A Unified View of Objective Interestingness Measures

Céline Hébert and Bruno Crémilleux

GREYC, CNRS - UMR 6072, Université de Caen
Campus Côte de Nacre
F-14032 Caen Cédex France
Forename.Surname@info.unicaen.fr

Abstract. Association rule mining often results in an overwhelming number of rules. In practice, it is difficult for the final user to select the most relevant rules. In order to tackle this problem, various interestingness measures were proposed. Nevertheless, the choice of an appropriate measure remains a hard task and the use of several measures may lead to conflicting information. In this paper, we give a unified view of objective interestingness measures. We define a new framework embedding a large set of measures called SBMs and we prove that the SBMs have a similar behavior. Furthermore, we identify the whole collection of the rules simultaneously optimizing all the SBMs. We provide an algorithm to efficiently mine a reduced set of rules among the rules optimizing all the SBMs. Experiments on real datasets highlight the characteristics of such rules.

1 Introduction

Exploring and analyzing correlations between features is on the core of KDD processes. Agrawal et al. [1] define association rules as the implications $X \rightarrow Y$ where X and Y represent one or several conjunctions of features (or attributes). However, among the overwhelming number of rules resulting from practical applications, it is difficult to determine the most relevant rules [10]. An essential task is to assist the user in selecting interesting rules.

Measuring the interestingness of discovered rules is an active and important area of data mining research. Interestingness measures are numerous and they are usually divided into two groups: subjective and objective measures. Whereas subjective measures take into account both the data and the user's expectations, objective measures are only based on raw data. In this paper, we focus on objective measures. Support and Confidence are probably the most famous ones [2], but there are more specific measures (e.g., Lift [6], Sebag and Schoenauer [18]). In practice, choosing a suitable measure and determining an appropriate threshold for its use is a challenge for the end user. Combining results coming from several measures is even much more difficult. Thus an important issue is to compare existing interestingness measures in order to highlight their similarities and differences and better understand their behaviors [17, 3]. The lack of generic

P. Perner (Ed.): MLDM 2007, LNAI 4571, pp. 533–547, 2007.

results about the characteristics captured by interestingness measures was the starting point of this work.

Contributions. This paper deals with the behavior of objective interestingness measures when applied to association rules. Our main objective is to make clear the choice of such a measure. For this purpose, we design an original framework which gives a unified view of a large set of measures, the Simultaneously lower Bounded Measures (SBM). We demonstrate that SBMs have similar behaviors so that choosing an appropriate measure among them becomes a secondary issue. This framework shows that three parameters (the minimal threshold for the antecedent frequency γ, the maximal consequent frequency η and the maximal number of exceptions δ) are on the core of many measures. This formalization provides lower bounds for the SBMs according to these parameters and thus guarantees a minimal quality for the rules. Moreover, we provide an efficient method to mine a reduced set of rules simultaneously optimizing all the SBMs, which ensures to produce the best rules according to these measures.

In a previous work [12], we addressed the specific case of the so-called classification rules (i.e., rules concluding on a class label). In this context, we showed that most of the usual interestingness measures only depend on the rule antecedent frequency and the rule number of exceptions and that they have a similar behavior. This paper is a generalization of [12] to any association rule. This generalization is not straightforward because one key point in [12] is the fact that the rule consequent is a class label and thus its frequency is known. This is obviously no longer true when considering any association rule and the major difficulty is the lack of information about the consequent frequency. We overcome it by bounding the consequent frequency. The fact that any attribute may appear in a rule consequent also requires to design a new algorithm to mine the rules simultaneously optimizing all the measures of the framework.

Organization. The rest of the paper is organized as follows. Section 2 discusses related work on rule selection and gives preliminary definitions. Section 3 introduces our framework and the SBMs. Section 4 shows how the SBMs can be simultaneously lower bounded and studies their behavior. Section 5 presents our algorithm to mine a reduced set among the most significant rules from a database. Section 6 gives experimental results about the quality of the discovered rules.

2 Preliminaries

2.1 Related Work

Lossless cover. It is well known that the whole set of association rules contains a lot of redundant rules [1]. So several approaches (see [13] for a survey) propose to restrict the mining to a rule cover [22] like the *informative rules* [15] or the *informative generic base* [9]. These rules have minimal antecedents and maximal consequents. They are lossless and informative since they enable to regenerate the whole set of valid association rules and their exact support and confidence

values. Our work is linked to this approach because we define informative SBM rules that have minimal antecedents and that simultaneously optimize the SBMs (see Section 5).

Selecting the most interesting rules with objective measures. As already told, researchers have proposed a lot of interestingness measures for various kinds of patterns. There is no widespread agreement on a formal definition of interestingness and several works attempt to define properties characterizing "good" interestingness measures [10, 16]. Piatetsky-Shapiro [16] proposes a framework with three properties and we set our work with respect to it. Other works compare interestingness measures to determine their differences and similarities, either in an experimental manner [20] or in a theoretical one [19, 8]. In [4], a visualization method is proposed to help the user in the rule exploration. There are also attempts to combine several measures to benefit from their joint qualities [7]. However choosing and using a measure remains a hard task. Our approach differs from these works : we argue that choosing the appropriate measure is a secondary issue because they all behave the same. We aim at analyzing the behavior of existing measures and showing their common features. We exhibit the minimal properties that a measure must satisfy to get a unified view of a lot of objective interestingness measures, the SBMs. Second, by simultaneously optimizing all the SBMs, our work combines the information brought by these measures.

2.2 Definitions

Basic definitions. A database $\mathcal{D}$ is a relation $\mathcal{R}$ between a set $\mathcal{A}$ of *attributes* and a set $\mathcal{O}$ of *objects*: for $a \in \mathcal{A}, o \in \mathcal{O}$, $a\,\mathcal{R}\,o$ if and only if the object o contains the attribute a. A *pattern* is a subset of $\mathcal{A}$. The frequency of a pattern X is the number of objects in $\mathcal{D}$ containing X; it is denoted by $\mathcal{F}(X)$. Table 1 shows an example of a database containing 8 attributes and 9 objects.

Table 1. An example of a database $\mathcal{D}$

$\mathcal{D}$	Attributes							
Objects	A	B	C	D	E	F	G	H
o_1	1	0	1	0	1	0	0	1
o_2	0	1	1	0	1	0	1	1
o_3	1	0	1	0	1	0	0	1
o_4	1	0	1	0	1	0	0	1
o_5	0	1	1	0	1	1	0	0
o_6	1	0	0	1	0	1	0	1
o_7	0	1	1	0	1	1	0	1
o_8	1	0	1	0	0	1	0	1
o_9	0	1	0	1	0	1	1	0

Association rules. An *association rule* $r : X \rightarrow Y$ is an implication where X and Y are patterns of $\mathcal{D}$. X is the *antecedent* of r and Y its *consequent*. $\mathcal{F}(XY)$

is the rule frequency, $\mathcal{F}(X)$ the antecedent frequency and $\mathcal{F}(Y)$ the consequent frequency. In Table 1, $r_1 : CG \rightarrow BEH$ and $r_2 : BCF \rightarrow E$ are association rules. The frequency of r_1 (resp. r_2) is equal to 1 (resp. 2), the frequency of its antecedent is 1 (resp. 2) and the frequency of its consequent is 2 (resp. 6).

Evaluating objective measures. An interestingness measure is a function which assigns a numerical value to an association rule according to its quality. A lot of interestingness measures are based on the rule, the antecedent and the consequent frequencies. We recall here the well-known Piatetsky-Shapiro's properties [16] which aim at specifying what a "good" measure is. In the next section, we will use properties P2 and P3 to define the SBMs.

Definition 1 (Piatetsky-Shapiro's properties). *Let $r : X \rightarrow Y$ be an association rule and M an interestingness measure.*

- *P1: $M(r) = 0$ if X and Y are statistically independent i.e. if $|\mathcal{D}| \times \mathcal{F}(XY) = \mathcal{F}(X) \times \mathcal{F}(Y)$;*
- *P2: When $\mathcal{F}(X)$ and $\mathcal{F}(Y)$ remain unchanged, $M(r)$ monotonically increases with $\mathcal{F}(XY)$;*
- *P3: When $\mathcal{F}(XY)$ and $\mathcal{F}(X)$ (resp. $\mathcal{F}(Y)$) remain unchanged, $M(r)$ monotonically decreases with $\mathcal{F}(Y)$ (resp. $\mathcal{F}(X)$).*

P2 ensures the increase of M according to the rule frequency and P3 the decrease of M according to the antecedent and the consequent frequencies. Most of usual measures satisfy P2 (e.g., support, confidence, interest, conviction). However, there are a few exceptions (e.g., J-measure, Goodman-Kruskal, Gini index). In [16], Piatetsky-Shapiro defines a measure called the Rule-Interest which satisfies the three properties P1, P2 and P3.

3 A Formal Framework for Objective Measures: The Set of Simultaneously Bounded Measures

This section presents our framework which gives a unified view of a large set of measures, the SBMs. The key idea is to express a measure according to variables which depend on frequencies in order to capture their joint effect. We will see in Section 4 that this rewriting provides lower bounds for the SBMs and highlights their behavior.

3.1 Measures as Functions

We rewrite any interestingness measure as a function according to the frequencies of a rule.

Definition 2 (Associated function). *Let M be an interestingness measure and $r : X \rightarrow Y$ an association rule. $\Psi_M(x, y, z)$ is the continuous function associated to M where $x = \mathcal{F}(X)$ and $y = \mathcal{F}(Y)$ and $z = \mathcal{F}(XY)$.*

For instance, the function associated to the Lift measure is: $\Psi_{Lift}(x, y, z) = \frac{z \times |\mathcal{D}|}{x \times y}$.

Let δ be the maximal authorized number of exceptions for a rule. Variables x, y and z are frequencies in the dataset and we only have to consider the case where they are greater than or equal to zero. Moreover, since the rules have less than δ exceptions, $z \geq 0$ implies $x \geq \delta$. Definition 3 underlines the influence of the rule number of exceptions.

Definition 3 (δ-dependent function). *Let M be an interestingness measure and $r : X \rightarrow Y$ an association rule. The δ-dependent function associated to M called $\Psi_{M,\delta}(x, y)$ is the two-variable function obtained by the change of variable $z = x - \delta$ in Ψ_M, i.e. $\Psi_{M,\delta}(x, y) = \Psi_M(x, y, x - \delta)$.*

Pursuing the Lift example, we obtain: $\Psi_{Lift,\delta}(x, y) = \frac{(x-\delta)\times|\mathcal{D}|}{x\times y}$.

3.2 Identifying Properties Shared by Measures

By using the previous definitions, we give now properties expressing basic characteristics of interestingness measures. These properties are on the core of our framework.

Property 1 (P2': weak P2). *Let M be an interestingness measure. Ψ_M increases with z.*

We call Property 1 *weak P2* since it is closely related to Piatetsky-Shapiro's property P2. The slight difference being it is not necessary that the measure monotonically increases.

Property 2 (P3': weak P3). *Let M be an interestingness measure. Ψ_M decreases with y.*

Property 2 is called *weak P3* since it corresponds to the first part of P3 (as well as P2, the definition does not require the *monotonical* decrease). Contrary to the Shapiro's set of properties, we do not make assumptions on the measure's behaviour according to the antecedent frequency. P3' only considers the consequent frequency and, unlike P3, does not require the symmetry between the antecedent and the consequent. As $\Psi_{Lift}(x, y, z)$ increases with z and decreases with y, it is immediate that the Lift satisfies P2' and P3'.

The link between the frequencies expressed by Definition 3 captures an important feature of an interestingness measure: its behavior with respect to the joint development of the antecedent and the consequent frequencies and the maximal rule number of exceptions. This characteristic is translated by Property 3 and we will use it in our framework.

Property 3 (P4: property of δ-dependent growth). *Let M be an interestingness measure. $\Psi_{M,\delta}$ increases with x.*

3.3 SBMs

Property 4 defines the SBMs. It establishes a powerful framework for analyzing the behavior of interestingness measures. Table 2 provides a sample of SBMs. The Rule-Interest measure (RI), which is a good measure according to Definition 1, belongs to this framework.

Property 4 (SBM). *An interestingness measure M is a simultaneously lower bounded measure (or SBM) if M satisfies P2', P3' and P4.*

Theorem 1 states that a linear combination of SBMs with positive coefficients is still a SBM. It also shows that the set of SBMs is infinite.

Table 2. A sample of SBMs

SBM	Definition
Support	$\frac{\mathcal{F}(XY)}{\lvert\mathcal{D}\rvert}$
Confidence	$\frac{\mathcal{F}(XY)}{\mathcal{F}(X)}$
Sensitivity	$\frac{\mathcal{F}(XY)}{\mathcal{F}(Y)}$
Specificity	$1 - \frac{\mathcal{F}(X) - \mathcal{F}(XY)}{\lvert\mathcal{D}\rvert - \mathcal{F}(Y)}$
Success Rate	$\frac{\lvert\mathcal{D}\rvert - \mathcal{F}(Y) - \mathcal{F}(X) + 2\mathcal{F}(XY)}{\lvert\mathcal{D}\rvert}$
Lift	$\frac{\lvert\mathcal{D}\rvert \times \mathcal{F}(XY)}{\mathcal{F}(Y) \times \mathcal{F}(X)}$
Rule-Interest [16]	$\mathcal{F}(XY) - \frac{\mathcal{F}(Y) \times \mathcal{F}(X)}{\lvert\mathcal{D}\rvert}$
Laplace (k=2)	$\frac{\mathcal{F}(XY) + 1}{\mathcal{F}(X) + 2}$
Odds ratio	$\frac{\mathcal{F}(XY)}{\mathcal{F}(X) - \mathcal{F}(XY)} \times \frac{\lvert\mathcal{D}\rvert - \mathcal{F}(Y) - \mathcal{F}(X) + \mathcal{F}(XY)}{\mathcal{F}(Y) - \mathcal{F}(XY)}$
Growth rate	$\frac{\mathcal{F}(XY)}{\mathcal{F}(X) - \mathcal{F}(XY)} \times \frac{\lvert\mathcal{D}\rvert - \mathcal{F}(Y)}{\mathcal{F}(Y)}$
Sebag & Schoenauer	$\frac{\mathcal{F}(XY)}{\mathcal{F}(X) - \mathcal{F}(XY)}$
Jaccard	$\frac{\mathcal{F}(XY)}{\mathcal{F}(Y) + \mathcal{F}(X) - \mathcal{F}(XY)}$
Conviction	$\frac{\lvert\mathcal{D}\rvert - \mathcal{F}(Y)}{\lvert\mathcal{D}\rvert} \times \frac{\mathcal{F}(X)}{\mathcal{F}(X) - \mathcal{F}(XY)}$
ϕ-coefficient	$\frac{\lvert\mathcal{D}\rvert \times \mathcal{F}(XY) - \mathcal{F}(Y) \times \mathcal{F}(X)}{\sqrt{\mathcal{F}(X) \times \mathcal{F}(Y) \times (\lvert\mathcal{D}\rvert - \mathcal{F}(X)) \times (\lvert\mathcal{D}\rvert - \mathcal{F}(Y))}}$
Added Value	$\frac{\mathcal{F}(XY)}{\mathcal{F}(X)} - \frac{\mathcal{F}(Y)}{\lvert\mathcal{D}\rvert}$
Certainty Factor	$\frac{\mathcal{F}(XY) \times \lvert\mathcal{D}\rvert - \mathcal{F}(X) \times \mathcal{F}(Y)}{\mathcal{F}(X) \times (\lvert\mathcal{D}\rvert - \mathcal{F}(Y))}$
Information Gain	$\log\left(\frac{\mathcal{F}(XY)}{\mathcal{F}(X)} \times \frac{\lvert\mathcal{D}\rvert}{\mathcal{F}(Y)}\right)$

Theorem 1. *Let $M_1, \ldots, M_n$ be SBMs and $\alpha_1, \ldots, \alpha_n$ be n positive real numbers. $\alpha_1 \times M_1 + \cdots + \alpha_n \times M_n$ is a SBM.*

The key idea of the proof relies on the fact that when multiplying a SBM M by a positive real number α, the associated function $\Psi_{\alpha M}$ and the δ-dependent function $\Psi_{\alpha M,\delta}$ behave like Ψ_M and $\Psi_{M,\delta}$.

Proof. We denote $\alpha_1 \times M_1 + \cdots + \alpha_n \times M_n$ by M. Let us show that M satisfies P2', P3' and P4. The following equalities hold: $\Psi_M = \alpha_1 \Psi_{M_1} + \cdots + \alpha_n \Psi_{M_n}$ and $\Psi_{M,\delta} = \alpha_1 \Psi_{M_1,\delta} + \cdots + \alpha_n \Psi_{M_n,\delta}$. Since $M_1, \ldots, M_n$ are SBMs, they satisfy P2', P3' and P4. $\Psi_{M_1}(x, y, z), \ldots, \Psi_{M_n}(x, y, z)$ increase with z and decrease with y, e.g., the partial derivatives of $\Psi_{M_1}, \ldots, \Psi_{M_n}$ w.r.t. z are positive and their partial derivatives w.r.t. y are negative. Thus the partial derivative of Ψ_M w.r.t. z remains positive and the partial derivative of Ψ_M w.r.t. y remains negative. We conclude that Ψ_M also increases with z and decreases with y. Thus M satisfies P2' and P3'. By the same reasoning, we prove that M satisfies P4 and we conclude that M is a SBM. □

Theorem 1 can be used to define new SBMs or to check if a candidate interestingness measure is a SBM. For instance, the Novelty [14] (defined by $Nov(r) = \dfrac{\mathcal{F}(XY) \times |\mathcal{D}| - \mathcal{F}(X) \times \mathcal{F}(Y)}{|\mathcal{D}|^2}$) can be expressed according to the Rule-Interest since $Nov = \alpha \times RI$ with $\alpha = \frac{1}{|\mathcal{D}|}$. As the Rule-Interest is a SBM and α is a positive real number, Theorem 1 ensures that Novelty is a SBM as well.

4 SBMs' Bounds and Behavior

This section provides lower bounds for the SBMs. We show that all the SBMs can be simultaneously lower bounded and behave in a similar way. Let γ be the minimal antecedent frequency and η the maximal consequent frequency. Except for Property 5, the values of the parameters γ, η and δ are fixed.

4.1 Lower Bounds

Theorem 2 provides for each SBM its lower bound according to γ, η and δ. Such a bound expresses the minimal quality of a rule according to γ, η and δ. Table 3 gives the lower bounds for SBMs quoted in Table 2. With $\gamma = 3$, $\delta = 1$ and $\eta = 5$ in Table 1, the Lift lower bound is 4.6 and the Rule-Interest lower bound is $\frac{1}{27}$. Note that Theorem 1 enables to calculate $\Psi_{Nov,\delta}(\gamma, \eta)$ with $\alpha \times \Psi_{RI,\delta}(\gamma, \eta)$ where $\alpha = \frac{1}{|\mathcal{D}|}$.

Theorem 2 (Lower bounds). *Let M be a SBM. If $r : X \rightarrow Y$ is an association rule such that $\mathcal{F}(X) \geq \gamma$, $\mathcal{F}(Y) \leq \eta$ and r admits less than δ exceptions, then $M(r)$ is greater than or equal to $\Psi_{M,\delta}(\gamma, \eta)$.*

Proof. According to P2', $\Psi_M(x, y, z)$ increases with the variable z. Since $X \rightarrow Y$ admits less than δ exceptions, $\mathcal{F}(XY) \geq \mathcal{F}(X) - \delta$ and consequently $\Psi_M(x, y, z) \geq \Psi_M(x, y, x - \delta) = \Psi_{M,\delta}(x, y)$. A lower bound for x is γ and a upper bound for y is η thus, since $\Psi_{M,\delta}$ increases with x and decreases with y (consequence of weak P3), a lower bound for $\Psi_{M,\delta}(x, y)$ is $\Psi_{M,\delta}(\gamma, \eta)$. □

Table 3. Lower bounds for SBMs defined in Table 2

SBM	Lower bound
Support	$\frac{\gamma - \delta}{\lvert\mathcal{D}\rvert}$
Confidence	$1 - \frac{\delta}{\gamma}$
Sensitivity	$\frac{\gamma - \delta}{\eta}$
Specificity	$1 - \frac{\delta}{\lvert\mathcal{D}\rvert - \eta}$
Success Rate	$1 + \frac{\gamma - 2\delta - \eta}{\lvert\mathcal{D}\rvert}$
Lift	$(1 - \frac{\delta}{\gamma}) \times \frac{\lvert\mathcal{D}\rvert}{\eta}$
Rule-Interest	$\gamma - \delta - \frac{\gamma\,\eta}{\lvert\mathcal{D}\rvert}$
Laplace (k=2)	$\frac{\gamma - \delta + 1}{\gamma + 2}$
Odds ratio	$[\frac{\gamma - \delta}{\eta - \gamma + \delta}] \times [\frac{\lvert\mathcal{D}\rvert - \eta - \delta}{\delta}]$
Growth rate	$\frac{\gamma - \delta}{\delta} \times \frac{\lvert\mathcal{D}\rvert - \eta}{\eta}$
Sebag & Schoenauer	$\frac{\gamma - \delta}{\delta}$
Jaccard	$\frac{\gamma - \delta}{\eta + \delta}$
Conviction	$\frac{\lvert\mathcal{D}\rvert - \eta}{\lvert\mathcal{D}\rvert} \times \frac{\gamma}{\delta}$
ϕ-coefficient	$\frac{\gamma \times (\lvert\mathcal{D}\rvert - \eta) - \delta \times \lvert\mathcal{D}\rvert}{\sqrt{\gamma \times (\lvert\mathcal{D}\rvert - \gamma) \times \eta \times (\lvert\mathcal{D}\rvert - \eta)}}$
Added Value	$\frac{\gamma - \delta}{\gamma} - \frac{\eta}{\lvert\mathcal{D}\rvert}$
Certainty Factor	$\frac{\gamma \times (\lvert\mathcal{D}\rvert - \eta) - \delta \times \lvert\mathcal{D}\rvert}{\gamma \times (\lvert\mathcal{D}\rvert - \eta)}$
Information Gain	$\log\left(\frac{\gamma - \delta}{\gamma} \times \frac{\lvert\mathcal{D}\rvert}{\eta}\right)$

As Theorem 2 is true for all the SBMs, we deduce that all the SBMs are simultaneously lower bounded. It means that the set of rules such that $\mathcal{F}(X) \geq \gamma$, $\mathcal{F}(Y) \leq \eta$ and admitting less than δ exceptions simultaneously satisfy minimal values according to all the SBMs. Thus, Theorem 2 enables to identify a set of "good" rules according to the SBMs because all the SBMs have high values,

at least greater than or equal to their lower bounds. In the following, we are interested in *all* the rules r such that $M(r) \geq \Psi_{M,\delta}(\gamma, \eta)$ for all the SBMs:

Definition 4 (SBM rule). *The set of rules satisfying $M(r) \geq \Psi_{M,\delta}(\gamma, \eta)$ for all the SBMs is denoted by $\mathcal{R}_{SBM}$. A SBM rule is a rule belonging to $\mathcal{R}_{SBM}$.*

4.2 SBMs' Behavior

Property 5 specifies the behavior of the lower bounds according to γ, η and δ.

Property 5. *$\Psi_{M,\delta}(\gamma, \eta)$ increases with γ and decreases with η and δ.*

Proof. As M is a SBM, it is obvious that $\Psi_{M,\delta}(\gamma, \eta)$ increases with γ (P4) and decreases with η (weak P3). From weak P2, it follows that $\Psi_M(x, y, z)$ increases with z. Hence assuming $\delta_1 \geq \delta_2$ we have $\Psi_M(x, y, x - \delta_2) \geq \Psi_M(x, y, x - \delta_1)$ and $\Psi_{M,\delta_2}(\gamma, \eta) \geq \Psi_{M,\delta_1}(\gamma, \eta)$. Consequently, $\Psi_{M,\delta}(\gamma, \eta)$ decreases with δ. □

Property 5 states that all the lower bounds behave in a similar way according to the parameters γ, η and δ. Consequently, it is possible to increase the rule quality according to the SBMs by increasing γ and decreasing η and δ.

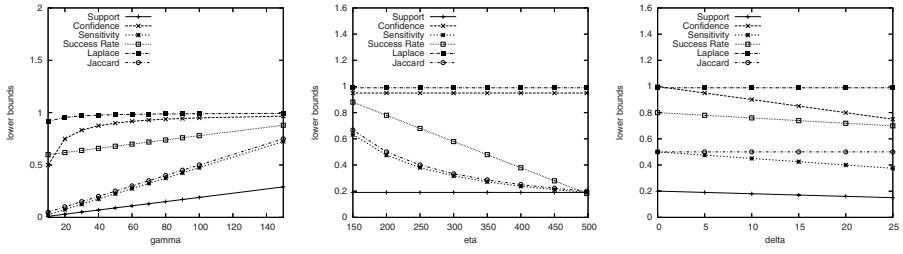

Fig. 1. Lower bounds according to γ, δ and η

For some usual measures, Figure 1 depicts the lowers bounds according to γ (with $\eta = 200$ and $\delta = 5$), η (with $\gamma = 100$ and $\delta = 5$) and δ (with $\eta = 200$ and $\gamma = 100$). These figures show the similar behavior of SBMs and that these measures can be simultaneously optimized.

5 Rule Mining

In this section, we start by characterizing $\mathcal{R}_{SBM}$. This characterization enables to infer an efficient rule mining algorithm.

5.1 Characterizing $\mathcal{R}_{SBM}$

Theorem 3 provides properties on the frequencies of a SBM rule.

Theorem 3. *If $r : X \rightarrow Y$ is a SBM rule then r satisfies the following conditions: $\mathcal{F}(X) \geq \gamma$, $\mathcal{F}(Y) \leq \eta$ and r admits less than δ exceptions.*

Proof. We define $M_1(r) = \mathcal{F}(X)$, $M_2(r) = \dfrac{1}{\mathcal{F}(Y)}$ and $M_3(r) = \dfrac{1}{\mathcal{F}(X) - \mathcal{F}(XY)}$. It is trivial to check that $M1$, $M2$ and $M3$ are SBMs. The inequalities $M_1(r) \geq \Psi_{M_1,\delta}(\gamma, \eta) = \gamma$, $M_2(r) \geq \Psi_{M_2,\delta}(\gamma, \eta) = \dfrac{1}{\eta}$ and $M_3(r) \geq \Psi_{M_3,\delta}(\gamma, \eta) = \dfrac{1}{\delta}$ immediately prove the result. □

Theorem 3 is the converse of Theorem 2. These two theorems prove that $\mathcal{R}_{SBM}$ is equal to the set of rules having a γ-frequent antecedent, an η-infrequent consequent and less than δ exceptions. Thus, even if the set of SBMs is infinite, this characterization of $\mathcal{R}_{SBM}$ makes feasible the mining of the rules optimizing all SBMs and ensures the completeness of the mining. The next section shows that we can only mine a reduced set of rules having minimal antecedents among $\mathcal{R}_{SBM}$.

5.2 Informative Rules of $\mathcal{R}_{SBM}$

Section 2.1 has introduced the rule cover based on informative rules [15]. Informative rules are build with minimal patterns (also called *free* [5] or key patterns [15]) as antecedents and one part of their closures (see [21] for a definition) as consequents. By analogy in Definition 5, we call an informative SBM rule a rule having a minimal pattern (i.e., free pattern) as antecedent and one part of its closure as consequent.

Definition 5 (Informative SBM rules). *An* informative SBM rule $r : X \rightarrow Y$ *is a SBM rule such that X is a free pattern and XY is a closed pattern. Thus r satisfies:*

- *X is γ-frequent and free*
- *Y is η-infrequent*
- *$X \cap Y = \emptyset$*
- *XY is closed*
- *r has less than δ exceptions*

The set of informative SBM rules is denoted by $\mathcal{I}nf(\mathcal{R}_{SBM})$.

This definition is precious in practice to mine the informative SBM rules because there are efficient algorithms to extract the free or key patterns and their closures [5]. The next section provides an algorithm which mines the whole set of informative SBM rules.

5.3 Algorithm Mining $\mathcal{I}nf(\mathcal{R}_{SBM})$

This section gives the main features of our algorithm for mining $\mathcal{I}nf(\mathcal{R}_{SBM})$. The basic principle is to associate the free and the closed patterns given in input to build the informative SBM rules. Definition 4 states that the SBM rules satisfy the following constraints: $\mathcal{F}(X) \geq \gamma$, $\mathcal{F}(Y) \leq \eta$ and $\mathcal{F}(X) - \mathcal{F}(XY) \leq \delta$. These constraints lead to Property 6 which provides pruning conditions:

Property 6. *The SBM rules satisfy:*

1. $\gamma \leq \mathcal{F}(X) \leq \eta + \delta$
2. $\gamma - \delta \leq \mathcal{F}(XY) \leq \mathcal{F}(Y) \leq \eta$

Proof.

1. Since $Y \subset XY$, $\mathcal{F}(XY) < \mathcal{F}(Y) \leq \eta$. $\mathcal{F}(XY) \leq \eta$ is obvious. Thus $\mathcal{F}(X) \leq \mathcal{F}(XY) + \delta$ and we have $\mathcal{F}(X) \leq \eta + \delta$.
2. Since $Y \subset XY$, $\mathcal{F}(XY) < \mathcal{F}(Y)$. We have $\gamma - \delta \leq \mathcal{F}(X) - \delta \leq \mathcal{F}(XY)$. Thus $\gamma - \delta \leq \mathcal{F}(Y)$.

Algorithm 1 considers each pattern X in $\mathcal{F}ree_{(\gamma,\eta+\delta)}$, i.e., each free pattern having a frequency between γ and $\eta + \delta$. Then the closed patterns containing X and having a frequency between $\gamma - \delta$ and η are determined. $\mathcal{I}$ is the set of discovered informative SBM rules. Note that the antecedent and the rule satisfy the frequency constraints of Property 6 by construction. Then, the number of exceptions and the consequent frequency are checked. This latter is obviously greater than $\gamma - \delta$ but not necessarily less than η. The consequent frequency is computed by finding the smallest closed pattern containing the rule consequent. When discovered, a valid rule is added to $\mathcal{I}$. The algorithm stops when all the patterns in $\mathcal{F}ree_{(\gamma,\eta+\delta)}$ have been considered.

Data: $\mathcal{F}ree$ the set of free patterns, $\mathcal{C}losed$ the set of closed patterns
Result: the informative SBM rules $\mathcal{I}nf(\mathcal{R}_{SBM})$

```
foreach X ∈ Free_(γ,η+δ) such that F(X) − F(Z) ≤ δ do
    foreach Z = XY ∈ Closed_(γ−δ,η) such that F(X) − F(Z) ≤ δ do
        if F(Y) ≤ η then
            I = I ∪ {X → Y}
        end
    end
end
return I
```

Algorithm 1. Mining $\mathcal{I}nf(\mathcal{R}_{SBM})$

6 Experiments

The aim of the experiments is twofold: first, we quantify the size of $\mathcal{I}nf(\mathcal{R}_{SBM})$ according to the parameters of our framework, and second we observe the quality of the informative SBM rules mined in practice. Experiments are performed on a real data set, the hepatitis data collected at the Chiba University Hospital (Japan). These data are used in discovery challenges [11]. They contain the examinations of 499 patients which are described with 168 attributes.

Number of informative SBM rules. Figures on the top of Figure 2 plot on a logarithmic scale the size of $\mathcal{I}nf(\mathcal{R}_{SBM})$ i.e. the number of informative SBM rules according to the minimal antecedent frequency threshold γ (on the left) and

the maximal number of exceptions δ (on the right) with $\eta = 200$. The figure on the bottom of Figure 2 plots the size of $\mathcal{I}nf(\mathcal{R}_{SBM})$ according to the maximal consequent frequency threshold η with $\gamma = 60$. As expected (cf. Property 5), the number of rules clearly decreases according to γ and increases both with η and δ. Nevertheless, these curves specify how these numbers vary.

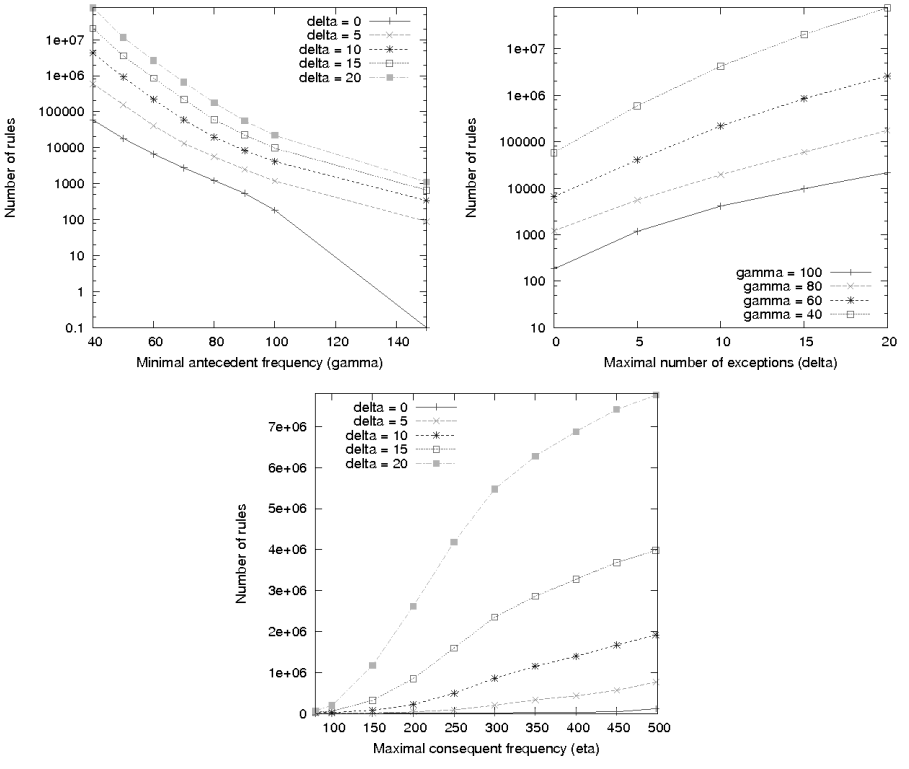

Fig. 2. Size of $\mathcal{I}nf(\mathcal{R}_{SBM})$ according to γ, δ and η

Figure 3 plots the number of rules with $\eta = 200$ and without maximal consequent frequency i.e., $\eta = 500$ (since the hepatitis data only contain 499 objects). It shows the reduction of the rule number due to η. For instance, with $\gamma = 80$, there are 5605 rules with $\eta = 200$ versus almost 200.000 rules with $\eta = 500$. Clearly, bounding the consequent frequency enables to drastically reduce the size of the output. This result is interesting because we know (thanks to Property 5) that the discarded rules have the worst values according to the set of SBMs.

Quality of the mined rules. We now focus on the quality of the informative SBM rules. With $\gamma = 60$, $\eta = 200$ and $\delta = 5$, $\mathcal{I}nf(\mathcal{R}_{SBM})$ includes 40697 rules. Table 4 indicates the minimal value, the lower bound (calculated with the expressions given in Table 3), the average value for the rules in $\mathcal{I}nf(\mathcal{R}_{SBM})$

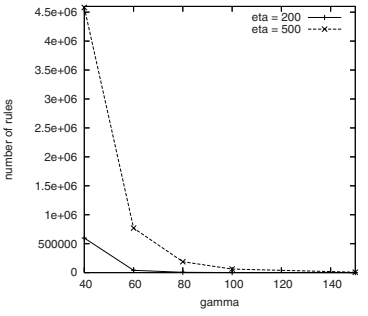

Fig. 3. Number of rules with and without a maximal consequent frequency

and the maximum value of a few SBMs. These results show for each SBM the minimal value guaranteed by our framework. Obviously, the average values are higher than the lower bounds and the difference between the average value and the lower bound depends on the measures. For instance, the Sensitivity ranges from 0 to 1. Its lower bound equals 0.275 and its average value is 0.411. For the Sebag & Schoenauer's measure (ranging from 0 to infinity), the lower bound is 11 while the average value is about 21.

Table 4. Minimum, lower bound, average value and maximum of a few SBMs

Measure	Support	Confidence	Sensitivity	Rule-Interest	Odds Ratio
Minimum	0	0	0	-0.25	0
Lower bound	0.1102	0.917	0.275	0.06203	22.303
Average value	0.134	0.962	0.411	0.089	74.038
Maximum	1	1	1	0.25	∞
Measure	GR	Sebag & Schoenauer	Jaccard	ϕ-Coefficient	Added Value
Minimum	0	0	0	-1	-0.5
Lower bound	16.45	11	0.300	0.452	0.599
Average value	43.616	21.274	0.406	0.540	0.626
Maximum	∞	∞	1	1	1

7 Conclusion and Future Work

Further work addresses the multi-criteria optimization of the SBMs. Theorem 1 shows that it is possible to combine several SBMs without loosing the properties of our framework. An approach is to get a lower bound for a weighted combination of SBMs in order to ensure a global quality for all SBMs. Another way is to automatically determine the parameters involved in the mining of the SBM rules in order to take into account the various semantics conveyed by the measures during the mining process.

Acknowledgements. The authors thank Nicolas Durand for preparing the hepatitis data. This work has been partially funded by the ACI "masse de données" (French Ministry of research), Bingo project (MD 46, 2004-2007).

References

[1] Agrawal, R., Imielinski, T., Swami, A.N.: Mining association rules between sets of items in large databases. In: Buneman, P., Jajodia, S. (eds.) SIGMOD'93 Conference, pp. 207–216. ACM Press, New York (1993)

[2] Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: Bocca, J.B., Jarke, M., Zaniolo, C. (eds.) VLDB'94, Proceedings of 20th International Conference on Very Large Data Bases, Santiago de Chile, Chile, September 12-15, 1994, pp. 487–499. Morgan Kaufmann, San Francisco (1994)

[3] Bayardo, J.R.J., Agrawal, R.: Mining the most interesting rules. In: KDD'99, pp. 145–154 (1999)

[4] Blanchard, J., Guillet, F., Briand, H.: A user-driven and quality-oriented visualization for mining association rules. In: the 3rd IEEE International Conference on Data Mining (ICDM 2003), pp. 493–496. IEEE Computer Society Press, Los Alamitos (2003)

[5] Boulicaut, J.-F., Bykowski, A., Rigotti, C.: Approximation of frequency queries by means of free-sets. In: Zighed, D.A., Komorowski, H.J., Zytkow, J.M. (eds.) PKDD 2000. LNCS (LNAI), vol. 1910, pp. 75–85. Springer, Heidelberg (2000)

[6] Brin, S., Motwani, R., Silverstein, C.: Beyond market baskets: Generalizing association rules to correlations. In: Peckham, J. (ed.) SIGMOD 1997, Proceedings ACM SIGMOD International Conference on Management of Data, Tucson, Arizona, USA, May 13-15, 1997, pp. 265–276. ACM Press, New York (1997)

[7] Francisci, D., Collard, M.: Multi-criteria evaluation of interesting dependencies according to a data mining approach. In: Congress on Evolutionary Computation, Canberra, Australia, 12, pp. 1568–1574. IEEE Computer Society Press, Los Alamitos (2003)

[8] Fürnkranz, J., Flach, P.A.: Roc 'n' rule learning-towards a better understanding of covering algorithms. Machine Learning 58(1), 39–77 (2005)

[9] Gasmi, G., Yahia, S.B., Nguifo, E.M., Slimani, Y.: Igb: A new informative generic base of association rules. In: Ho, T.-B., Cheung, D., Liu, H. (eds.) PAKDD 2005. LNCS (LNAI), vol. 3518, pp. 81–90. Springer, Heidelberg (2005)

[10] Hilderman, R.J., Hamilton, H.J.: Measuring the interestingness of discovered knowledge: A principled approach. Intell. Data Anal. 7(4), 347–382 (2003)

[11] Hirano, S., Tsumoto, S.: Guide to the hepatitis data. In: ECML/PKDD'05 Discovery Challenge on hepatitis data co-located with the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'05), Porto, Portugal, October 2005, pp. 120–124 (2005)

[12] Hébert, C., Crémilleux, B.: Optimized rule mining through a unified framework for interestingness measures. In: Tjoa, A.M., Trujillo, J. (eds.) DaWaK 2006. LNCS, vol. 4081, pp. 238–247. Springer, Heidelberg (2006)

[13] Kryszkiewicz, M.: Concise representations of association rules. In: Hand, D.J., Adams, N.M., Bolton, R.J. (eds.) Pattern Detection and Discovery. LNCS (LNAI), vol. 2447, pp. 92–109. Springer, Heidelberg (2002)

[14] Lavrac, N., Flach, P., Zupan, B.: Rule evaluation measures: a unifying view. In: Džeroski, S., Flach, P.A. (eds.) Inductive Logic Programming. LNCS (LNAI), vol. 1634, pp. 174–185. Springer, Heidelberg (1999)
[15] Pasquier, N., Bastide, Y., Taouil, R., Lakhal, L.: Discovering frequent closed itemsets for association rules. In: Beeri, C., Bruneman, P. (eds.) ICDT 1999. LNCS, vol. 1540, pp. 299–312. Springer, Heidelberg (1998)
[16] Piatetsky-Shapiro, G.: Discovery, analysis, and presentation of strong rules. In: Knowledge Discovery in Databases, pp. 229–248. AAAI/MIT Press, Cambridge (1991)
[17] Plasse, M., Niang, N., Saporta, G., Leblond, L.: Une comparaison de certains indices de pertinence des règles d'association. In: Ritschard, G., Djeraba, C. (eds.) EGC. Revue des Nouvelles Technologies de l'Information, vol. RNTI-E-6, pp. 561–568. Cépaduès-Éditions (2006)
[18] Sebag, M., Schoenauer, M.: Generation of rules with certainty and confidence factors from incomplete and incoherent learning bases. In: Boose, M. L. J., Gaines, B. (eds.)European Knowledge Acquisistion Workshop, EKAW'88, pages 28–1–28–20 (1988)
[19] Tan, P.-N., Kumar, V., Srivastava, J.: Selecting the right interestingness measure for association patterns. In: KDD, pp. 32–41. ACM, New York (2002)
[20] Vaillant, B., Lenca, P., Lallich, S.: A clustering of interestingness measures. In: The 7th International Conference on Discovery Science, 10, 2004, pp. 290–297 (2004)
[21] Wille, R.: chapter Restructuring lattice theory: an approach based on hierachies of concepts. In: Ordered sets, pp. 445–470. Reidel, Dordrecht (1982)
[22] Zaki, M.J.: Generating non-redundant association rules. In: KDD'00, pp. 34–43 (2000)