Minimal Jumping Emerging Patterns: Computation and Practical Assessment

Bamba Kane, Bertrand Cuissart[(✉)], and Bruno Crémilleux

GREYC - CNRS UMR 6072, University of Caen Basse-Normandie,
14032 Caen Cedex 5, France
{bamba.kane,bertrand.cuissart,bruno.cremilleux}@unicaen.fr

Abstract. Jumping Emerging Patterns (JEP) are patterns that only occur in objects of a single class, a *minimal JEP* is a JEP where none of its proper subsets is a JEP. In this paper, an efficient method to mine the *whole set* of the minimal JEPs is detailed and fully proven. Moreover, our method has a larger scope since it is able to compute the essential JEPs and the top-k minimal JEPs. We also extract minimal JEPs where the absence of attributes is stated, and we show that this leads to the discovery of new valuable pieces of information. A performance study is reported to evaluate our approach and the practical efficiency of minimal JEPs in the design of rules to express correlations is shown.

Keywords: Pattern mining · Emerging patterns · Minimal jumping emerging patterns · Ruled-based classification

1 Introduction

Contrast set mining is a well established data mining area [14] which aims at discovering conjunctions of attributes and values that differ meaningfully in their distributions across groups. This area gathers many techniques such as subgroup discovery [17] and emerging patterns [2]. Because of their discriminative power, contrast sets are highly useful in supervised tasks to solve real world problems in many domains [1,7,12].

Let us consider a dataset of objects partitioned into several classes, each object being described by binary attributes. Initially introduced in [2], emerging patterns (EPs) are patterns whose frequency strongly varies between two datasets. A *Jumping Emerging Pattern* (JEP) is an EP which has the notable property to occur only in a single class. JEPs are greatly valuable to obtain highly accurate rule-based classifiers [8,9]. They are used in many domains like chemistry [12], knowledge discovery from a database of images [7], predicting or understanding diseases [3], or DNA sequences [1]. A *minimal JEP* designates a JEP where none of its proper subsets is a JEP. Minimal JEPs are of great interest because they capture the vital information that cannot be skipped to characterize a class. Using more attributes may not help and even add noise in a classification purpose. Mining minimal JEPs is a challenging task because it is

a time consuming process. Current methods require either a frequency threshold [4] or a given number of expected patterns [16]. On the contrary, one of the results of this paper is to be able to compute the *whole set* of minimal JEPs.

The contribution of this paper can be summarized as follows. First, we introduce an efficient method to obtain *all minimal JEPs*. A key idea of our method is to introduce an alternative definition of a minimal JEP which stems from the differences between pairs of objects, each of a different class. A backtrack algorithm for computing all minimal JEPs is detailed and the related proofs are provided. Our method does not require either a frequency threshold or a number of patterns to extract. It provides a general approach and its scope encompasses the essential JEPs [4] (i.e., JEPs satisfying a given minimal frequency threshold) and the k most supported minimal JEPs [16] which constitute the state of the art in this field. Second, taking into account the absence of attributes may provide interesting pieces of knowledge to build more accurate classifiers as experimentally shown by Terlecki and Walczak [15]. We address this issue. Our method integrates the absence of attributes in the process by adding their negation. It produces the whole set of minimal JEPs both with the present and absent attributes. Practical results advocate in favor of this addition of negated attributes in the description of the objects. Third, the results of an experimental study are given. We analyze the computation of the minimal JEPs, including the absence of attributes and comparisons with essential JEPs and top-k minimal JEPs. Finally, we experimentally assess the quality of minimal JEPs, essential JEPs and top-k minimal JEPs as correlations between a pattern and a class.

Section 2 gives the preliminaries. The description of our method is provided in Section 3. Section 4 presents the experiments. We review related work in Section 5 and we round up with conclusions and perspectives in Section 6.

2   Preliminaries

Let $\mathcal{G}$ be a *dataset*, a multiset consisting of n elements, an element of $\mathcal{G}$ is named an *object*. The *description* of an object is given by a set of attributes, an *attribute* being an atomic proposition which may hold or not for an object. The finite *set of all the attributes occurring in $\mathcal{G}$* is denoted by $\mathcal{M}$. In the remainder of this text, for the sake of simplicity, the word "object" is also used to designate the description of an object.

A *pattern* denominates a set of attributes, an element of the power set $\mathcal{M}$, denoted $\mathcal{P}(\mathcal{M})$. A pattern is *included in* the object g if p is a subset of the description of g: $p \subseteq g$. The *extent of a pattern p* in $\mathcal{G}$, denoted $p'_{\mathcal{G}}$, corresponds to the set of the objects that include p: $p'_{\mathcal{G}} = \{g \in \mathcal{G} : p \subseteq g\}$. A pattern is *supported* if it is included in at least one object of the dataset. Moreover, we define a relation, I, on $\mathcal{G} \times \mathcal{P}(\mathcal{M})$ as follows: for any object g and any pattern p, $gIp \iff p \subseteq g$.

Usual data mining methods only consider the presence of attributes. With binary descriptions, the absence of an attribute can be explicitly denoted by adding the negation of this attribute in order to build patterns conveying this

Table 1. A dataset of 6 objects

Objects \ Attributes		1	¬1	2	¬2	3	¬3	4	¬4
$\mathcal{G}_+$	g_1	x			x	x		x	
	g_2		x	x		x			x
$\mathcal{G}_-$	g_3		x	x			x	x	
	g_4		x	x		x		x	
	g_5		x		x	x		x	
	g_6	x		x		x			x

Table 2. Differences from the dataset in Table 1

	g_3	g_4	g_5	g_6
g_1	1,3,¬2	1,¬2	1	¬2,4
g_2	3,¬4	¬4	2,¬4	¬1
$\mathcal{D}_{\bullet j}$	1,3,¬2,¬4	1,¬2,¬4	1,2,¬4	¬1,¬2,4

information. We integrate this idea in this paper by adding the negation of absent attributes and thus the description of an object always mentions every attribute either positively or negatively. In other words, $\mathcal{M}$ explicitly contains the negation of any of its attributes, the symbol ¬ is used to denote the negation of an attribute (cf. Table 1 as an example).

Minimal Jumping Emerging Pattern. We now suppose that the dataset $\mathcal{G}$ is partitioned into two subsets $\mathcal{G}_+$ and $\mathcal{G}_-$, every subset of such a partition is usually named a *class* of the dataset. We call an object of $\mathcal{G}_+$ a *positive object* and an object of $\mathcal{G}_-$ a *negative object*. We say that a supported pattern p is a *JEP* if it is never included in any negative object: $p'_\mathcal{G} \neq \emptyset$ and $p'_\mathcal{G} \subseteq \mathcal{G}_+$.

A JEP is *minimal* if it does not contain another JEP as a proper subset. The set of the minimal JEPs is a subset of the set of the JEPs which groups all the most general JEPs. As a JEP contains at least one minimal JEP, when an object includes a JEP then it includes a minimal JEP.

Table 1 displays a dataset of 6 objects partitioned in two datasets: $\mathcal{G}_+ = \{g_1, g_2\}$ and $\mathcal{G}_- = \{g_3, g_4, g_5, g_6\}$. The pattern $p = \{1, \neg2\}$ is a JEP as $p'_{\mathcal{G}_+} = \{g_1\}$ and $p'_{\mathcal{G}_-} = \emptyset$ and $\{1\}$ and $\{\neg2\}$ are not JEPs, p is thus a minimal JEP.

3   Contribution

Section 3.1 introduces the key notion of a *difference* between two objects, it provides a new definition of a minimal JEP. The latter is the support of our algorithm for extracting minimal JEPs which is detailed and proven in Section 3.2.

3.1   A Relation Between the Minimal JEPs and the Differences Between Objects

Let $\mathcal{G}$ be a dataset partitioned into two subsets $\mathcal{G}_+$ and $\mathcal{G}_-$. The *difference between an object i and an object j* groups the attributes of i that are not satisfied by j: $\mathcal{D}_{i,j} = i \setminus j = \{m \in \mathcal{M} : i \ I \ m \text{ and } \neg j \ I \ m\}$. When one focuses on a negative object j, the *gathering of the differences for a negative object j* corresponds to the union of the differences between i and j, for any *positive object i*: $\mathcal{D}_{\bullet j} = \cup_{i \in \mathcal{G}_+} \mathcal{D}_{i,j}$. In Table 2, the *gathering of the differences* for the negative object 4 is $\mathcal{D}_{\bullet 4} = \mathcal{D}_{1,4} \cup \mathcal{D}_{2,4} = \{1, \neg2\} \cup \{\neg4\} = \{1, \neg2, \neg4\}$.

The following lemma is a direct consequence of the definition of the gathering of the differences for a negative object.

Lemma 1. *Let j be a negative object and p be a pattern. If $\mathcal{D}_{\bullet j} \cap p \neq \emptyset$ then p is not included in $j : \neg(j \ I \ p)$.*

It follows that, if a supported pattern p intersects with every gathering of the differences for a negative object and, thanks to Lemma 1, p cannot be included in any negative object, thus p is a JEP. We now reason by contraposition and we suppose that a supported pattern p does not intersect with the gathering of the differences for one negative object j_0: $\mathcal{D}_{\bullet j_0} \cap p = \emptyset$. If p is supported by a positive object i_0, as $\mathcal{D}_{\bullet j_0} \cap p = \emptyset$ implies $\mathcal{D}_{i_0,j_0} \cap p = \emptyset$, then p is supported by j_0. Thus p cannot be a JEP.

A JEP corresponds to a supported pattern which has at least one attribute in every $\mathcal{D}_{\bullet j}$, for j a negative object. Proposition 1 follows:

Proposition 1. *A supported pattern p is a JEP if $\mathcal{D}_{\bullet j} \cap p \neq \emptyset, \ \forall j \in \mathcal{G}_-$*

On the example, the JEP $p = \{1, \neg 2\}$ intersects with every $\mathcal{D}_{\bullet j}$ (see Table 2): $\mathcal{D}_{\bullet g_3} \cap p = \{1, \neg 2\}, \mathcal{D}_{\bullet g_4} \cap p = \{1, \neg 2\}$, $\mathcal{D}_{\bullet g_5} \cap p = \{1\}$ and $\mathcal{D}_{\bullet g_6} \cap p = \{\neg 2\}$.

We now establish a relation between the gathering of the differences and the minimal JEPs.

Proposition 2. *A JEP p is a minimal JEP if, for every attribute a of p, $\exists j \in \mathcal{G}_-$ such that $p \cap \mathcal{D}_{\bullet j} = \{a\}$.*

On the example, the JEP $p = \{3, 1, \neg 2\}$ is not a minimal JEP since it contains the JEP $\{1, \neg 2\}$. Proposition 2 gives another point of view: since no intersection between p and a $\mathcal{D}_{\bullet j}$ (for j a negative object) corresponds to $\{3\}$, the attribute $\{3\}$ does not play a necessary part in the discriminative power of p, thus p is not a minimal JEP.

Proof (of Proposition 2). Let p be a JEP.
Suppose p is not minimal: there exists a JEP q, different from p, such that $q \subsetneq p$. Consider an attribute a such that $a \in p \setminus q$. As q is a JEP, Prop. 1 imposes that $\forall j \in \mathcal{G}_-, \ q \cap \mathcal{D}_{\bullet j} \neq \emptyset$, it ensues that $\forall j \in \mathcal{G}_-, \ p \cap \mathcal{D}_{\bullet j} \neq \{a\}$. One now can state that, if p is not minimal, then p contains one attribute a such that $\forall j \in \mathcal{G}_-, \ p \cap \mathcal{D}_{\bullet j} \neq \{a\}$.
Conversely, suppose there exists an attribute a in p such that $\forall j \in \mathcal{G}_-, \ p \cap \mathcal{D}_{\bullet j} \neq \{a\}$. As p is a JEP, Prop. 1 ensures that $\mathcal{D}_{\bullet j} \cap p \neq \emptyset, \ \forall j \in \mathcal{G}_-$. It follows that, $\forall j \in \mathcal{G}_-, \mathcal{D}_{\bullet j} \cap p \setminus \{a\} \neq \emptyset$. By applying Prop. 1, $p \setminus \{a\}$ is a JEP and p cannot be minimal. $\square$

Prop. 2 states that a minimal JEP is a supported pattern that excludes all the negative objects and where every attribute is necessary to exclude (at least one) object. It follows:

Consequence of Prop. 2. Let p be a minimal JEP for the dataset $\mathcal{G}_+ \cup \mathcal{G}_-$ and $g_- \in \mathcal{G}_-$. If p is not a minimal JEP for the dataset $\mathcal{G}_+ \cup \mathcal{G}_- \setminus \{g_-\}$ then there exists a unique attribute a, $a \in p$, such that $p \setminus \{a\}$ is a minimal JEP for the dataset $\mathcal{G}_+ \cup \mathcal{G}_- \setminus \{g_-\}$.

3.2   Calculation of the Minimal JEPs

We now introduce a structure designed to generate all the minimal JEPs for a dataset: a rooted tree whose "valid" leaves are in a one-to-one correspondence with the minimal JEPs. We suppose here that for $\forall j \in \mathcal{G}_-$, $\mathcal{D}_{\bullet j} \neq \emptyset$, as it follows from Prop. 1 that this condition is a necessity for the existence of at least one minimal JEP. We also assume that an arbitrary order is given on the negative objects: for two negative objects j and j', $j \prec j'$ if j is accounted before j'.

Rooted Tree. A *rooted tree* (T, r) is a tree in which one node, the root r, is distinguished. In a rooted tree, any node of degree one, unless it is the root, is called a *leaf*. If $\{u, v\}$ is an edge of a rooted tree such that u lies on the path from the root to v, then v is a *child* of u. An *ancestor* of u is any node of the path from the root to u. If u is an ancestor of v, then v is a *descendant* of u, and we write $u \leqslant v$; if $u \neq v$, we write $u < v$.

A Tree of the Minimal JEPs. We create the tree (T, r) as a rooted tree in which each node x, except the root r, holds two labels: an attribute, $l_{attr}(x) \in \mathcal{M}$, and a negative object $l_{obj}(x) \in \mathcal{G}_-$. For a node x of (T, r), $Br(x)$ gathers the attributes that occur along the path from the root to x: $Br(x) = \{l_{attr}(y), y \leqslant x\}$; $Br(x)$ indicates the pattern considered at x. For any node x of T and any attribute a, $a \in Br(x)$, $crit(a, x)$ gathers the negative objects already considered at the level of x and whose exclusion is due to the sole presence of a in $Br(x)$: $crit(a, x) = \{j \preceq l_{obj}(x) : \mathcal{D}_{\bullet j} \cap Br(x) = \{a\}\}$.

Definition 1 (A tree of the minimal JEPs (ToMJEPs)). *A rooted tree* (T, r) *is a tree of the minimal JEPs for* $\mathcal{G}$ *if:*

  i) *any node x, except the root r, holds two labels: an attribute label, $l_{attr}(x) \in \mathcal{M}$, and a negative object label, $l_{obj}(x) \in \mathcal{G}_-$.*
 ii) *if x is an internal node then:*
    a) *the children of x hold the same negative object label: $l_{obj}(y) = min\{j \in \mathcal{G}_- : \mathcal{D}_{\bullet j} \cap Br(x) = \emptyset\}, \forall y$ a child of x,*
    b) *every child of x holds a different attribute label,*
    c) *the union of the attribute labels of the children y of x corresponds to $\mathcal{D}_{\bullet l_{obj}(y)}$.*
iii) *x is a leaf if it satisfies one of the following conditions:*
    a) *$\exists z \preceq x$ such that $crit(l_{attr}(z), x) = \emptyset$,*
    b) *$\forall j \in \mathcal{G}_-$, $\mathcal{D}_{\bullet j} \cap Br(x) \neq \emptyset$.*

A leaf which satisfies the criteria iii)a) is named *dead-end leaf*, otherwise it is named a *candidate leaf*.

Figure 1 depicts a ToMJEPs for the dataset of Tables 1 and 2. The nodes with a dashed line are the dead-end leaves, the nodes surrounded by a solid line the candidate leaves. A candidate leaf surrounded by a bold plain line is associated to a *supported* pattern: it represents a minimal JEP. For example, the node x such that $Br(x) = \{1, \neg 2\}$ is associated to a minimal JEP while the node
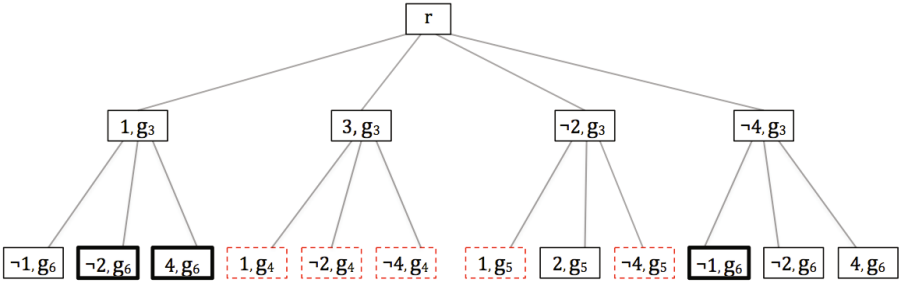
Fig. 1. Example of a tree for minimal JEPs

y such that $Br(y) = \{\neg 4, \neg 2\}$ is associated to a pattern which is not supported by the dataset. The node z such that $Br(z) = \{3, \neg 2\}$ is a dead-end leaf: since $\forall j \in \{\ g_3, g_4\}, \{3, \neg 2\} \cap \mathcal{D}_{\bullet j} \neq \{3\}$, the attribute 3 does not fulfill the constraint raised by Prop. 2, thus $crit(3, z) = \emptyset$.

We will now demonstrate that there is a one-to-one mapping between the "supported" candidate leaves of a ToMJEPs and the minimal JEPs. The following lemma is an immediate consequence of the definition of a ToMJEPs, together with the application of Prop. 1 and 2.

Lemma 2. *Let (T, r) be a ToMJEPs and x be a node of T, different from a dead-end leaf. If there exists $i \in \mathcal{G}_+$ such that $i\ \mathcal{I}\ Br(x)$ then $Br(x)$ is a minimal JEP for the dataset $\mathcal{G}' = \mathcal{G}_+ \cup \{j \leq l_{obj}(x)\}$.*

Proof. By definition of a ToMJEPs, for a node x, we have $Br(x) \cap \mathcal{D}_{\bullet j} \neq \emptyset, \forall j \leq l \leq l_{obj}(x)$. Thanks to Prop. 1, it follows that $Br(x)$ is a JEP for $\mathcal{G}_+ \cup \{j \leq l_{obj}(x)\}$.

If x is not a dead-end leaf, by definition of a ToMJEPs, we have $\forall z \leq x,\ crit(l_{attr}(z), x) \neq \emptyset$, thus $\forall a \in Br(x), \exists j \in \cup\{j \leq l_{obj}(x)\}$ such that $Br(x) \cap \mathcal{D}_{\bullet j} = \{a\}$. Prop. 2 ensures that $Br(x)$ is a minimal JEP for the dataset $\mathcal{G}_+ \cup \{j \leq l_{obj}(x)\}$. □

Lemma 3. *Let (T, r) be a ToMJEPs. Let p be pattern. If p is a minimal JEP for the dataset $\mathcal{G}_+ \cup \mathcal{G}_-$ then there exists a unique candidate leaf x such that $Br(x) = p$.*

Proof. The proof reasons inductively on $\mathcal{G}_-$. For a sake of simplicity, we denote here the set of the negative objects as $\{1, \dots, k\}$ with $k = |\mathcal{G}_-|$ and $\forall 1 \leq j \leq k - 1,\ j \prec j + 1$.

Definition 1 implies that the children of the root r deal with 1 (the first negative object), we have $\mathcal{D}_{\bullet 1} = \{l_{attr}(x) :\ x$ is a child of $r\}$. Moreover, as by definition of a ToMJEPs, $crit(l_{attr}(x), x) \neq \emptyset$, no child of r is a dead-end leaf. Thus, associated to any pattern p which is a minimal JEP for the dataset $\mathcal{G}_+ \cup \{1\}$, there is a unique node x, different from a dead-end leaf such that $Br(x) = p$.

Let us now suppose that, considering any minimal JEP p for $\mathcal{G}_+ \cup \{1, \ldots, l\}$ with $l < k$, there exists a unique node x, different from a dead-end leaf, such that $Br(x) = p$. When we consider a pattern q, minimal JEP for the dataset $\mathcal{G}_+ \cup \{1, \ldots, l, l+1\}$, two cases arise:

- If q is a minimal JEP for $\mathcal{G}_+ \cup \{1, \ldots, l\}$, then, thanks to the induction hypothesis, there exists a unique node x_q such that $Br(x_q) = q$.
- Otherwise, thanks to the consequence of Prop. 2, there exists one attribute a such that $\mathcal{D}_{\bullet l+1} \cap q = \{a\}$ and $\mathcal{D}_{\bullet j} \cap a \neq \{a\}, \forall j \preceq l$. Prop 2 ensures that $q \setminus \{a\}$ is minimal JEP for $\mathcal{G}_+ \cup \{1, \ldots, l\}$. Thanks to the induction hypothesis, there exists a unique node x, different from a dead-end leaf, such that $Br(x) = q \setminus \{a\}$. By definition of a ToMJEPs, there exists a unique child of x, such that $Br(q) = x$. As q is a *minimal* JEP, x is not a dead-end leaf. $\square$

Prop. 3 is a consequence of Lemmas 2 and 3:

Proposition 3 (One-To-One correspondence). *Let (T, r) be a ToMJEPs. There is a one-to-one correspondence between the set of the candidate leaves x such that $Br(x)$ is a supported pattern and the set of the minimal JEPs.*

Prop. 3 ensures that we can generate the minimal JEPs by simply performing a depth first traversal of a ToMJEPs and output the candidate leaves such that $Br(x)$ is a supported pattern. Note that it is not necessary to compute and store the entire ToMJEPs. A depth first traversal only requires to store the path from the root to the node currently visited.

The sketch of implemention provided in Section 4.1 gives information about the calculation of the extent, the calculation of the essential JEPs and the top-k minimal JEPs that are inferred from a ToMJEPs.

4   Experimental Evaluation

This section provides and comments results from a study conducted on 13 benchmark datasets. We investigate the computation of the JEPs according to running time, setting a minimum frequency threshold. It also indicates the reliability of correlation between a JEP and a class. In the following, a *JEP* denominates a supported pattern with respect to any class.

4.1   Material and Methods

The datasets. The study is conducted on 13 usual datasets described in Table 3. All the datasets are available from the UCI Machine Learning repository [10]. We selected these datasets because they have been used, at least once, in an experimental assessment of JEPs [3, 4, 16]. Non binary attributes were converted into a binary valued format by applying a sanctioned method [6, 11] which is available at Frans Coenen's website[1].

---

[1] http://cgi.csc.liv.ac.uk/~frans/KDD/Software/LUCS-KDD-DN/exmpleDNnotes.html

Table 3. The datasets and their characteristics

Datasets	Objects	Attributes	Classes	Datasets	Objects	Attributes	Classes
breast	699	20	2	mushroom	8124	90	2
congres	435	34	2	pima	768	38	2
ecoli	336	34	8	tic-tac-toe	958	29	2
glass	214	48	7	waveform	5000	101	3
heart	303	52	5	wine	178	68	3
hepatitis	155	56	2	zoo	101	42	7
iris	150	19	3				

Implementation. Our algorithm partially explores a ToMJEPs in a depth first manner, it outputs every candidate leaf whose associated pattern is a supported one. We implemented two solutions to ensure to only output supported patterns. The first one, called *post-filtering* solution, generates all the candidate leaves and then checks whether their extent is empty or not. The second one, named *maintaining_extent* solution, integrates the computation of the extents with the calculation of the child of an internal node of a ToMPJEPs. It enables to backtrack as soon as the extent is empty.

Moreover, when a minimum frequency threshold is provided, the *maintaining_extent* solution is straightforwardly adapted to improve the computing of the essential JEPs. Indeed, the frequency of candidate essential JEPs [4] is directly derived from the cardinality of the extent. For the same reason, this solution also enables to compute the top-k minimal JEPs [16] when a value for k is provided. Moreover, the pruning strategy becomes more and more efficient during the mining step because the minimal frequency threshold to belong to the top-k minimal JEP only increases during the mining.

Protocol. In order to compute *all* the minimal JEPs whatever the positive class is, we successively consider each class (of the dataset) as the positive class while the union of the others classes constitutes the negative class. Computations were performed on a server using Ubuntu 12.04 with 2 processors Intel Xeon 2.80 GHz and 512 gigabytes of RAM.

4.2   Results and Discussions

Computation of the Minimal JEPs. We computed all the minimal JEPs on the 13 selected datasets, by using the *post-filtering* and *maintaining_extent* solutions. Moreover, essential JEPs are computed with two minimum frequency thresholds (1% and 5%), and the top-k JEPs with $k = 10$ and $k = 20$. Table 4 gives the cardinalities of the sets of the minimal JEPs and the running times. For computing all the minimal JEPs, the *maintaining_extent* solution always operates faster than the *post-filtering* solution, by a factor varying from 1.6 to 3. By observing the results for the essential JEPs and top-k minimal JEPs, one notes that the running time decreases significantly when a minimal threshold is set for the cardinality of the extent. The use of a frequency constraint related to the cardinality of the extent is efficient, obviously there is the risk to miss interesting patterns.

Minimal JEPs as Rules to Express Correlations. A JEP expresses a correlation between the occurrence of a pattern and one class of objects. This part provides

Table 4. Computation of minimal JEP including negation of attributes

		All minimal JEPs		Essential JEPs		Top-K minimal JEPs	
		post-filtering	maintaining_extent	1%	5%	10	20
Datasets	Min.JEPs	Time	Time	Time	Time	Time	Time
iris	40	70.564 ms	24.348 ms	14.316 ms	9.783 ms	13.043 ms	17.303 ms
breast	38	924.998 ms	347.572 ms	190.432 ms	79.198 ms	95.212 ms	119.213 ms
ecoli	200	842.345 ms	353.734 ms	173.658 ms	98.982 ms	134.314 ms	136.712 ms
zoo	3323	1339.008 ms	579.208 ms	232.023 ms	101.032 ms	67.178 ms	79.032 ms
pima	1443	7.323 s	3.093 s	895.053 ms	532.123 ms	1.009 s	1.694 s
glass	59747	27.172 s	12.418 s	6.927 s	3.241 s	1.439 s	2.081 s
congres	55449	89.396 s	38.077 s	19.145 s	8.380 s	3.107 s	4.929 s
hepatitis	410404	123.520 s	53.706 s	25.576 s	14.419 s	2.978 s	3.097 s
heart	122865	3.351 mn	1.194 mn	29.560 s	15.201 s	9.432 s	8.921 s
tic-tac-toe	109949	5.664 mn	2.797 mn	55.860 s	13.182 s	4.541 s	6.325 s
wine	1353996	200.321 mn	99.366 mn	58.053 mn	36.324 mn	8.342 mn	11.821 mn
mushroom	17345228	673.563 mn	423.116 mn	192.743 mn	101.765 mn	27.545 mn	50.325 mn
waveform	23895434	1845.431 mn	954.190 mn	421.813 mn	238.425 mn	47.342 mn	59.175 mn

experimental results to assess the interest of such rules: do these rules cover a large part of the objects? Are they confident enough? We have also performed experiments to evaluate the usefulness of the explicit description of the absent attributes by adding their negations.

The study has been conducted by using a leave-one-out framework: every object has been successively discarded from the dataset. For every object g, the minimal JEPs have been extracted by considering $\mathcal{G} \setminus \{g\}$ as the dataset and the resulting rules have been applied on g.

Table 5 provides results obtained by applying minimal JEPs, essential JEPs, or top-k minimal JEPs as association rules. *No Negated attributes* designates the descriptions which do not explicitly take into account the absence of attributes whereas *With Negated attributes* points the descriptions that explicitly consider the absence of attributes. The column *Cov* denotes the *coverage* of the set of association rules (the part of the objects for which at least one association rule has applied). The column *Con* refers to the *average confidence* (i.e., the ratio between the number of correct applications of the rules over the whole number of applications of the rules). For example, if we consider the dataset named *breast*, whith the *No Negated attributes* description, 47.78% of the objects contain at least one minimal JEP, this coverage raises to 49.33% of the objects when the descriptions *With Negated attributes* are accounted. With the same dataset, by using the *No Negated attributes* description, 98.19% of the rules resulting from a minimal JEP apply on an object of the proper class ; this average confidence slightly decreases to 96.13% when the *No Negated attributes* description is used.

First of all, the JEPs often apply on a large portion of the objects: for 7 datasets among the 13 datasets, more than 80% of the objects contain at least one JEP. Note that this coverage increases when the description turns from *No Negated attributes* to *With Negated attributes*, up to 8% for the *hepatitis* dataset.

The average confidences indicate that minimal JEPs often point a reliable association between a pattern and a class, even when no frequency constraint is set.

Table 5. Evaluation of minimal JEPs as rules to express correlations

Datasets	No negated attributes										With Negated attributes									
	All Min. JEPs		eJEP				top-k				All Min. JEPs		eJEP				top-k			
			1%		5%		10		20				1%		5%		10		20	
	Cov	Con	Cov	Con	Cov	Con	Cov	Con	Cov	Con	Cov	Con	Cov	Con	Cov	Con	Cov	Con	Cov	Con
breast	47.78	98.19	45.2	98.32	44.81	99.10	39.16	98.5	43.29	97.01	49.33	96.13	47.42	97.49	45.25	98.51	38.21	97.92	40.21	95.85
congres	99.08	89.7	91.29	91.98	80.37	92.31	71.32	93.72	78.53	91.67	99.00	88.10	92.28	90.31	85.42	92.13	60.43	95.23	70.14	94.23
ecoli	34.52	63.79	30.71	66.38	25.74	71.63	20.32	75.73	25.92	68.97	37.32	71.23	32.12	73.98	26.43	74.21	25.92	75.43	27.13	74.01
glass	90.18	62.17	80.32	70.38	77.90	78.12	70.45	70.32	75.27	68.43	94.13	64.15	87.12	66.53	80.23	68.43	75.43	70.29	78.58	65.23
heart	97.02	58.5	83.32	61.90	80.32	64.87	70.32	71.43	74.23	69.25	98.02	56.42	86.32	57.82	82.55	60.43	61.21	80.32	67.01	78.43
hepatitis	82.42	78.39	72.63	81.30	69.32	84.01	69.91	85.90	72.72	83.94	91.03	75.38	73.23	79.17	70.31	81.83	62.62	80.72	70.39	77.38
iris	88.67	90.98	92.23	75.43	88.43	94.61	96.10	74.38	94.67		91.33	89.40	85.32	95.39	83.21	96.37	78.62	97.23	82.76	96.37
mushroom	70.15	77.12	52.20	79.44	47.53	88.43	58.18	81.38	60.06	79.92	74.76	78.43	57.92	80.01	53.91	82.54	63.74	80.32	66.34	77.98
pima	17.05	54.2	14.47	58.12	13.54	60.01	13.01	62.43	14.43	58.91	18.54	61.20	16.03	63.01	14.92	65.53	10.74	63.74	13.59	59.54
tic-tac-toe	81.62	86.57	65.32	88.10	61.93	89.21	60.43	80.21	62.32	78.32	82.12	82.34	70.64	83.21	65.47	84.24	72.31	69.43	75.32	67.30
waveform	45.28	72.19	32.93	74.91	29.99	75.10	33.19	73.19	35.92	72.01	48.09	76.98	40.67	78.01	39.52	82.94	37.91	79.63	40.47	74.78
wine	72.91	67.43	56.72	71.93	41.73	83.01	51.84	70.84	53.42	62.81	73.54	65.89	61.92	67.81	59.02	69.67	55.81	69.51	61.90	67.39
zoo	89.11	92.22	72.19	94.44	60.91	95.10	68.92	93.78	73.24	87.32	93.24	91.43	79.32	92.47	77.19	93.93	60.33	92.48	67.39	90.03

By paying the price of a lower coverage, setting a minimum frequency threshold – as it is done for the essential JEPs or, indirectly, for the top-k minimal JEPs – causes an increase of the average confidence, depending on the dataset. The average confidence levels reached by the two descriptions, *No Negated attributes* and *With Negated attributes*, are very comparable.

As a conclusion, both description families, *With Negated attributes* and *No Negated attributes*, lead to minimal JEPs reaching a similar level of confidence. However, the minimal JEPs extracted with the *With Negated attributes*

descriptions cover a wider range of objects than the minimal JEPs extracted with the *No Negated attributes* descriptions, but with a longer running time.

5   Related Work

Since the key paper of Dong and Li [2], subsequent research has focused on mining emerging patterns and contrast sets. However, there are very few attempts to tackle the discovery of minimal JEPs. Fan and Ramamohanarao have proposed an algorithm extracting the minimal JEPs whose frequency of occurrence is greater than a given threshold, such JEPs are called *essential JEPs* [4]. Terlecki and Walczak have designed a computational method based on a CP-Tree to get the k most supported minimal JEPs, named *top-k minimal JEPs* [16]. These methods require either a frequency threshold or a given number of expected patterns. On the contrary, our method is free from these parameters and computes the whole set of minimal JEPs. Terlecki and Walczak [15] have experimentally shown that taking into account the absence of attributes may provide interesting pieces of knowledge to build more accurate classifiers. We have dealt with this issue since our method extracts minimal JEPs including the negation of the attributes which are absent.

In addition, JEPs can be associated to *version space* [13]. A *version space* gathers the descriptions that match all objects of one class and no object of the other class. Therefore a version space corresponds to the JEPs that match all objects of one class. JEPs are also related to the concept of *disjunctive version space* since a JEP corresponds to all descriptions of objects that match at least one object of one class and no object for the other classes. In Formal Concept Analysis, a JEP is also named "hypothesis" [5] (a *hypothesis* brings together the descriptions of objects that match at least one object in one class and no object in others).

6   Conclusion

We have introduced an efficient method to extract the whole set of minimal JEPs. To the best of our knowledge, it is the first method which does not require either a frequency threshold or a given number of expected patterns. Our method is also able to straightforwardly extract the essential JEPs and the k most supported minimal JEPs. Moreover it enables the integration of negated attributes that can be precious for a classification purpose. We have experimentally analyzed the computation of these JEPs, together with the reliability of the correlations between a JEP and a class.

The structure of tree of the minimal JEPs constitutes a framework for designing and expressing algorithms to compute the minimal JEPs from a dataset. In order to speed up the calculation, this framework will be used to seek for efficient orderings on the attributes or on the objects. Another direction is to produce patterns correlated to one class to a lesser extent and mine emerging patterns with high growth-rate values. Beyond this work, we plan to use minimal JEPs in the design of an advanced rule-based classifier.

References

1. Chen, X., Chen, J.: Emerging patterns and classification algorithms for dna sequence. JSW **6**(6), 985–992 (2011)
2. Dong, G., Li, J.: Efficient mining of emerging patterns: discovering trends and differences. In: KDD, pp. 43–52 (1999)
3. Dong, G., Li, J.: Mining border descriptions of emerging patterns from dataset pairs. Knowl. Inf. Syst. **8**(2), 178–202 (2005)
4. Fan, H., Ramamohanarao, K.: An efficient single-scan algorithm for mining essential jumping emerging patterns for classification. In: Chen, M.-S., Liu, B., Yu, P.S. (eds.) PAKDD 2002. LNCS (LNAI), vol. 2336, pp. 456–462. Springer, Heidelberg (2002)
5. Ganter, B., Kuznetsov, S.O.: Hypotheses and version spaces. In: Ganter, B., de Moor, A., Lex, W. (eds.) ICCS 2003. LNCS (LNAI), vol. 2746, pp. 83–95. Springer, Heidelberg (2003)
6. Kerber, R.: Chimerge: discretization of numeric attributes. In: Swartout, W.R. (ed.) AAAI, pp. 123–128. AAAI Press / The MIT Press (1992)
7. Kobyliński, Ł., Walczak, K.: Spatial emerging patterns for scene classification. In: Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2010, Part I. LNCS (LNAI), vol. 6113, pp. 515–522. Springer, Heidelberg (2010)
8. Kobylinski, L., Walczak, K.: Efficient mining of jumping emerging patterns with occurrence counts for classification. T. Rough Sets **13**, 73–88 (2011)
9. Li, J., Dong, G., Ramamohanarao, K.: Making use of the most expressive jumping emerging patterns for classification. In: Terano, T., Liu, H., Chen, A.L.P. (eds.) PAKDD 2000. LNCS (LNAI), vol. 1805, pp. 220–232. Springer, Heidelberg (2000)
10. Lichman, M.: UCI machine learning repository (2013). http://archive.ics.uci.edu/ml
11. Liu, H., Setiono, R.: Feature selection via discretization. IEEE Trans. Knowl. Data Eng. **9**(4), 642–645 (1997)
12. Lozano, S., Poezevara, G., Halm-Lemeille, M.P., Lescot-Fontaine, E., Lepailleur, A., Bissell-Siders, R., Cremilleux, B., Rault, S., Cuissart, B., Bureau, R.: Introduction of Jumping Fragments in Combination with QSARs for the Assessment of Classification in Ecotoxicology. J. Chem. Inf. Model. **50**(8), 1330–1339 (2010)
13. Mitchell, T.M.: Generalization as search. Artif. Intell. **18**(2), 203–226 (1982)
14. Novak, P.K., Lavrac, N., Webb, G.I.: Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. Journal of Machine Learning Research **10**, 377–403 (2009)
15. Terlecki, P., Walczak, K.: Jumping emerging patterns with negation in transaction databases classification and discovery. Inf. Sci. **177**(24), 5675–5690 (2007)
16. Terlecki, P., Walczak, K.: Efficient discovery of top-K minimal jumping emerging patterns. In: Chan, C.-C., Grzymala-Busse, J.W., Ziarko, W.P. (eds.) RSCTC 2008. LNCS (LNAI), vol. 5306, pp. 438–447. Springer, Heidelberg (2008)
17. Wrobel, S.: An algorithm for multi-relational discovery of subgroups. In: Komorowski, J., Żytkow, J.M. (eds.) PKDD 1997. LNCS, vol. 1263, pp. 78–87. Springer, Heidelberg (1997)