Mining a New Fault-Tolerant Pattern Type as an Alternative to Formal Concept Discovery

Jérémy Besson[1,2] and Céline Robardet[3], and Jean-François Boulicaut[1]

[1] INSA Lyon, LIRIS CNRS UMR 5205, F-69621 Villeurbanne cedex, France
[2] UMR INRA/INSERM 1235, F-69372 Lyon cedex 08, France
[3] INSA Lyon, PRISMa, F-69621 Villeurbanne cedex, France
Firstname.Name@insa-lyon.fr

Abstract. Formal concept analysis has been proved to be useful to support knowledge discovery from boolean matrices. In many applications, such 0/1 data have to be computed from experimental data and it is common to miss some one values. Therefore, we extend formal concepts towards fault-tolerance. We define the DR-bi-set pattern domain by allowing some zero values to be inside the pattern. Crucial properties of formal concepts are preserved (number of zero values bounded on objects and attributes, maximality and availability of functions which "connect" the set components). DR-bi-sets are defined by constraints which are actively used by our correct and complete algorithm. Experimentation on both synthetic and real data validates the added-value of the DR-bi-sets.

1 Introduction

Many application domains can lead to possibly huge boolean matrices whose rows denote objects and columns denote attributes. Mining such 0/1 data has been studied extensively and quite popular data mining techniques have been designed for set pattern extraction (e.g., frequent sets or association rules which capture some regularities among the one values within the data). We are interested in bi-set mining, i.e., the computation of sets of objects and sets of attributes which are somehow "associated". An interesting case concerns Conceptual Knowledge Discovery [8,9,10,11,6]. It is based on the formal concepts contained in the data, i.e., the maximal bi-sets of one values [17]. Examples of formal concepts in $\mathbf{r}_1$ (Table 1) are $(\{o_1, o_2, o_3, o_4\}, \{a_1, a_2\})$ and $(\{o_4\}, \{a_1, a_2, a_3, a_4\})$. Formal concept discovery is related to the popular frequent (closed) set computation. Efficient algorithms can nowadays compute complete collections of constrained formal concepts (see, e.g., [15,2]).

In this paper, we address one fundamental limitation of Knowledge Discovery processes based on formal concepts. Within such local patterns, the strength of the association of the two set components is often too strong in real-life data. Indeed, errors of measurement and boolean encoding techniques may lead to erroneous zero values which will give rise to a combinatorial explosion of the number of formal concepts. Assume that K_1 represents a real phenomenon but

H. Schärfe, P. Hitzler, and P. Øhrstrøm (Eds.): ICCS 2006, LNAI 4068, pp. 144–157, 2006.

Table 1. A formal context K_1 (left), K_2 with 17% of noise (right)

	a_1	a_2	a_3	a_4
o_1	1	1	0	0
o_2	1	1	0	0
o_3	1	1	0	0
o_4	1	1	1	1
o_5	0	0	1	1
o_6	0	0	1	1

	a_1	a_2	a_3	a_4
o_1	1	1	0	0
o_2	1	0	1	0
o_3	1	1	0	1
o_4	1	1	1	1
o_5	0	0	1	0
o_6	0	0	1	1

that data collection and preprocessing lead to the data K_2. The number of formal concepts in K_2 is approximately twice larger than in K_1. Based on our expertise in real-life data mining, it is now clear that the extraction of formal concepts, their post-processing and their interpretation is not that relevant in noisy data which encode measured and/or computed boolean relationships. Our hypothesis is that mining formal concepts with some zero values might be useful and should be considered as a valuable alternative to formal concept discovery. For example, the bi-set $(\{o_1, o_2, o_3, o_4\}, \{a_1, a_2\})$ appears to be relevant in K_2: its objects and attributes are strongly associated (only one zero value) and the outside objects and attributes contain more zero values.

Therefore, we propose to extend formal concepts towards such fault-tolerant patterns by specifying a new type of bi-sets, the so-called DR-bi-sets. The main challenge is to preserve important properties of formal concepts which have been proved useful during pattern interpretation:

- The numbers of zero values are bounded on objects and attributes.
- These bi-sets are maximal on both dimensions.
- It does not exist an outside pattern object (resp. attribute) which is identical to an inside pattern object (resp. attribute). It increases pattern relevancy.
- There exist two functions, one which associates to a set of objects (resp. attributes) a unique set of attributes (resp. objects). Such functions ensure that every DR-bi-set captures a relevant association between the two set components. As such it provides powerful characterization mechanisms.

Section 2 discusses related work. Section 3 is a formalization of our new pattern domain. It is shown that DR-bi-sets are a fairly natural extensions of formal concepts. Section 4 sketches our correct and complete algorithm which computes every DR-bi-set. Section 5 provides experimental results on both synthetic and real data. Section 6 concludes.

2 Related Work

Looking for fault-tolerant pattern has been already studied. To the best of our knowledge, most of the related work has concerned mono-dimensional patterns and/or the use of heuristic techniques. In [18], the frequent set mining task is extended towards fault-tolerance. A level-wise algorithm is proposed but their fault-tolerant property is not anti-monotonic while this is needed to achieve tractability. Therefore, [18] provides a greedy algorithm leading to an incomplete computation. [14] revisits this work and it looks for an anti-monotonic constraint such that a level-wise algorithm can provide every set whose density of one values is greater than δ in at least σ situations. Anti-monotonicity is obtained by enforcing that every subset of extracted sets satisfies the constraint as well. The extension of such dense sets to dense bi-sets is difficult: the connection which associates objects to properties and vice-versa is not decreasing while this is an appreciated property of formal concepts. Instead of using a relative density definition, [12] considers an absolute threshold to define fault-tolerant frequent patterns: given a threshold δ, a set of attributes P, such that $\sharp P > \delta$, holds in an object X iff $\sharp(X \cap P) \geq \sharp P - \delta$ where $\sharp X$ denotes the size of X. To ensure that the support is significant for each attribute, they use a minimum support threshold per attribute beside the classical minimum support. Thus, each object of an extracted pattern contains less than δ zero values and each attribute contains more one values than the given minimum support for each attribute. This definition is not symmetrical on the object and attribute dimension, and the more the support increases, the less the patterns are relevant. In [7], the authors are interested in geometrical tiles (i.e., dense bi-sets which involve contiguous elements given orders on both dimensions). Their local optimization algorithm is not deterministic and thus can not guarantee the global quality of the extracted patterns. Furthermore, the hypothesis on built-in orders can not be accepted on many data.

Some fault-tolerant extensions of formal concepts have been recently proposed as well. In [1], available formal concepts are merged while checking for a bounded number of exceptions on both dimensions. The proposed technique is however incomplete, and the mapping between set components of the extracted bi-sets is not guaranteed. The proposal in [13] concerns an extension which can be computed efficiently but none of the appreciated properties are available. This research is also related to condensed representations of concept lattices or dense bi-sets. [16] introduces a “zooming” approach on concept lattices. The so-called α-Galois lattices exploit a partition on the objects to reduce the collection of the extracted bi-sets: a situation s is associated to a set G if α% of the objects which have the same class value than s are associated to elements from G and if s is associated to G as well. Our context is different since we want to preserve the duality between objects and attributes as far as possible.

3 Formalization

Let G and M be sets, called the set of objects and attributes respectively. Let I be a relation $I \subseteq G \times M$ between objects and attributes: for $g \in G, m \in M$, $(g,m) \in I$ holds iff the object g has the attribute m. The triple $K = (G, M, I)$ is called a (formal) context.

A bi-set (X, Y) is a couple of sets from $2^G \times 2^M$. Some specific types of bi-sets have been extensively studied. This is the case of formal concepts which can be defined thanks to Galois connection [17]:

Definition 1. *Given $X \subseteq G$ and $Y \subseteq M$, the Galois connection on K is the couple of functions (ϕ, ψ) s.t. $\psi(X) = \{m \in M \mid \forall g \in X,\ (g,m) \in I\}$ and $\phi(Y) = \{g \in G \mid \forall m \in Y,\ (g,m) \in I\}$. A bi-set (X,Y) is a formal concept with extent X and intent Y iff $X = \phi(Y)$ and $Y = \psi(X)$.*

We now give a new way to define formal concepts which will be generalised to DR-bi-sets.

Definition 2. *Let us denote by $\mathcal{Z}_o(x,Y)$ the number of zero values of an object x on the attributes in Y: $\mathcal{Z}_o(x,Y) = \sharp\{y \in Y | (x,y) \notin I\}$. Similarly $\mathcal{Z}_a(y,X) = \sharp\{x \in X | (x,y) \notin I\}$ denotes the number of zero values of an attribute y on the objects in X.*

Formal concepts can now be characterized by the following lemma:

Lemma 1. *A bi-set (X,Y) is a formal concept of the context K iff:*
$\forall x \in X,\ \mathcal{Z}_o(x,Y) = 0$ *or similarly,* $\forall y \in Y,\ \mathcal{Z}_a(y,X) = 0$ *(1)*
$(\forall x \in G \setminus X,\ \mathcal{Z}_o(x,Y) \geq 1)$ *and* $(\forall y \in M \setminus Y,\ \mathcal{Z}_a(y,X) \geq 1)$ *(2)*

It introduces constraints which can be used to compute formal concepts [2]. Interestingly, these constraints ensure the maximality (w.r.t. set inclusion) of the bi-sets which satisfy them. It is well-known that constraint monotonicity properties are extremely important for a clever exploration of the associated search space. These properties are related to a specialization relation. Let us consider an unusual specialization relation for building concept lattices.

Definition 3. *Our specialization relation $\preceq$ on bi-sets is defined as follows: $(X_1,Y_1) \preceq (X_2,Y_2)$ iff $X_1 \subseteq X_2$ and $Y_1 \subseteq Y_2$. A constraint $\mathcal{C}$ is said anti-monotonic w.r.t. $\preceq$ iff $\forall D,E \in 2^G \times 2^M$ s.t. $D \preceq E$, $\mathcal{C}(E) \Rightarrow \mathcal{C}(D)$. Dually, $\mathcal{C}$ is said monotonic w.r.t. $\preceq$ iff $\mathcal{C}(D) \Rightarrow \mathcal{C}(E)$. Notice that $\mathcal{C}(D)$ denotes that the constraint $\mathcal{C}$ is satisfied by the bi-set D.*

For instance, we might use a minimal size constraint $\mathcal{C}_{ms}(\sigma_1, \sigma_2, (X,Y)) \equiv \sharp X \geq \sigma_1 \wedge \sharp Y \geq \sigma_2$. Such a constraint is monotonic w.r.t. $\preceq$.

3.1 Dense Bi-sets

We want to compute bi-sets with a strong association between the two sets and such that its number of zero values can be controlled. We can decide to bound the number of zero values per object/attribute or on the whole bi-set (strong density vs. weak density). We can also look at relative or absolute density, i.e., to take into account the density w.r.t. the size of the whole bi-set or not. If we use the weak density, we can obtain bi-sets containing objects or attributes with only zero values. In this case, these objects (resp. attributes) are never associated to the bi-set attributes (resp. objects). We decided to use an absolute strong density constraint that enforces an upper bound for the number of zero values per object and per attribute. Using strong density enables to get the important monotonicity property.

Definition 4. *Given* $(X,Y) \in 2^G \times 2^M$ *and a positive integer value* α*,* (X,Y) *is said dense iff it satisfies the anti-monotonic constraint* $\mathcal{C}_d(\alpha,(X,Y)) \equiv (\forall x \in X,\ \mathcal{Z}_o(x,Y) \leq \alpha)$ *and* $(\forall y \in Y,\ \mathcal{Z}_a(y,X) \leq \alpha)$*.*

3.2 Relevant Bi-sets

We want to extract bi-sets (X,Y) such that the objects of X (resp. the attributes of Y) have a larger density of one values on the attributes from Y (resp. on the objects from X) than on the other attributes, i.e., $M \setminus Y$ (resp. objects, i.e., $G \setminus X$). It leads to the formalisation of a relevancy constraint where the parameter δ is used to enforce the difference of zero values inside and outside the bi-set.

Definition 5. *Given* $(X,Y) \in 2^G \times 2^M$*, and a positive integer value* δ*,* (X,Y) *is said relevant iff it satisfies the following constraint:*

$$\begin{aligned} \mathcal{C}_r(\delta,(X,Y)) \quad \equiv \quad & (\forall g \in G \setminus X,\ \forall x \in X,\ \mathcal{Z}_o(g,Y) \geq \mathcal{Z}_o(x,Y) + \delta) \\ \textit{and} \quad & (\forall m \in M \setminus Y,\ \forall y \in Y,\ \mathcal{Z}_a(m,X) \geq \mathcal{Z}_a(y,X) + \delta) \end{aligned}$$

3.3 DR-Bi-sets

The bi-sets which satisfy both $\mathcal{C}_d$ and $\mathcal{C}_r$ constraints are a new type of fault-tolerant patterns. Dense and relevant bi-sets are indeed a generalisation of formal concepts (bi-sets with $\alpha = 0$ and $\delta = 1$). $\mathcal{C}_d$ is a straightforward generalisation of the first equation in Lemma 1. $\mathcal{C}_r$ generalizes the second equation in Lemma 1 by enforcing that all outside elements of the bi-set contain at least δ zero values in addition to the one of every inside element. Parameter α controls the density of the bi-sets whereas the parameter δ enforces a significant difference with the outside elements. $\mathcal{C}_d$ is anti-monotonic w.r.t. $\preceq$ (see Definition 3) and can give rise to efficient pruning. $\mathcal{C}_r$ is neither monotonic nor anti-monotonic but we explain in Section 4 how to exploit this constraint efficiently. Fig. 1 shows the collection of bi-sets in K_3 which satisfy $\mathcal{C}_d \wedge \mathcal{C}_r$ when $\alpha = 5$ and $\delta = 1$ ordered w.r.t. $\preceq$. Each level indicates the maximal number of zero values per object and per attribute. For instance, if $\alpha = 1$, a sub-collection containing five bi-sets is

	a_1	a_2	a_3	a_4
o_1	1	1	1	1
o_2	1	1	0	0
o_3	0	1	0	0
o_4	0	0	1	0
o_5	0	0	0	0

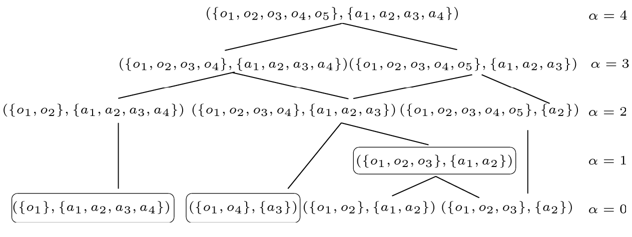

Fig. 1. A formal context K_3 and the bi-sets satisfying $\mathcal{C}_d \wedge \mathcal{C}_r$ with $\alpha = 5$ and $\delta = 1$

extracted, four of them being formal concepts ($\alpha = 0$). Density and relevancy constraints do not ensure maximality which is a desired property. For instance, in Fig. 1, if B denotes $(\{o_1, o_2, o_3\}, \{a_1, a_2\})$, we have $(\{o_1, o_2\}, \{a_1, a_2\}) \preceq B$ and $(\{o_1, o_2, o_3\}, \{a_2\}) \preceq B$. As a result, to increase bi-set relevancy, we finally consider the so-called DR-bi-sets which are the maximal dense and relevant bi-sets.

Definition 6. *Let $(X, Y) \in 2^G \times 2^M$ be a dense and relevant bi-set (i.e., satisfying $\mathcal{C}_d \wedge \mathcal{C}_r$). (X, Y) is called a DR-bi-set iff it is maximal w.r.t. $\preceq$, i.e. it does not exist $(X', Y') \in 2^G \times 2^M$ s.t. (X', Y') satisfies $\mathcal{C}_d \wedge \mathcal{C}_r$ and $(X, Y) \preceq (X', Y')$.*

This collection is denoted $\mathrm{DR}_{\alpha\delta}$. For example, DR_{11} on K_3 contains the three circled bi-sets of Fig. 1. It is important to notice that different threshold values might be considered on objects/attributes (say α/α' for the density constraint and δ/δ' for the relevancy constraint).

3.4 Properties

Let us first emphasize that the DR-bi-set size increases with parameter α.

Property 1. Given $0 \le \alpha_1 \le \alpha$, $\forall (X_1, Y_1) \in \mathrm{DR}_{\alpha_1\delta}$, $\exists (X, Y) \in \mathrm{DR}_{\alpha\delta}$ such that $(X_1, Y_1) \preceq (X, Y)$.

Proof. $\forall (X, Y)$ satisfying $\mathcal{C}_d(\alpha_1, (X, Y)) \wedge \mathcal{C}_r(\delta, (X, Y))$ then (X, Y) satisfies $\mathcal{C}_d(\alpha, (X, Y)) \wedge \mathcal{C}_r(\delta, (X, Y))$. $\mathrm{DR}_{\alpha\delta}$ contains (X, Y) or a bi-set (X', Y') s. t. $(X, Y) \preceq (X', Y')$. □

The larger α is, the more the size of each extracted bi-set from $\mathrm{DR}_{\alpha\delta}$ increases while extracted associations with smaller α value are preserved. In practice, an important reduction on the size of the extracted collections is observed when the parameters are well chosen (see Section 5). As a result, a zooming effect is obtained when α is varying. Parameter δ enables to select more relevant patterns. For example, when $\delta = 2$ and $\alpha \le 1$ the collection in K_3 is reduced to the DR-bi-set $(\{o_1\}, \{a_1, a_2, a_3, a_4\})$.

The following property ensures that DR-bi-sets are actually a generalisation of formal concepts, i.e., they are related by two functions.

Property 2. For $\delta > 0$, there exists two functions called ψ_{DR} and ϕ_{DR} such that $\psi_{DR} : 2^G \to 2^M$ and $\phi_{DR} : 2^M \to 2^G$ such that (X, Y) is a DR-bi-set iff $X = \phi_{DR}(Y)$ and $Y = \psi_{DR}(X)$.

Proof. Let $(S_1, S_2), (S_1, S_3) \in \mathrm{DR}_{\alpha\delta}$ such that $S_2 \neq S_3$. Let $\mathcal{M}ax\mathcal{Z}_a(X, Y) \equiv \max_{m \in X} \mathcal{Z}_a(m, Y)$ and $\mathcal{M}in\mathcal{Z}_a(X, Y) \equiv \min_{m \in X} \mathcal{Z}_a(m, Y)$

As $\mathrm{DR}_{\alpha\delta}$ contains maximal bi-sets, $S_2 \not\subseteq S_3$ and $S_3 \not\subseteq S_2$. We have

$$\begin{aligned}\mathcal{M}ax\mathcal{Z}_a(S_1, S_3) &\le \mathcal{M}in\mathcal{Z}_a(S_1, M \setminus S_3) - \delta \ (\mathcal{C}_r \text{ constraint}) \\ &\le \mathcal{M}in\mathcal{Z}_a(S_1, S_2 \setminus S_3) - \delta \ (\text{set inclusion}) \\ &< \mathcal{M}in\mathcal{Z}_a(S_1, S_2 \setminus S_3) \ (\delta > 0) \le \mathcal{M}ax\mathcal{Z}_a(S_1, S_2 \setminus S_3) \\ &\le \mathcal{M}ax\mathcal{Z}_a(S_1, S_2)\end{aligned}$$

Then, we have $\mathcal{M}ax\mathcal{Z}_a(S_1, S_3) < \mathcal{M}ax\mathcal{Z}_a(S_1, S_2)$ and similarly we can derive $\mathcal{M}ax\mathcal{Z}_a(S_1, S_2) < \mathcal{M}ax\mathcal{Z}_a(S_1, S_3)$ which leads to a contradiction.

Thus, we have a function between 2^G and 2^M. The existence of a function between 2^M and 2^G can be proved in a similar way. □

These functions are extremely useful to support pattern interpretation: to a set of objects X corresponds at most one set of attributes. Typically, they were missing in previous approaches for fault-tolerance extensions of formal concepts [1,12]. Unfortunately, we do not have an explicit definition of these functions. This remains an open problem.

4 A Complete Algorithm

The whole collection of bi-sets ordered by $\preceq$ forms a lattice whose bottom is $(\bot_G, \bot_M) = (\emptyset, \emptyset)$ and top is $(\top_G, \top_M) = (G, M)$. Let us note by $\mathcal{B}$ the set of sublattices[1] of $((\emptyset, \emptyset), (G, M))$, $\mathcal{B} = \{((X_1, Y_1), (X_2, Y_2))$ s.t. $X_1, X_2 \in 2^G, Y_1, Y_2 \in 2^M$ and $X_1 \subseteq X_2$, $Y_1 \subseteq Y_2\}$, where the first (resp. the second) bi-set is the bottom (resp. the top) element. The algorithm DR-Miner explores some of the sublattices of $\mathcal{B}$ built by means of three mechanisms: enumeration, pruning and propagation.

Table 2. DR-Miner pseudo-code

$K = (G, M, I)$ is a formal context, $\mathcal{C}$ a conjunction of monotonic and anti-monotonic constraints on $2^G \times 2^M$ and α, δ are positive integer values.

```
DR-Miner
        Generate((∅,∅),(G,M))
End DR-Miner

Generate(L)
      Let L = ((⊥_G,⊥_M),(⊤_G,⊤_M))
      L ← Prop(L)
      If Prune(L) then
         If (⊥_G,⊥_M) ≠ (⊤_G,⊤_M) then
            (L_1,L_2) ← Enum(L,Choose(L))
            Generate(L_1)
            Generate(L_2)
         Else Store (⊥_G,⊥_M)
         End if
      End if
End Generate
```

[1] X is a sublattice of Y if Y is a lattice, X is a subset of Y and X is a lattice with the same join and meet operations as Y.

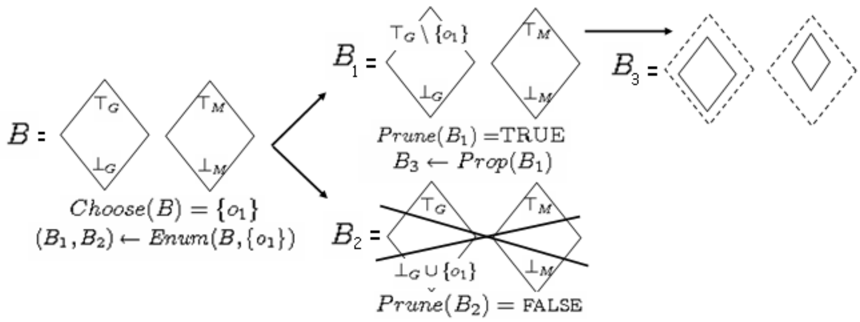

Fig. 2. Example of DR-Miner execution

DR-Miner starts with the complete lattice $((\emptyset, \emptyset), (G, M))$ and then recursively propagates the constraints using *Prop* function, check the consistency of the obtained sublattice with *Prune* function and then generates two new sublattices thanks to *Enum* function (see Table 2). The Figure 2 shows an example of DR-Miner execution.

- **Enumeration:** Let $Enum : \mathcal{B} \times G \cup M \to \mathcal{B}^2$ such that

$$Enum(((\bot_G, \bot_M), (\top_G, \top_M)), e)$$
$$= \begin{cases} (((\bot_G \cup \{e\}, \bot_M), (\top_G, \top_M)), ((\bot_G, \bot_M), (\top_G \setminus \{e\}, \top_M))) \text{ if } e \in G \\ (((\bot_G, \bot_M \cup \{e\}), (\top_G, \top_M)),\ ((\bot_G, \bot_M), (\top_G, \top_M \setminus \{e\}))) \text{ if } e \in M \end{cases}$$

  where $e \in \top_G \setminus \bot_G$ or $e \in \top_M \setminus \bot_M$. *Enum* generates two new sublattices which are a partition of its input parameter.
  Let $Choose : \mathcal{B} \to G \cup M$ be a function which returns (one of) the element $e \in \top_G \setminus \bot_G \cup \top_M \setminus \bot_M$ containing the largest number of zero values on $\top_M$ if $e \in G$ or on $\top_G$ if $e \in M$. It is an heuristic which tends to increase the efficiency of propagation mechanisms by reducing the search space as soon as possible.
- **Pruning:** We prune a sublattice if we are sure that none of its bi-sets satisfies the constraint. Let $Prune^m_{\mathcal{C}} : \mathcal{B} \to \{$TRUE,FALSE$\}$ be a function which returns True iff the monotonic constraint $\mathcal{C}$ (w.r.t. $\preceq$) is satisfied by the top of the sublattice: $Prune^m_{\mathcal{C}}((\bot_G, \bot_M), (\top_G, \top_M)) \equiv \mathcal{C}(\top_G, \top_M)$

  Let $Prune^{am}_{\mathcal{C}} : \mathcal{B} \to \{$TRUE,FALSE$\}$ be a function which returns True iff the anti-monotonic constraint $\mathcal{C}$ (w.r.t $\preceq$) is satisfied by the bottom of the sublattice: $Prune^{am}_{\mathcal{C}}((\bot_G, \bot_M), (\top_G, \top_M)) \equiv \mathcal{C}(\bot_G, \bot_M)$

  $\mathcal{C}_d$ is anti-monotonic and thus it can be used as $Prune^{am}_{\mathcal{C}_d}$. Nevertheless, $\mathcal{C}_r$ is neither monotonic nor anti-monotonic. The $\mathcal{C}_r$ constraint is adapted to ensure that the elements which do not belong to the sublattice might contain more zero values on the top (the elements that can be included in the bi-sets) than the inside ones do on the bottom (the elements that belong to each bi-set). Let $Prune_{\mathcal{C}_r} : \mathcal{B} \rightarrow \{\text{TRUE,FALSE}\}$ be a function such that

$$Prune_{\mathcal{C}_r}((\bot_G, \bot_M), (\top_G, \top_M)) \equiv$$
$$\forall s \in G \setminus \top_G, \forall t \in \bot_G, \mathcal{Z}_o(s, \top_M) \geq \mathcal{Z}_o(t, \bot_M) + \delta \text{ and}$$
$$\forall s \in M \setminus \top_M, \forall t \in \bot_M, \mathcal{Z}_a(s, \top_G) \geq \mathcal{Z}_a(t, \bot_G) + \delta$$

If $Prune^m_{\mathcal{C}_1}(\mathcal{L})$ (resp. $Prune^{am}_{\mathcal{C}_2}(\mathcal{L})$ and $Prune_{\mathcal{C}_r}(\mathcal{L})$) is FALSE, then any bi-set contained in $\mathcal{L}$ does not satisfy $\mathcal{C}_1$ (resp. $\mathcal{C}_2$ and $\mathcal{C}_r$).
In DR-MINER, we use $Prune : \mathcal{B} \rightarrow \{\text{TRUE,FALSE}\}$ which is such that $Prune(\mathcal{L}) \equiv Prune^m_{\mathcal{C}_1}(\mathcal{L}) \wedge Prune^{am}_{\mathcal{C}_2}(\mathcal{L}) \wedge Prune_{\mathcal{C}_r}(\mathcal{L}) \wedge Prune^m_{\mathcal{C}_d}(\mathcal{L})$

– **Propagation:** $\mathcal{C}_d$ and $\mathcal{C}_r$ can be used to reduce the size of the sublattices by moving objects of $\top_G \setminus \bot_G$ into $\bot_G$ or outside $\top_G$. The fonctions $Prop_{in} : \mathcal{B} \rightarrow \mathcal{B}$ and $Prop_{out} : \mathcal{B} \rightarrow \mathcal{B}$ are used to do it as follow:

$$Prop_{in}((\bot_G, \bot_M), (\top_G, \top_M)) = \{((\bot'_G, \bot'_M), (\top_G, \top_M)) \in \mathcal{B} \mid$$
$$\bot'_G = \bot_G \cup \{x \in \top_G \setminus \bot_G \mid \exists t \in \bot_G,\ \mathcal{Z}_o(x, \top_M) < \mathcal{Z}_o(t, \bot_M) + \delta\}$$
$$\bot'_M = \bot_M \cup \{x \in \top_M \setminus \bot_M \mid \exists t \in \bot_M,\ \mathcal{Z}_a(x, \top_G) < \mathcal{Z}_a(t, \bot_G) + \delta\}\}$$

$$Prop_{out}((\bot_G, \bot_M), (\top_G, \top_M)) = \{((\bot_G, \bot_M), (\top'_G, \top'_M)) \in \mathcal{B} \mid$$
$$\top'_G = \top_G \setminus \{x \in \top_G \setminus \bot_G \mid \mathcal{Z}_o(x, \bot_M) > \alpha\}$$
$$\top'_M = \top_M \setminus \{x \in \top_M \setminus \bot_M \mid \mathcal{Z}_a(x, \bot_G) > \alpha\}\}$$

$Prop : \mathcal{B} \rightarrow \mathcal{B}$ is defined as $Prop(\mathcal{L}) = Prop_{in}(Prop_{out}(\mathcal{L}))$. It is recursively applied as long as its result changes.

To prove the correctness and completeness of DR-MINER, a sublattice $\mathcal{L} = ((\bot_G, \bot_M), (\top_G, \top_M))$ is called a leaf when it contains only one bi-set i.e., $(\bot_G, \bot_M) = (\top_G, \top_M)$. DR-bi-sets are these maximal bi-sets. To extract only maximal dense and relevant ones, we have adapted the DUAL-MINER strategy for pushing maximality constraints [4].

DR-Miner correctness: Every bi-set (X, Y) belonging to leaf $\mathcal{L}$ satisfies $\mathcal{C}_d \wedge \mathcal{C}_r$ according to $Prune^{am}_{\mathcal{C}_d}$ and $Prune_{\mathcal{C}_r}$.

DR-Miner completeness: Let $T_1 = ((\bot^1_G, \bot^1_M), (\top^1_G, \top^1_G))$ and $T_2 = ((\bot^2_G, \bot^2_M), (\top^2_G, \top^2_G))$. Let $\sqsubseteq$ be a partial order on $\mathcal{B}$ defined as $T_1 \sqsubseteq T_2$ iff $(\bot^2_G, \bot^2_M) \preceq (\bot^1_G, \bot^1_M)$ and $(\top^1_G, \top^1_G) \preceq (\top^2_G, \top^2_G)$ (see Definition 3). $\sqsubseteq$ is the partial order used to generate the sublattices.

We show that for each bi-set (X, Y) satisfying $\mathcal{C}_d \wedge \mathcal{C}_r$, it exists a leaf $\mathcal{L} = ((X, Y), (X, Y))$ which is generated by the algorithm.

Property 3. If $\mathcal{F}$ is a sublattice such that $\mathcal{L} \sqsubseteq \mathcal{F}$ then among the two sublattices obtained by the enumeration of $\mathcal{F}$ $(Enum(\mathcal{F}, Choose(\mathcal{F})))$ one and only one is a super-set of $\mathcal{L}$ w.r.t. $\sqsubseteq$. This property is conserved by function *Prop*.

Proof. Let $\mathcal{F} = ((\bot_G, \bot_M), (\top_G, \top_M)) \in \mathcal{B}$ such that $\mathcal{L} \sqsubseteq \mathcal{F}$. Assume that the enumeration is done on objects (it is similar on attributes) and that the two sublattices generated by the enumeration of $o \in \top_G \setminus \bot_G$ are $\mathcal{L}_1$ and $\mathcal{L}_2$. If $o \in X$ then $\mathcal{L} \sqsubseteq \mathcal{L}_1$ and $\mathcal{L} \not\sqsubseteq \mathcal{L}_2$, otherwise $\mathcal{L} \sqsubseteq \mathcal{L}_2$ and $\mathcal{L} \not\sqsubseteq \mathcal{L}_1$.

Let us now show that constraint propagation (function *Prop*) on any sublattice $\mathcal{F} = ((\bot_G, \bot_M), (\top_G, \top_M))$ such that $\mathcal{L} \sqsubseteq \mathcal{F}$ preserves this order. More precisely, no element of X is removed of $\top_G$ due to $Prop_{out}$ (Case 1) and no element of $G \setminus X$ is moved to $\bot_G$ due to $Prop_{in}$ (Case 2).

- Case 1: (X, Y) satisfies $\mathcal{C}_r$ then $\forall p \in \top_G \setminus \bot_G$ s.t. $p \in G \setminus X$ and $\forall t \in \bot_G$, we have $\mathcal{Z}_o(p, Y) \geq \mathcal{Z}_o(t, Y) + \delta$. But $\bot_M \subseteq Y \subseteq \top_M$, and thus $\mathcal{Z}_o(p, \top_M) \geq \mathcal{Z}_o(p, Y) \geq \mathcal{Z}_o(t, Y) + \delta \geq \mathcal{Z}_o(t, \bot_M) + \delta$. Consequently, $\mathcal{Z}_o(p, \top_M) < \mathcal{Z}_o(t, \bot_M) + \delta$ is false. Consequently, p is not moved to $\bot_G$.
- Case 2: (X, Y) satisfies $\mathcal{C}_d$ then $\forall p \in \top_G \setminus \bot_G$ s.t. $p \in X$, we have $\mathcal{Z}_o(p, Y) \leq \alpha$. But $\bot_M \subseteq Y$, and thus $\mathcal{Z}_o(p, \bot_M) \leq \mathcal{Z}_o(p, Y) \leq \alpha$. Consequently, p is not removed from $\top_G$. □

Since DR-Miner starts with $((\emptyset, \emptyset), (G, M))$ which is a super-set of $\mathcal{L}$, given that $\mathcal{B}$ is finite and that recursively it exists always a sublattice which is an super-set of $\mathcal{L}$ w.r.t. $\sqsubseteq$ even after the propagation has been applied, then we can affirm that every bi-set satisfying $\mathcal{C}_d \wedge \mathcal{C}_r$ is extracted by DR-miner.

5 Experimentation

5.1 Robustness on Synthetic Data

Let us first illustrate the added-value of DR-bi-set mining in synthetic data. Our goal is to show that the extraction of these patterns in noisy data sets enables to find some originally built-in formal concepts blurred by some random noise. Our raw synthetic data is a matrix 30×15 in which three disjoint formal concepts of size 10×5 hold. Then, we introduced a uniform random noise on the whole matrix and 5 different data sets have been produced for each level of noise, i.e., from 1% to 30% (each zero or one value has a probability of X% to be changed).

To compare the extracted collections with the three original built-in formal concepts, we used a measure which tests the presence of a subset of the original pattern collection in the extracted ones. This measure σ associates to each pattern of one collection $\mathcal{C}_1$ the closest pattern of the other one $\mathcal{C}_2$ (and reciprocally). It is based on a distance measure taking into account their shared area:

$$\sigma(\mathcal{C}_1, \mathcal{C}_2) = \frac{\rho(\mathcal{C}_1, \mathcal{C}_2) + \rho(\mathcal{C}_2, \mathcal{C}_1)}{2}$$
$$\rho(\mathcal{C}_1, \mathcal{C}_2) = \frac{1}{\sharp \mathcal{C}_1} \sum_{(X_i, Y_i) \in \mathcal{C}_1} \max_{(X_j, Y_j) \in \mathcal{C}_2} \frac{\sharp (X_i \cap X_j) * \sharp (Y_i \cap Y_j)}{\sharp (X_i \cup X_j) * \sharp (Y_i \cup Y_j)}$$

when $\rho(\mathcal{C}_1, \mathcal{C}_2) = 1$, each pattern of $\mathcal{C}_1$ has an identical instance in $\mathcal{C}_2$, and when $\sigma = 1$, the two collections are identical. High values of σ mean that (a) we can

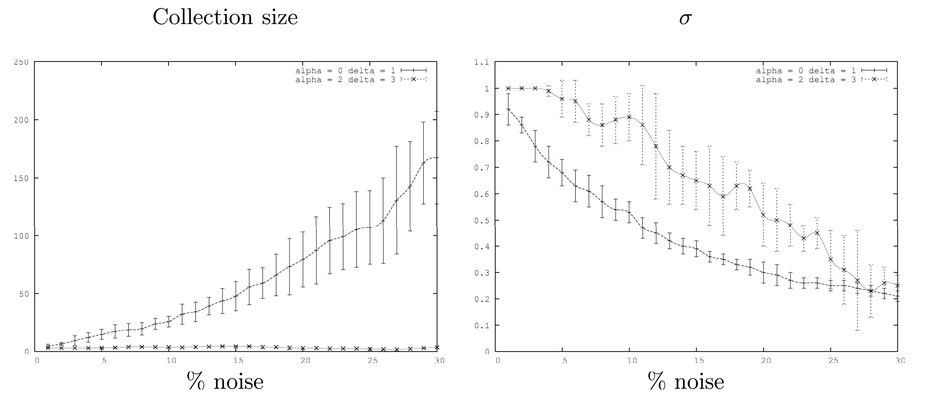

Fig. 3. Mean and standard deviation of the number of bi-sets (5 trials) (left) and of σ (right) w.r.t. the percentage of noise

find all the formal concepts of the reference collection within the noised matrix, and (b) the collection extracted from noised matrices does not contain many bi-sets that are too different from the reference ones.

Figure 3 presents the mean and the standard deviation of the number of extracted bi-sets (left) and the mean and standard deviation of σ (right) for each level of noise. Two collections are represented: one for $\alpha = 0$ and $\delta = 1$ (i.e., the case of formal concepts), and the second one for $\alpha = 2$ and $\delta = 3$. On both collections, a minimal size constraint is added which enforces that each pattern contains at least 3 elements on each dimension (i.e., satisfying $\mathcal{C}_{ms}(3,3)$). It avoids the computation of the smallest bi-sets which can indeed be due to noise.

We can observe that when the noise level increases, the number of extracted formal concepts (i.e., $\alpha = 0$ and $\delta = 1$) increases drastically, whereas σ decreases drastically as well. For $\alpha = 2$ and $\delta = 3$, we observe an important reduction of the number of extracted DR-bi-sets and an important increase of the DR-bi-set quality: for 10 % of noise the collection is similar to the built-in formal concept collection. These graphics emphasize the difference between the use of formal concepts and DR-bi-sets in noisy data: the first one constitutes a large collection (tens to hundreds of patterns) of poorly relevant patterns, whereas the second one is clearly closer to the three built-in patterns. Indeed, we get between 2 and 4 patterns with higher σ values. When the level of noise is very high (say over 20%), the DR-bi-sets are not relevant any more. Indeed, with such level of noise, the data turns to be random.

5.2 Impact of Parameters α and δ

To study the influence of the α parameter, we performed several mining tasks on the UCI data set Internet Advertisements which is large on both dimensions (matrix 3 279 $\times$ 1 555) [3].

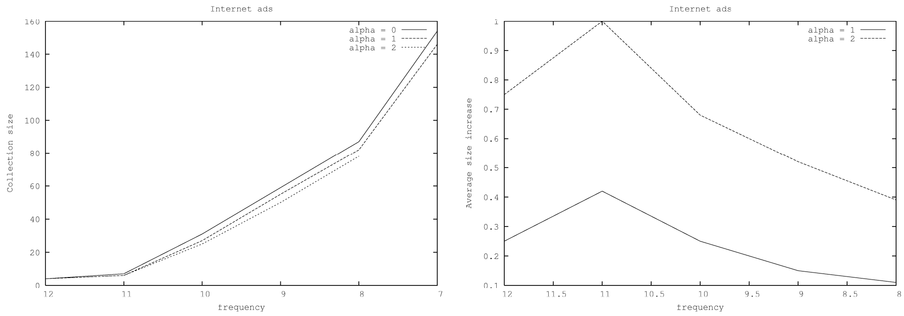

Fig. 4. Number of extracted DR-bi-sets (left) and average increase of bi-set size w.r.t. formal concepts (right) for several frequency thresholds on both dimensions (‰), with $\delta = 1$ and $\alpha \in 0..2$

We have extracted DR-bi-set collections with a minimal size constraint on both dimensions varying between 12‰ and 7‰, where $\delta = 1$ and α varying between 0 and 2. Figure 4 (left) shows the size of DR-bi-set collections. In this data set, the collection sizes decrease with α. Figure 4 (right) shows the average number of added objects and attributes of each formal concept. More formally, if $\mathcal{C}_0$ denotes the collection of formal concepts and if $\mathcal{C}_\alpha$ denotes the collection of DR-bi-sets obtained with parameter α, the measure is computed as follow:

$$\frac{1}{\sharp\mathcal{C}_0} \sum_{(X_0,Y_0)\in\mathcal{C}_0} \max_{(X_\alpha,Y_\alpha)\in\mathcal{A}(X_0,Y_0)} \sharp(X_\alpha \setminus X_0) \, * \, \sharp(Y_\alpha \setminus Y_0)$$

where $\mathcal{A}(X_0, Y_0) = \{(X, Y) \in \mathcal{C}_\alpha \, such \, that (X_0, Y_0) \preceq (X, Y)\}$ and $\preceq$ is the order of Definition 3. As proved in Property 1, the average sizes of the extracted bi-sets increase with α. But we can observe that this increase is quite important: for example, for $\alpha = 2$ and $frequency = 11$, one element has been added to each formal concept in average.

To study the influence of the δ parameter, we have also performed experiments on the UCI data set Mushroom (matrix 8 124 × 128) [3] and on the real world medical data set Meningitis [5]. Meningitis data have been gathered from children hospitalized for acute meningitis. The pre-processed Boolean data set is composed of 329 patients described by 60 Boolean properties.

A straightforward approach to avoid some irrelevant patterns and to reduce the pattern collection size is to use size constraints on bi-set components. For these experiments, we use the constraint $\mathcal{C}_{ms}(500, 10)$ on Mushroom and $\mathcal{C}_{ms}(10, 5)$ on Meningitis. Using D-Miner [2], we have computed the collection of such large enough formal concepts and we got more than 1 000 formal concepts on Mushroom and more than 300 000 formal concepts on Meningitis (see Table 3). We used different values of δ on G (denoted δ) and on M (denoted δ').

Table 3 gathers the results obtained on the two data sets. For Mushroom, α is fixed to 0 and $\delta = \delta'$ are varying between 2 and 6. We can observe that the collection sizes drastically decrease with δ and δ'. On Meningitis, α is set to 1 and δ' is varying between 2 and 6 whereas δ is set to 1. We use different values for δ and δ' because the pattern sizes were greater on the object set components and thus we wanted to enforce the difference with the outside elements on these components. For this data set, not only the collection sizes, but also the computational times are considerably reduced when δ' increases. Notice that $\delta' = 1$ leads to an intractable extraction but, with $\delta' = 2$, the resulting collection is 80% smaller than the related formal concept collection. Such decreases are observed when considering higher δ' values.

Table 3. DR-bi-set collection sizes and extraction time when δ' is varying from 1 to 6 on Mushroom and Meningitis

Mushroom ($\mathcal{C}_{ms}(500, 10)$, $\alpha = 0$)							
$\delta = \delta'$	Concepts	1	2	3	4	5	6
size	1 102	1 102	11	6	2	1	0
time	1.6s	10s	4s	4s	3s	2s	2s
Meningitis ($\mathcal{C}_{ms}(10, 5)$, $\alpha = 1$, $\delta = 1$)							
δ'	Concepts	1	2	3	4	5	6
size	354 366	-	75 376	22 882	8 810	4 164	2 021
time	5s	-	693s	327s	181s	109s	70s

6 Conclusion

We have considered the challenging problem of computing fault-tolerant bi-sets. Formal concepts fail to emphasize relevant associations when the data is intrinsically noisy. We have formalized a new task, maximal dense and relevant bi-set mining, within the constraint-based data mining framework. We propose a complete algorithm DR-Miner which computes every DR-bi-set by pushing these constraints during an enumeration process. Density refers to the bounded number of zero values and relevancy refers to the specificities of the elements involved in the extracted bi-sets when considering the whole data set. We experimentally validated the added-value of this approach on both synthetic and real data. Fixing the various parameters might appear difficult (it is often driven by tractability issues) but this is balanced by the valuable counterpart of completeness: the user knows exactly which properties are satisfied by the extracted collections.

Acknowledgements. This research is partially funded by ACI Masse de Données Bingo (CNRS STIC MD 46) and the EU contract IQ FP6-516169 (FET arm of the IST programme). We thank Ruggero G. Pensa for his contribution to the experimental validation and Jean-Marc Petit for his comments.

References

1. J. Besson, C. Robardet, and J.-F. Boulicaut. Mining formal concepts with a bounded number of exceptions from transactional data. In *Post-Workshop KDID'04*, volume 3377 of *LNCS*, pages 33–45. Springer, 2005.
2. J. Besson, C. Robardet, J.-F. Boulicaut, and S. Rome. Constraint-based bi-set mining for biologically relevant pattern discovery in microarray data. *IDA journal*, 9(1):59–82, 2005.
3. C. Blake and C. Merz. UCI repository of machine learning databases, 1998.
4. C. Bucila, J. E. Gehrke, D. Kifer, and W. White. Dualminer: A dual-pruning algorithm for itemsets with constraints. In *ACM SIGKDD*, pages 42–51, 2002.
5. P. François, C. Robert, B. Cremilleux, C. Bucharles, and J. Demongeot. Variables processing in expert system building: application to the aetiological diagnosis of infantile meningitis. *Med. Inf.*, 15(2):115–124, 1990.
6. B. Ganter, G. Stumme, and R. Wille, editors. *Formal Concept Analysis, Foundations and Applications*, volume 3626 of *LNCS*. springer, 2005.
7. A. Gionis, H. Mannila, and J. K. Seppänen. Geometric and combinatorial tiles in 0-1 data. In *PKDD'04*, volume 3202 of *LNAI*, pages 173–184. Springer, 2004.
8. A. Guenoche and I. V. Mechelen. Galois approach to the induction of concepts. *Categories and concepts : Theorical views and inductive data analysis*, pages 287–308, 1993.
9. J. Hereth, G. Stumme, R. Wille, and U. Wille. Conceptual knowledge discovery and data analysis. In *ICCS'00*, pages 421–437, 2000.
10. S. O. Kuznetsov and S. A. Obiedkov. Comparing performance of algorithms for generating concept lattices. *JETAI*, 14 (2-3):189–216, 2002.
11. E. M. Nguifo, V. Duquenne, and M. Liquiere. Concept lattice-based knowledge discovery in databases. *JETAI*, 14((2-3)):75–79, 2002.
12. J. Pei, A. K. H. Tung, and J. Han. Fault-tolerant frequent pattern mining: Problems and challenges. In *DMKD*. Workshop, 2001.
13. R. G. Pensa and J.-F. Boulicaut. Towards fault-tolerant formal concept analysis. In *AI*IA'05*, volume 3673 of *LNAI*, pages 212–223. Springer-Verlag, 2005.
14. J. K. Seppänen and H. Mannila. Dense itemsets. In *ACM SIGKDD'04*, pages 683–688, 2004.
15. G. Stumme, R. Taouil, Y. Bastide, N. Pasqier, and L. Lakhal. Computing iceberg concept lattices with TITANIC. *DKE*, 42:189–222, 2002.
16. V. Ventos, H. Soldano, and T. Lamadon. Alpha galois lattices. In *ICDM IEEE*, pages 555–558, 2004.
17. R. Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. In I. Rival, editor, *Ordered sets*, pages 445–470. Reidel, 1982.
18. C. Yang, U. Fayyad, and P. S. Bradley. Efficient discovery of error-tolerant frequent itemsets in high dimensions. In *ACM SIGKDD*, pages 194–203. ACM Press, 2001.