Towards Fault-Tolerant Formal Concept Analysis

Ruggero G. Pensa and Jean-François Boulicaut

INSA Lyon, LIRIS CNRS UMR 5205,
F-69621 Villeurbanne cedex, France
{Ruggero.Pensa, Jean-Francois.Boulicaut}@insa-lyon.fr

Abstract. Given Boolean data sets which record properties of objects, Formal Concept Analysis is a well-known approach for knowledge discovery. Recent application domains, e.g., for very large data sets, have motivated new algorithms which can perform constraint-based mining of formal concepts (i.e., closed sets on both dimensions which are associated by the Galois connection and satisfy some user-defined constraints). In this paper, we consider a major limit of these approaches when considering noisy data sets. This is indeed the case of Boolean gene expression data analysis where objects denote biological experiments and attributes denote gene expression properties. In this type of intrinsically noisy data, the Galois association is so strong that the number of extracted formal concepts explodes. We formalize the computation of the so-called δ-bi-sets as an alternative for capturing strong associations between sets of objects and sets of properties. Based on a previous work on approximate condensed representations of frequent sets by means of δ-free itemsets, we get an efficient technique which can be applied on large data sets. An experimental validation on both synthetic and real data is given. It confirms the added-value of our approach w.r.t. formal concept discovery, i.e., the extraction of smaller collections of relevant associations.

1 Introduction

Formal Concept Analysis has been developed for more than two decades [1]. It supports knowledge discovery (e.g., clustering, association rule mining) in contexts where a number of Boolean properties hold or not for a collection of objects. For instance, Table 1 is a toy example data set $\mathbf{r}$ where we see that attributes p_2 and p_5 are true for object o_2. Informally, formal concepts are maximal rectangle[1] of true values. For instance, $(\{o_1, o_3\}, \{p_1, p_3, p_4\})$ is a formal concept in $\mathbf{r}$.

Among others, formal concepts can be considered as overlapping clusters which are intrinsically characterized: the reason why o_1 and o_3 are in the same cluster is that they all share properties p_1, p_3, and p_4. Such a conceptual clustering [2] is crucially needed in many application domains. For this purpose, co-clustering (also called bi-clustering) has been proposed [3,4,5,6]. The goal is to identify bi-partitions, i.e., associated partitions on both dimensions. When applied on Boolean matrices, these techniques tend to provide rectangles with

[1] Rectangle has to be understood modulo arbitrary permutations of lines and columns.

S. Bandini and S. Manzoni (Eds.): AI*IA 2005, LNAI 3673, pp. 212–223, 2005.

Table 1. A Boolean context **r**

	p_1	p_2	p_3	p_4	p_5
o_1	1	0	1	1	0
o_2	0	1	0	0	1
o_3	1	0	1	1	0
o_4	0	0	1	1	0
o_5	1	1	0	0	1
o_6	0	1	0	0	1
o_7	0	0	0	0	1

mainly true values. Notice however that they are based on heuristic techniques (i.e., local optimization) and that they generally compute collections of non overlapping bi-clusters. Instead, the strength of Formal Concept Analysis is that, when tractable, it is based on complete collections of formal concepts which are overlapping clusters. The state-of-the-art is that we can compute collections of formal concepts in many practical applications. First, some algorithms are dedicated to formal concept discovery (see [7] for a survey). Then, for tackling very large contexts, constraint-based mining of formal concepts has been studied (see, e.g., [8,9]). In this case, we still compute complete collections containing every formal concept which satisfies some other user-defined constraints (e.g., a minimal size for their set components).

The application domain which motivates our research is Boolean gene expression data analysis, i.e., knowledge discovery from data sets which encode gene expression properties (e.g., over-expression) in various biological situations or experiments. Given Table 1, we might say that, e.g., genes denoted by p_1, p_3, p_4 are considered over-expressed in situation o_1. Interestingly, in such a context, formal concepts can be considered as putative transcription modules, i.e., maximal sets of genes that are co-regulated associated to the maximal sets of experiments which seem to trigger this co-regulation. Notice that bi-cluster overlapping makes sense from the biological point of view (i.e., the same gene can be involved in various biological functions). Transcription module discovery is an important step towards the understanding of gene regulation and we address a severe limitation of putative transcription module discovery from formal concepts[2]. Within a formal concept, we have a maximal set of objects (i.e., a closed set) which are in relation with all the elements of a maximal set of properties and vice versa. The strength of such an association is often too strong in real-life data. Assume that, e.g., $c_1 = (\{o_1, o_3, o_4\}, \{p_1, p_3, p_4\})$ is a "valid" association in the application domain. Let us now consider that, like in **r**, we do not record that p_1 is true for o_4. As a result, we do not get c_1 but instead the two formal concepts $(\{o_1, o_3, o_4\}, \{p_3, p_4\})$ and $(\{o_1, o_3\}, \{p_1, p_3, p_4\})$. In fact, the presence of false values which have been set "by error" leads to an explosion of the number of formal concepts. We have problems with values inappropriately set to true as well. Such noisy data is quite common, e.g., in life science domains, where we can not avoid errors of measurement but also further problems with Boolean

[2] More generally, we consider the search for interesting bi-clusters from Boolean data.

property encoding. For instance, encoding a gene expression property, say over-expression, from typical numerical microarray data relies on the definition of a threshold whose value enables to encode true or false [10]. This intrinsically introduces noise. As a result, the number of formal concepts which hold in real-life Boolean gene expression data sets can be huge, e.g., up to several millions. Even though the extraction might remain tractable, the needed post-processing phases turn to be tedious or even impossible.

These observations have motivated a new direction of research where interesting bi-clusters are considered as dense rectangles of true values (see, e.g., [11,12,13]). Such rectangles look like formal concepts with a number of exceptions, i.e., hopefully, a bounded number of false values per line and per column. To the best of our knowledge, previous attempts are not really satisfactory for our application domain. Either they rely on very expensive algorithms (e.g., [11] which is based on formal concept merging) or they assume quite strong hypothesis on the data (e.g., [12] which assumes a built-in order on both dimensions). Instead of looking for such fault-tolerant formal concepts, we would like to revisit a previous work on the so-called δ-free itemsets, i.e., one of the few approximate condensed representations of frequent itemsets [14]. The idea was to consider specific itemsets, the δ-free ones, whose frequency have to be counted in order to infer without counting and with a bounded error the frequency of many others. We consider the bi-sets which can be built on δ-free sets of properties and their δ-closures (i.e., associated attributes which are almost always true) on one hand, on the sets of objects which support the δ-free set on the properties on another hand. As a result, δ-bi-sets contain a bounded number of exceptions per column. An example in the data set $\mathbf{r}$ is that $\{p_1\}$ is a 1-free set whose 1-closure (the properties which are almost always true with p_1, i.e., with at least 1 exception) is $\{p_3, p_4\}$. It means that the bi-set $(\{o_1, o_3, o_5\}, \{p_1, p_3, p_4\})$ is a 1-bi-set. Indeed, it has at most 1 exception per column. The extraction of δ-bi-sets can be extremely efficient thanks to δ-freeness anti-monotonicity. Such collections can be computed in many data sets, including huge ones. Our intuition is that, in real data sets, the distribution of these exceptions among the lines will be acceptable such that δ-bi-sets capture really strong associations between sets of objects and sets of properties. By considering synthetic data sets but also real-life data sets, we illustrate that formal concept extraction can be hard and/or useless in noisy data sets. We also demonstrate the added-value of the δ-bi-set extraction method in order to get an a priori interesting collections of overlapping bi-clusters.

Section 2 provides the needed definitions for the formalization and the use of the δ-bi-set mining task. Section 3 provides experimental results on synthetic or benchmark data when various levels of noise are added. Section 4 considers several experiments on real-life bio-medical data sets which are intrinsically noisy. Section 5 is a short conclusion.

2 An Alternative to Formal Concepts in Noisy Data Sets

Assume a set of objects $\mathcal{O} = \{o_1, \ldots, o_m\}$ and a set of Boolean attributes $\mathcal{P} = \{p_1, \ldots, p_n\}$. The Boolean context to be mined is $\mathbf{r} \subseteq \mathcal{O} \times \mathcal{P}$, where $r_{ij} = 1$ if the

attribute p_j is true for the object o_i. In Boolean gene expression data sets, if o_i is a biological sample, p_j denotes an expression property of a gene, e.g., $r_{ij} = 1$ means that gene associated to p_j is over-expressed in o_i.

It is interesting to look for associations between sets of objects and sets of properties, i.e., bi-sets. An obvious measure which quantifies the strength of such associations is the density of true values within the bi-set. Formal concepts are maximal bi-sets with only true values. The problem is that the number of formal concepts in noisy data sets explodes and that it makes sense to relax the associated closeness constraint to capture less but relevant strong associations, i.e., some kind of fault-tolerant formal concepts.

Formally, a bi-set (T, G) is a couple of sets from $2^{\mathcal{O}} \times 2^{\mathcal{P}}$. T is called an objectset and G is called an itemset. Let us first recall the basic definition of our Galois connection (see, e.g., [1]).

Definition 1 (Galois connection). *If $T \subseteq \mathcal{O}$ and $G \subseteq \mathcal{P}$, ϕ and ψ constitute a Galois connection when $\phi(T, \mathbf{r}) = \{p \in \mathcal{P} \mid \forall o \in T,\ (o, p) \in \mathbf{r}\}$ and $\psi(G, \mathbf{r}) = \{o \in \mathcal{O} \mid \forall p \in G,\ (o, p) \in \mathbf{r}\}$. $h = \phi \circ \psi$ and $h' = \psi \circ \phi$ denote the closure operators. A set $T \subseteq \mathcal{O}$ (resp. $G \subseteq \mathcal{P}$) is said closed in $\mathbf{r}$ iff $T = h'(T, \mathbf{r})$ (resp. $G = h(G, \mathbf{r})$).*

We can now formalize some usual pattern types.

Definition 2 (1-rectangles, formal concepts, supporting sets). *A bi-set (T, G) is a 1-rectangle in $\mathbf{r}$ iff $\forall o \in T$ and $\forall p \in G$, $(o, p) \in \mathbf{r}$. A bi-set (T, G) is a formal concept in $\mathbf{r}$ iff $T = \psi(G, \mathbf{r})$ and $G = \phi(T, \mathbf{r})$. It is equivalent to $T = h'(T, \mathbf{r})$ and $G = \phi(T, \mathbf{r})$ or to $G = h(G, \mathbf{r})$ and $T = \psi(G, \mathbf{r})$. An important property is indeed that each closed set on one of the two dimensions is associated to a unique closed set on the other dimension. We say that the support of an itemset G (resp. an objectset T) in $\mathbf{r}$ is $\psi(G, \mathbf{r})$ (resp. $\phi(T, \mathbf{r})$).*

For example, $\{\{o_1, o_3\}, \{p_1, p_3\}\}$ is a 1-rectangle in $\mathbf{r}$ (see Table 1) but it is not maximal. $(\{o_1, o_3\}, \{p_1, p_3, p_4\})$,$(\{o_1, o_3, o_4\}, \{p_3, p_4\})$, and $(\{o_2, o_5, o_6\}$,$\{p_2, p_5\})$ are examples of formal concepts among the 8 ones which hold in $\mathbf{r}$. $\{o_1, o_3, o_5\}$ is the supporting set of $\{p_1\}$. $\{p_1, p_3, p_4\}$ is the supporting set of $\{o_1\}$.

Sections 3 and 4 illustrate on concrete examples that, even in small matrices, the number of formal concepts can be huge. In fact, the size of the collection of formal concepts in a given matrix is exponential in its smallest dimension. Formalizing the δ-bi-set mining task, we want to compute smaller collections which still capture important associations within the data. Collections are smaller because a given δ-bi-set can always be described as a merge of some formal concepts.

Let us first recall the popular association rule mining task [15] since it is needed to understand the δ-freeness property.

Definition 3 (association rule, frequency, confidence). *An association rule R in a data set $\mathbf{r}$ is an expression of the form $X \Rightarrow Y$, where $X, Y \subseteq \mathcal{P}$, $Y \neq \emptyset$ and $X \cap Y = \emptyset$. The frequency of R is $|\psi(X \cup Y, \mathbf{r})|$ and the confidence of R is $|\psi(X \cup Y, \mathbf{r})|/|\psi(X, \mathbf{r})|$.*

In an association rule $X \Rightarrow Y$ with high confidence, the properties in Y are almost always true for an object when the properties in X are true. Intuitively, the itemset $X \cup Y$ associated to the set of object $T = \psi(X, \mathbf{r})$ is then a dense bi-set. Moreover, the more the rule is frequent, the larger the bi-set will be.

We now consider our technique for computing association rules with high confidence, the so-called δ-strong rules [14].

Definition 4 (δ-strong rule). *Given an integer value δ, a δ-strong rule in a Boolean context $\mathbf{r}$ is an association rule $X \Rightarrow Y$ ($X, Y \subset \mathcal{P}$) such that $|\psi(X, \mathbf{r})| - |\psi(X \cup Y, \mathbf{r})| \leq \delta$, i.e., the rule is violated in no more than δ objects.*

Interesting collections of δ-strong rules with minimal left-hand side can be computed efficiently from the so-called δ-free-sets [14] and their δ-closures.

Definition 5 (δ-free set, δ-closure). *Let δ be an integer and an $X \subset \mathcal{P}$ be an itemset, X is a δ-free-set w.r.t. $\mathbf{r}$ if and only if there is no δ-strong rule which holds between two of its own and proper subsets. The δ-closure of X in $\mathbf{r}$, $h_\delta(X, \mathbf{r})$, is the maximal (w.r.t. set inclusion) superset Y of X s.t. for every item $p \in Y \setminus X$, $|\psi(X \cup \{p\})|$ is at least $|\psi(X, \mathbf{r})| - \delta$. In other terms, the frequency of the δ-closure of X in $\mathbf{r}$ is almost the same than the frequency of X when δ is small w.r.t. the number of objects. It means also that $\forall p \in h_\delta(X) \setminus X$, $X \Rightarrow p$ is an association rule with high confidence, more precisely a δ-strong rule.*

For example, in Table 1, the 1-free itemsets are $\{p_1\}$, $\{p_2\}$, $\{p_3\}$, $\{p_4\}$, $\{p_5\}$, $\{p_1, p_2\}$, and $\{p_1, p_5\}$. An example of 1-closure for $\{p_1\}$ is $\{p_1, p_3, p_4\}$. The association rules $\{p_1\} \Rightarrow \{p_3\}$ and $\{p_1\} \Rightarrow \{p_4\}$ have only one exception.

δ-freeness is an anti-monotonic property such that it is possible to compute δ-free sets (eventually combined with a minimal frequency constraint) in very large data sets. Notice than when $\delta = 0$, $h_0 = h$, i.e., the classical closure operator. Looking for a 0-free itemset, say X, and its 0-closure provides a closed itemset Y. When a closed set is computed by this technique, we get easily the formal concept (T, Y) by associating its supporting set of objects $T = \psi(Y, \mathbf{r})$. We can now use the properties of δ-free-sets and δ-strong rules to extract a collection of dense bi-sets with a bounded number of exceptions per column.

Definition 6 (δ-bi-set). *A δ-bi-set (T, G) in $\mathbf{r}$ is built on each δ-free-set $X \subset \mathcal{P}$ and we have $G = h_\delta(X, \mathbf{r})$ and $T = \psi(X, \mathbf{r})$.*

It is clear that for a δ-bi-set (T, G), when $\delta << |T|$, (T, G) denotes a strong association between T and G. In Table 1, itemsets $\{p_3\}$ and $\{p_5\}$ are examples of 1-free-sets. The related 1-bi-sets are $\{\{o_1, o_3, o_4\}, \{p_1, p_3, p_4\}\}$ and $\{\{o_2, o_5, o_6, o_7\}, \{p_2, p_5\}\}$. Obviously, when $\delta = 0$, each δ-bi-set is a formal concept.

An algorithm to extract δ-bi-sets. For the experimentations, we have been using a straightforward extension of the MIN-EX implementation described in [14]. Indeed, we just added the automatic generation of the supporting set for each extracted δ-free-set. MIN-EX is a typical instance of the levelwise search

algorithm presented in [16]. Thanks to the antimonotonicity of the conjunction of δ-freeness and a minimal frequency constraint, it explores the itemset lattice (w.r.t. the inclusion) levelwise, starting from the empty set and stopping at the level of the largest frequent δ-free-set. More precisely, the collection of candidates is initialized with the empty set as single member (the only set of size 0) and then the algorithm iterates on candidate evaluation (i.e., checking both δ-freeness and minimal frequency) and larger candidate generation. At iteration i, it scans the data to find out which candidates of size i are frequent δ-free-sets and it computes their δ-closure as well. Then it generates candidates for the next iteration, taking every set of size $i+1$ such that all their proper subsets are frequent δ-free-sets. The algorithm stops when there is no more candidate. The needed δ-free sets can thus be extracted by setting the frequency threshold to 1. Also, our implementation outputs each supporting set of lines for each discovered δ-free set of columns.

3 Experiments on Data Plus Noise

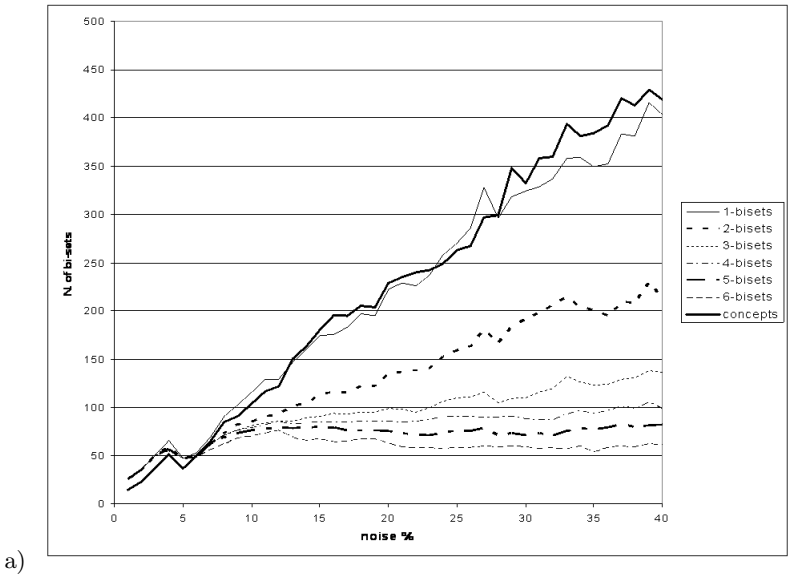
First, we study the relevancy of δ-bi-sets w.r.t. formal concepts when considering the addition of noise to a synthetical data set and to a benchmark data set from UCI Macine Learning Repository [17]. Let us first discuss the evaluation method.

Hereafter, $\mathbf{r}$ denotes a reference data set, i.e., a data set which is assumed to be noise-free. We use it to generate noisy data sets by adding a given quantity of uniform random noise (for a X% noise level, each value is randomly changed with a probability of X%). Then, we compare the collection of formal concepts which are "built-in" within $\mathbf{r}$ with various collections of bi-sets (i.e., formal concepts and δ-bi-sets) extracted from the noised matrices. To measure the relevancy of each extracted collection w.r.t the reference one, we look for a subset of the reference collection in each of them. Since the objectset and the itemset of each formal concept can be changed when adding noise to the data, we identify those having the largest area in common with the original ones, and we compute a measure called σ which takes into account the common area:

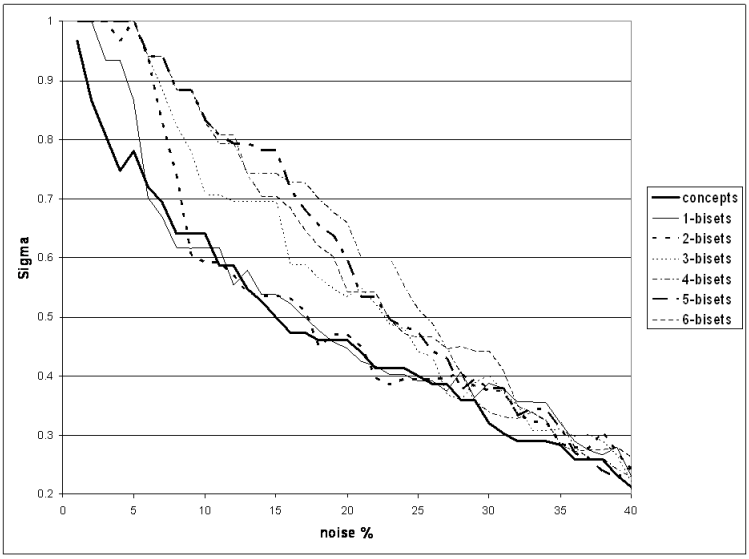
$$\sigma(\mathcal{C}_r, \mathcal{C}_a) = \frac{1}{N_r} \sum_{i=1}^{N_r} max_j \left(\frac{|(T_i, G_i)_r \cap (T_j, G_j)_a|}{|(T_i, G_i)_r \cup (T_j, G_j)_a|} \right)$$

where $\mathcal{C}_r$ is the collection of concepts computed on the reference $\mathbf{r}$, $\mathcal{C}_a$ is a noised collection of bi-sets, $(T_i, G_i)_r$ and $(T_j, G_j)_a$ are bi-sets belonging to $\mathcal{C}_r$ and $\mathcal{C}_a$ respectively, and N_r is the size of the reference collection of formal concepts. When $\sigma(\mathcal{C}_r, \mathcal{C}_a) = 1$, all the bi-sets belonging to $\mathcal{C}_r$ have identical instances in the collection $\mathcal{C}_a$.

In the experiment, $\mathbf{r}$ has 30 objects and 15 properties and it contains 3 formal concepts of the same size which are pair-wise disjoints. In other terms, the formal concepts are $(\{o_1, \ldots, o_{10}\}, \{p_1, \ldots, p_5\})$, $(\{o_{11}, \ldots, o_{20}\}, \{p_6, \ldots, p_{10}\})$, and $(\{o_{21}, \ldots, o_{30}\}, \{p_{11}, \ldots, p_{15}\})$. We generated 40 different data sets by adding increasing quantities of noise (from 1% to 40% of the matrix). The idea is that a robust technique should be able to capture the three associations despite the



a)



b)

Fig. 1. Size of collections of bi-sets (a) and related values of σ (b) w.r.t. noise level

introduced noise. Therefore, for each data set, we have extracted a collection of formal concepts and different collections of δ-bi-sets with increasing values of δ (from 1 to 6). Then, we looked for the occurrence of the three concepts in each of these extracted collections by using our σ measure. Results are collected in Fig. 1b.

The σ measure obviously decreases when the noise level increases. Interestingly, its values for δ-bi-set collections are always greater or similar to the values

for the collection of formal concepts. In particular, for $\delta = 1$, the collections of 1-bi-sets behave better than the collection of formal concepts until noise level is greater than 5%. When $\delta = 2$, this noise threshold is 10%. Finally, for higher values of δ (3,4 and 5), the noise threshold for which δ-bi-sets perform better is quite high (30%). Then, we computed the number of extracted patterns in each collection (Fig. 1a). The collections of δ-bi-sets contain always less patterns than the collection of formal concepts (for a noise level greater than 7%). For $\delta = 2$, the size is halved. For greater values of δ, noise does not influence the size of the collections of δ-bi-sets.

This experiment confirms that δ-bi-sets are more robust to noise than formal concepts. Furthermore, we can reduce significantly the size of the extracted collections and this is crucial to support the interpretation process by data owners.

We applied the same experimental methodology to the voting-records data set from UCI Machine Learning repository. We generated the reference boolean matrix by encoding each variable-modality pair (except the class variable) into a single Boolean attribute. We obtained a matrix with 435 objects and 48 properties. Then, we generated 40 data sets by adding increasing quantities of uniform random noise (from 1% to 40% of the matrix). δ-bi-sets have been extracted with three values of δ (5, 7 and 10), and with a minimal frequency constraint of 7% (i.e., the minimal δ-free-set support size is greater than 30 objects).

The collections of formal concepts have been extracted from the noised matrices with a minimal objectset size constraints set to 30 by using DMiner [9]. The reference collection is the set of all formal concepts with at least 30 objects and 10 attributes in the original matrix. This is motivated by our goal which is to look for rather large associations because too small formal concepts are not relevant in noisy data sets. Using these constraints, we obtained a collection of 4114 formal concepts.

We have computed the values of σ for the different collections of bi-sets and we obtained the results collected in Fig. 3a. The advantages of using δ-bi-sets are visible as soon as the noise level reaches about 5%. It is even more obvious when looking at another reference collection of formal concepts with a minimal itemset size constraint set to 13. In this case, we obtained 24 rather large formal concepts and the benefit of δ-bi-sets starts with a noise level of about 2%. Moreover, if we look at the number of extracted bi-sets (see Fig. 2), we see that the collection of formal concepts is huge w.r.t. any δ-bi-set collection. Notice also that starting from a noise level of 20%, all the sizes are almost the same. Again, mining dense bi-sets as δ-bi-sets enable to get a significantly smaller collection of more relevant patterns.

Until now, we added some noise to a priori noise-free data sets. We tried to identify a subset of the formal concepts which holds in these reference data sets within various collections of bi-set extracted from the matrices after noise introduction. Let us now consider a comparison between formal concepts and δ-bi-sets in three "real world" intrinsically noisy data sets. Two of them (drodophila [18] and malaria [19]) are gene expression data sets. The last one (meningitis) is a medical data set. For the gene expression data sets, the techniques used for

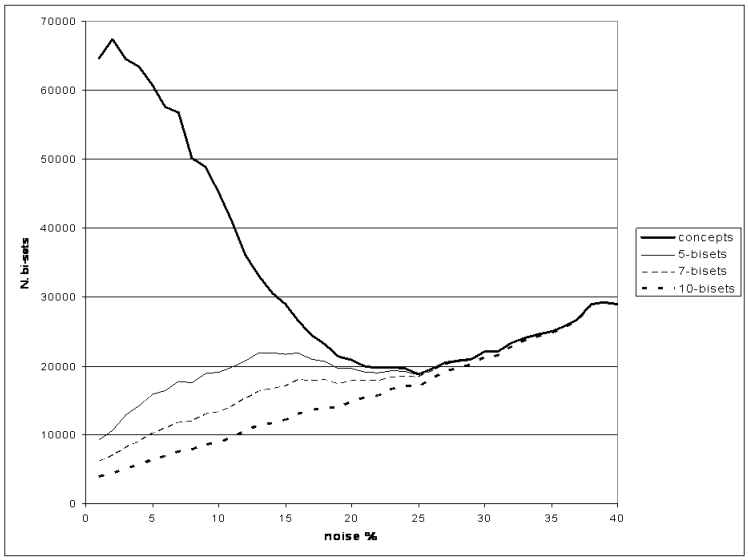


Fig. 2. Size of collections of bi-sets w.r.t. noise level in voting-record

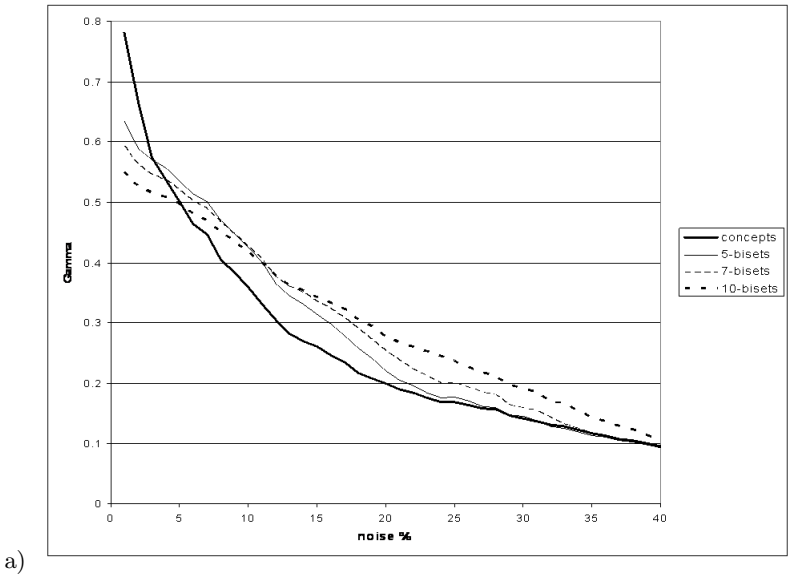
Table 2. Sizes of different bi-set collections for three "real world" data sets

dataset	lines	columns	concepts	1-bisets	2-bisets	3-bisets
drosophila	81	3030	2,288,850	1,801,369	778,526	443,668
malaria	46	3719	3,768,135	844,245	377,739	215,821
meningitis	329	60	689,943	329,834	132,703	69,494

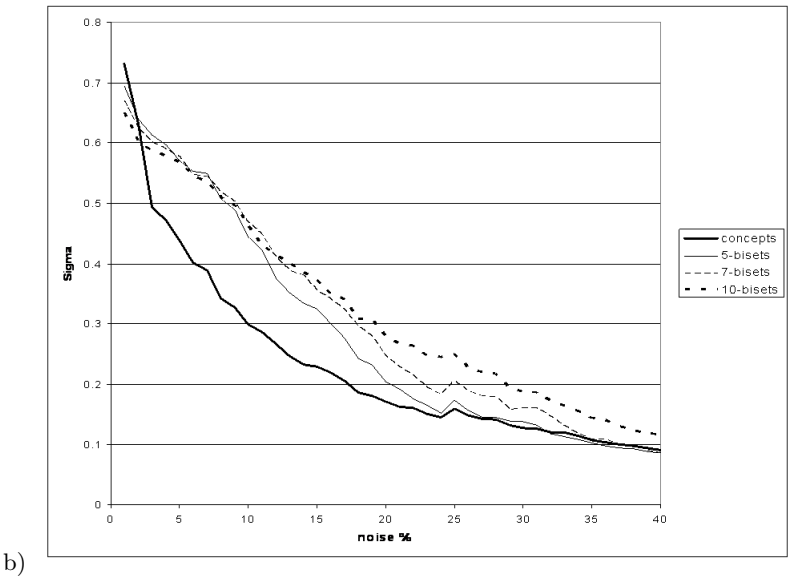
measuring the expression level of genes are unlikely to be fault-free. Moreover, encoding Boolean expression properties from continuous values introduces noise in the data as well. We used the encoding technique described in [10]. For the medical data set, both the discretization of continuous measures and the missing values lead to noisy data. This data set has been preprocessed and provided by Bruno Crémilleux from the University of Caen (France).

For each of these data sets, we extracted a collection of formal concepts and three collections of δ-bi-sets, with rather low values of δ (from 1 to 3), and we compared the number of extracted patterns. Results are collected in Table 2.

For the drosophila and meningitis data sets, the number of pattern is approximatively halved at each incrementation of the δ value. A quite interesting result is the important reduction of the pattern collection size when shifting from formal concept to 1-bi-set mining within malaria (see Table 2). Then, we tried to identify in this malaria data set a group of 135 genes which are known to take part in the same biological function (i.e., cytoplasmic translation machinery) in association with the group of 17 samples in which these genes are known to be active (see [19] for details). We used our σ measure, that, in this case, is



a)



b)

Fig. 3. Values of σ w.r.t. noise level for two reference collections in voting-record

equal to the normalized intersection of the previously described bi-set with the extracted bi-set which better matches it. We found that, for the collection of formal concepts, the value of σ is 0.142, while, for the 1-bi-set collection, its value is 0.146. This enforces our conviction that mining fault-tolerant formal concepts in intrinsically noisy data is a relevant method to reduce the workload for the interpretation by data owners and to provide more relevant patterns.

4 Conclusion

Looking for strong associations between sets of objects and sets of properties in eventually large and noisy Boolean data sets, we have discussed a fundamental limitation of formal concept mining. Computing fault-tolerant formal concepts, e.g., with a bounded number of exceptions on both dimensions is known to be computationally very hard. We proposed a solution based on an efficient technique for mining association rules with few exceptions. The δ-bi-set mining task has been experimentally evaluated on both noised data sets and real data sets. The results are quite promising because we get smaller collections of more relevant patterns. This is crucial for the needed post-processing phases like, e.g., the tedious process of bi-set interpretation by molecular biologists when they are looking for putative transcription modules within Boolean gene expression data sets. The relationship to other fault-tolerant formal concepts must be studied. If δ-bi-sets can indeed be extracted efficiently, it would be much more relevant to ensure a bounded number of exceptions on both lines and columns like in [11]. It means that other classes of fault-tolerant formal concepts might be more relevant than δ-bi-sets but also probably much harder to extract. Another related work in artificial intelligence concern the fuzzy concept analysis framework (see, e.g., [20]). It is an attempt to manage uncertainty and it is clearly related to noisy data analysis. Further investigation is needed in this direction. An interesting perspective on which we are currently working is to use collections of fault-tolerant formal concepts for building relevant bi-partitions from noisy data. The challenge here is to enable a user-driven control for bi-cluster overlapping and to look at the opportunities for constraint-based mining of such models.

Acknowledgements. The authors wish to thank Bruno Crémilleux who provided the data on child's meningitis. We also thanks Christophe Rigotti, Jérémy Besson and Céline Robardet for exciting discussions. This research is partially funded by ACI MD 46 (CNRS STIC 2004-2007) Bingo (Bases de Données Inductives pour la Génomique).

References

1. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In Rival, I., ed.: Ordered sets. Reidel (1982) 445–470
2. Fisher, D.H.: Knowledge acquisition via incremental conceptual clustering. Machine Learning **2** (1987) 139–172
3. Cheng, Y., Church, G.M.: Biclustering of expression data. In: Proceedings ISMB 2000, San Diego, USA, AAAI Press (2000) 93–103
4. Robardet, C., Feschet, F.: Efficient local search in conceptual clustering. In: Proceedings DS'01. Number 2226 in LNCS, Springer-Verlag (2001) 323–335
5. Dhillon, I., Mallela, S., Modha, D.: Information-theoretic co-clustering. In: Proceedings ACM SIGKDD 2003, Washington, USA, ACM Press (2003) 89–98
6. Madeira, S.C., Oliveira, A.L.: Biclustering algorithms for biological data analysis: A survey. IEEE/ACM Trans. Comput. Biol. Bioinf. **1** (2004) 24–45

7. Kuznetsov, S.O., Obiedkov, S.A.: Comparing performance of algorithms for generating concept lattices. Journal of Experimental and Theoretical Artificial Intelligence **14** (2002) 189–216
8. Stumme, G., Taouil, R., Bastide, Y., Pasqier, N., Lakhal, L.: Computing iceberg concept lattices with TITANIC. Data & Knowledge Engineering **42** (2002) 189–222
9. Besson, J., Robardet, C., Boulicaut, J.F.: Constraint-based mining of formal concepts in transactional data. In: Proceedings PaKDD'04. Volume 3056 of LNAI., Sydney (Australia), Springer-Verlag (2004) 615–624
10. Pensa, R.G., Leschi, C., Besson, J., Boulicaut, J.F.: Assessment of discretization techniques for relevant pattern discovery from gene expression data. In: Proceedings ACM BIOKDD'04, Seattle, USA (2004) 24–30
11. Besson, J., Robardet, C., Boulicaut, J.F.: Mining formal concepts with a bounded number of exceptions from transactional data. In: Proceedings KDID'04. Volume 3377 of LNCS., Springer-Verlag (2004) 33–45
12. Gionis, A., Mannila, H., Seppänen, J.K.: Geometric and combinatorial tiles in 0-1 data. In: Proceedings PKDD'04. Volume 3202 of LNAI., Pisa, Italy, Springer-Verlag (2004) 173–184
13. Geerts, F., Goethals, B., Mielikäinen, T.: Tiling databases. In: Proceedings DS'04, Padova, Italy, Springer (2004) 278–289
14. Boulicaut, J.F., Bykowski, A., Rigotti, C.: Free-sets: a condensed representation of boolean data for the approximation of frequency queries. Data Mining and Knowledge Discovery **7** (2003) 5–22
15. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings ACM SIGMOD'93, Washington, D.C., USA, ACM Press (1993) 207–216
16. Mannila, H., Toivonen, H.: Levelwise search and borders of theories in knowledge discovery. Data Mining and Knowledge Discovery **1** (1997) 241–258
17. Blake, C., Merz, C.: UCI repository of machine learning databases (1998)
18. Arbeitman, M., Furlong, E., Imam, F., Johnson, E., Null, B., Baker, B., Krasnow, M., Scott, M., Davis, R., White, K.: Gene expression during the life cycle of drosophila melanogaster. Science **297** (2002) 2270–2275
19. Bozdech, Z., Llinás, M., Pulliam, B.L., Wong, E., Zhu, J., DeRisi, J.: The transcriptome of the intraerythrocytic developmental cycle of plasmodium falciparum. PLoS Biology **1** (2003) 1–16
20. Huynh, V.N., Nakamori, Y., Ho, T.B., Resconi, G.: A context model for fuzzy concept analysis based upon modal logic. Inf. Sci. **160** (2004) 111–129