Condensed Representation of EPs and Patterns Quantified by Frequency-Based Measures

Arnaud Soulet, Bruno Crémilleux, and François Rioult

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

Abstract. Emerging patterns (EPs) are associations of features whose frequencies increase significantly from one class to another. They have been proven useful to build powerful classifiers and to help establishing diagnosis. Because of the huge search space, mining and representing EPs is a hard and complex task for large datasets. Thanks to the use of recent results on condensed representations of frequent closed patterns, we propose here an *exact* condensed representation of EPs (i.e., all EPs and their growth rates). From this condensed representation, we give a method to provide interesting EPs, in fact those with the highest growth rates. We call strong emerging patterns (SEPs) these EPs. We also highlight a property characterizing the jumping emerging patterns. Experiments quantify the interests of SEPs (smaller number, ability to extract longer and less frequent patterns) and show their usefulness (in collaboration with the Philips company, SEPs successfully enabled to identify the failures of a production chain of silicon plates). These concepts of condensed representation and "strong patterns" with respect to a measure are generalized to other interestingness measures based on frequencies.

Keywords: Emerging patterns, condensed representations, closed patterns, characterization of classes, frequency-based measures.

1 Introduction

The characterization of classes and classification are significant fields of research in data mining and machine learning. Initially introduced in [13], emerging patterns (EPs) are patterns whose frequency strongly varies between two datasets (i.e., two classes). EPs characterize the classes in a quantitative and qualitative way. Thanks to their capacity to emphasize the distinctions between classes, EPs enable to build classifiers or to propose a help for diagnosis. They are at the origin of varied works and they are also used in the realization of powerful classifiers [14, 16]. From an applicative point of view, we can quote various works on the characterization of biochemical properties or medical data [18].

B. Goethals and A. Siebes (Eds.): KDID 2004, LNCS 3377, pp. 173–189, 2005.

Nevertheless, mining EPs in large datasets remains a challenge because of the very high number of candidate patterns. The pruning property used by the level-wise algorithms [20] and often used in data mining cannot be directly applied. Usual methods use handlings of borders [13] in order to find version spaces.

In this paper, we are interested in the extraction of emerging patterns and the definition and characterization of useful kinds of emerging patterns. One originality of our approach is to take advantage of recent progress on the condensed representations of patterns and more precisely on closed patterns [22, 5]. By synthesizing sets of patterns and making easier a process in which users can query data and patterns, condensed representations are an important concept in inductive databases. A brief overview of the condensed representation based on closed pattern is given in Section 2.3.

This paper mainly proposes four contributions. Firstly, we define an exact condensed representation of the emerging patterns for a dataset. Contrary to the borders approach (Section 2.2) which provides the emerging patterns with a lower bound of their growth rate, this condensed representation easily enables to know the *exact* growth rate for each emerging pattern. Moreover, there are efficient algorithms to extract this condensed representation. Secondly, we highlight a new property characterizing a particular kind of emerging patterns, the jumping emerging patterns which make up an active research topic. Thirdly, we propose a new kind of emerging patterns, we call them " strong emerging patterns " (SEPs): these EPs have the best growth rates and we think that they are of a great interest. Furthermore, we show that SEPs are easily obtained from the exact condensed representation of the emerging patterns. This work is also justified by requests from providers of data. Experiments quantify the interests of SEPs (smaller number, ability to extract longer and less frequent patterns). We also give the results achieved by the use of the strong emerging patterns for characterizing patients with respect to atherosclerosis and for successfully identifying the failures of a production chain of silicon plates in collaboration with the Philips company. Lastly, we show that these concepts of condensed representation and "strong patterns" with respect to a measure can be generalized to other interestingness measures based on frequencies.

This paper is an extension of a preliminary work presented in [29]: new contributions are a property characterizing the jumping emerging patterns, the ability to easily obtain the *exact* growth rate for each emerging pattern, the proofs of the properties, in-depth experiments (qualitative results coming from our collaboration with the Philips company, atherosclerosis dataset, influences of the minimal frequency threshold) and the generalization to other interestingness measures based on frequencies.

The paper is organized in the following way. Section 2 introduces the context, the required notations and the works related to this field. Section 3 proposes a new characterization of the jumping emerging patterns. It defines an exact condensed representation of the emerging patterns and also the strong emerging patterns, which are easily achieved from this condensed representation. Section 4 presents the experimental evaluations which quantify the interests of SEPs and

their successful use within a collaboration with the Philips company. Finally, Section 5 extends results highlighted in the case of EPs to other measures based on frequencies.

2 Context and Related Works

2.1 Notations and Definitions

Let $\mathcal{D}$ be a dataset (Table 1), which is an excerpt of the data used for the search for failures in a production chain (cf. Section 4). This table (which is a simplification of the real problem) is used as an elementary example to present the concepts throughout this paper.

Each line (or *transaction*) of Table 1 represents a batch (noted $B_1, \ldots, B_8$) described by features (or *items*) : $A, \ldots, E$ denote the advance of the batch within the production chain and C_1, C_2 the class values. $\mathcal{D}$ is partitioned here into two datasets $\mathcal{D}_1$ (the right batches) and $\mathcal{D}_2$ (the defective batches). The transactions having item C_1 (resp. C_2) belong to $\mathcal{D}_1$ (resp. $\mathcal{D}_2$). A *pattern* is a set of items (e.g., $\{A, B, C\}$) noted by the string ABC. A transaction t contains the pattern X if and only if $X \subseteq t$. Lastly, $|\mathcal{D}|$ (as usual $|.|$ denotes the cardinality of a set) is the number of transactions of $\mathcal{D}$.

The concept of emerging patterns is related to the notion of frequency. The frequency of a pattern X in a dataset $\mathcal{D}$ (noted $\mathcal{F}(X, \mathcal{D})$) is the number of transactions of $\mathcal{D}$ which contain X (for example, $\mathcal{F}(ABC, \mathcal{D}) = 4$). X is *frequent* if its frequency is at least the frequency threshold fixed by the user. From the absolute frequency, we can compute the relative frequency which is $\mathcal{F}(X, \mathcal{D})/|\mathcal{D}|$. Unless otherwise indicated, we use in this paper the absolute frequency. Let us note that by the definition of the partial sets $\mathcal{D}_i$ associated to the class identifiers C_i, we have the relation $\mathcal{F}(X, \mathcal{D}_i) = \mathcal{F}(XC_i, \mathcal{D})$.

Intuitively, an emerging pattern is a pattern whose frequency increases significantly from one class to another. The capture of contrast between classes brought by a pattern is measured by its growth rate. The *growth rate* of a pat-

Table 1. Example of a transactional dataset

$\mathcal{D}$

Batch	Items
B_1	C_1 A B C D
B_2	C_1 A B C D
B_3	C_1 A B C
B_4	C_1 A D E
B_5	C_2 A B C
B_6	C_2 B C D E
B_7	C_2 B C E
B_8	C_2 B E

tern X from $\mathcal{D}_2$ to $\mathcal{D}_1$, noted $GR_1(X)$, is defined as :

$$\begin{cases} 0, & \text{if } \mathcal{F}(X,\mathcal{D}_1) = 0 \text{ and } \mathcal{F}(X,\mathcal{D}_2) = 0 \\ \infty, & \text{if } \mathcal{F}(X,\mathcal{D}_1) \neq 0 \text{ and } \mathcal{F}(X,\mathcal{D}_2) = 0 \\ \frac{|\mathcal{D}_2| \times \mathcal{F}(X,\mathcal{D}_1)}{|\mathcal{D}_1| \times \mathcal{F}(X,\mathcal{D}_2)}, & \text{otherwise} \end{cases}$$

Thus, the definition of an emerging pattern (EP in summary) is given by :

Definition 1 (Emerging Pattern). *Given a threshold $\rho > 1$, a pattern X is said to be an* emerging pattern *from $\mathcal{D}_2$ to $\mathcal{D}_1$ if $GR_1(X) \geq \rho$.*

Let us give some examples from Table 1. With $\rho = 3$, A, ABC, and $ABCD$ are EPs from $\mathcal{D}_2$ to $\mathcal{D}_1$. Indeed, $GR_1(A) = 4/1 = 4$, $GR_1(ABC) = 3/1 = 3$ and $GR_1(ABCD) = 2/0 = \infty$. Conversely, BCD is not an EP: $GR_1(BCD) = 2/1 = 2$ ($< \rho$). When the pattern X is not present in $\mathcal{D}_2$ (i.e. $\mathcal{F}(X,\mathcal{D}_2) = 0$), we get $GR_1(X) = \infty$ and such a pattern is called *jumping emerging pattern* (JEP). For instance, $ABCD$ is a JEP for $\mathcal{D}_1$ and $BCDE$ is a JEP for $\mathcal{D}_2$. Unless otherwise indicated, we consider that the growth rate of a pattern X must be higher than 1 in order that X is an EP.

2.2 Related Works

Efficient computation of all EPs in high dimensional datasets remains a challenge because the number of candidate patterns is exponential according to the number of items. The naive enumeration of all patterns with their frequencies fails quickly. In addition, the definition of EPs does not provide anti-monotonous (e.g., BCD is an EP for $\mathcal{D}_1$, not BC) constraints to apply a powerful pruning of the search space for methods stemming from the framework of level-wise algorithms [20]. Thus, various authors proposed other ways.

The approach of handling borders, introduced by Dong and al. [13], mines multiple couples of maximal and minimal borders from the datasets. The interval described by these two borders corresponds to EPs. Each couple provides an interval giving a concise description of emerging patterns. Unfortunately, the computation of the intervals must be repeated very often and for all the $\mathcal{D}_i$ and this process does not provide for each EP its growth rate. This technique is particularly effective for the search of JEPs due to the convexity of their search space [17]. Nevertheless, Bailey and al. [2] propose a new tree-based data structure for storing the dataset. Their approach is 2-10 times faster than the technique of handling borders.

Other approaches exist. Zhang et al. [32] introduce an anti-monotonous constraint to be able to apply a level-wise algorithm. But this one eliminates many EPs and loses the completeness of the search. In a more general way, this problem can be seen as the search for the patterns checking the conjunction of an anti-monotonous constraint and a monotonous constraint [12, 11], this work drawing its origins from version spaces [21].

2.3 Condensed Representation Based on Closed Patterns

As indicated in the introduction, this paper revisits the search and the characterization of EPs by taking advantage of recent progress on the condensed representations of patterns. We briefly point out below the main concepts required to understand the rest of this paper.

A condensed representation of patterns provides a synthesis of large data sets highlighting the correlations embedded in the data. There is a twofold advantage to use condensed representations. First, such an approach enables powerful pruning criteria during the extraction which greatly improve the efficiency of algorithms [5, 22]. Second, the synthesis of the data provided by a condensed representation is at the core of relevant and multiple uses of patterns (e.g., redundant or informative rules [31], rules with minimal body [9], clustering [15], classification,...), which are key points in many practical applications. There are several kinds of condensed representations of patterns [22, 5]. The most current ones are based on closed patterns, free (or key) patterns or δ-free. A general framework is presented in [7].

For the rest of the paper, we focus on the condensed representation based on closed patterns. A *closed* pattern in $\mathcal{D}$ is a maximal set of items (with respect to the set inclusion) shared by a set of transactions. This concept is related to the lattice theory [3] and the Galois connection. In Table 1, ABC is a closed pattern because B_1, B_2, B_3 and B_5 do not share another item. The notion of *closure* is linked to the one of closed pattern.

Definition 2 (Closure). *The* closure *of a pattern X in $\mathcal{D}$ is $h(X, \mathcal{D}) = \bigcap\{transaction\ t\ in\ \mathcal{D} | X \subseteq t\}$.*

An important property on the frequency stems from this definition. An item A belongs to the closure of X in $\mathcal{D}$ if and only if $\mathcal{F}(XA, \mathcal{D}) = \mathcal{F}(X, \mathcal{D})$. The closure of X is a closed pattern and $\mathcal{F}(X, \mathcal{D}) = \mathcal{F}(h(X, \mathcal{D}), \mathcal{D})$. In our example, $h(AB, \mathcal{D}) = ABC$ and $\mathcal{F}(AB, \mathcal{D}) = \mathcal{F}(ABC, \mathcal{D})$. Thus, the set of the closed patterns is a condensed representation of all patterns because the frequency of any pattern can be inferred from its closure.

3 Condensed Representation and Strong Emerging Patterns

This section highlights a new property to characterize jumping emerging patterns and defines an exact condensed representation of the emerging patterns. Lastly, it proposes the strong emerging patterns.

3.1 Characterization of JEPs

Let us start by generalizing the definition of EPs to data having more than two classes. In Section 2.1, we have $\mathcal{D}_2 = \mathcal{D} \backslash \mathcal{D}_1$. So, $\mathcal{F}(X, \mathcal{D}_2) = \mathcal{F}(X, \mathcal{D}) - \mathcal{F}(X, \mathcal{D}_1)$ (and, similarly, $|\mathcal{D}_2| = |\mathcal{D}| - |\mathcal{D}_1|$). So, the generalization of the growth rate (see its definition in Section 2.1) and thus the definition of EPs, are straightforward.

Let $\mathcal{D}$ be a dataset partitioned into k parts denoted $\mathcal{D}_1, \ldots, \mathcal{D}_k$ ($\mathcal{D} = \bigcup_i \mathcal{D}_i$). The items $C_1, \ldots, C_k$ respectively indicate the membership of a transaction to a dataset $\mathcal{D}_1, \ldots, \mathcal{D}_k$. $\forall i \in \{1, \ldots, k\}$, the growth rate of $\mathcal{D} \backslash \mathcal{D}_i$ in $\mathcal{D}_i$ is:

$$GR_i(X) = \underbrace{\frac{|\mathcal{D}| - |\mathcal{D}_i|}{|\mathcal{D}_i|}}_{noted\ \alpha_i} \times \frac{\mathcal{F}(X, \mathcal{D}_i)}{\mathcal{F}(X, \mathcal{D}) - \mathcal{F}(X, \mathcal{D}_i)} \quad (1)$$

We are now able to provide a new characterization of JEPs for data having any number of classes. An item A belongs to the closure of X in $\mathcal{D}$ if and only if $\mathcal{F}(XA, \mathcal{D}) - \mathcal{F}(X, \mathcal{D}) = 0$ (Definition 2). Then, Property 1 shows how to characterize JEPs:

Property 1 (Characterization of JEPs Based on Closed Patterns).

$$X\ is\ a\ JEP\ of\ \mathcal{D}_i \iff C_i \in h(X, \mathcal{D})$$

Proof. $C_i \in h(X, \mathcal{D}) \iff \mathcal{F}(XC_i, \mathcal{D}) = \mathcal{F}(X, \mathcal{D})$. By definition of $\mathcal{D}_i$, $\mathcal{F}(X, \mathcal{D}_i) = \mathcal{F}(XC_i, \mathcal{D})$. Then $\mathcal{F}(X, \mathcal{D}) = \mathcal{F}(X, \mathcal{D}_i)$ and the denominator of $GR_i(X)$ is null (cf. Equation 1) and X is a JEP.

This property is helpful: it enables to easily obtain JEPs from the closures. Indeed, for each closed pattern XC_i, it is enough to check if X is contained in the condensed representation. If X does not belong to the condensed representation, it means that its closure is XC_i (because XC_i is a closed pattern) and X is a jumping emerging pattern of $\mathcal{D}_i$.

3.2 Exact Condensed Representation of Emerging Patterns

Let us move now how to get the growth rate of any pattern X. Equation 1 shows that it is enough to compute $\mathcal{F}(X, \mathcal{D})$ and $\mathcal{F}(X, \mathcal{D}_i)$. These frequencies can be obtained from the condensed representation of frequent closed patterns. Indeed, $\mathcal{F}(X, \mathcal{D}) = \mathcal{F}(h(X, \mathcal{D}), \mathcal{D})$ (closure property) and by definition of the partial bases $\mathcal{D}_i$, $\mathcal{F}(X, \mathcal{D}_i) = \mathcal{F}(XC_i, \mathcal{D}) = \mathcal{F}(h(XC_i, \mathcal{D}), \mathcal{D})$. Unfortunately, these relations require the computation of two closures ($h(X, \mathcal{D})$ and $h(XC_i, \mathcal{D})$), which it is not efficient. The following properties solve this disadvantage:

Property 2. *Let X be a pattern and $\mathcal{D}_i$ a dataset, $\mathcal{F}(X, \mathcal{D}_i) = \mathcal{F}(h(X, \mathcal{D}), \mathcal{D}_i)$.*

Proof. The properties of the closure operator ensure that for any transaction t, $X \subseteq t \iff h(X, \mathcal{D}) \subseteq t$. In particular, the transactions of $\mathcal{D}_i$ containing X are identical to those containing $h(X, \mathcal{D})$ and we have the equality of the frequencies.

It is now simple to show that the growth rate of every pattern X is obtained thanks to the only knowledge of the growth rate of $h(X, \mathcal{D})$:

Property 3. *Let X be a pattern, we have $GR_i(X) = GR_i(h(X, \mathcal{D}))$.*

Proof. Let X be a pattern. By replacing $\mathcal{F}(X, \mathcal{D})$ with $\mathcal{F}(h(X, \mathcal{D}), \mathcal{D})$ and $\mathcal{F}(X, \mathcal{D}_i)$ with $\mathcal{F}(h(X, \mathcal{D}), \mathcal{D}_i)$ in Equation 1, we immediately recognize the growth rate of $h(X, \mathcal{D})$.

For instance, $h(AB, \mathcal{D}) = ABC$ and $GR_1(AB) = GR_1(ABC) = 3$. The closed patterns with their growth rates are enough to synthesize the whole set of EPs with their growth rates. So, we obtain an *exact* condensed representation of the EPs (i.e. the growth rate of each emerging pattern is exactly known). Let us recall that the borders technique (cf. Section 2.2) only gives a lower bound of the growth rate. This property is significant because the number of closed patterns is lower (and, in general, much lower) than that of all patterns [6]. In practice, $h(X, \mathcal{D})$ is directly obtained by the minimal (with respect to the set inclusion) closed pattern containing X of the condensed representation.

3.3 Strong Emerging Patterns

The number of emerging patterns of a dataset can be crippling for their use. In practice, it is judicious to keep only the most frequent EPs having the best growth rates. But thoughtlessly raising these two thresholds may be problematic. On the one hand, if the minimal growth rate threshold is too high, the EPs found tend to be too specific (i.e. too long). On the other hand, if the minimal frequency threshold is too high, EPs have a too low growth rate.

We define here the strong emerging patterns which are the patterns having the best possible growth rates. They are a trade-off between the frequency and the growth rate.

Definition 3 (Strong Emerging Pattern). *A strong emerging pattern X (SEP in summary) for $\mathcal{D}_i$ is an emerging pattern such that XC_i is a closed pattern in $\mathcal{D}_i$.*

A great interest of SEPs concerns their growth rate: the following property indicates that the SEPs have the best possible growth rates.

Property 4 (SEPs: EPs with Maximum Growth Rate). *Let X be a pattern not containing the item C_i. Then the SEP coming from $h(X, \mathcal{D}_i)$ has a better growth rate than X, i.e. one has $GR_i(X) \leq GR_i(h(X, \mathcal{D}_i) \backslash \{C_i\})$.*

Proof. Let $Y = h(X, \mathcal{D}_i) \backslash \{C_i\}$. Thanks to the closure property, $\mathcal{F}(X, \mathcal{D}_i) = \mathcal{F}(Y, \mathcal{D}_i)$. We can then write (Equation 1) $GR_i(Y) = \alpha_i \times \frac{\mathcal{F}(X, \mathcal{D}_i)}{\mathcal{F}(Y, \mathcal{D}) - \mathcal{F}(X, \mathcal{D}_i)}$. The extensivity of the closure operator makes it possible to write $X \subseteq h(X, \mathcal{D}_i)$ and $C_i \notin X$ thus $X \subseteq Y$ and $\mathcal{F}(X, \mathcal{D}) \geq \mathcal{F}(Y, \mathcal{D})$ due to the property of frequency, which shows that $GR_i(X) \leq GR_i(Y)$.

Let us illustrate Property 4 on the elementary example. The pattern BC is not a SEP for class 1 (because $h(BC, \mathcal{D}_1) \backslash \{C_1\} = ABC$), its growth rate is 1, one has $GR_1(BC) \leq GR_1(ABC) = 3$ and we notice that $\mathcal{F}(BC, \mathcal{D}_1) = \mathcal{F}(ABC, \mathcal{D}_1)$. Let us note that Property 4 enables to highlight an alternative definition of SEPs: an emerging pattern X is said to be a SEP in $\mathcal{D}_i$ when

$GR_i(X) > GR_i(Y)$ for all supersets Y of X such that $\mathcal{F}(X, \mathcal{D}_i) = \mathcal{F}(Y, \mathcal{D}_i)$. This new definition is based on two key points. First, the condition on frequency (i.e. $\mathcal{F}(X, \mathcal{D}_i) = \mathcal{F}(Y, \mathcal{D}_i)$) indicates that we choose a particular pattern for each equivalence class of frequency. Second, this pattern must maximize the growth rate in this equivalence class and Property 4 shows that this pattern corresponds to the closed one.

As for EPs, the property of "being a SEP" is neither monotonous (e.g., B is a SEP for $\mathcal{D}_2$, not BC), nor convertible [23] because no ordering relation over items allows to get a pruning criterion for prefixes. Nevertheless, SEPs are efficiently mined thanks to the properties of the condensed representations (see Section 2.3) and the simple post-processing step to get them. The second advantage of the strong emerging patterns is that their growth rates are immediately known (cf. Property 5). We start by giving Lemma 1 which facilitates the understanding of this property.

Lemma 1. *If XC_i is closed in $\mathcal{D}_i$, then XC_i is closed in $\mathcal{D}$.*

Proof. No transaction of $\mathcal{D} \backslash \mathcal{D}_i$ contains item C_i. If XC_i is closed in $\mathcal{D}_i$, the only transactions of $\mathcal{D}$ containing XC_i are in $\mathcal{D}_i$ and $h(XC_i, \mathcal{D}) = XC_i$, therefore XC_i is closed in $\mathcal{D}$.

Property 5 indicates that the growth rate of SEPs is immediately obtained.

Property 5 (SEPs: Computing Their Growth Rate). *If X is a strong emerging pattern for $\mathcal{D}_i$, then $GR_i(X)$ can be obtained directly with the frequencies of the condensed representation based on the frequent closed patterns of $\mathcal{D}$.*

Proof. Let X be a SEP, therefore XC_i is closed in $\mathcal{D}_i$ (Definition 3). To calculate $GR_i(X)$, it is necessary to calculate $\mathcal{F}(X, \mathcal{D}_i)$ and $\mathcal{F}(X, \mathcal{D})$. By definition of $\mathcal{D}_i$, $\mathcal{F}(X, \mathcal{D}_i) = \mathcal{F}(XC_i, \mathcal{D})$ and Lemma 1 ensures that XC_i is closed in $\mathcal{D}$, thus, its frequency is provided by the condensed representation of the closed patterns of $\mathcal{D}$. To calculate $\mathcal{F}(X, \mathcal{D})$, two cases arise: if X is closed in $\mathcal{D}$, its frequency is directly available. If not, XC_i being closed in $\mathcal{D}$, Property 1 indicates that X is a JEP: its growth rate is infinite.

SEPs are computed thanks to the condensed representation of closed patterns in $\mathcal{D}$ by filtering the closed patterns containing a class value C_i. For each of them, we simply deduce $GR_i(X)$ by considering the pattern X as indicated in the proof above.

Compared to EPs, Properties 4 and 5 show two meaningful advantages of SEPs: on the one hand, they have the best possible growth rates, on the other hand, they are easy to discover from the condensed representation of frequent closed patterns of $\mathcal{D}$ (Lemma 1 ensures that we only have to filter frequent closed patterns containing C_i). Let us note that the EPs based on X and $h(X, \mathcal{D}_i)$ have the same frequency, thus they have the same quality according to this criterion. However, the SEP coming from $h(X, \mathcal{D}_i)$ has a stronger (i.e. higher) growth rate and thus offers a better compromise between frequency and growth rate.

4 Experiments

Experiments provide both quantitative and qualitative results. Quantitative results address the number of SEPs with regard to other kinds of EPs, according to the frequency threshold, etc. and qualitative results deal with the successful use of SEPs to identify the failures of a production chain of silicon plates within a collaboration with the Philips company. Even if some overall results are expected (for instance, the number of SEPs can be only smaller than the number of EPs), we think that it is interesting to quantify them (following our example on the number of SEPs versus those of EPs, is it a drastic reduction or not?).

We use the MVMINER prototype [26] to produce the condensed representation of frequent closed patterns which enables to provide SEPs (see the previous section). In order to compare quantitative results achieved by SEPs with regard to EPs, it is necessary to obtain EPs. For that, we used an APRIORI-like prototype, which computes frequent patterns and selects those having a growth rate greater than a threshold (let us recall that the use of borders does not allow to get the exact growth rate of each pattern [13], so we cannot compare straightforwardly this approach with results stemming from the exact condensed representation of EPs). We did not perform run-time experiments about the efficiency of the extraction of the condensed representation of closed patterns because this efficiency has been shown by several authors [5, 22, 24].

4.1 Data Overview

Experiments were carried out on two real datasets. This first dataset $\mathcal{D}_{athero}$ comes from the STULONG project[1]. These data address a twenty-year longitudinal study of the risk factors of atherosclerosis in a population of 1417 men in former Czechoslovakia. We are interested in characterizing patients according to whether they die or not due to atherosclerosis. From this available data base, we prepare a dataset constituted of 748 rows (divided into 2 classes) described by 119 items (details are in [10]).

The second dataset $\mathcal{D}_{Philips}$ comes from a collaboration with the Philips company. The industrial aim is to identify mistaken tools in a silicon plate production chain. Data are composed of batches, a batch gathers several silicon plates. Briefly speaking, a batch is described by the equipment used at each stage of the flow-chart which is followed during the production. The quality test leads to three quasi-homogeneous classes corresponding to three quality levels. Finally, the characterization is performed on a dataset made up of 44 items (i.e. stage/equipment) and comprising 84 lines (i.e. 84 batches).

4.2 Quantitative Results About SEPs Versus Other Kinds of EPs

Numbers of EPs, Closed EPs and SEPs. We compare here the numbers of EPs, closed EPs (which stemmed from closed patterns) and SEPs. The number of closed EPs is a measure of the size of the condensed representation. Figure 1

[1] Euromise data, `http://lisp.vse.cz/challenge/ecmlpkdd2003/`

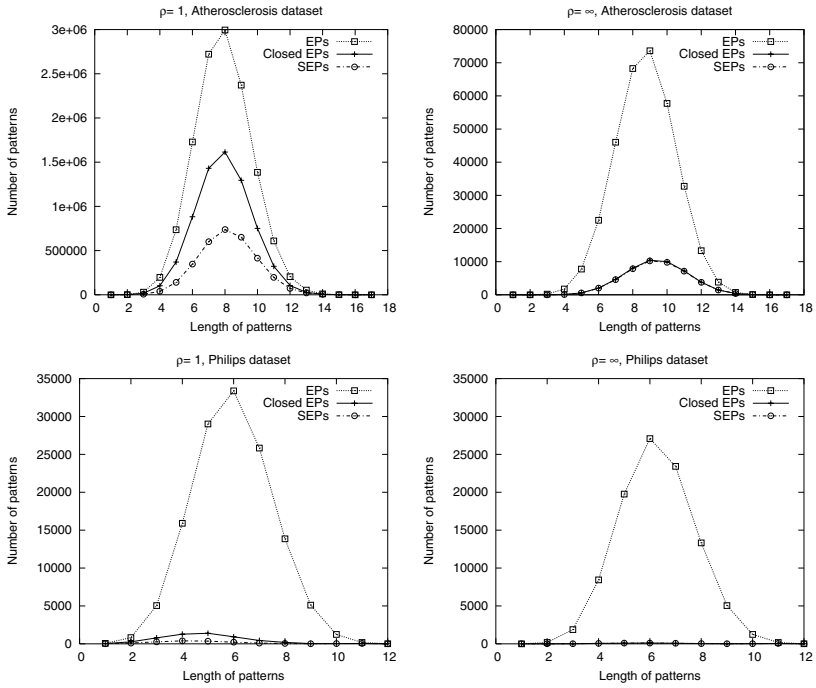

Fig. 1. Comparison between the different kinds of emerging patterns

depicts the distributions of EPs according to the length of patterns for a minimal frequency threshold of 4.0% in $\mathcal{D}_{athero}$ and 1.2% in $\mathcal{D}_{Philips}$. Two threshold values of the minimal growth rate (1 and ∞) are used. This figure shows that the number of EPs is very high compared to the number of closed EPs or SEPs. In $\mathcal{D}_{Philips}$, this disproportion does not decrease in spite of the rise of the minimal growth rate. These too large numbers of EPs cannot be presented to an expert for his analysis task.

Influences of the Minimal Frequency Threshold. Let us see now the role of the minimal frequency threshold. Figure 2 compares the number of EPs with a minimal growth rate of 1 according to the minimal frequency thresholds. We see that the numbers of closed EPs and SEPs increase less quickly than the number of EPs when the frequency decreases. It means that the search for SEPs can be carried out with a smaller minimum frequency. In other words, as the number of SEPs and the size of the exact condensed representation are small compared to the number of EPs, it is possible to examine longer and less frequent patterns.

Figure 3 indicates the variations of the number of EPs, closed EPs and SEPs according to the length of patterns on $\mathcal{D}_{athero}$ (the minimal frequency threshold is 2.3%). We note that the number of SEPs and the size of the exact condensed representation of the EPs increase less quickly when the minimal frequency decreases. For searching long emerging patterns, the combinatory explosion is con-

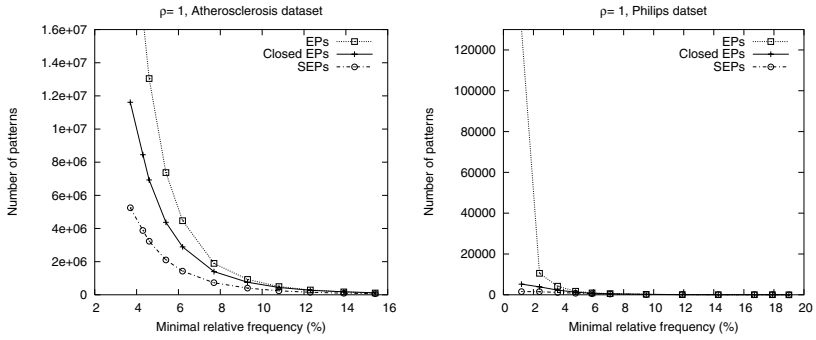

Fig. 2. Number of patterns according to the frequency threshold

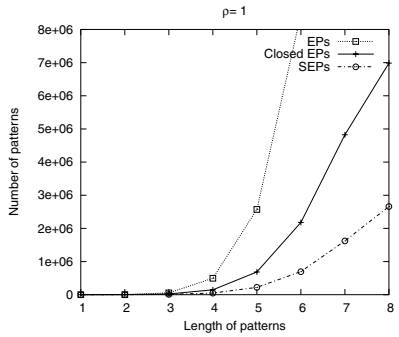

Fig. 3. Number of patterns according to their length ($\mathcal{D}_{athero}$)

trolled in the case of the exact condensed representation of the EPs and the SEPs, but the extraction of EPs fails for patterns longer than 6 items. Again, it allows to mine less frequent and longer patterns.

4.3 Results on Applications

Let us say a few words on the applicative results brought by these experiments. On $\mathcal{D}_{athero}$, we have proposed SEPs to distinguish the patients who die or not due to atherosclerosis. Experiments highlighted SEPs with a quite high growth rate and frequency, and physicians are interested in continuing this work. Furthermore, experts have a strong interest in the quantification of the results (growth rate, frequency).

In our collaboration with Philips (dataset $\mathcal{D}_{Philips}$), experts were the most interested by the confrontation of SEPs having the strongest growth rates and a length equals 1 or 2. Table 2 indicates the most useful SEPs. There is no reliable characteristic SEP of length 1. For instance, the pattern `E=727` has a growth rate close to 1 and it is present both in `Low` and `High`. On the contrary, SEPs of length 2 appeared relevant. The contrast between the pattern `E=727 A=284`

Table 2. Examples of strong emerging patterns

SEPs with a length of 1			
Class	Pattern	GR	Frequency
`Low`	`E=727`	1.01	100% (45)
`Medium`	`F=232`	1.03	100% (37)
`High`	`E=727`	1.01	100% (45)
SEPs with a length of 2 and $GR > 1.5$			
Class	Pattern	GR	Frequency
`Low`	`E=727 A=284`	3.64	75.6 % (34)
`Medium`	`I=504 F=232`	1.84	91.9 % (34)
`Medium`	`L=490 F=232`	1.62	54.0 % (20)
`High`	`E=727 B=288`	2.92	71.1 % (32)
`High`	`E=727 A=222`	2.33	91.1 % (41)

(for `Low`) and the pattern `E=727 A=222` (for `High`) enabled to suspect a problem on the stage `A` (since `E=727` is not a discriminant item). Moreover, the stage `A` comprises only two kinds of equipment (the 222 and the 284). This result tends to show the need for modifying the adjustments of equipment 284 in order that they are similar to those of the equipment 222. After talks with the experts, they have confirmed that the stage suspected by the SEPs was the real cause of the failures (an equipment was badly tuned). This experiment shows the practical contribution of SEPs on real-world data. In other contexts [10], longer SEPs were proved useful to establish diagnostic and the brute force did not allow to obtain these patterns.

Let us recall that SEPs have the advantage of giving a precise growth rate contrary to EPs which would be found by handlings of borders. This quantification is useful at the same time for the selection of EPs and the judgment of the experts. Lastly, thanks to their fewer number, they provide a more understanding characterization of the data than ordinary EPs.

5 Generalization to Frequency-Based Measures

In this section, we generalize the concepts of exact condensed representations and "strong patterns" with respect to other usual interestingness measures based on frequencies. As for the growth rate, which can be seen as a particular measure, the closure operator provides good properties to extend these concepts.

5.1 Exact Condensed Representation of a Frequency-Based Measure

We consider various interestingness measures based on frequencies proposed in statistics, machine learning and data mining. Metrics such as support, confidence, lift, correlation and collective strength are useful to evaluate the quality of classification rules $X \rightarrow C_i$ [27, 8, 28, 1, 19].

Let us define a frequency-based measure M_i which enables to estimate the quality of the premise of the rule $X \rightarrow C_i$ to characterize the class i. For instance, such a measure can be the growth rate. More formally:

Definition 4 (Frequency-Based Measure). *Let $\mathcal{D}$ be a dataset partitioned into k parts denoted $\mathcal{D}_1, \ldots, \mathcal{D}_k$, a frequency-based measure M_i to characterize $\mathcal{D}_i$ is a function of frequencies $\mathcal{F}(X, \mathcal{D}_1), \ldots, \mathcal{F}(X, \mathcal{D}_k)$ i.e. $M_i(X) = F(\mathcal{F}(X, \mathcal{D}_1), \ldots, \mathcal{F}(X, \mathcal{D}_k))$.*

A frequency-based measure is limited to a combination of frequencies of $\mathcal{D}_i$. In particular, such a measure cannot contain other parameters (e.g., the length of a pattern). Some frequency-based measures are indicated in Table 3. Notice that all these measures are expressed in term of frequencies while the literature about interestingness measures often writes these measures by using probabilities (e.g., $P(A|C_i)$ corresponding to $\mathcal{F}(X, \mathcal{D}_i)/|\mathcal{D}_i|$). Some measures (e.g., lift, J-Measure) use frequencies non restricted to datasets $\mathcal{D}_1, \ldots, \mathcal{D}_k$ but these frequencies can be computed from $\mathcal{F}(X, \mathcal{D}_1), \ldots, \mathcal{F}(X, \mathcal{D}_k)$. For example, the frequency $\mathcal{F}(X, \mathcal{D})$ corresponds to $\sum_{j=1}^{k} \mathcal{F}(X, \mathcal{D}_j)$. Thus, these measures respect Definition 4.

As for the emerging patterns, we can know the value of a frequency-based measure on any pattern X from its closure in $\mathcal{D}$:

Theorem 1. *Let X be a pattern, we have $M_i(X) = M_i(h(X, \mathcal{D}))$.*

Proof. Let X be a pattern. For each i, Property 2 allows to replace $\mathcal{F}(X, \mathcal{D}_i)$ by $\mathcal{F}(h(X, \mathcal{D}), \mathcal{D}_i)$. $M_i(X) = F(\mathcal{F}(X, \mathcal{D}_1), \ldots, \mathcal{F}(X, \mathcal{D}_k)) = F(\mathcal{F}(h(X, \mathcal{D}), \mathcal{D}_1), \ldots, \mathcal{F}(h(X, \mathcal{D}), \mathcal{D}_k)) = M_i(h(X, \mathcal{D}))$.

For instance, the closure of AB in $\mathcal{D}$ is ABC and we have $lift_1(AB) = lift_1(ABC) = 3/2$. In the same way, $h(CDE, \mathcal{D}) = BCDE$ and $L_2(CDE) = L_2(BCDE) = 0.529$ with $k = 2$.

The closed patterns with their measure M_i are enough to synthesize the whole set of patterns according to M_i. In practice, the number of closed patterns is lower (and often, much lower) than that of all patterns [4]. Thus, the closed patterns with their measure M_i are an *exact* condensed representation of the measure M_i.

5.2 Strong Frequency-Based Measure

In large datasets, the number of a priori interestingness patterns satisfying a given threshold for a measure M_i can be too huge for their use. As for the SEPs, the notion of strength can be extended to select the patterns which maximalize a measure M_i.

Definition 5 (Strong Frequency-Based Measure). *A frequency-based measure M_i which decreases with $\mathcal{F}(X, \mathcal{D})$, when $\mathcal{F}(X, \mathcal{D}_i)$ remains unchanged, is a strong frequency-based measure.*

For instance, the lift is $\frac{|\mathcal{D}| \times \mathcal{F}(X, \mathcal{D}_i)}{|\mathcal{D}_i| \times \mathcal{F}(X, \mathcal{D})}$. When $\mathcal{F}(X, \mathcal{D}_i)$ remains unchanged and $\mathcal{F}(X, \mathcal{D})$ increases, the lift decreases because the denominator increases. Thus,

Table 3. Examples of frequency-based measures to characterize $\mathcal{D}_i$

Frequency-based measure	Formula	Strong	P_3
J-Measure (J) [28]	$\frac{\mathcal{F}(X,\mathcal{D}_i)}{\lvert\mathcal{D}\rvert} \times log(\frac{\mathcal{F}(X,\mathcal{D}_i)\times\mathcal{D}}{\lvert\mathcal{D}_i\rvert\times\mathcal{F}(X,\mathcal{D})})$ $+\frac{\mathcal{F}(X,\mathcal{D}\setminus\mathcal{D}_i)}{\lvert\mathcal{D}\rvert} \times log(\frac{\mathcal{F}(X,\mathcal{D}\setminus\mathcal{D}_i)\times\mathcal{D}}{\mathcal{F}(X,\mathcal{D})\times\lvert\mathcal{D}\setminus\mathcal{D}_i\rvert})$	no	no
Support [1]	$\mathcal{F}(X,\mathcal{D}_i)/\lvert\mathcal{D}\rvert$	yes	no
Confidence [1]	$\mathcal{F}(X,\mathcal{D}_i)/\mathcal{F}(X,\mathcal{D})$	yes	no
Sensitivity	$\mathcal{F}(X,\mathcal{D}_i)/\lvert\mathcal{D}_i\rvert$	yes	no
Success rate	$\frac{\mathcal{F}(X,\mathcal{D}_i)}{\lvert\mathcal{D}\rvert} + \frac{\lvert\mathcal{D}\setminus\mathcal{D}_i\rvert-\mathcal{F}(X,\mathcal{D}\setminus\mathcal{D}_i)}{\lvert\mathcal{D}\rvert}$	yes	yes
Specificity	$\frac{\lvert\mathcal{D}\setminus\mathcal{D}_i\rvert-\mathcal{F}(X,\mathcal{D}\setminus\mathcal{D}_i)}{\lvert\mathcal{D}\rvert}$	yes	yes
Piatetsky-Shapiro's (PS) [25]	$\frac{\mathcal{F}(X,\mathcal{D}_i)}{\lvert\mathcal{D}\rvert} - \frac{\mathcal{F}(X,\mathcal{D})}{\lvert\mathcal{D}\rvert} \times \frac{\lvert\mathcal{D}_i\rvert}{\lvert\mathcal{D}\rvert}$	yes	yes
Lift [19]	$\frac{\lvert\mathcal{D}\rvert\times\mathcal{F}(X,\mathcal{D}_i)}{\lvert\mathcal{D}_i\rvert\times\mathcal{F}(X,\mathcal{D})}$	yes	yes
Odds ratio (α)	$\frac{\mathcal{F}(X,\mathcal{D}_i)\times(\lvert\mathcal{D}\setminus\mathcal{D}_i\rvert-\mathcal{F}(X,\mathcal{D}\setminus\mathcal{D}_i))}{(\mathcal{F}(X,\mathcal{D})-\mathcal{F}(X,\mathcal{D}_i))\times(\lvert\mathcal{D}_i\rvert-\mathcal{F}(X,\mathcal{D}_i))}$	yes	yes
Laplace (L) [8]	$\frac{\mathcal{F}(X,\mathcal{D}_i)/\lvert\mathcal{D}\rvert+1}{\mathcal{F}(X,\mathcal{D})/\lvert\mathcal{D}\rvert+k}$ with $k>1$	yes	yes
Growth rate (GR) [27]	$\frac{\lvert\mathcal{D}\rvert-\lvert\mathcal{D}_i\rvert}{\lvert\mathcal{D}_i\rvert} \times \frac{\mathcal{F}(X,\mathcal{D}_i)}{\mathcal{F}(X,\mathcal{D})-\mathcal{F}(X,\mathcal{D}_i)}$	yes	yes

the lift is a strong frequency-based measure. In the same way, for the growth rate (Equation 1), when $\mathcal{F}(X,\mathcal{D})$ increases and $\mathcal{F}(X,\mathcal{D}_i)$ is unchanged, the numerator is constant and the denominator increases. So, the growth rate decreases and it is also a strong frequency-based measure.

We link now Definition 5 and the framework defining a good measure given by Piatetsky-Shapiro [25]. The latter has proposed three key properties which have to be satisfied to get a good measure. On a formal point of view, Definition 5 is almost similar to the third property P_3 given by Piatetsky-Shapiro: *M_i monotonically decreases with $P(X)$ when the rest of the parameters (i.e. $P(X,C_i)$ and $P(C_i)$) remain unchanged*. Indeed, we can observe that $P(X) = \mathcal{F}(X,\mathcal{D})/\lvert\mathcal{D}\rvert$, $P(X,C_i) = \mathcal{F}(X,\mathcal{D}_i)/\lvert\mathcal{D}\rvert$ and $P(C_i) = \lvert\mathcal{D}_i\rvert/\lvert\mathcal{D}\rvert$. In comparison with Definition 5, the only slight difference is that M_i must strictly decrease when $\mathcal{F}(X,\mathcal{D})$ increases whereas, in our definition, M_i may remain unchanged. In practice, most of usual measures are strong frequency-based measure because most of them check the property P_3. A survey [30] is carried out on the property P_3 about twenty one interestingness measures. Table 3 gives, for several measures, these ones satisfying or not Definition 5 and property P_3.

Theorem 2. *Let M_i be a strong frequency-based measure and X be a pattern, we have $M_i(X) \leq M_i(h(X, \mathcal{D}_i) \backslash \{C_i\})$. $h(X, \mathcal{D}_i) \backslash \{C_i\}$ is called a strong pattern in class i.*

Proof. Let M_i be a strong measure of frequencies and X be a pattern. If we note $Y = h(X, \mathcal{D}_i) \backslash \{C_i\}$, X and Y have the same frequency in dataset $\mathcal{D}_i$ (property of the closure operator) i.e. $\mathcal{F}(X, \mathcal{D}_i) = \mathcal{F}(Y, \mathcal{D}_i)$. As $X \subseteq Y$, we obtain that $\mathcal{F}(X, \mathcal{D}) \geq \mathcal{F}(Y, \mathcal{D})$. Thus, Definition 5 allows to conclude that $M_i(X) \leq M_i(Y)$.

Let us illustrate Theorem 2 on the running example. The pattern CD is not a strong pattern for class 1 (because $h(CD, \mathcal{D}_1) \backslash \{C_1\} = ABCD$), its Piatetsky-Shapiro's measure is 0.0625 and one has $PS_1(CD) \leq PS_1(ABCD) = 0.125$ as well.

The pattern X and its corresponding strong pattern $h(X, \mathcal{D}_i) \backslash \{C_i\}$ have the same frequency in dataset $\mathcal{D}_i$ and the strong pattern coming from X has an higher value of the measure. Thus, the strong patterns are a good choice to reduce the number of patterns and preserve the best patterns with respect to the measure.

Let us note that as for the SEPs, only $\mathcal{F}(X, \mathcal{D}_i)$ and $\mathcal{F}(X, \mathcal{D})$ are necessary to compute any measure M_i. The same filtering proposed in Section 3.3 can be applied to efficiently mine strong patterns with respect to M_i thanks to the condensed representation of frequent closed patterns.

6 Conclusion

Based on recent results in condensed representations, we have revisited the field of emerging patterns. We have defined an exact condensed representation of the emerging patterns and a new characterization of the jumping emerging patterns. We have proposed a new kind of emerging patterns, the strong emerging patterns which are the EPs with the highest growth rates. We have provided an efficient method to extract SEPs from the exact condensed representation of EPs.

In addition to the simplicity of their extraction, this approach produces only few SEPs which are particularly useful for helping to diagnosis. So, it is easier to use SEPs than search relevant EPs among a large number of EPs. Dealing with our collaboration with the Philips company, SEPs enabled to successfully identify the failures of a production chain of silicon plates. These promising results encourage the use of SEPs in many practical domains.

Finally, we have extended the main ideas to frequency-based measures. We have proven that any frequency-based measure can be exactly and concisely represented in the condensed representation of the closed patterns. This result stems from the properties of the closure operator. As for the SEPs, the concept of strength allows to select less patterns, called strong patterns, which maximalize most of the interestingness measures. Further work is the use of the exact condensed representation and strong patterns for classification tasks.

Acknowledgements. The authors wish to thank the Philips company and in particular, G. Ferru for having provided data and many valuable comments. F. Rioult is supported by the IRM department (University Hospital of Caen France) and the "Comité de la Ligue contre le Cancer de la Manche" and the "Conseil Régional de Basse-Normandie". This work has been partially funded by the AS "Discovery Challenge" supported by the French research organism (CNRS).

References

[1] R. Agrawal, T. Imielinsky, and A Swami. Mining associations rules between sets of items in large databases. In *In Proceedings of the ACM SIGMOD'93*, pages 207–216, 1993.

[2] J. Bailey, T. Manoukian, and K. Ramamohanarao. Fast algorithms for mining emerging patterns. In *Sixth European Conference on Principles Data Mining and Knowledge Discovery, PKDD'02*, pages 39–50, Helsinki, Finland, 2002. Springer.

[3] G. Birkhoff. Lattices theory. *American Mathematical Society, vol. 25*, 1967.

[4] E. Boros, V.r Gurvich, L. Khachiyan, and K. Makino. On the complexity of generating maximal frequent and minimal infrequent sets. In *Symposium on Theoretical Aspects of Computer Science*, pages 133–141, 2002.

[5] J. F. Boulicaut, A. Bykowski, and C. Rigotti. Free-sets: a condensed representation of boolean data for the approximation of frequency queries. *Data Mining and Knowledge Discovery journal*, 7(1):5–22, 2003. Kluwer Academics Publishers.

[6] T. Calders and B. Goethals. Mining all non-derivable frequent itemsets. In T. Elomaa, H. Mannila, and H. Toivonen, editors, *proceedings of the 6th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD'02)*, pages 74–85. Springer, 2002.

[7] T. Calders and B. Goethals. Minimal k-free representations of frequent sets. In *In proceedings of the 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'03)*, pages 71–82. Springer, 2003.

[8] P. Clark and R. Boswell. Rule induction with CN2: Some recent improvements. In *Proc. Fifth European Working Session on Learning*, pages 151–163, Berlin, 1991. Springer.

[9] B. Crémilleux and J. F. Boulicaut. Simplest rules characterizing classes generated by delta-free sets. In *22nd Int. Conf. on Knowledge Based Systems and Applied Artificial Intelligence*, pages 33–46, Cambridge, UK, December 2002.

[10] B. Crémilleux, A. Soulet, and F. Rioult. Mining the strongest emerging patterns characterizing patients affected by diseases due to atherosclerosis. In *proceedings of the workshop Discovery Challenge, PKDD'03*, pages 59–70, 2003.

[11] L. De Raedt, M. Jäger, S. D. Lee, and H. Mannila. A theory of inductive query answering. In *proceedings of the IEEE Conference on Data Mining*, pages 123–130, Maebashi, Japan.

[12] L. De Raedt and S. Kramer. The levelwise version space algorithm and its application to molecular fragment finding. In *IJCAI*, pages 853–862, 2001.

[13] G. Dong and J. Li. Efficient mining of emerging patterns: Discovering trends and differences. In *Knowledge Discovery and Data Mining*, pages 43–52, 1999.

[14] G. Dong, X. Zhang, W. Wong, and J. Li. CAEP: Classification by aggregating emerging patterns. In *Discovery Science*, pages 30–42, 1999.

[15] E-H. Han, G. Karypis, V. Kumar, and B. Mobasher. Clustering based on association rule hypergraphs. In *proceedings of the workshop on Research Issues on Data Mining And Knowledge Discovery, SIGMOD 97*, 1997.

[16] J. Li, G. Dong, and K. Ramamohanarao. Making use of the most expressive jumping emerging patterns for classification. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 220–232. Morgan Kaufmann, San Francisco, CA, 2000.

[17] J. Li and K. Ramamohanarao. The space of jumping emerging patterns and its incremental maintenance algorithms. In *Proc. 17th International Conf. on Machine Learning*, pages 551–558. Morgan Kaufmann, San Francisco, 2000.

[18] J. Li and L. Wong. Emerging patterns and gene expression data. In *Genome Informatics 12*, pages 3–13, 2001.

[19] International Business Machines. IBM intelligent miner, user's guide, version 1, release 1, 1996.

[20] H. Mannila and H. Toivonen. Levelwise search and borders of theories in knowledge discovery. *Data Mining and Knowledge Discovery*, 1(3):241–258, 1997.

[21] T. Mitchell. Generalization as search. *Artificial Intelligence, vol. 18*, pages 203–226, 1980.

[22] N. Pasquier, Y. Bastide, T. Taouil, and L. Lakhal. Discovering frequent closed itemsets for association rules. *Lecture Notes in Computer Science*, 1540:398–416, 1999.

[23] J. Pei, J. Han, and L. V. S. Lakshmanan. Mining frequent item sets with convertible constraints. In *ICDE*, pages 433–442, 2001.

[24] J. Pei, J. Han, and Mao R. CLOSET: An efficient algorithm for mining frequent closed itemsets. In *ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, pages 21–30, 2000.

[25] G. Piatetsky-Shapiro. Discovery, analysis and presentation of strong rules. In G. Piatetsky-Shapiro and W. Frawley, editors, *Knowledge Discovery in Databases*, pages 229–248, Cambridge, MA, 1991. AAAI/MIT Press.

[26] F. Rioult and B. Crémilleux. Condensed representations in presence of missing values. In *5th International Conference on Intelligent Data Analysis (IDA'03)*, 2003.

[27] M. Sebag and M Schoenauer. Generation of rules with certainty and confidence factors from incomplete and incoherent learning bases. In G. Piatetsky-Shapiro and W. Frawley, editors, *in proceedings pf the European Knowledge Acquisition Workshop, EKAW'88*, 1988.

[28] P. Smyth and R. M. Goodman. Rule induction using information theory. In G. Piatetsky-Shapiro and W. Frawley, editors, *Knowledge Discovery in Databases*, pages 159–176, Cambridge, MA, 1991. AAAI/MIT Press.

[29] A. Soulet, B. Crémilleux, and F. Rioult. Condensed representation of emerging patterns. In *8th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Lecture Notes in Computer Science, pages 127–132, Sydney, 2004.

[30] P. Tan, V. Kumar, and J. Srivastava. Selecting the right interestingness measure for association patterns. In *In proceedings The Eighth ACM Special Interest Group on Knowledge Discovery in Data and Data Mining (SIGKDD'02)*, Edmonton, Alberta, Canada, 2002.

[31] M. Zaki. Generating non-redundant association rules. In *In proceedings The 6th ACM Special Interest Group on Knowledge Discovery in Data and Data Mining (SIGKDD'00)*, pages 34–43, 2000.

[32] X. Zhang, G. Dong, and K. Ramamohanarao. Exploring constraints to efficiently mine emerging patterns from large high-dimensional datasets. In *Knowledge Discovery and Data Mining*, pages 310–314, 2000.