rank.

Theorem 1 *For every Boolean matrix I, there exists $\mathcal{F} \subseteq \mathcal{B}(X, Y, I)$ such that $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$ and $|\mathcal{F}| = \rho(I)$.*

Example 1 As an illustration, consider the 4×5 Boolean matrix I depicted below. Using the described method one can find a decomposition of the matrix into 4×3 and 3×5 Boolean matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ with the number of factors k (and also the Schein rank $\rho(I)$) equal to 3:

$$
\begin{array}{ccccc}
I & = & A_{\mathcal{F}} & \circ & B_{\mathcal{F}} \\
\begin{pmatrix} 1 1 0 0 0 \\ 1 1 0 0 1 \\ 1 1 1 1 0 \\ 1 0 0 0 1 \end{pmatrix} & = & \begin{pmatrix} 1 0 0 \\ 1 1 0 \\ 1 0 1 \\ 0 1 0 \end{pmatrix} & \circ & \begin{pmatrix} 1 1 0 0 0 \\ 1 0 0 0 1 \\ 1 1 1 1 0 \end{pmatrix}
\end{array}
$$

Considering the input Boolean matrix as a formal context $\langle X, Y, I \rangle$, where $X = \{1, \ldots, 4\}$ and $Y = \{1, \ldots, 5\}$, the following three formal concepts of $\langle X, Y, I \rangle$ are utilized as factors:

$$\mathcal{F} = \{\langle\{1, 2, 3\}, \{1, 2\}\rangle, \langle\{2, 4\}, \{1, 5\}\rangle, \langle\{3\}, \{1, 2, 3, 4\}\rangle\}.$$

As has been demonstrated in [3], a useful feature of using formal concepts as factors is the fact that formal concepts may easily be interpreted. Namely, every factor, i.e. a formal concept $\langle C_l, D_l \rangle$, consists of a set C_l of objects (objects are measures of interestingness in our case) and a set D_l of attributes (properties of measures in our case). C_l contains just the objects to which all the attributes from D_l apply and D_l contains all attributes shared by all objects from C_l. From a clustering point of view, the factors $\langle C_l, D_l \rangle$ may thus be seen as clusters C_l with their descriptions by attributes from D_l. The factors thus have a natural, easy to understand meaning. Since the problem of computing the smallest set of factors is NP-hard, an approximation algorithm was proposed in [3, Algorithm 2]. This algorithm is utilized below in our paper (and also in the above example). The algorithm is based on a greedy strategy to select in every step a large formal concept that covers most of the yet uncovered data. The algorithm uses a heuristic way of selecting formal concepts that makes it possible to avoid the computation of the whole concept lattice, making it much faster yet with comparable quality when compared to a direct greedy strategy performing an exhaustive search, which is represented by [3, Algorithm 1].

3 Results of clustering of interestingness measures by means of BFA

In this section, we present three case studies of clustering interestingness measures using BFA. The input in each case study is a Boolean matrix I in which the rows correspond to interestingness measures and the columns to their properties. Using Algorithm 2 from [3] we compute a set $\mathcal{F}$ of factors, i.e. formal concepts, which decompose I, i.e. for which $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$. The formal concepts $\langle C_i, D_i \rangle \in \mathcal{F}$, cf. (3), provide us with clusters the following way. Each formal concept $\langle C_i, D_i \rangle$ consists of a set C_i of measures and a set D_i of their properties. C_i may be thought of as a cluster of measures and D_i as its description. In the next section, we denote the factors $\langle C_i, D_i \rangle$ by F_i. For convenience, we sometimes identify the set C_i, i.e. the measures covered by F_i, with F_i. That is, we say e.g. $F_i \subseteq C$ instead of the proper $C_i \subseteq C$, to indicate that the measures covered by factor F_i are included in some set C. Note also that according to the nature of BFA, the first factors account for most data and are thus the most important, while the last ones may usually be disregarded. Therefore, we considered the first factors as candidates for important clusters of measures.

3.1 Case study 1

3.1.1 Input: measures and their properties

In case study 1, we analyzed the interestingness measures and their properties basically as described in [14]. Table 1 lists the properties with references to the

literature where one can find details about these properties. The properties are also described in detail in [14]. We adopt the terminology regarding association rules (even though the usage is somewhat inconsistent and the terminology objectionable from a mathematical viewpoint). In particular, when speaking of a rule $A \Rightarrow B$, an example of $A \Rightarrow B$ is an object (row) in which both A and B are valid (i.e. the object has all attributes from $A \cup B$); one speaks of independence when the probability of having B is independent of the fact that A is true or not, i.e. $P(B|A) = P(B)$; logical implication when the confidence of $A \Rightarrow B$ is 1; equilibrium or indetermination when having maximum uncertainty of B given that A is true, i.e. $P(B|A) = 0.5$ (or, equivalently, $P(A \cap B) = P(A)/2$); attraction when having A increases the chances of having B, i.e., when $P(B|A) > P(B)$; repulsion when having A decreases the chances of having B i.e., when $P(B|A) < P(B)$.

The authors in [14] proposed an evaluation of 61 interestingness measures according to 19 properties, namely P_3 to P_{21}. Properties P_1 and P_2 were not taken into account in this study because of their subjective character and for this reason we disregard them as well. The measures and their properties result in a Boolean 61×19 measure-property matrix that is used for clustering the measures according to their properties. The clustering performed in [14] using the agglomerative hierarchical method and the K-means method revealed 7 clusters of measures which are used in the next section for comparison with the results obtained by BFA from the same measure-property matrix.

The measure-property matrix that we used for clustering by BFA is depicted in Table 2. It consists of 61 measures (taken from [14]) described by 21 properties (18 binary properties and one three-valued property, namely P_{14}, which is represented by three yes-no properties $P_{14.1}$, $P_{14.2}$, and $P_{14.3}$).

3.1.2 Output: clustering using Boolean factors

We computed the decomposition of the matrix using Algorithm 2 from [3] and obtained 28 factors. Several of them may be disregarded as not very important, cf. Section 2.3. In addition, we extended the original 61×21 Boolean matrix by adding for every property its negation, and obtained a 61×42 Boolean matrix. The reason for adding negated properties is due to our goal to compare the results with the two clustering methods mentioned above and the particular role of the properties and their negations in these clustering methods. From the 61×42 matrix, we obtained 38 factors, denoted $F_1, \ldots, F_{38}$. The factors are presented in Tables 3 and 4. Table 3 depicts the object-factor matrix describing the interestingness measures by factors, i.e. the matrix $A_{\mathcal{F}}$, see Section 2.3. Table 4 depicts the factor-property matrix explaining factors by properties of measures, i.e. the matrix $B_{\mathcal{F}}$, see Section 2.3. The factors are sorted from the most important to the least important, where the importance is determined by the number of 1s in the input measure-property matrix covered by the factor [3]. The first factors cover a large part of the matrix, while the last ones cover only a small part and may thus be omitted [3], see also the graph of cumulative cover of the matrix by the factors in Fig. 1.

3.1.3 Interpretation and comparison

The aim of this section is to provide an interpretation of the results described in the previous section and compare them to the results already reported in the literature,

Table 2 Input Boolean matrix describing interestingness measures by their properties

	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14.1	P15	P16	P17	P18	P19	P20	P21	P14.2	P14.3
Correlation		1	1	1	1	1	1			1	1			1	1	1			1	1	
Cohen		1	1	1	1	1	1			1	1					1			1	1	
Confidence	1	1	1	1				1	1										1	1	
Causal confidence	1	1	1	1	1	1		1											1	1	
Pavillon	1	1	1	1	1		1			1	1				1				1	1	
Ganascia	1	1	1	1				1	1						1				1	1	
Causal confirmation	1	1	1	1	1	1													1	1	
Descriptive confirmation	1	1		1					1						1						
Conviction	1	1	1	1	1	1	1			1	1								1	1	1
Cosine	1	1	1	1		1													1	1	
Coverage	1	1		1																1	
Dependency	1	1					1														
Causal dependency	1	1		1	1	1															
Gray Orlowska	1	1			1	1															
Bayes factor	1	1	1	1	1	1	1	1		1	1		1						1	1	1
Loevinger	1	1	1	1	1	1	1			1	1								1	1	1
Collective strength	1	1		1	1	1	1			1	1					1					
Fukuda	1	1		1								1						1	1		
Information gain	1	1		1	1	1	1			1	1			1	1				1	1	
Goodman	1	1	1	1			1	1		1	1			1	1	1		1	1	1	
Implication index	1	1	1	1			1												1	1	

Table 2 (continued)

	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14.1	P15	P16	P17	P18	P19	P20	P21	P14.2	P14.3
IPEE	1	1	1	1					1			1					1	1	1		
IP3E	1	1	1	1					1			1					1	1	1		
PDI	1	1	1	1	1	1	1			1		1	1				1	1			
II	1	1	1	1	1	1				1	1	1	1				1	1	1		
EII	1	1	1	1	1	1				1	1	1					1	1	1		
REII	1	1	1	1	1	1				1		1					1	1	1		
Likelihood index	1	1		1	1	1						1					1	1		1	
Interest		1	1	1		1	1			1	1								1		1
Jaccard		1		1		1															1
Jmeasure	1																				
Klosgen	1	1	1	1	1					1											
Laplace	1	1	1	1	1				1	1	1									1	
Mgk	1	1	1	1	1			1	1						1					1	
Least contradiction	1	1	1	1	1					1	1									1	
Pearl			1	1			1			1	1					1			1	1	
Piatetsky-Shapiro		1	1	1		1				1				1		1		1		1	1
Precision		1	1	1	1	1				1				1	1					1	1
Prevalence	1	1	1	1						1										1	
YuleQ			1	1			1	1		1	1	1	1	1	1	1			1	1	
Recall	1	1		1	1	1	1			1	1					1					
Gini	1	1		1	1	1					1										
Relative risk	1	1	1	1					1	1				1							1
Sebag	1	1	1	1					1	1										1	1
Support	1	1	1	1																	1
One way support	1	1	1		1	1	1			1	1										1
Two way support	1	1	1		1	1	1			1	1								1		1
Examples rate	1	1	1	1				1	1			1									

Table 2 (continued)

	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14.1	P15	P16	P17	P18	P19	P20	P21	P14.2	P14.3
VT100		1	1	1	1	1				1				1	1	1	1		1	1	1
Variation support		1	1																1		1
YuleY		1	1	1	1		1	1		1	1		1	1	1	1			1	1	
Zhang	1	1	1	1	1		1	1		1	1		1		1				1	1	
Causal confirm-confidence	1	1	1	1	1	1	1	1				1							1	1	
Czekanowski-dice		1		1	1	1													1		
Negative reliability	1	1	1	1	1	1		1											1		
Mutual information	1	1														1			1		
Kulczynski	1	1		1	1	1													1	1	1
Leverage	1	1		1	1	1				1	1			1	1	1			1	1	1
Novelty		1	1	1	1	1	1			1	1					1			1		
Odds ratio		1	1	1	1	1	1			1	1								1	1	
Specificity	1	1		1	1	1													1	1	1

Table 3 Interestingness measures described by factors obtained by decomposition of the input matrix from Table 2 extended by negated properties

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	F22	F23	F24	F25	F26	F27	F28	F29	F30	F31	F32	F33	F34	F35	F36	F37	F38
Correlation	1					1			1				1								1					1					1							1
Cohen	1					1							1								1										1							
Confidence		1		1		1				1		1						1												1								
Causal confidence		1		1						1		1						1												1								
Pavillon	1			1		1				1	1									1											1		1					
Ganascia		1		1		1					1	1														1				1								
Causal confirmation		1		1						1	1							1								1				1								
Descriptive confirmation		1		1										1												1							1					
Conviction	1														1																		1					
Cosine		1			1		1						1	1																		1	1				1	
Coverage			1							1				1						1					1					1								
Dependency		1	1							1	1			1				1						1						1			1					
Causal dependency		1			1	1				1	1																1			1			1					
Gray			1								1						1				1	1							1	1			1				1	
Orlowska													1							1									1									
Bayes factor	1				1										1				1		1										1		1				1	
Loevinger	1					1						1							1		1										1							
Collective strength	1						1													1	1						1											1
Fukuda	1										1									1		1						1		1	1		1	1				
Information gain	1												1				1			1												1	1	1				
Goodman	1		1														1									1				1	1		1					
Implication index			1														1											1			1		1				1	

Table 3 (continued)

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	F22	F23	F24	F25	F26	F27	F28	F29	F30	F31	F32	F33	F34	F35	F36	F37	F38
IPEE	1							1														1			1			1		1				1				
IP3E				1				1										1				1						1		1				1			1	
PDI								1			1											1						1						1	1		1	
II								1														1								1				1	1			
EII								1									1	1				1												1	1			
REII								1									1	1				1												1	1			
Likelihood index													1						1		1									1	1							
Interest					1	1	1				1		1							1		1									1	1	1					
Jaccard					1	1	1						1							1											1	1	1					
Jmeasure			1																					1														
Klosgen	1								1		1				1		1			1	1								1				1					
Laplace		1									1							1												1								
Mgk	1	1								1	1	1	1													1				1	1							
Least contradiction		1					1		1		1			1																								
Pearl			1										1			1					1		1			1					1							
Piatetsky-Shapiro						1			1																									1				
Precision											1		1	1										1														
Prevalence			1							1		1							1		1	1	1		1													
YuleQ	1													1						1										1			1			1		
Recall										1	1			1	1		1						1				1											
Gini											1		1				1	1			1	1								1							1	
Relative risk	1			1	1		1				1			1		1																	1				1	1
Sebag		1		1																				1														
Support																																1	1					
One way support	1																			1									1			1	1	1			1	

Table 3 (continued)

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	F22	F23	F24	F25	F26	F27	F28	F29	F30	F31	F32	F33	F34	F35	F36	F37	F38
Two way support	1				1		1													1									1	1		1	1				1	
Examples rate				1														1				1			1					1			1					
VT100			1			1	1						1			1							1		1		1								1		1	1
Variation support									1					1	1		1										1									1	1	
YuleY	1					1				1		1						1	1			1								1						1		
Zhang	1										1	1		1	1									1			1					1	1			1		1
Causal confirm-confidence	1	1				1				1	1	1						1											1									
Czekanowski-dice	1	1									1		1	1																		1	1					
Negative reliability	1	1				1	1			1	1	1		1				1					1	1			1			1								
Mutual information			1				1				1		1	1									1															
Kulczynski					1	1	1			1	1		1	1													1					1	1					
Leverage					1	1	1		1	1		1	1	1		1		1			1					1	1			1								
Novelty	1	1				1					1										1										1							
Odds ratio	1	1			1		1			1		1					1				1						1			1			1				1	1
Specificity	1	1	1		1	1				1	1		1									1	1							1			1					1

Table 4 Factors obtained by decomposition of the input matrix from Table 2 extended by negated properties

	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14.1	P15	P16	P17	P18	P19	P20	P21	P14.2	P14.3	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14.1	P15	P16	P17	P18	P19	P20	P21	P14.2	P14.3
F1	1	1	1		1		1			1	1								1									1	1		1	1	1	1	1		1	1	1			1
F2	1	1	1	1	1		1			1	1								1	1		1			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
F3	1	1		1					1													1			1		1	1						1	1	1	1	1				
F4		1		1	1	1													1	1		1	1		1	1	1			1			1		1		1	1	1	1	1	1
F5		1	1	1	1	1													1	1	1			1	1			1			1		1	1	1	1	1	1		1		1
F6		1	1	1	1	1															1				1	1	1	1		1					1	1	1	1				
F7		1	1	1	1																				1				1						1	1		1		1		
F8	1	1	1	1	1						1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1				1	1	1	1	1	1	1	1	1		1
F9	1	1	1	1	1		1					1													1		1	1					1							1		
F10	1	1					1															1			1	1		1	1	1	1	1	1	1	1	1	1	1	1	1		1
F11		1	1		1														1			1			1	1	1		1	1	1	1	1	1	1	1	1	1	1	1		1
F12		1	1	1				1											1				1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1		
F13		1	1	1	1	1															1			1	1	1	1	1						1	1				1			1
F14		1	1	1															1			1			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1
F15	1	1	1		1					1	1								1			1		1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1		1
F16		1	1		1	1	1			1	1			1		1	1		1				1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1		1
F17		1			1	1	1						1																1	1				1	1	1	1			1		
F18		1	1		1	1																1		1	1	1	1	1		1					1			1	1			
F19		1	1	1								1	1											1		1	1															
F20		1	1	1	1	1	1			1						1		1	1			1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1
F21	1	1		1	1	1	1			1	1	1										1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1
F22		1																				1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
F23		1		1												1										1				1		1		1	1		1	1	1	1	1	1
F24	1	1													1		1											1										1		1		
F25		1		1																	1			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
F26		1																	1		1			1			1		1	1	1	1	1	1	1	1	1	1	1	1	1	1
F27		1																	1	1	1			1					1		1	1	1	1	1	1	1	1	1	1	1	1

Table 4 (continued)

	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14.1	P15	P16	P17	P18	P19	P20	P21	P14.2	P14.3	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14.1	P15	P16	P17	P18	P19	P20	P21	P14.2	P14.3
F28		1	1																							1					1	1			1	1	1					1
F29		1			1														1					1					1		1	1			1	1	1	1				
F30	1																																									
F31		1					1																		1					1			1		1		1	1	1		1	
F32		1																	1	1		1		1																		
F33		1																	1				1	1					1							1	1				
F34		1	1	1														1										1	1												
F35		1	1	1	1									1			1										1														
F36		1	1	1			1	1			1		1		1				1										1	1					1						1
F37							1			1							1		1	1	1									1		1					1	1	1		1
F38																1													1	1			1		1	1	1	1	1		

The factors are described in terms of the original and negated properties

Fig. 1 Cumulative cover of input matrix from Table 2 extended by negated properties by factors obtained by decomposition of the matrix
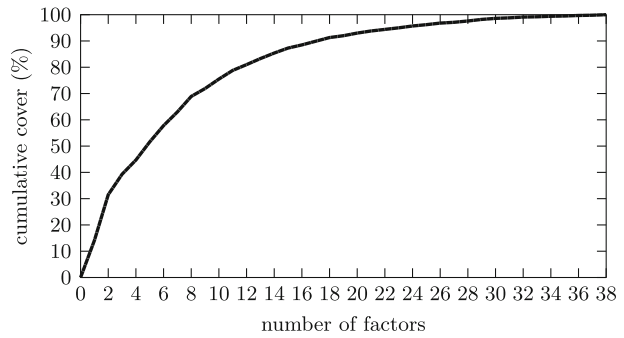
focusing mainly on [14]. As was described in the previous section, 38 factors were obtained. The first 21 of them cover 94 % of the input measure-property matrix (1s in the matrix), the first nine cover 72 %, and the first five cover 52 %. Another remark is that the first ten factors cover the whole set of measures.

Note first that the Boolean factors represent overlapping clusters, contrary to the clustering using the agglomerative hierarchical method and the K-means method performed in [14]. Namely, the clusters are depicted in Fig. 2 describing the Venn diagram of the first five Boolean factors (plus the eighth and part of the sixth and
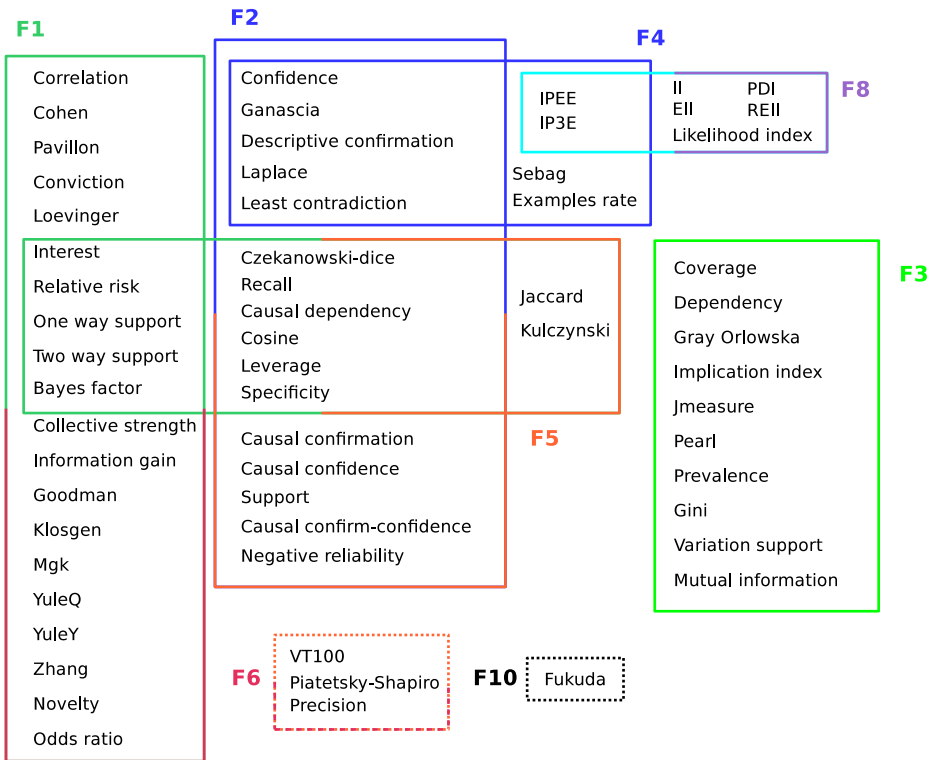
Fig. 2 Venn diagram of selected factors of the input matrix from Table 2 extended by negated properties
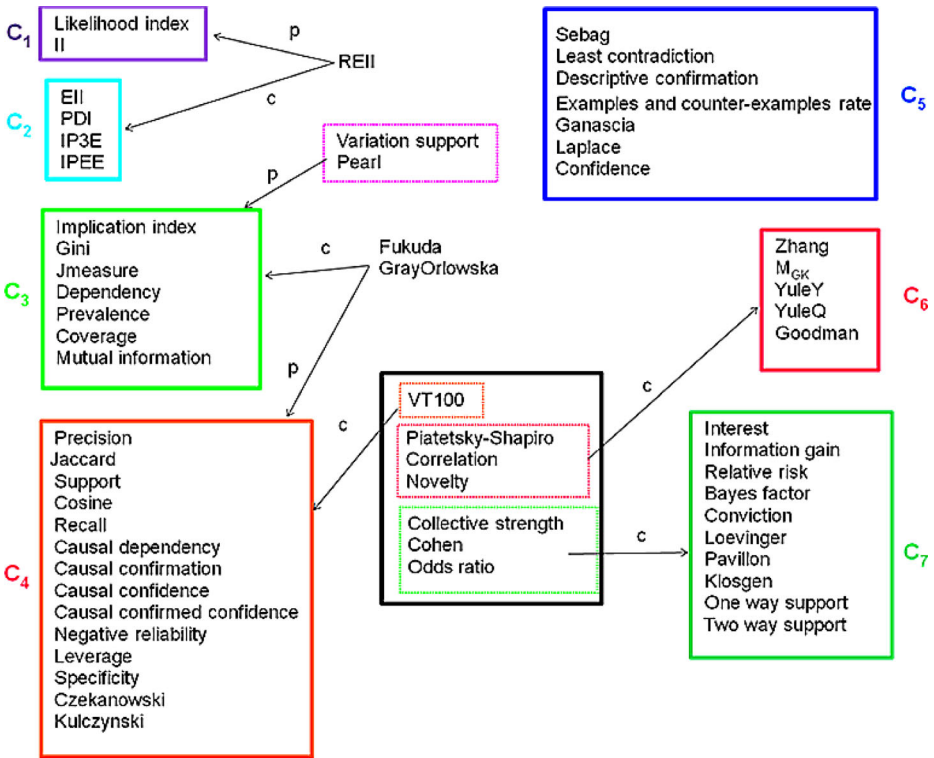
Fig. 3 Classes of measures obtained by the hierarchical and K-means clusterings

tenth to cover the whole set of measures). Figure 3, which is borrowed from [14], demonstrates the consensus of the clusterings obtained by both hierarchical and K-means clustering methods. Classes C_1 to C_7 are collections of measures that were identified to belong to the same cluster by both methods. Some other measures were also found to belong to one of these classes by each method, i.e. by one but not the other. This is indicated by the arrows labeled c (hierarchical clustering) and p (K-means clustering). For example, Fukuda was found to be in C_3 by the hierarchical clustering and in C_4 by the K-means clustering. Given these 7 classes, we can then make a comparison with the factors revealed by the BFA method. A comparison with the classes revealed by the hierarchical and K-means methods follows.

Factor 1 The first factor F_1 applies to 20 measures, see Table 3: Correlation, Cohen, Pavillon, Conviction, Bayes factor, Loevinger, Collective strength, Information gain, Goodman, Interest, Klosgen, Mgk, YuleQ, Relative risk, One way support, Two way support, YuleY, Zhang, Novelty, and Odds ratio. These measures share the following 9 properties: P4, P7, P9, not P11, P12, P13, not P19, not P20, P21, see Table 4.

Interpretation The factor applies to measures whose value increases w.r.t. the number of examples and have a fixed point in the case of independence (this allows

one to identify the attractive and repulsive area of a rule). The factor also applies only to descriptive and discriminant measures that are not based on a certain probabilistic model.

Comparison When looking at the classification results reported in [14], F_1 covers two classes from [14]: C_6 and C_7, which together contain 15 measures. Those classes are closely related according to the dendrogram obtained with the agglomerative hierarchical clustering, which is shown in Fig. 4 taken from [14]. The 5 missing measures form a class obtained with K-means method in [14] with Euclidian distance.
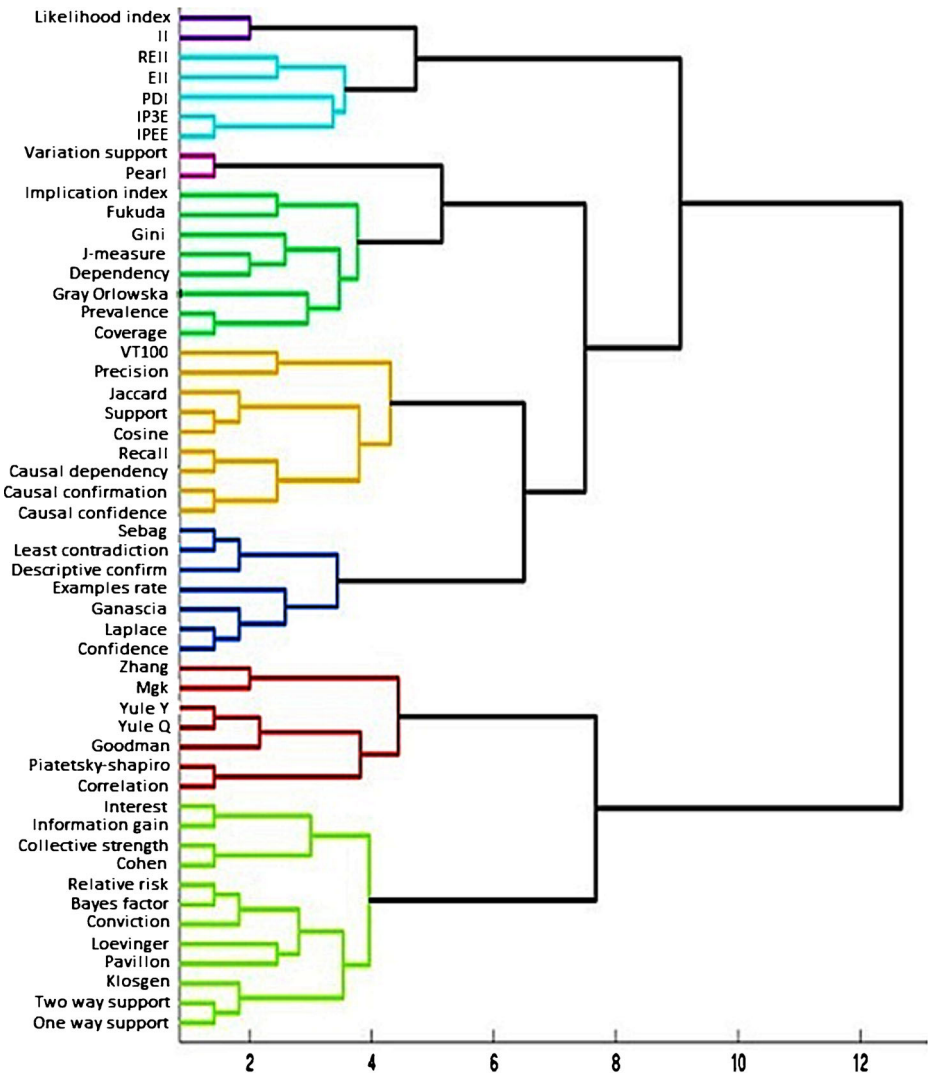
Fig. 4 Dendrogram of the agglomerative hierarchical clustering using Ward criterion

Factor 2 F_2 applies to 16 measures, namely: Confidence, Causal confidence, Ganascia, Causal confirmation, Descriptive confirmation, Cosine, Causal dependency, Laplace, Least contradiction, Recall, Support, Causal confirm-confidence, Czekanowski-dice, Negative reliability, Leverage and Specificity. These measures share the following 12 properties: P4, P6, not P9, not P12, not P13, P14.2, not P15, not P16, not P18, not P19, not P20, P21.

Interpretation   The factor applies to measures whose value increases w.r.t. the number of examples and has a variable point in the case of independence, which implies that the attractive and repulsive areas of a rule are not identifiable. The factor also applies only to measures that are discriminant, are indifferent to a few counter-examples, and are not based on a certain probabilistic model.

Comparison F_2 corresponds to two classes, C_4 and C_5 reported in [14]. The dendrogram exposed in Fig. 4 confirms their closeness. $C_4 \cup C_5$ contains 21 measures. The measures present in $C_4 \cup C_5$ and missing in F_2 are: Jaccard, Kulczynski, Accuracy, Examples rate, and Sebag. Those measures are not covered by F_2 since they are not indifferent to a few counter-examples.

Factor 3 F_3 applies to 10 measures, namely: Coverage, Dependency, Gray Orlowska, Implication index, Jmeasure, Pearl, Prevalence, Gini, Variation support, and Mutual information. These measures share the following 10 properties: not P6, not P8, not P10, not P11, not P13, not P14.1, not P15, not P16, not P17, not P19.

Interpretation   The factor applies to measures that share only negated properties and whose value does not increase w.r.t. the number of examples. Their common behavior makes them unique compared to the other studied measures and explains the absence of factor that overlaps with F_3. Factor F_3 represents the only factor among those presented in the Venn diagram which only gather measures that do not share any original (non-negated) property.

Comparison F_3 corresponds to class C_3 reported in [14], which contains 8 measures. The two missing measures, Variation support and Pearl, belong to C_3 within K-means method and are closely related to it according to the dendogram (Fig. 4). We see a strong correspondence between the results obtained by using Boolean factors and the ones reported in [14].

Factor 4 F_4 applies to 9 measures, namely: Confidence, Ganascia, Descriptive confirmation, IPEE, IP3E, Laplace, Least contradiction, Sebag, and Examples rate. These measures share the following 12 properties: P3, P4, P6, P11, not P7, not P8, not P9, not P12, not P13, not P15, not P16, not P18.

Interpretation   The factor applies to measures whose value increases w.r.t. the number of examples and has a fixed value in the equilibrium case.

Comparison F_4 mainly applies to measures of class C_5 obtained in [14]. In fact, 7 measures from C_5 are in F_4. The only two missing measures, IPEE and IP3E, which belong to a different class C_2, are represented in F_8 which overlaps with C_5. This can be explained by the fact that IPEE and IP3E share property P11 with other measures of the same factor F_4.

Factor 5 F_5 applies to 13 measures, namely: Interest, Relative risk, One way support, Two way support, Bayes factor, Czekanowski-dice, Recall, Causal dependency, Cosine, Leverage, Specificity, Jaccard and Kulczynski. These measures share the following 13 properties: P4, not P5, P8, not P10, not P11, not P14.1, not P15, not P16, not P17, not P18, not P19, not P20, P21.

Interpretation   The factor applies to descriptive and discriminant measures whose value decreases w.r.t. the increasing number of items satisfying the consequent. It applies also to measures having a variable point in the equilibrium case and logical implication case and which are not based on a certain probabilistic model.

Comparison   Factor F_5 covers principally two classes C_4 and C_7 from the classification results reported in [14]. $C_4 \cup C_7$ contain 24 measures. We then have the following 11 measures missing from F_5: Precision, Support, Causal confirmation, Causal confidence, Causal confirm-confidence, Negative reliability which belong to C_4, and Information gain, Conviction, Loevinger, Pavillon, Klosgen which represent C_7 missing measures.

Factor 6 F_6 applies to VT100, Piatetsky-Shapiro, Precision, and, moreover, to Correlation, Cohen, Pavillon, Loevinger, Novelty, Interest, Causal dependency, Leverage, Specificity, Causal confirmation, Causal confidence, Causal confirm-confidence, and Negative reliability, which belong to classes C_4, C_7 and G_{p8} revealed by [14]. These measures share the following 8 properties: P4, P6, P7, P8, not P11, P14.2, not P15, P21.

Interpretation   The factor applies to measures whose value increases w.r.t. the number of examples and the data size but decreases with increasing number of items satisfying the consequent. The factor also applies only to discriminant measures that are indifferent to a few counter-examples and have a variable point in the equilibrium case.

Comparison   The clustering resulted from K-means method in [14] revealed the presence of the class G_{p8} which gather Piatetsky-Shapiro and VT100 measures from F_6. On the other hand, according to the agglomerative hierarchical clustering method, we find that VT100 is also close to Precision and to other additional measures from both C_4 and F_6, e.g., Negative reliability, Cosine, or Recall. Furthermore, we mention that G_{p8} measures are divided between F_1 (Correlation, Cohen, Collective strength, Odds ratio and Novelty) and F_6 (Piatetsky-Shapiro and VT100).

Factor 8 F_8 applies to 7 measures, namely: IPEE, IP3E, II, PDI, EII, Likelihood index, REII. These measures share the following 10 properties: P4, P5, P6, not P10, P14.1, not P16, not P17, not P18, P19, P20.

Interpretation   The factor applies to measures whose value increases w.r.t. the number of examples and does not assign a fixed value to a rule with confidence equal to 1. It also applies to statistical measures which are based on a certain probabilistic model.

Comparison F_8 corresponds to two classes, C_1 and C_2 reported in [14], which contain 6 measures and represent the statistical measures. Only one measure is missing, REII, which according to the partitioning method belongs to C_1, and to C_2 with the agglomerative hierarchical method.

Factor 10   The only remaining measure not covered by previous factors is Fukuda. F_{10} revealed its presence in addition to the following 15 measures: Causal confidence, Causal confirm-confidence, Pavillon, Causal confirmation, Coverage, Dependency, Causal dependency, Loevinger, Implication index, Mgk, Prevalence, Recall, Negative reliability, Leverage and Specificity. These measures share the following 7 properties: P3, not P11, P14.2, not P15, not P16, not P18, not P19.

Interpretation   The factor applies to asymmetric measures having a variable point in the equilibrium case. It also applies to measures which are not based on a certain probabilistic model and are indiferent to a few counter-examples.

Comparison F_{10} covers classes C_3, C_4, C_6 and C_7 described in Fig. 3 in addition to Fukuda measure for which no consensus was found. It belongs to C_3 according to the agglomerative hierarchical method and to C_4 according to K-means. The union of C_3, C_4, C_6 and C_7 unveils 36 measures among which about twenty are missing from F_{10}. With seven shared properties, F_{10} represents the factor having the minimal number of properties common to all its measures. This explains the mutual dissimilarity of the measures belonging to this factor.

3.2 Case study 2

3.2.1 Input: measures and their properties

In this study we analyzed the measures studied in [35] which provides a study of 20 measures described by 9 properties. From these 9 properties, we considered 7 in case study 1, namely, P_3, P_7, P_8, P_9, P_{10}, P_{11}, P_{14}. The two other properties, namely, "comprehensibility of the measure" and "easiness to set a threshold of acceptance" were considered too subjective and thus not taken into account in the analysis. The authors in [35] then obtained the 20×15 measure-property matrix shown in Table 5 to cluster the measures by the agglomerative hierarchical method using

Table 5 Input Boolean measure-property matrix from [35]

	P3	$\overline{P3}$	$\overline{P8}$	P8	$\overline{P10}$	P10	$\overline{P9}$	P9	$\overline{P11}$	P11	P14.3	P14.2	P14.1	$\overline{P7}$	P7
Support	1		1		1		1		1			1		1	
Confidence		1	1			1		1		1		1		1	
Correlation	1			1	1			1	1			1		1	
Pavillon		1		1	1			1	1			1		1	
Piatetsky-Shapiro	1			1	1			1	1			1		1	
Loevinger		1		1		1		1	1			1		1	
Zhang		1		1		1		1	1				1	1	
Implication index		1		1	1			1	1			1		1	
Interest	1			1	1			1	1			1		1	
Least contradiction		1	1		1		1			1		1		1	
Sebag		1	1			1		1		1	1			1	
Bayes factor		1	1			1		1		1	1			1	
Conviction		1	1			1		1		1	1			1	
Examples rate		1	1			1		1		1			1	1	
Cohen	1			1	1			1		1		1		1	
Information gain	1			1	1			1		1			1	1	
II		1	1			1		1		1		1			1
REII		1	1			1		1		1		1			1
PDI		1		1	1			1		1		1			1
Laplace		1	1		1		1			1		1		1	

Ward criterion and Manhattan distance. The authors point out that when using other criteria, they obtained similar results. They identified the five following classes:

$ClBV_1$ = {Support, Least contradiction, Laplace},

$ClBV_2$ = {Confidence, Sebag, Examples rate},

$ClBV_3$ = {Correlation, Piatetsky-Shapiro, Pavillon, Interest, Implication index,

Cohen, Information gain},

$ClBV_4$ = {Loevinger, Bayes factor, Conviction},

$ClBV_5$ = {Zhang, REII, II, PDI}.

3.2.2 Output: clustering using Boolean factors

We applied BFA to the matrix I from Table 5 and obtained a set $\mathcal{F}$ of 15 factors which decompose I, i.e. $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$. The factors are denoted $F_v 1, \ldots, F_v 15$ and are shown in Table 6 (object-factor matrix $A_{\mathcal{F}}$) and Table 7 (factor-attribute matrix $B_{\mathcal{F}}$). The graph of the cumulative cover of the matrix by the factors depicted in Fig. 5. The first eight factors cover 91 % of the input-measure property matrix, the first four cover 69 % and the first three cover 57 %, more than a half of the matrix. However, the first five cover the whole set of measures, in that every measure is covered by some of the first five factors. Figure 6 provides the Venn diagram of the first five factors.

Table 6 Object-factor matrix $A_\mathcal{F}$ of the matrix in Table 5

	F_v1	F_v2	F_v3	F_v4	F_v5	F_v6	F_v7	F_v8	F_v9	F_v10	F_v11	F_v12	F_v13	F_v14	F_v15
Support				1				1			1	1	1		1
Confidence		1					1				1			1	1
Correlation	1			1				1				1	1		1
Pavillon	1					1		1					1		1
Piatetsky-Shapiro	1			1				1				1	1		1
Loevinger	1		1										1		1
Zhang	1		1							1			1		
Implication index	1					1		1					1		1
Interest	1			1				1				1	1		1
Least contradiction		1				1					1			1	1
Sebag		1					1		1					1	
Bayes factor	1		1						1			1			
Conviction	1		1						1			1			
Examples rate		1					1			1				1	
Cohen	1			1				1				1	1		1
Information gain	1							1		1		1	1		
II			1		1					1			1		
REII			1		1		1			1				1	
PDI					1			1		1			1		
Laplace		1				1		1			1		1		1

3.2.3 Interpretation and comparison

In this section, we interpret and describe the characteristics of the five obtained factors $F_v1, \ldots, F_v5$. Then we compare each factor firstly with the factors depicted in Fig. 2 and secondly with the clusters obtained in [35].

Table 7 Factor-attribute matrix $B_\mathcal{F}$ of the matrix in Table 5

	P3	$\overline{P3}$	$\overline{P8}$	P8	$\overline{P10}$	P10	$\overline{P9}$	P9	$\overline{P11}$	P11	P14.3	P14.2	P14.1	$\overline{P7}$	P7
F_v1				1				1	1					1	
F_v2		1	1				1							1	
F_v3		1		1		1	1								
F_v4	1			1					1			1		1	
F_v5		1		1			1						1		1
F_v6		1				1						1		1	
F_v7		1				1				1					
F_v8				1					1						
F_v9		1				1					1			1	
F_v10													1		
F_v11			1				1					1		1	
F_v12	1			1					1					1	
F_v13									1						
F_v14		1								1					
F_v15												1		1	

**Fig. 5** Cumulative cover by factors of the input matrix from Table 5
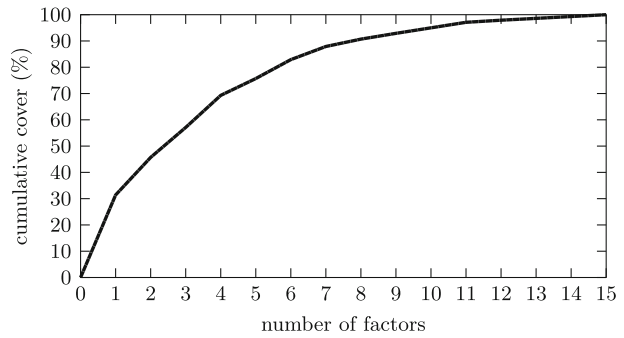
Factor 1   The first factor F_v1 applies to 11 measures, namely: Correlation, Piatetsky-Shapiro, Interest, Cohen, Pavillon, Implication index, Information gain, Loevinger, Zhang, Bayes factor and Conviction. These measures share the following 4 properties: P8, P9, not P7 and not P11.

Interpretation   The factor applies to measures whose value decreases w.r.t. the data and with the number of items matching the consequent and has a fixed point in the case of independence and a variable point in the case of equilibrium.

Factor 2 F_v2 applies to 5 measures, namely: Confidence, Least contradiction, Sebag, Examples rate and Laplace. These measures share the following 3 properties: not P3, not P7 and not P8.

Interpretation   The factor applies to symmetric measures whose evolutionary curve decreases w.r.t. the data size and increases w.r.t. the consequent size.

Factor 3 F_v3 applies to 6 measures, namely: Loevinger, Zhang, Bayes factor, Conviction, II and REII. These measures share the following 4 properties: not P3, P8, P9 and P10.

Fig. 6 Venn diagram of the first five factors of the matrix from Table 5
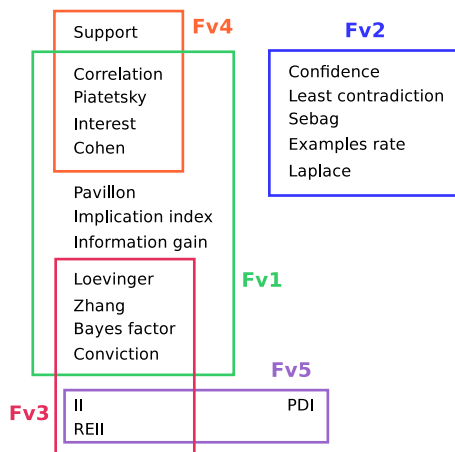
Interpretation   The factor applies to symmetric measures whose evolutionary curve decreases w.r.t. the consequent size and has a fixed point in the independence and logical implication case.

Factor 4 F_v4 applies to 5 measures, namely: Support, Correlation, Piatetsky-Shapiro, Interest and Cohen. These measures share the following 5 properties: P3, not P7, not P10, not P11 and P14.2.

Interpretation   The factor applies to asymmetric measures whose value decreases w.r.t. the data size and does not assign a fixed value to a rule with confidence equal to 1 and in the equilibrium case. The factor also applies only to measures that are indifferent to a few counter-examples.

Factor 5 F_v5 applies to 3 measures, namely: II, REII and PDI. These measures share the following 5 properties: not P3, P7, P8, P9 and P14.1.

Interpretation   The factor applies to symmetric measures whose value increases w.r.t. the data size and decreases w.r.t. the number of examples matching the consequent. The factor also applies only to measures that have a fixed value in the independence case and tolerate a few counter-examples.

Comparison: Comparing the Venn diagram from Fig. 2 with the one presented in Fig. 6 we see that (cf. the remark at the beginning of Section 3: we identify factors with the sets of measures covered by the factors)

$$F_v2 \subset F4 \quad \text{and} \quad F_v5 \subset F8,$$

and, moreover, one may observe the following interesting relationship revealing similarity of the factors:

$$F_v1 - \{\text{Piatetsky-Shapiro, Implication index}\} \subset F1,$$
$$F_v4 - \{\text{Piatetsky-Shapiro}\} \subset F1,$$
$$F_v3 - \{\text{II, REII}\} \subset F1.$$

Note also that $F1$ contains half of the measures studied in [35], cf. Fig. 2. Further similarities between the factors may be seen from the Venn diagrams in Figs. 2 and 6.

Regarding a comparison of factors $F_v1,\dots,F_v5$ with clusters $ClBV_1,\dots,ClBV_5$ obtained in [35], we get the following relationships:

$$F_v1 = ClBV_3 \cup ClBV_4 \cup ClBV_5 - \{\text{PDI}\},$$
$$F_v2 = ClBV_1 - \{\text{Support}\} \cup ClBV_2,$$
$$F_v3 = ClBV_4 - \{\text{Zhang, II, REII}\},$$
$$F_v4 - \{\text{Support}\} = ClBV_3 - \{\text{Pavillon, Implication index, Information gain}\},$$
$$F_v5 = ClBV_5 - \{\text{Zhang}\}.$$

We also observed that BFA was able to reveal similarities among measures found in [35]. In addition the first five factors allow to cover the whole 20 measures, and show some relationships among clusters found in [35].

Table 8 Input Boolean measure-property matrix obtained from [13]

	P3	$\overline{P3}$	P6	$\overline{P6}$	P7	$\overline{P7}$	P8	$\overline{P8}$	P9	$\overline{P9}$	P10	$\overline{P10}$	P14.1	P14.2	P14.3
Support	1		1		1			1	1			1		1	
Confidence		1	1		1			1	1		1			1	
Prevalence		1		1	1			1	1			1		1	
Recall		1	1		1			1	1			1	1		
Accuracy	1		1		1	1			1			1		1	
Lift	1		1		1	1			1			1	1		
Leverage		1	1		1	1			1			1		1	
Pavillon		1	1		1	1	1					1		1	
Relative Risk		1	1		1	1			1			1		1	
Loevinger		1	1		1	1	1					1			1
Odds ratio	1		1		1	1			1		1				1
Yule's Q	1		1		1	1	1				1				1
Yule's Y	1		1		1	1	1				1				1
Klosgen		1	1		1	1	1					1			1
Conviction		1	1		1		1		1		1				1
Weighted dependency		1	1		1			1	1			1			
Collective strength	1		1		1	1			1			1			1
Laplace		1	1		1			1	1			1		1	
Gini		1		1	1			1	1			1			1
Goodman	1			1	1			1	1			1			
Mutual information		1	1		1	1			1			1			
J-measure		1		1	1		1		1		1				1
One way support		1	1		1	1			1			1			1
Two way support	1		1		1	1			1			1			1
Correlation	1		1		1	1			1			1			1
Piatetsky-Shapiro	1		1		1	1			1			1		1	
Cosine	1		1		1	1				1		1	1		
Information gain	1		1		1	1				1		1	1		
Sebag		1	1		1	1			1	1					1
Least contradiction		1	1		1	1			1			1	1		
Bayes factor		1	1		1	1			1	1					1
Examples rate		1	1		1	1			1	1			1		
Zhang		1		1	1		1	1				1			1
Jaccard	1		1		1	1			1			1		1	
Variation support	1			1	1		1	1				1			1
Coverage		1		1	1			1	1			1			
Specificity		1		1	1			1	1			1			

3.3 Case study 3

3.3.1 Input: measures and their properties

In this study, we analyzed the measures and their properties studied in [13]. As in the previous case studies, we apply BFA to obtain factors and then interpret and compare these factors with clusters obtained in the two previous case studies, as the

Table 9 Object-factor matrix $A_{\mathcal{F}}$ of the matrix in Table 8

	F_h1	F_h2	F_h3	F_h4	F_h5	F_h6	F_h7	F_h8	F_h9	F_h10	F_h11	F_h12	F_h13
Support			1			1	1		1				
Confidence			1				1	1	1				
Prevalence		1		1			1		1				
Recall		1	1							1	1		
Accuracy	1		1			1	1						
Lift	1		1			1					1		
Leverage	1	1	1				1						
Pavillon	1	1					1					1	
Relative Risk	1	1	1				1						
Loevinger	1	1			1							1	
Odds ratio	1		1									1	1
Yule's Q	1				1						1	1	1
Yule's Y	1				1						1	1	1
Klosgen	1	1			1						1		
Conviction			1					1	1			1	
Weighted dependency		1	1						1				
Collective strength	1		1			1							1
Laplace		1	1				1		1				
Gini		1		1	1						1		
Goodman				1		1					1		
Mutual information	1	1									1		
J-measure				1	1			1			1	1	
One way support	1	1			1						1		
Two way support	1				1	1					1		1
Correlation	1				1	1					1		1
Piatetsky-Shapiro	1					1	1				1		
Cosine	1		1			1				1			
Information gain	1					1				1	1		
Sebag	1		1					1				1	
Least contradiction	1	1	1								1		
Bayes factor	1		1					1				1	
Examples rate	1		1					1		1			
Zhang		1		1	1						1		
Jaccard	1		1			1	1						
Variation support				1	1	1					1		1
Coverage		1		1					1				
Specificity		1		1					1				

authors of [13] did not provide any clustering. Geng and Hamilton [13] analyzed 38 interestingness measures using 11 properties. From these, we only retain the measures and properties that also appear in our two previous studies, which means that we retain 37 measures and 7 properties. The 7 properties retained are P_3, P_6, P_7, P_8, P_9, P_{10}, and P_{14}. Nevertheless, the authors of [13] considered slight variations of P_8 and P_9 that we denote $P_{8.1}$ and $P_{9.1}$, respectively. For the case of $P_{8.1}$, they consider that an interestingness measure must decrease with increasing number of items matching the antecedent as well as the consequent, but in our study (P_8) we only consider the consequent as in [14]. Concerning $P_{9.1}$, Geng and Hamilton focus on the value of an interestingness measure in the independence case considering that the value must be 0, while we (P_9) only ask that the measure has a fixed value as in [14]. According to these definitions, $P_{8.1}$ is stronger than P_8; however, $P_{9.1}$ is a generalization of P_9. The strong relationships existing between these properties encourage us to apply them in order to obtain the clusters of measures. From these measures and properties, we obtained a Boolean matrix I shown in Table 8.

3.3.2 Output: clustering using Boolean factors

We applied BFA to the matrix I from Table 8 and obtained a set $\mathcal{F}$ of 13 factors which decompose I, i.e. $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$. The factors are denoted $F_h1,\ldots,F_h13$ and are shown in Table 9 (object-factor matrix $A_{\mathcal{F}}$) and Table 10 (factor-attribute matrix $B_{\mathcal{F}}$). The graph of the cumulative cover of the matrix by the factors is depicted in Fig. 7. The graph reveals that the first nine factors cover 91 % of the input-measure property matrix, the first five cover 70 % and the first four cover 61 %. However, the first four cover the whole set of measures. The first four factors are visualized using their Venn diagram in Fig. 8, which reveals the overlap between the factors.

Note that the authors in [13] focus only on the assessment of interestingness measures according to their properties and their purpose is in no case the clustering of these measures. However, even if no classification of measures is performed, we think that it is quite important to apply BFA on the measure-property matrix obtained by [13] to see the factors resulting from it and to compare them with previous works.

Table 10 Factor-attribute matrix $B_{\mathcal{F}}$ of the matrix in Table 8

	P3	$\overline{P3}$	P6	$\overline{P6}$	P7	$\overline{P7}$	P8	$\overline{P8}$	P9	$\overline{P9}$	P10	$\overline{P10}$	P14.1	P14.2	P14.3
F_h1			1			1	1								
F_h2		1				1					1				
F_h3			1			1				1					
F_h4				1		1		1							
F_h5						1			1						1
F_h6	1					1						1			
F_h7						1								1	
F_h8		1				1					1				
F_h9						1		1		1					
F_h10			1			1							1		
F_h11						1			1						
F_h12						1					1				1
F_h13	1					1									1

Fig. 7 Cumulative cover by factors of the input matrix from Table 8
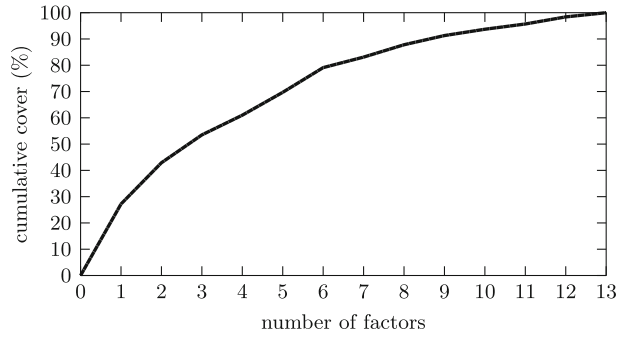
3.3.3 Interpretation and comparison

An interpretation of the four obtained factors $F_h1,\ldots,F_v4$ and their comparison with the factors depicted in Fig. 2 follow.

Factor 1   The first factor F_h1 applies to 23 measures, namely: Pavillon, Loevinger, Klosgen, Mutual information, One way support, Leverage, Relative risk, Least
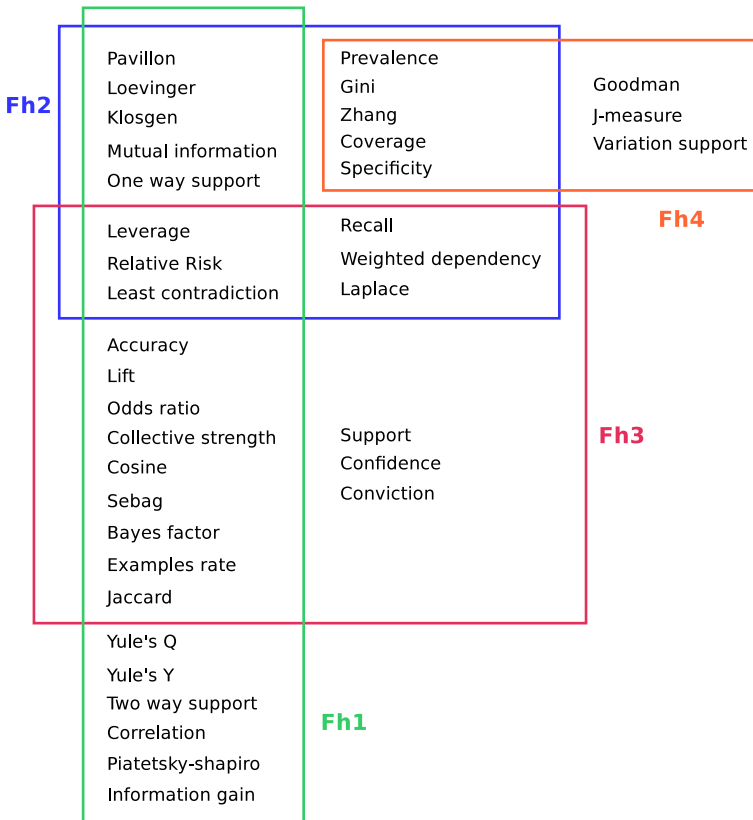

Fig. 8 Venn diagram of the first five factors of the matrix from Table 8

contradiction, Accuracy, Lift, Odds ratio, Collective strength, Cosine, Sebag, Bayes factor, Examples rate, Jaccard, Yule's Q, Yule's Y, Two way support, Correlation, Piatetsky-Shapiro and Information gain. These measures share the following 3 properties: P6, not P7 and P8.

Interpretation The factor applies to measures whose value increases w.r.t. the number of examples and decreases w.r.t. the data and the consequent size.

Factor 2 F_h2 applies to 16 measures, namely: Pavillon, Loevinger, Klosgen, Mutual information, One way support, Leverage, Relative risk, Least contradiction, Prevalence, Gini, Zhang, Coverage, Specificity, Recall, Weighted dependency and Laplace. These measures share the following 3 properties: not P3, not P7 and not P10.

Interpretation The factor applies to symmetric measures whose value decreases w.r.t. the data size and does not assign a fixed value to a rule with confidence equal to 1.

Factor 3 F_h3 applies to 18 measures, namely: Leverage, Relative risk, Least contradiction, Accuracy, Lift, Odds ratio, Collective strength, Cosine, Sebag, Bayes factor, Examples rate, Jaccard, Recall, Weighted dependency, Laplace, Support, Confidence and Conviction. These measures share the following 3 properties: P6, not P7 and not P9.

Interpretation The factor applies to measures whose value increases w.r.t. the number of examples and decreases w.r.t. the data size and has a variable point in the independence case.

Factor 4 F_h4 applies to 8 measures, namely: Prevalence, Gini, Zhang, Coverage, Specificity, Goodman, Jmeasure and Variation support. These measures share the following 3 properties: not P6, not P7 and not P8.

Interpretation The factor applies to measures whose value decreases w.r.t. the number of examples and the data size and increases w.r.t. the number of examples matching the consequent.

Comparison: The comparison of these clusters with those from Fig. 2 reveals the following similarities between the factors from case study 1 and the factors of the present case study:

$$F1 - \{\text{Conviction, Goodman, Zhang}\} \subset F_h1,$$

$$F_h2 \subset F1 \cup F2 \cup F3,$$

$$F_h3 \subset F1 \cup F2 \cup F3 \cup F4 \cup F5 \cup F6,$$

$$F_h4 - \{\text{Zhang, Specificity, Goodman}\} \subset F3.$$

We can see that in fact, F_h4 is very similar to $F3$, the $F1$ measures are included in F_h1 and that the F_h2 and F_h3 measures are scattered among $F1$, $F2$, $F3$ and $F1$, $F2$, $F3$, $F4$, $F5$, $F6$, respectively. Note also that looking at Tables 8 and 2 tells us that we

disagree with the authors of [13] on the properties of some measures, namely of P_8 and P_9, because the authors in [13] use slightly modified version of these properties compared to us. Some differences between the clusterings obtained therefore result from this fact.

3.4 Discussion

In this section we discuss the outcome of our experiments mainly by comparing the results of the three cases with the measures and their properties taken from [13, 35], and [14]. We applied BFA on the corresponding Boolean measure-property matrices and visualized the most important clusters thus obtained by means of Venn diagrams. Such clusters overlap and correspond to subgroups of measures with a common behavior.

An examination of the Venn diagrams depicted in Figs. 2, 6 and 8 reveals three stable groups of interestingness measures:

{Sebag, Examples rate, Laplace, Confidence, Least contradiction},
{Interest, Information gain, Relative risk},
{Bayes factor, Conviction, Loevinger, Pavillon}.

These new groups of measures were not found in the previous studies. These groups represent particular kinds of measures with similar behavior and may be added to those mentioned in [14]. From this viewpoint, BFA as a clustering method is complementary to other clustering techniques.

A work of Suzuki [32] mentions four pitfalls to avoid when studying categorization of interestingness measures: rule bias, data bias, expert bias, and search bias. In our work none of this bias is present, as we focus on formal characteristics of measures and the Boolean factors are generated using Formal concept analysis. No additional parameters are used to compute Boolean factors. From this viewpoint, we provide a study that fits most of the requirements of categorization described in [32].

Recall from Section 1 that the clusters of measures obtained may be seen as representing certain kinds of measures and may be used by a user to select measures, in particular when the user intends to select various measures of different kinds as in Bouker et al. [6]. For the particular form of clusters used in the present study, such scenario is particularly suitable because the clusters are represented by formal concepts and have thus a natural meaning. Namely, the meaning of each cluster derives from the intent of the corresponding formal concept, i.e. from the collection of attributes which uniquely determines the cluster.

4 Conclusions and further issues

We analyzed by means of clustering three datasets of interestingness measures for association rules that were described in the literature in terms of their binary properties. For this purpose, we used a recently developed method of Boolean factor analysis. We interpreted the Boolean factors obtained by this method as clusters, i.e. classes of measures described by their properties. We demonstrated that Boolean factors provide us with clearly interpretable, meaningful clusters of measures. Contrary to other clustering methods, Boolean factors represent overlapping clusters. We

argued that this is an advantage because overlapping clusters are a natural phenomenon in human classification in general and in the case when objects are represented by overlapping attributes in particular. We performed a detailed comparison of the clusterings obtained by Boolean factor analysis and the ones reported in the literature. It turns out that among the clusters obtained from Boolean factors, the ones corresponding to the first couple of factors seem natural and are highly similar to other clusters of measures reported in the literature. The clustering results obtained in this paper may therefore be regarded as cluster validation results as well and may be taken as reference points for future studies on classification of interestingness measures.

An interesting topic for future research is a further exploration of Boolean factor analysis as a clustering technique and the empirical and theoretical study of such method including its relationships to other methods of clustering. Another topic of future research is to exploit the possibility to utilize the clusters in helping a user select measures with desired properties. As mentioned in Sections 1 and 3.4, an approach based on selecting and aggregating the scores of several measures was suggested in [6]. A suitable clustering of measures, with clusters representing various kinds of measures, may help the user select measures representative of the intended kinds.

References

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proc. VLDB Conf., pp. 478–499 (1994)
2. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proc. ACM SIGMOD, pp. 207–216 (1993)
3. Belohlavek, R., Vychodil, V.: Discovery of optimal factors in binary data via a novel method of matrix decomposition. J. Comput. Syst. Sci. **76**(1), 3–20 (2010)
4. Blanchard, J., Guillet, F., Briand, H., Gras, R.: Assessing rule with a probabilistic measure of deviation from equilbrium. In: Proc. of 11th International Symposium on Applied Stochastic Models and Data Analysis ASMDA, pp. 191–200. Brest, France (2005)
5. Blanchard, J., Guillet, F., Briand, H., Gras, R.: IPEE: Indice Probabiliste d'Écart à l'Équilibre pour l'évaluation de la qualité des règles. In: Dans l'Atelier Qualité des Données et des Connaissances, pp. 26–34 (2005)
6. Bouker, S., Saidi, R., Ben, Yahia S., Mephu, Nguifo E.: Ranking and selecting association rules based on dominance relationship. In: IEEE ICTAI, pp. 658–665 (2012)
7. Brin, S., Motwani, R., Silverstein, C.: Beyond market baskets: generalizing association rules to correlations. In: Proc. of the ACM SIGMOD Conference, pp. 265–276. Tucson, Arizona (1997)
8. Carpineto, C., Romano, G.: Concept data analysis. Theory and Applications. J. Wiley (2004)
9. Delgado, M., Ruiz, D.-L., Sanchez, D.: Studying interest measures for association rules through a logical model. Int. J. Uncertain. Fuzz. Knowl Based Syst. **18**(1), 87–106 (2010)
10. Feno, D.R.: Mesures de qualité des règles d'association: normalisation et caractérisation des bases. Ph.D. thesis, Université de La Réunion (2007)
11. Ganter, B., Wille, R.: Formal concept analysis. Mathematical Foundations. Springer, Berlin (1999)

12. Geng, L., Hamilton, H.J.: Interestingness measures for data mining: a survey. ACM Comput. Surv. **38**(3), 1–31 (2006)
13. Geng, L., Hamilton, H.J.: Choosing the right lens: finding what is interesting in data mining. Qual. Meas. Data Min. **43**, 3–24 (2007)
14. Guillaume, S., Grissa, D., Mephu Nguifo, E.: Categorization of interestingness measures for knowledge extraction. ArXiv e-prints 1206.6741. cs.IT (2012)
15. Grissa, D., Guillaume, S., Mephu Nguifo, E.: Combining clustering techniques and formal concept analysis to characterize interestingness measures. ArXiv e-prints 1008.3629. cs.IT. (2010)
16. Gras, R., Kuntz, P., Couturier, R., Guillet, F.: Une version entropique de l'intensité d'implication pour les corpus volumineux. In: EGC, pp. 69–80 (2001)
17. Hájek, P., Havránek, T.: Mechanizing Hypotheses Formation. Springer (1978)
18. Hájek, P., Holeňa, Rauch, J.: The GUHA method and its meaning for data mining. J. Comput. Syst. Sci. **76**, 34–48 (2010)
19. Heravi, M.J., Zaïane, O.R.: A study on interestingness measures for associative classifiers. In: ACM SAC, pp. 1039–1046 (2010)
20. Hilderman, R.J., Hamilton, H.J.: Knowledge discovery and measures of interest. In: The International Series in Engineering and Computer Science, vol. 638, p. 2, 81. Kluwer (2001)
21. Huynh, X.-H., Guillet, F., Briand, H.: Clustering interestingness measures with positive correlation. ICEIS **2**, 248–253 (2005)
22. Lallich, S., Teytaud, O.: Évaluation et validation de mesures d'intérêt des règles d'association. In: RNTI-E-1, numéro spécial, pp. 193–217 (2004)
23. Lallich, S., Teytaud, O., Prudhomme, E.: Association rule interestingness: measure and statistical validation. Qual. Meas. Data Min. **43**, 251–275 (2007)
24. Lenca, P., Meyer, P., Vaillant, B., Picouet P.: Aide multicritére à la décision pour évaluer les indices de qualité des connaissances—modélisation des préférences de l'utilisatieur. EGC **1**, 271–282 (2003)
25. Lenca, P., Meyer, P., Picouet, P., Vaillant, B., Lallich, S.: Critères d'évaluation des mesures de qualité en ecd. RNTI (Entreposage et Fouille de données) **1**(1), 123–134 (2003)
26. Lenca, P., Meyer, P., Vaillant, B., Lallich, S.: A multicriteria decision aid for interestingness measure selection. Technical Report LUSSI-TR-2004-01-EN, (chap. 1). Dpt. LUSSI, ENST Bretagne (2004)
27. Lenca, P., Vaillant, B., Meyer, P., Lallich, S.: Association rule interestingness measures: experimental and theoretical studies. Qual. Meas. Data Min. **43**, 51–76 (2007)
28. Maddouri, M., Gammoudi, J.: On semantic properties of interestingness measures for extracting rules from data. Lect. Notes Comput. Sci. **4431**, 148–158 (2007)
29. Piatetsky-Shapiro, G.: Discovery, analysis and presentation of strong rules. In: Piatetsky-Shapiro, G., Frawley, W.J. (eds.) Knowledge Discovery in Databases, pp. 229–248. AAAI Press (1991)
30. Sese, J., Morishita, S.: Answering the most correlated n association rules efficiently. In: Proceedings of the 6th European Conf on Principles of Data Mining and Knowledge Discovery, pp. 410–422. Springer-Verlag (2002)
31. Surana, A., Kiran R.U., Reddy P.K.: Selecting a right interestingness measure for rare association rules. In: Proceedings of the 16th International Conference on Management of Data (COMAD 2010), pp. 115–124. Nagpur, India (2010)
32. Suzuki, E.: Pitfalls for categorizations of objective interestingness measures for rule discovery. In: Statistical Implicative Analysis, pp. 383–395. Springer-Verlag (2008)
33. Tan, P.-N., Kumar, V., Srivastava J.: Selecting the right objective measure for association analysis. Inf. Syst. **29**(4), 293–313 (2004)
34. Tan, P.-N., Steinbach, M., Kumar, V.: Introduction to Data Mining. Addison-Wesley (2005)
35. Vaillant, B.: Mesurer la qualité des règles d'association: études formelles et expérimentales. Ph.D. thesis, ENST Bretagne (2006)
36. Vaillant, B., Lenca, P., Lallich, S.: A clustering of interestingness measures. In: DS'04, The 7th International Conference on Discovery Science LNAI, vol. 3245, pp. 290–297 (2004)
37. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In: Rival, I. (ed.) Ordered Sets, pp. 445–470. Reidel, Dordrecht, Boston (1982)