



# How Your Supporters and Opponents Define Your Interestingness

Bruno Crémilleux<sup>1</sup>, Arnaud Giacometti<sup>2</sup>, and Arnaud Soulet<sup>2</sup>(✉)

<sup>1</sup> Normandie Univ, UNICAEN, ENSICAEN, CNRS – UMR GREYC, Caen, France  
bruno.cremilleux@unicaen.fr

<sup>2</sup> Université de Tours – LIFAT EA 6300, Blois, France  
{arnaud.giacometti,arnaud.soulet}@univ-tours.fr

**Abstract.** How can one determine whether a data mining method extracts interesting patterns? The paper deals with this core question in the context of unsupervised problems with binary data. We formalize the quality of a data mining method by identifying patterns – the *supporters* and *opponents* – which are related to a pattern extracted by a method. We define a typology offering a global picture of the methods based on two complementary criteria to evaluate and interpret their interests. The quality of a data mining method is quantified via an evaluation complexity analysis based on the number of supporters and opponents of a pattern extracted by the method. We provide an experimental study on the evaluation of the quality of the methods.

## 1 Introduction

In contrast to a lot of data analysis methods where the goal is to describe all the data with one model, pattern mining focuses on information describing only parts of the data. However, in practice, the number of discovered patterns is huge and patterns have to be filtered or ranked according to additional quality criteria in order to be used by a data analyst. As surveyed by Vreeken and Tatti [26], there exist numerous methods for evaluating the interestingness of extracted patterns, e.g. based on simple measures or the use of statistical testing. However, it remains difficult to clearly identify the advantages and limitations of each approach.

How can one determine whether a data mining method extracts interesting patterns? How can one know and evaluate if a data mining method is better than another for a given task? Our work addresses these core questions. Completely answering those questions is clearly out of the scope of this (or any single) paper but we propose major improvements in these directions in the context of unsupervised problems with binary data.

Our goal is to propose an interestingness theory independent of any assumption about the data such as a model of the data or an expectation using statistical tests. Our key principle to assess the quality of a data mining method extracting a pattern  $X$  is to study the relationships between  $X$  and the other patterns when  $X$  is selected. Roughly speaking, the higher the number of necessary comparisons between  $X$  and the other patterns to select  $X$ , the higher the quality

of the method. As an example, let us consider correlation measures such as the lift [26] and the productive itemset [27]. While calculating lift involves only the individual items contained in  $X$  to select  $X$ , the productive itemset involves all subsets of  $X$ . A pattern selected by the productive itemset must satisfy more tests and the productive itemset is a more effective selector for mining correlated itemsets. Our framework addresses methods to select patterns such as interestingness measures [25], the constraint-based pattern mining imposing constraints on a single pattern [20] or several patterns such as condensed representations of patterns [5] or *top-k* patterns [9]. We call *selector* a data mining method providing patterns. The goal of our framework is to evaluate the quality of a selector and therefore the interestingness of the patterns extracted by the selector. For that purpose, we introduce the notions of *supporter* and *opponent*. A supporter  $Y$  of  $X$  is a pattern which increases the interestingness of  $X$  when only the support of  $Y$  increases while all other patterns' support remains unchanged. In other words, when the support of  $Y$  increases, it raises the likelihood of  $X$  to be selected and therefore  $Y$  supports  $X$  to be selected. Analogously, an opponent  $Y$  of  $X$  is a pattern which decreases the interestingness of  $X$  when only the support of  $Y$  decreases. We show that the number of supporters and opponents and their relations with  $X$  ( $Y$  is a generalization or a specialization of  $X$ ,  $Y$  and  $X$  are incomparable) provide meaningful information about the quality of the selector at hand. Notably, this approach evaluates the quality of a selector only based on the relationships between the patterns from the data without assuming a model or any hypothesis on the data.

This paper formalizes the relationships between patterns to evaluate the quality of a selector through the new notions of supporters and opponents. We present a typology of selectors defined by formal properties and based on two complementary criteria to evaluate and interpret the quality of a selector. This typology offers a global picture of selectors and clarifies their interests and limitations. Highlighting the kinds of patterns' relationships required by a selector helps to compare selectors to each other. We quantify the quality of a selector via an evaluation complexity analysis based on its number of supporters and opponents. This analysis enables us to contrast the quality of a selector with its computing cost. Finally, we conduct an experimental study in the context of correlation measures to evaluate the quality of selectors according to their complexity.

This paper is structured as follows. Section 2 discusses related work. Section 3 introduces preliminaries and defines what a selector is. We present the key notions of supporters and opponents in Sect. 4 and the typology of selectors in Sect. 5. We continue with the analysis of the complexity of the selectors in Sect. 6. Section 7 provides an experimental study on the evaluation of the quality of a few selectors. We round up with discussion and conclusion in Sect. 8.

## 2 Related Work

As we focus on formal approaches on interestingness, experimental protocols to evaluate the quality of a method like rediscovery [29] or randomization [14]

are out of the scope of this related work. The proposal of a general theory of interestingness was already indicated as a challenge for the past decade [8, 16].

Several approaches have been proposed in the literature to analyze pattern discovery methods. Regarding condensed representations of patterns, the size of a condensed representation is often used as an objective measure to assess its quality [5]. As a condensed representation based on closed patterns is always more compact than a condensed representation based on free (or key) patterns, closed patterns are deemed most interesting. However one of the most compact condensed representations – non-derivable itemsets (NDI) [4] – is little used. The semantics of NDI, which appears complex, may explain this unpopularity. In this paper, we propose a measure to formally identify the complexity of a selector (cf. Sect. 6). The survey of Vreeken and Tatti [26] presents interestingness measures on patterns by dividing them into two categories, absolute measures and advanced ones. An absolute measure is informally defined as follows: “score patterns using only the data at hand, without contrasting their calculations over the data to any expectation using statistical tests”. Advanced measures were introduced to limit redundancy in the results. They are based on statistical models (independence model, partition models, MaxEnt models) having different complexities. Our formalization of complexity classes based on relationships between patterns clarifies the distinction between absolute measures and advanced ones.

A lot of works [12, 18, 23, 24] propose axioms that should be satisfied by an interestingness measure for association rule in order that the measure is considered relevant. These methods state what should be the expected variations of a well-behaved measure under certain conditions (e.g., when the frequency of the body or the head of the association rule increases). More recently, these works were extended to itemsets [15, 27] but only by considering their subsets. Our proposal systematizes this approach by taking into account all patterns of the lattice. Besides, the axioms previously introduced in the literature are mainly focused on correlation measures and there are not such axioms for constraints. In this paper, we generalize these principles to constraints (cf. Sect. 5).

There are very few attempts to define interactions between patterns when evaluating an interestingness measure. The concept of global constraints has been informally defined in [6]. This notion has been formalized in [13] by defining a relational algebra extended to pattern discovery. Our framework provides a broader and more precise formal definition, especially to better analyze the interrelationships between the patterns.

### 3 Preliminaries

Let  $\mathcal{I}$  be a set of distinct literals called *items*, an itemset (or pattern) is a subset of  $\mathcal{I}$ . The language of itemsets corresponds to  $\mathcal{L} = 2^{\mathcal{I}}$ . A transactional dataset is a multi-set of itemsets of  $\mathcal{L}$ . Each itemset, usually called *transaction*, is a dataset entry. For instance, Table 2 gives three transactional datasets with 5 transactions  $t_1, \dots, t_5$  each described by 3 items  $A, B$  and  $C$ . Note that the transaction  $t_5$  in  $\mathcal{D}_1$  is empty.  $\mathcal{D}$  denotes a dataset and  $\Delta$  all datasets. The *frequency* of an

itemset  $X$ , denoted by  $freq(X, \mathcal{D})$ , is the number of transactions of  $\mathcal{D}$  containing  $X$ . For simplicity, we write  $freq(X)$  when there is no ambiguity.

Constraint-based pattern mining [19] aims at enumerating all patterns occurring at least once in a dataset  $\mathcal{D}$  and satisfying a user-defined selection predicate  $q$ . A well-known example is the minimal support constraint, based on the frequency measure, which provides the patterns having a support greater than a given minimal threshold. Despite the filtering performed by a constraint, the collection of mined patterns is often too large to be managed and interestingness measures are additionally used to rank patterns and focus on the most relevant ones. There are numerous measures [24], several of which (support, bond, lift, all-confidence) are given in Table 1. In this paper, we consider constraint-based pattern mining imposing constraints on a single pattern or several patterns such as condensed representations of patterns or top-k patterns. Table 1 depicts several examples of constraints. The productive itemset is here defined as a constraint.

**Table 1.** Itemset mining approaches based on frequency

Name	Definition
Correlation	
Support [1]	$freq(X)/freq(\emptyset)$
All-confidence [21]	$freq(X)/\max_{i \in X} freq(\{i\})$
Bond [21]	$freq(X)/ \{t \in \mathcal{D} : X \cap t \neq \emptyset\} $
Lift [26]	$supp(X)/\prod_{i \in X} supp(i)$
Productive itemset [27]	$(\forall Y \subset X)(prod(X) \Rightarrow supp(X) > supp(Y) \times supp(X \setminus Y))$
Condensed Representation (CR)	
Maximal itemset [19]	$(\forall Y \supset X)(max(X) \Rightarrow freq(X) \geq \gamma \wedge freq(Y) < \gamma)$
Free itemset [2]	$(\forall Y \subset X)(free(X) \Rightarrow freq(X) < freq(Y))$
Closed itemset [22]	$(\forall Y \supset X)(closed(X) \Rightarrow freq(X) > freq(Y))$
Non-derivable itemset [4]	$(\forall X \in \mathcal{L})(ndi(X) \Leftrightarrow LB(X, \mathcal{D}) \neq UB(X, \mathcal{D}))$ where $LB(X, \mathcal{D})$ and $UB(X, \mathcal{D})$ are respectively lower and upper bounds derived from subsets of $X$ in $\mathcal{D}$
Other	
Top- $k$ freq. itemset [9]	$(\forall X \in \mathcal{L})(top_k(X) \Leftrightarrow  \{Y \in \mathcal{L} : freq(Y) > freq(X)\}  < k)$
FPOF [17]	$-\sum_{Y \subseteq X} freq(Y)/\sum_{Z \in \mathcal{L}} freq(Z)$

Let  $\mathbb{S}$  be a poset. We formally define the notion of *selector* as follows.

**Definition 1 (Interestingness Selector).** *An interestingness selector  $s$  is a function defined from  $\mathcal{L} \times \Delta$  to  $\mathbb{S}$  that increases when  $X$  is more interesting.*

$\mathbb{S}$  is the set of reals  $\mathbb{R}$  if the selector is an interestingness measure<sup>1</sup> and booleans  $\mathbb{B}$  (i.e. *true* or *false*) with the order *false* < *true* if the selector is a constraint. Clearly, selectors define very different views on what should be a relevant pattern. Relevance may highlight correlations between items (regularity), correlations with a class of the dataset (contrast), removing redundancy (condensed representation), complementarity between several patterns (top- $k$ ), outlier detection such as the FPOF measure (cf. Table 1).

## 4 Framework of Supporters and Opponents

### 4.1 Fundamental Definitions and Notations

Deciding if a pattern is interesting (and why) generally depends on its frequency, but also on the frequencies of some other patterns. In our framework, we show how the knowledge of those patterns for a given selector makes it possible to qualify this selector and evaluate its quality.

More precisely, in order to isolate the impact of the change in frequency of a pattern  $Y$  on the evaluation of the interestingness of a pattern  $X$ , we propose to compare the interestingness of the assessed pattern  $X$  with respect to two very similar datasets  $\mathcal{D}$  and  $\mathcal{D}'$ , where only the frequency of itemset  $Y$  varies. Therefore, we introduce the following definition.

**Definition 2 (Increasing at a Point).** *Compared to  $\mathcal{D}$ , a dataset  $\mathcal{D}'$  is increasing at a point  $Y$ , denoted by  $\mathcal{D} <_Y \mathcal{D}'$ , iff  $\text{freq}(Y, \mathcal{D}) < \text{freq}(Y, \mathcal{D}')$  and  $\text{freq}(X, \mathcal{D}) = \text{freq}(X, \mathcal{D}')$  for all patterns  $X \neq Y$ .*

For instance, the first two datasets provided by Table 2 satisfy  $\mathcal{D}_1 <_{ABC} \mathcal{D}_2$ . It means that patterns  $\emptyset, A, B, C, AB, AC, BC$  have the same frequency in both datasets, while the frequency of  $ABC$  is greater in  $\mathcal{D}_2$ . Indeed, we have  $\text{freq}(ABC, \mathcal{D}_1) = 1$ , whereas  $\text{freq}(ABC, \mathcal{D}_2) = 2$ . In the same way, we can easily see that  $\mathcal{D}_1 <_A \mathcal{D}_3$  due to the addition in  $\mathcal{D}_3$  (compared to  $\mathcal{D}_1$ ) of an item  $A$  in the fifth transaction  $t_5$ . Thus, we have  $\text{freq}(A, \mathcal{D}_1) = 3$ , whereas  $\text{freq}(A, \mathcal{D}_3) = 4$ , and  $\text{freq}(X, \mathcal{D}_1) = \text{freq}(X, \mathcal{D}_3)$  for all other patterns  $X \neq A$ .

Intuitively, given a selector  $s$ , a *supporter*  $Y$  of an assessed pattern  $X$  is a pattern that increases the interestingness of  $X$  when only the support of  $Y$  increases (while all other patterns' support remains unchanged). In other words, when the support of  $Y$  increases, it raises the likelihood of  $X$  to be selected using  $s$ . Conversely, if  $Y$  is an *opponent* of  $X$ , when the support of  $Y$  increases, it reduces the likelihood of  $X$  to be selected. Using Definition 2, the following definition formalizes these notions of supporter and opponent.

**Definition 3 (Supporters and Opponents).** *Given a selector  $s$ , let  $X$  be a pattern in  $\mathcal{L}$ .  $Y$  is a supporter of  $X$  for  $s$ , denoted by  $Y \in s^+(X)$ , iff there exist two datasets  $\mathcal{D}$  and  $\mathcal{D}'$  such that  $\mathcal{D}' >_Y \mathcal{D}$  and  $s(X, \mathcal{D}') > s(X, \mathcal{D})$ .*

<sup>1</sup> The choice of the order  $<_{\mathbb{S}}$  has an impact. For instance, in the case of support,  $<_{\mathbb{S}} = <_{\mathbb{R}}$  (resp.  $<_{\mathbb{S}} = >_{\mathbb{R}}$ ) enables us to select the positive (resp. negative) correlations.

**Table 2.** Three toy datasets with slight variations

$\mathcal{D}_1$			$\mathcal{D}_2$			$\mathcal{D}_3$		
Trans.	Items		Trans.	Items		Trans.	Items	
$t_1$	A	B C	$t_1$	A	B C	$t_1$	A	B C
$t_2$	A	B	$t_2$	A	B C	$t_2$	A	B
$t_3$	A	C	$t_3$	A		$t_3$	A	C
$t_4$		B C	$t_4$		B	$t_4$		B C
$t_5$			$t_5$		C	$t_5$	A	

Conversely,  $Y$  is an opponent of  $X$  for  $s$ , denoted by  $Y \in s^-(X)$ , iff there exist two datasets  $\mathcal{D}$  and  $\mathcal{D}'$  such that  $\mathcal{D}' >_Y \mathcal{D}$  and  $s(X, \mathcal{D}') < s(X, \mathcal{D})$ .

Given a selector, the strength of the notions of *supporter* and *opponent* is to clearly identify the patterns actually involved in the evaluation of an assessed pattern. Moreover, it is important to note that the set of supporters and opponents of a pattern (given a selector) is not dependent on a specific dataset. They are a property of a given selector.

Considering the datasets given in Table 2, let us illustrate Definition 3 with the all-confidence selector. We already noted that  $\mathcal{D}_1 <_{ABC} \mathcal{D}_2$ . Additionally, we have  $all-conf(ABC, \mathcal{D}_1) = \frac{freq(ABC, \mathcal{D}_1)}{\max_{i \in ABC} freq(i, \mathcal{D}_1)} = \frac{1}{3}$ , whereas  $all-conf(ABC, \mathcal{D}_2) = \frac{freq(ABC, \mathcal{D}_2)}{\max_{i \in ABC} freq(i, \mathcal{D}_2)} = \frac{2}{3}$ . Therefore, we have  $\mathcal{D}_1 <_{ABC} \mathcal{D}_2$  and  $all-conf(ABC, \mathcal{D}_1) < all-conf(ABC, \mathcal{D}_2)$ , which means that  $ABC$  is a supporter of itself for the all-confidence measure. On the other hand, we have  $\mathcal{D}_1 <_A \mathcal{D}_3$ ,  $all-conf(ABC, \mathcal{D}_1) = \frac{1}{3} > all-conf(ABC, \mathcal{D}_3) = \frac{1}{4}$ . Thus, by Definition 3,  $A$  is an opponent of  $ABC$  for the all-confidence measure.

More generally, it is possible to show that for all patterns  $X$ ,  $all-conf^+(X)$  is equal to  $\{X\}$ , i.e.  $X$  has no supporters other than itself, and that  $all-conf^-(X) = \{i : i \in X\}$ . In the following Sect. 4.2, we give the set of supporters and opponents for a representative set of usual selectors.

## 4.2 Supporters and Opponents of Usual Selectors

In this section, we give the sets of supporters  $s^+$  and opponents  $s^-$  for a representative set of selectors  $s$ . These sets are presented in Table 3 both for interestingness measures (support, all-confidence, bond, lift, etc.) and boolean constraints (productive, free, closed itemset, etc.). Due to lack of space, we do not present a proof for every selector considered in Table 3. Nevertheless, we provide a proof for two examples: the lift measure (see Property 1) and the free constraint (see Property 2). Note that the schema of these proofs could be easily adapted to identify the supporters and opponents of other correlation measures (support, all-confidence, bond, etc.) and other condensed representation constraints (maximal, closed, etc.).

Before detailing the proofs of Properties 1 and 2, given an itemset  $X$ , Lemma 1 stresses that it is always possible to build two transactional datasets  $\mathcal{D}$

**Table 3.** Analysis of methods based on supporters and opponents

Selector $s$	$s^+(X)$	$s^-(X)$	$ s^\pm(X) $	Notation	Formula
support	$\{X\}$	$\{\emptyset\}$	$O(1)$	$k$	$ X $
all-confidence	$\{X\}$	sing.	$O(k)$	$n$	$ I $
bond	$\{X\}$	sing.	$O(k)$	<i>singletons</i>	$\{\{i\} : i \in X\}$
lift	$\{X\}$	sing.	$O(k)$	<i>direct sub.</i>	$X^\perp \{X \setminus \{i\} : i \in X\}$
prod. itemset	$\{X\}$	$X^\perp$	$O(2^k)$	<i>direct sup.</i>	$X^\top \{X \cup \{i\} : i \in I \setminus X\}$
max. itemset	$\{X\}$	$X^\top$	$O(n - k)$	<i>subsets</i>	$X^\downarrow 2^X \setminus \{X\}$
free itemset	$X^\perp$	$\{X\}$	$O(k)$	<i>supersets</i>	$X^\uparrow \{Y \in \mathcal{L} : X \subset Y\}$
closed itemset	$\{X\}$	$X^\top$	$O(n - k)$	<i>incomp.</i>	$X^\leftrightarrow \{Y \in \mathcal{L} : Y \not\subseteq X\}$
NDI	$X^\downarrow$	$X^\downarrow$	$O(2^k)$		$\wedge X \not\subseteq Y$
top- $k$ frequent	$\{X\}$	$X^\leftrightarrow$	$O(2^n - 2^k)$	<i>lattice</i>	$\mathcal{L} \setminus \{X\}$
FPOF	$X^\uparrow \cup X^\leftrightarrow$	$X^\downarrow \cup \{X\}$	$O(2^n)$		

and  $\mathcal{D}'$  such that  $\mathcal{D}'$  is increasing at  $X$  in comparison to  $\mathcal{D}$ , i.e.  $\mathcal{D}' >_X \mathcal{D}$ . Note that the sets of transactions  $\mathcal{D}_X^-$  and  $\mathcal{D}_X^+$  introduced in this lemma are crucial to identify the sets of supporters and opponents of a selector.

**Lemma 1.** *Given an itemset  $X \subseteq \mathcal{I}$ , let  $\mathcal{D}_X^-$  and  $\mathcal{D}_X^+$  be the datasets defined by  $\mathcal{D}_X^+ = \{Y \subseteq X : |X \setminus Y| \text{ is even}\}$  and  $\mathcal{D}_X^- = 2^X \setminus \mathcal{D}_X^+$ . We have  $\mathcal{D}_X^- <_X \mathcal{D}_X^+$ .*

For instance, using datasets shown in Table 2, it is easy to see that  $\mathcal{D}_2 = \{ABC\} \cup \mathcal{D}_{ABC}^+$  with  $\mathcal{D}_{ABC}^+ = \{ABC, A, B, C\}$ , and  $\mathcal{D}_1 = \{ABC\} \cup \mathcal{D}_{ABC}^-$  with  $\mathcal{D}_{ABC}^- = \{AB, AC, BC, \emptyset\}$ . Thus, Lemma 1 implies that  $\mathcal{D}_1 <_{ABC} \mathcal{D}_2$ . Given  $\mathcal{D}_0 = \{ABC, AB, AC, BC\}$ , we can also check that  $\mathcal{D}_1 = \mathcal{D}_0 \cup \mathcal{D}_A^-$  with  $\mathcal{D}_A^- = \emptyset$ , and  $\mathcal{D}_3 = \mathcal{D}_0 \cup \mathcal{D}_A^+$  with  $\mathcal{D}_A^+ = \{A\}$ . Thus, Lemma 1 implies that  $\mathcal{D}_1 <_A \mathcal{D}_3$ .

*Proof.* First, it is easy to see that  $\text{freq}(X, \mathcal{D}_X^-) = 0$  and  $\text{freq}(X, \mathcal{D}_X^+) = 1$ , which shows that  $\text{freq}(X, \mathcal{D}_X^+) > \text{freq}(X, \mathcal{D}_X^-)$ . Then, for all itemsets  $Y \neq X$ , if  $Y \not\subseteq X$ , we have  $\text{freq}(Y, \mathcal{D}_X^-) = \text{freq}(Y, \mathcal{D}_X^+) = 0$ . Otherwise, if  $Y \subseteq X$ , we can see that  $\text{freq}(Y, \mathcal{D}_X^-) = \text{freq}(Y, \mathcal{D}_X^+) = \text{freq}(Y, 2^X)/2$  where  $\text{freq}(Y, 2^X) = |\{t \in 2^X : Y \subseteq t\}| = |\{Y \cup t : t \in 2^{X \setminus Y}\}| = 2^{|X| - |Y|}$ . Thus, for all  $Y \neq X$ ,  $\text{freq}(Y, \mathcal{D}_X^-) = \text{freq}(Y, \mathcal{D}_X^+)$ , which completes the proof that  $\mathcal{D}_X^- <_X \mathcal{D}_X^+$ .  $\square$

Using Lemma 1, we now prove Property 1, which defines the supporters and opponents of the lift measure.

**Property 1.** For all itemsets  $X$  such that  $|X| > 1$ ,  $\text{lift}^+(X) = \{X\}$  and  $\text{lift}^-(X) = \{\{i\} : i \in X\}$ .

*Proof.* Given an itemset  $X$  such that  $|X| > 1$ , we distinguish three cases:

1. Let  $\mathbf{Y} = \mathbf{X}$  and two datasets  $\mathcal{D}'$  and  $\mathcal{D}$  such that  $\mathcal{D}' >_{\mathbf{Y}} \mathcal{D}$ . By definition we have:  $\text{freq}(X, \mathcal{D}') > \text{freq}(X, \mathcal{D})$  and  $\text{freq}(Z, \mathcal{D}') = \text{freq}(Z, \mathcal{D})$  for all  $Z \neq (Y = X)$ . Because  $|X| > 1$ , we also have  $\{i\} \neq X$  for all  $i \in X$ . Therefore,  $\text{freq}(\{i\}, \mathcal{D}') = \text{freq}(\{i\}, \mathcal{D})$  for all  $i \in X$ , which implies that the denominators of  $\text{lift}(X, \mathcal{D}')$  and  $\text{lift}(X, \mathcal{D})$  are

equal. Finally, we have  $\text{lift}(X, \mathcal{D}') = \frac{\text{supp}(X, \mathcal{D}')}{\prod_{i \in X} \text{supp}(\{i\}, \mathcal{D}')} > \text{lift}(X, \mathcal{D}) = \frac{\text{supp}(X, \mathcal{D})}{\prod_{i \in X} \text{supp}(\{i\}, \mathcal{D})}$ , which shows that  $X \in \text{lift}^+(X)$ , whereas  $X \notin \text{lift}^-(X)$ .

**2.** Let  $Y$  be an itemset such that  $\mathbf{Y} \neq \mathbf{X}$  and  $|\mathbf{Y}| > 1$ , and two datasets  $\mathcal{D}'$  and  $\mathcal{D}$  such that  $\mathcal{D}' >_Y \mathcal{D}$ . By definition, we have:  $\text{freq}(X, \mathcal{D}') = \text{freq}(X, \mathcal{D})$  since  $Y \neq X$ , and  $\text{freq}(\{i\}, \mathcal{D}') = \text{freq}(\{i\}, \mathcal{D})$  for all  $i \in X$  since  $Y \neq \{i\}$  (indeed, we assume that  $|Y| > 1$ ). Thus, we necessarily have  $\text{lift}(X, \mathcal{D}') = \text{lift}(X, \mathcal{D})$  for all datasets  $\mathcal{D}'$  and  $\mathcal{D}$  such that  $\mathcal{D}' >_Y \mathcal{D}$ . It implies that for all itemsets  $Y$  such that  $Y \neq X$  and  $|Y| > 1$ ,  $Y \notin \text{lift}^+(X)$  and  $Y \notin \text{lift}^-(X)$ .

**3.** Let  $Y$  be an itemset such that  $\mathbf{Y} \neq \mathbf{X}$  and  $|\mathbf{Y}| = 1$ , and two datasets  $\mathcal{D}'$  and  $\mathcal{D}$  such that  $\mathcal{D}' >_Y \mathcal{D}$ . Using the same reasoning as before, it is easy to see that if  $Y \not\subseteq X$ , we necessarily have  $\text{lift}(X, \mathcal{D}') = \text{lift}(X, \mathcal{D})$ , which implies that  $Y \notin \text{lift}^+(X)$  and  $Y \notin \text{lift}^-(X)$ . Dually, if  $Y \subset X$ , because  $|Y| = 1$ , there exists  $j \in X$  such that  $Y = \{j\}$ . Since  $\mathcal{D}' >_Y \mathcal{D}$  and  $X \neq Y$ , we have  $\text{freq}(X, \mathcal{D}') = \text{freq}(X, \mathcal{D})$  and  $\prod_{i \in X} \text{supp}(\{i\}, \mathcal{D}') > \prod_{i \in X} \text{supp}(\{i\}, \mathcal{D})$  because  $\text{freq}(\{j\}, \mathcal{D}') > \text{freq}(\{j\}, \mathcal{D})$  and  $j \in X$ . Thus, we have  $\text{lift}(X, \mathcal{D}') = \frac{\text{supp}(X, \mathcal{D}')}{\prod_{i \in X} \text{supp}(\{i\}, \mathcal{D}')} < \text{lift}(X, \mathcal{D}) = \frac{\text{supp}(X, \mathcal{D})}{\prod_{i \in X} \text{supp}(\{i\}, \mathcal{D})}$ , which shows that  $Y = \{j\} \subset X$  is an opponent of  $X$  for the lift measure (and not a supporter).  $\square$

We now consider the case of a condensed representation selector, and prove Property 2, which defines the supporters and opponents of the free constraint.

*Property 2.* For all itemsets  $X$  such that  $|X| > 1$ ,  $\text{free}^+(X) = \{X \setminus \{i\} : i \in X\}$  and  $\text{free}^-(X) = \{X\}$ .

*Proof.* Let  $X$  be an itemset such that  $|X| > 1$ . We first show that  $X^\perp \subseteq \text{free}^+(X)$ , i.e. that for all  $k \in X$ ,  $Y = X \setminus \{k\} \in \text{free}^+(X)$ . By definition, we have to find two datasets  $\mathcal{D}$  and  $\mathcal{D}'$  such that  $\mathcal{D}' >_Y \mathcal{D}$ ,  $\text{free}(X, \mathcal{D}) = \text{false}$ , whereas  $\text{free}(X, \mathcal{D}') = \text{true}$ . Let  $\mathcal{D} = \{X\} \cup \{X \setminus \{i\} : i \in Y\} \cup \mathcal{D}_Y^-$  and  $\mathcal{D}' = \{X\} \cup \{X \setminus \{i\} : i \in Y\} \cup \mathcal{D}_Y^+$ . First, it is easy to see that  $\mathcal{D}' >_Y \mathcal{D}$ . Moreover, we have  $\text{freq}(X, \mathcal{D}) = 1$  and  $\text{freq}(Y, \mathcal{D}) = 1$  since  $Y \subseteq X$  and  $Y \notin \mathcal{D}_Y^-$ . Thus,  $X$  is not a free itemset in  $\mathcal{D}$ , i.e.  $\text{free}(X, \mathcal{D}) = \text{false}$ . Then, we can see that  $\text{freq}(X, \mathcal{D}') = 1$  and  $\text{freq}(X \setminus \{i\}, \mathcal{D}') = 2$  for all  $i \in X$  (in particular, note that  $Y = (X \setminus \{k\}) \in \mathcal{D}_Y^+$ ). Thus,  $X$  is a free itemset in  $\mathcal{D}'$ , i.e.  $\text{free}(X, \mathcal{D}') = \text{true}$ , which completes the proof that  $Y = (X \setminus \{k\}) \in \text{free}^+(X)$ .

We now show that  $X \in \text{free}^-(X)$ . We have to find two datasets  $\mathcal{D}$  and  $\mathcal{D}'$  such that  $\mathcal{D}' >_X \mathcal{D}$ ,  $\text{free}(X, \mathcal{D}) = \text{true}$ , whereas  $\text{free}(X, \mathcal{D}') = \text{false}$ . Let  $\mathcal{D} = \{X\} \cup \mathcal{D}_X^-$  and  $\mathcal{D}' = \{X\} \cup \mathcal{D}_X^+$ . By construction (see the definitions of  $\mathcal{D}_X^-$  and  $\mathcal{D}_X^+$  in the proof of Lemma 1), it is clear that  $\mathcal{D}' >_X \mathcal{D}$ . Moreover, we have  $\text{freq}(X, \mathcal{D}) = 1$  and  $\text{freq}(X \setminus \{k\}, \mathcal{D}) = 2$  for all  $k \in X$  (because  $X \setminus \{k\} \subseteq X \in \mathcal{D}$ , and  $X \setminus \{k\} \in \mathcal{D}_X^-$ ). Therefore,  $X$  is a free itemset in  $\mathcal{D}$ , i.e.  $\text{free}(X, \mathcal{D}) = \text{true}$ . Then, we can also check that  $\text{freq}(X, \mathcal{D}') = 2$  and  $\text{freq}(X \setminus \{k\}, \mathcal{D}') = 2$  for all  $k \in X$ . Thus,  $X$  is not a free itemset in  $\mathcal{D}'$ , which completes the proof that  $X \in \text{free}^-(X)$ .

To complete the proof, we have to show that any other pattern  $Y \notin \{X\} \cup X^\perp$  cannot be a supporter or an opponent of  $X$ . In particular, we have to show that for all  $Y \subset X$ , if  $Y \notin X^\perp$ , then  $Y \notin \text{free}^+(X)$ , i.e. that for all databases  $\mathcal{D}$  and  $\mathcal{D}'$  such that  $\mathcal{D}' >_Y \mathcal{D}$ , it is impossible to have  $\text{free}(X, \mathcal{D}) = \text{false}$  and  $\text{free}(X, \mathcal{D}') = \text{true}$ . If  $\text{free}(X, \mathcal{D}) = \text{false}$ , it means that there exists  $k \in X$  such that  $\text{freq}(X, \mathcal{D}) = \text{freq}(X \setminus \{k\}, \mathcal{D})$ . Moreover, because  $\mathcal{D}' >_Y \mathcal{D}$  and  $Y \not\subseteq X^\perp$ , we have  $\text{freq}(X, \mathcal{D}') = \text{freq}(X \setminus \{k\}, \mathcal{D}')$ , which shows that  $X$  cannot be free in  $\mathcal{D}'$ , i.e.  $\text{free}(X, \mathcal{D}') = \text{false}$ , and contradicts the hypothesis. Thus, we have shown that only the direct subsets of  $X$  are supporters of  $X$  for the free constraint. The rest of the proof is omitted for lack of space.



To conclude this section, we stress that the strength of the concept of supporters and opponents is to clearly identify the patterns actually involved in the evaluation of a selector. For instance, whereas the definition of free itemsets given in Table 1 involves *all strict subsets* of  $X$  (with  $\forall Y \subset X$ ), we can see that *only direct subsets* of  $X$  are supporters. In the following sections, we show how supporters and opponents can be used to compare selectors (see Sect. 5), and how the number of supporters and opponents of a selector is related to its effectiveness to select interesting patterns (see Sect. 6).

## 5 Typology of Interestingness Selectors

### 5.1 Polarity of Interestingness Selectors

We distinguish two broad categories of selectors according to whether they aim at discovering over-represented phenomena in the data (e.g., positive correlation) or under-represented phenomena in the data (e.g., outlier detection). Naturally, the characterization of these categories is related to the evaluation of the frequency on the pattern to assess. For instance, it is well-known that the interestingness of a pattern  $X$  increases with its frequency for finding correlations between items. In order that an interestingness selector  $s$  will be sensitive to this variation, it is essential that  $s$  increases with the frequency of  $X$ . This principle has first been proposed for association rules [23] (Property P2) and after, extended to correlated itemsets [15, 27]. Conversely, a selector for outlier detection will favor patterns whose frequency decreases. Indeed, a pattern is more likely to be abnormal as it is not representative of the dataset i.e., its frequency is low.

We formalize these two types of patterns thanks to reflexivity property:

**Definition 4 (Positive and Negative Reflexive).** *An interestingness selector  $s$  is positive (resp. negative) reflexive iff any pattern is its own supporter i.e.,  $(\forall X \in \mathcal{L})(X \in s^+(X))$  (resp. opponent i.e.,  $(\forall X \in \mathcal{L})(X \in s^-(X))$ ).*

As  $all-conf^+(X) = \{X\}$ , the all-confidence selector is positive reflexive. Conversely, the free selector is negative reflexive because  $free^-(X) = \{X\}$  (when frequency of  $X$  increases,  $X$  is less likely to be free because its frequency becomes closer to that of its subsets).

This clear separation based on reflexive property constitutes the first analysis axis of our selector typology. Table 4 schematizes this typology where the polarity is the vertical axis of analysis. The horizontal axis (semantics) will be described in the next section. Note that the correlation measures and the closed itemsets are in the same column. Several works in the literature have shown that closed itemsets maximize classification measures [11] and correlation measures [10]. For instance, the lift of a closed pattern has the highest value of its equivalence class because the frequency of  $X$  remains the same (numerator) while the denominator decreases.

**Table 4.** Typology of interestingness selectors

		POLARITY	
		Positive $X \in s^+(X)$	Negative $X \in s^-(X)$
SEMANTICS	<b>Subsets</b> $X^\downarrow$	C1: $X^\downarrow \cap s^-(X) \neq \emptyset$ (all-confidence, bond, productive itemsets, NDI, FPOF, lift)	C2: $X^\downarrow \cap s^+(X) \neq \emptyset$ (free itemset, NDI, negative border)
	<b>Supersets</b> $X^\uparrow$	C3: $X^\uparrow \cap s^-(X) \neq \emptyset$ (closed itemsets, maximal itemsets)	$X^\uparrow \cap s^+(X) \neq \emptyset$ (FPOF)
	<b>Incomparable sets</b> $X^\leftrightarrow$	$X^\leftrightarrow \cap s^-(X) \neq \emptyset$ (top-k frequent itemsets)	$X^\leftrightarrow \cap s^+(X) \neq \emptyset$ (FPOF)

QC 2: *Compléteness*  
 $s^+(X) \cup s^-(X) = \mathcal{L}$

QC 1: *Soundness*  
 $s^+(X) \cap s^-(X) = \emptyset$

Of course, it should not be possible for an interestingness selector to both isolate over-represented phenomena (i.e., positive) and under-represented phenomena (i.e., negative). For this reason, a selector should never be both positive and negative. Besides, the behavior of an interestingness selector is easier to understand for the end user if the change in frequency of a pattern  $Y$  still impacts  $s(X)$  in the same way. In other words, the increase of  $freq(X)$  should not decrease  $s(X)$  in some cases and increase  $s(X)$  in others.

**Quality Criterion 1 (Soundness).** *An interestingness selector  $s$  is sound iff no pattern is at the same time a supporter and an opponent of another pattern:  $\forall X \in \mathcal{L}, s^+(X) \cap s^-(X) = \emptyset$ .*

When Quality Criterion 1 is violated, it makes difficult to interpret a mined pattern. For instance, frequent free itemset mining is not sound. There are two opposite reasons for explaining that a pattern is not extracted: its frequency is too low (non-frequent rejection), or its frequency is too high (non-free rejection). Conversely, for frequent closed patterns, a pattern is not extracted if and only if its frequency is too low (whatever the underlying cause: the pattern is not frequent or non-closed). It means that frequent closed pattern mining is sound. We therefore think that the violation of Quality Criterion 1 (where  $s^+(X) = s^-(X) = X^\downarrow$ ) could partly explain the failure of NDI (non-derivable itemsets) even if they form an extremely compact condensed representation.

**Recommendation.** A well-behaving pattern mining method should not mix interestingness selectors with opposite polarities or make possible the existence of patterns that are supporters and opponents of the same pattern.

Before describing the semantics axis of our typology, Table 4 classifies all the selectors presented in Table 1. As expected, all selectors seeking to isolate over-represented phenomena are in the Positive column.

## 5.2 Semantics of Interestingness Selectors

This section presents three complementary criteria to identify the nature of an interestingness selector. The key idea is to focus on the relationships between patterns to qualify the semantics of the selector. More precisely, the meaning of a positive selector (whose primary objective is to find over-represented patterns) depends strongly on the set of opponents that can lead to the rejection of the assessed pattern. Conversely, a negative reflexive selector relies often on supporters to better isolate under-represented phenomena. For this reason, the positive (resp. negative) column of Table 4 involves opponents  $s^-(X)$  (resp. supporters  $s^+(X)$ ).

Furthermore, for two selectors of the same polarity, it is possible to distinguish their goals (e.g., correlation or condensed representation) according to the opponents/supporters that they involve. Thus, we break down the semantics axis into three parts: subsets  $X^\downarrow = \{Y \subset X\}$ , supersets  $X^\uparrow = \{Y \supset X\}$  and incomparable sets  $X^\leftrightarrow = \{Y \in \mathcal{L} : Y \not\subseteq X \wedge Y \not\supseteq X\}$ . This decomposition of the lattice of the opponents and the lattice of the supporters is useful to redefine coherent classes of usual selectors (these classes are indicated in Table 4):

**Definition 5 (Selector Classes).** *An interestingness selector  $s$  belongs to:*

- **C1 (Positive correlation)** iff  $(\forall X \in \mathcal{L})(X^\downarrow \cap s^-(X) \neq \emptyset)$
- **C2 (Minimal condensed representation)** iff  $(\forall X \in \mathcal{L})(X^\downarrow \cap s^+(X) \neq \emptyset)$
- **C3 (Maximal condensed representation)** iff  $(\forall X \in \mathcal{L})(X^\uparrow \cap s^-(X) \neq \emptyset)$

Intuitively a pattern is a set of correlated items (or correlated in brief) when its frequency is higher than what was expected by considering the frequency of some of its subsets (this set of opponents varies depending on the statistical model). This means that the increase of the frequency of one of these subsets may lead to the rejection of the assessed pattern. In other words, a correlation measure is based on subsets as opponents. This observation has already been made in the literature for association rules [23] (with Property P3) and itemsets [15, 27]. Table 4 shows that most of correlation measures in the literature satisfy  $(\forall X \in \mathcal{L})(X^\downarrow \cap s^-(X) \neq \emptyset)$ . The extraction of NDI, classified as a condensed representation, also meets this criterion. It is intriguingly since the NDI selector is not usually used as a correlation measure.

A condensed representation is a reduced collection of patterns that can regenerate some properties of the full collection of patterns. Typically, frequent closed patterns enable to retrieve the exact frequency of any frequent pattern. Most approaches are based on the notion of equivalence class where two patterns are equivalent if they have the same value for a function  $f$  and if they are comparable. The equality for  $f$  and the comparability result in an interrelation between the assessed pattern and its subsets/supersets. Class C3 (i.e., maximal condensed

representations) includes the measures that remove the assessed pattern when a more specific pattern provides more information. Closed patterns and maximal patterns satisfy this criterion:  $(\forall X \in \mathcal{L})(X^\uparrow \cap s^-(X) \neq \emptyset)$ . Minimal condensed representations are in the dual class (i.e., Class C2).

Unlike the polarity that opposes two types of irreconcilable patterns, the three parts of the semantics axis (i.e., subsets, supersets and incomparable sets) are simultaneously satisfiable. We think that an ideal pattern extraction method should always belong to these three parts:

**Quality Criterion 2 (Completeness).** *A selector  $s$  is complete iff all patterns are either supporter or opponent:  $\forall X \in \mathcal{L}, s^+(X) \cup s^-(X) = \mathcal{L}$ .*

Let us illustrate the principle behind this quality criterion by considering an ideal pattern mining method that isolates correlations. Of course, this method relies on a selector  $s$  that belongs to the class of correlations (for example, the lift). At equal frequency, the longer pattern will be preferred because it will maximize lift. This property corresponds to the criterion  $X^\uparrow \cap s^-(X) \neq \emptyset$ . At this stage, two incomparable patterns can cover the same set of transactions. To retain only one, we must add a new selection criterion that verifies the criterion  $X^\leftrightarrow \cap s^-(X) \neq \emptyset$ . This approach is at the heart of many proposals in the literature [3, 10, 28]: (i) use of a correlation measure, (ii) elimination of non-closed patterns, (iii) elimination of incomparable redundant patterns.

**Recommendation.** All patterns should be either supporters or opponents in a well-behaving pattern mining method. It is often necessary to combine a measure with local and global redundancy reduction techniques.

## 6 Evaluation Complexity of Interestingness Selectors

As Quality Criterion 2 is often violated, we propose to measure its degree of satisfaction to evaluate and compare interestingness selectors. More precisely, we measure the quality of an interestingness selector considering its degree of satisfaction of the semantics criterion. Let us consider the correlation family, it is clear that to detect correlations, support is a poorer measure than lift which is itself less effective than productivity. Whatever the part of the lattice, the more numerous the opponents/supporters of a selector, the better its quality. In other words, a selector is more effective to assess the interestingness of a pattern  $X$  when the number of supporters and opponents of  $X$  is very high.

**Definition 6 (Evaluation Complexity).** *The evaluation complexity of an interestingness selector  $s$  is the asymptotic behavior of the cardinality of its supporters/opponents.*

The evaluation complexity of a selector usually depends on the cardinality of the assessed pattern (denoted by  $k = |X|$ ) and the cardinality of the set of items  $\mathcal{I}$  (denoted by  $n = |\mathcal{I}|$ ). For instance,  $|all-conf^\pm(X)| = |all-conf^+(X)| \cup |all-conf^-(X)| = 1 + k$ . Therefore, the behavior of the number of evaluations

of all-confidence is linear with respect to itemset size. Similarly, the evaluation complexity of productive itemsets is exponential with respect to the size of the assessed pattern since all subsets are involved in the evaluation of this constraint. According to the evaluation complexity, we say that the quality of the constraint of productive itemsets is better than that of the all-confidence, because the opponents are more numerous. More generally, this complexity allows to compare several interestingness selectors to each other. The column  $|s^\pm(X)|$  (where  $s^\pm(X)$  is the total number of supporters and opponents of  $X$ ) in Table 3 indicates the evaluation complexity of each measure or constraint defined in Table 1. Three main complexity classes emerge: constant, linear and exponential. Although Table 1 is an extremely small sample of measures, we observe that the evaluation complexity of pattern mining methods has increased over the past decades. Interestingly, we also note that the evaluation complexity of global constraints [6, 13] (or advanced measures [26]) is greater than those of local constraints (or absolute measures).

For Classes C2 and C3, the most condensed representations (among those that enable to regenerate the frequency of each pattern) are also those with the greatest evaluation complexity. Indeed, the free itemsets are more numerous than the closed ones, themselves more numerous than the NDIs. For Class C1, it is clear that measures based on more sophisticated statistical models require more relationships [26]. They have therefore a higher evaluation complexity. We will also experimentally verify this hypothesis in the next section.

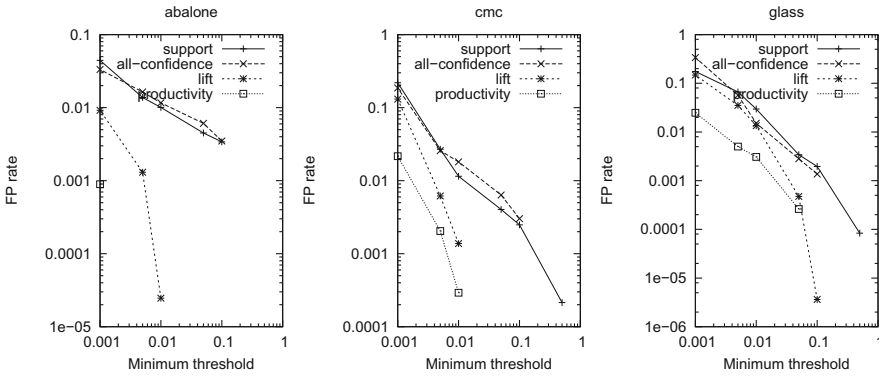
## 7 Experimental Study

Our goal is to verify whether the quality of the correlated pattern selectors follow the evaluation complexity. In other words, if a correlation measure has a greater evaluation complexity than another measure, it is expected to be more effective.

To verify this hypothesis, we rely on the experimental protocol inspired by [14]. The idea is to compare the extracted patterns in an original dataset  $\mathcal{D}$  with the same randomized dataset  $\mathcal{D}^*$ . Specifically, in the randomized dataset  $\mathcal{D}^*$ , a large number of items are randomly swapped two by two in order to clear any correlation. Nevertheless, this dataset  $\mathcal{D}^*$  retains the same characteristics (transaction length and frequency of each item). So, if a pattern  $X$  extracted in the original dataset  $\mathcal{D}$  is also extracted in the randomized dataset  $\mathcal{D}^*$ ,  $X$  is said to be false positive (FP). Its presence in  $\mathcal{D}^*$  is not due to the correlation between items but due to the distribution of items in data. Then we evaluate for each selector how many false positive patterns are extracted on average by repeating the protocol on 10 randomized datasets. Experiments were conducted on datasets coming from the UCI ML Repository [7]. Given a minimum support threshold, we compare 4 selectors: **Support** (all frequent patterns); **All-confidence** (all frequent patterns having at least 5 as all-confidence); **Lift** (all frequent patterns having at least 1.5 as lift); and **Productivity** (all frequent patterns having at least 1.5 as productivity).

Even if arbitrary thresholds are used for the last three selectors, the results are approximately the same with other thresholds because we use the FP rate as

evaluation measure. This normalized measure is a ratio, it returns the proportion of FP patterns among all the mined patterns.



**Fig. 1.** FP rate with minimum support threshold

Figure 1 plots the FP rate of each selector on **abalone**, **cmc** and **glass**. Since lift and productivity measures sometimes do not return FP patterns, there are missing points because the scale is logarithmic. For each dataset, we observe that the FP rate increases when the minimum support threshold decreases regardless of the measure. The evolution of the FP rate for the all-confidence is very similar to that of the support even if the all-confidence has a greater complexity in evaluation. For the other selectors, there is a clear ranking from worst to best: support, lift and productivity. This ranking also corresponds to the classes of complexity from worst to best: constant, linear, exponential. Our framework could be refined to consider a set of patterns as an opponent (or a supporter). Then, the relation between a pattern and its supporters (or opponents) would become a relation between a pattern and a *set* of supporters (or opponents). That would make possible to capture refinements between selectors as follows: the all-confidence depends only on one item at once (due to the maximum of its denominator) whereas the lift can vary according to a set of items (due to the multiplication). Nevertheless, our experiments show that on the whole the correlation measures with the highest evaluation complexity are also the best ones according to the FP rate.

## 8 Conclusion and Discussion

In this paper, we have addressed the question of the quality of a data mining method in the context of unsupervised problems with binary data. A key concept is to study the relationships between a pattern  $X$  and the other patterns when  $X$  is selected by a method. These relationships are formalized through the notions of supporters and opponents. We have presented a typology of methods defined by

formal properties and based on two complementary criteria. This typology offers a global picture and a methodology helping to compare methods to each other. Besides, if a new method is proposed, its quality can be immediately compared to the quality of the other methods according to our framework. Finally, the quality of a method is quantified via an evaluation complexity analysis based on the number of supporters and opponents of a pattern extracted by the method.

Two recommendations can be drawn from this work. We think that the result of a data mining operation should be understandable by the user. So, our first recommendation is a data mining method should not simultaneously extract over-represented phenomena and under-represented phenomena because mixing these two kinds of phenomena obstructs the understandability of the extracted patterns. This recommendation is formally defined by our soundness criterion. Most of methods satisfy this property, but there are a few exceptions such as the constraints extracting NDI and frequent free patterns. The violation of this recommendation might explain why these patterns are of little use.

Another recommendation is a data mining method should extract patterns for which all patterns contribute to the quality of an extracted pattern. This recommendation is formalized by our completeness criterion stating that all patterns must be either supporters or opponents of a pattern extracted by a method. In practice, this recommendation is not satisfied by a lot of methods. However, a few methods are endowed with this behavior, such as [3,10,28] in the context of the correlations. We think that a goal of pattern mining should be to design methods following this recommendation which is more often reached by pattern sets [3] as illustrated by the previous examples [3,10,28].

A perspective of this work is to study an interestingness theory for methods producing pattern sets. Pattern sets are a promising avenue since the interestingness of a pattern also depends on the interestingness of the other patterns of the pattern set, thus providing a global quality of the method. Finally, it is important to note that our framework can be generalized to other pattern languages (sequence, graph, etc.) and other basic functions, e.g., observing the variation of support in a target class would extend our approach to the supervised context.

**Acknowledgements.** The authors thank Albrecht Zimmermann for highly valuable discussions. This work has been partly supported by the QCM-BioChem project (CNRS Mastodons).

## References

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: VLDB, pp. 487–499. Morgan Kaufmann, Burlington (1994)
2. Boulicaut, J.F., Bykowski, A., Rigotti, C.: Free-sets: a condensed representation of boolean data for the approximation of frequency queries. *Data Min. Knowl. Discov.* **7**(1), 5–22 (2003)
3. Bringmann, B., Zimmermann, A.: The chosen few: on identifying valuable patterns. In: ICDM, pp. 63–72. Omaha, NE (2007)
4. Calders, T., Goethals, B.: Non-derivable itemset mining. *Data Min. Knowl. Discov.* **14**(1), 171–206 (2007)

5. Calders, T., Rigotti, C., Boulicaut, J.-F.: A survey on condensed representations for frequent sets. In: Boulicaut, J.-F., De Raedt, L., Mannila, H. (eds.) *Constraint-Based Mining and Inductive Databases*. LNCS (LNAI), vol. 3848, pp. 64–80. Springer, Heidelberg (2006). [https://doi.org/10.1007/11615576\\_4](https://doi.org/10.1007/11615576_4)
6. Crémilleux, B., Soulet, A.: Discovering knowledge from local patterns with global constraints. In: Gervasi, O., Murgante, B., Laganà, A., Tanar, D., Mun, Y., Gavrilova, M.L. (eds.) *ICCSA 2008, Part II*. LNCS, vol. 5073, pp. 1242–1257. Springer, Heidelberg (2008). [https://doi.org/10.1007/978-3-540-69848-7\\_99](https://doi.org/10.1007/978-3-540-69848-7_99)
7. Dheeru, D., Taniskidou, E.K.: UCI machine learning repository (2017). <http://archive.ics.uci.edu/ml>
8. Fayyad, U.M., Piatetsky-Shapiro, G., Uthurusamy, R.: Summary from the kdd-03 panel: data mining: the next 10 years. *ACM SIGKDD Explor.* **5**(2), 191–196 (2003)
9. Wai-chee Fu, A., Wang-wai Kwong, R., Tang, J.: Mining  $N$ -most interesting itemsets. In: Raš, Z.W., Ohsuga, S. (eds.) *ISMIS 2000*. LNCS (LNAI), vol. 1932, pp. 59–67. Springer, Heidelberg (2000). [https://doi.org/10.1007/3-540-39963-1\\_7](https://doi.org/10.1007/3-540-39963-1_7)
10. Gallo, A., De Bie, T., Cristianini, N.: MINI: mining informative non-redundant itemsets. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) *PKDD 2007*. LNCS (LNAI), vol. 4702, pp. 438–445. Springer, Heidelberg (2007). [https://doi.org/10.1007/978-3-540-74976-9\\_44](https://doi.org/10.1007/978-3-540-74976-9_44)
11. Garriga, G.C., Kralj, P., Lavrač, N.: Closed sets for labeled data. *J. Mach. Learn. Res.* **9**, 559–580 (2008)
12. Geng, L., Hamilton, H.J.: Interestingness measures for data mining: a survey. *ACM Comput. Surv.* **38**(3), 9 (2006). <https://doi.org/10.1145/1132960.1132963>
13. Giacometti, A., Marcel, P., Soulet, A.: A relational view of pattern discovery. In: Yu, J.X., Kim, M.H., Unland, R. (eds.) *DASFAA 2011, Part I*. LNCS, vol. 6587, pp. 153–167. Springer, Heidelberg (2011). [https://doi.org/10.1007/978-3-642-20149-3\\_13](https://doi.org/10.1007/978-3-642-20149-3_13)
14. Gionis, A., Mannila, H., Mielikäinen, T., Tsaparas, P.: Assessing data mining results via swap randomization. *TKDD* **1**(3), 14 (2007)
15. Hämmäläinen, W.: Efficient search for statistically significant dependency rules in binary data. Ph.D. thesis, University of Helsinki (2010)
16. Han, J., Cheng, H., Xin, D., Yan, X.: Frequent pattern mining: current status and future directions. *Data Min. Knowl. Discov.* **15**(1), 55–86 (2007)
17. He, Z., Xu, X., Huang, Z.J., Deng, S.: FP-outlier: frequent pattern based outlier detection. *Comput. Sci. Inf. Syst.* **2**(1), 103–118 (2005)
18. Lenca, P., Meyer, P., Vaillant, B., Lallich, S.: On selecting interestingness measures for association rules: user oriented description and multiple criteria decision aid. *EJOR* **184**(2), 610–626 (2008)
19. Mannila, H., Toivonen, H.: Levelwise search and borders of theories in knowledge discovery. *Data Min. Knowl. Discov.* **1**(3), 241–258 (1997)
20. Morik, K., Boulicaut, J.-F., Siebes, A. (eds.): *Local Pattern Detection*. LNCS (LNAI), vol. 3539. Springer, Heidelberg (2005). <https://doi.org/10.1007/b137601>
21. Omiecinski, E.: Alternative interest measures for mining associations in databases. *IEEE Trans. Knowl. Data Eng.* **15**(1), 57–69 (2003)
22. Pasquier, N., Bastide, Y., Taouil, R., Lakhal, L.: Efficient mining of association rules using closed itemset lattices. *Inf. Syst.* **24**(1), 25–46 (1999)
23. Piatetsky-Shapiro, G.: Discovery, analysis, and presentation of strong rules. In: *Knowledge Discovery in Databases*, pp. 229–248. AAAI/MIT Press, Cambridge (1991)
24. Tan, P.N., Kumar, V., Srivastava, J.: Selecting the right objective measure for association analysis. *Inf. Syst.* **29**(4), 293–313 (2004)



25. Tew, C.V., Giraud-Carrier, C.G., Tanner, K.W., Burton, S.H.: Behavior-based clustering and analysis of interestingness measures for association rule mining. *Data Min. Knowl. Discov.* **28**(4), 1004–1045 (2014)
26. Vreeken, J., Tatti, N.: Interesting patterns. In: Aggarwal, C.C., Han, J. (eds.) *Frequent Pattern Mining*, pp. 105–134. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-07821-2\\_5](https://doi.org/10.1007/978-3-319-07821-2_5)
27. Webb, G.I., Vreeken, J.: Efficient discovery of the most interesting associations. *ACM Trans. Knowl. Discov. Data* **8**(3), 15:1–15:31 (2013)
28. Xin, D., Cheng, H., Yan, X., Han, J.: Extracting redundancy-aware top-k patterns. In: *KDD*, pp. 444–453. ACM (2006)
29. Zimmermann, A.: Objectively evaluating condensed representations and interestingness measures for frequent itemset mining. *J. Intell. Inf. Syst.* **45**(3), 299–317 (2015)