Intelligent Data Analysis 10(5):457-472, 2006. IOS Press.

Supporting bi-cluster interpretation in 0/1 data by means of local patterns

Ruggero G. Pensa, Céline Robardet, Jean-François Boulicaut
INSA Lyon
LIRIS CNRS UMR 5205
F-69621 Villeurbanne cedex, France
{ruggero.pensa,celine.robardet,jfboulicaut}@insa-lyon.fr

June 26, 2006

Abstract

Clustering or co-clustering techniques have been proved useful in many application domains. A weakness of these techniques remains the poor support for grouping characterization. As a result, interpreting clustering results and discovering knowledge from them can be quite hard. We consider potentially large Boolean data sets which record properties of objects and we assume the availability of a bi-partition which has to be characterized by means of a symbolic description. Our generic approach exploits collections of local patterns which satisfy some user-defined constraints in the data, and a measure of the accuracy of a given local pattern as a bi-cluster characterization pattern. We consider local patterns which are bi-sets, i.e., sets of objects associated to sets of properties. Two concrete examples are formal concepts (i.e., associated closed sets) and the so-called δ-bi-sets (i.e., an extension of formal concepts towards fault-tolerance). We introduce the idea of characterizing query which can be used by experts to support knowledge discovery from bi-partitions thanks to available local patterns. The added-value is illustrated on benchmark data and three real data sets: a medical data set and two gene expression data sets.

1 Introduction

Exploratory data analysis processes are often based on clustering techniques to get insights about global patterns within the data. Clustering has been studied extensively, including for the special case of Boolean data which record properties of objects (see a toy example in Tab. 1). Its main goal is to identify a partition of objects and/or properties such that an objective function which specifies its quality is optimized (e.g., maximizing intra-cluster similarity and inter-cluster dissimilarity) [16]. Looking for optimal solutions is intractable but

heuristic local search optimizations can be performed. As a result, many efficient algorithms which compute good partitions are available and widely used. In this paper, we assume that clustering results are available and we are interested in knowledge discovery from such results. For example, in our running example $\mathbf{r}$, we could get $\{\{\{o_1, o_3, o_4\}, \{o_2, o_5, o_6, o_7\}\}$ as a partition on objects.

Table 1: A Boolean context $\mathbf{r}$

	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

Our thesis is that expert users need symbolic descriptions to characterize the computed groups. Indeed, it is well-known that using various settings for a given clustering algorithm and/or using different algorithms can provide quite different clustering results. The interpretation phase is then tedious. In fact, many clustering approaches suffer from the lack of an explicit cluster characterization. It has motivated the research on conceptual clustering [12]. Among others, it has been studied in the context of co-clustering (see [19] for a survey), including for the special case of categorical or Boolean data. The goal is to identify bi-clusters or bi-partitions in the data, i.e., a mapping between a partition of objects and a partition of properties. For instance, an algorithm like Cocluster [11] can compute in $\mathbf{r}$ the interesting bi-partition $\{\{\{o_1, o_3, o_4\}, \{p_1, p_3, p_4\}\}, \{\{o_2, o_5, o_6, o_7\}, \{p_2, p_5\}\}\}$. The first bi-cluster indicates that the characterization for objects from $\{o_1, o_3, o_4\}$ is that they almost always share properties $\{p_1, p_3, p_4\}$. Also, properties $\{p_2, p_5\}$ are characteristics for objects $\{o_2, o_5, o_6, o_7\}$.

Our experience is that this first step towards characterization is not sufficient, especially in high dimensional data sets in which global patterns like bi-partitions do not reflect unexpected but strong local associations between some sets of objects and some sets of properties. Our proposal is to combine bi-clustering with a characterization phase based on collections of local patterns. We assume that a bi-partition on a Boolean data set is available (e.g., computed using Cocluster [11]). Our contribution to bi-partition characterization is as follows. First, we introduce an original and generic cluster characterization technique which is based on constraint-based bi-set mining, i.e., mining bi-sets whose set components satisfy some constraints. We show how to measure that a given bi-set is an accurate characterization pattern for a given bi-cluster. Thanks to such accuracy measures, it is possible to consider characterizing queries which can support knowledge discovery from co-clustering results. The method is illus-

trated on two kinds of bi-sets, the well-known formal concepts (i.e., associated closed sets [25]) and a new class, the so-called δ-bi-sets. This later pattern type is new and it is based on a previous work about approximate condensed representations for frequent patterns [7]. Intuitively, a formal concept is a maximal rectangle of true values modulo arbitrary permutations of rows and columns. Following that perspective, a δ-bi-set is a fault-tolerant extension of a formal concept for which a bounded number of exceptions (i.e., 0 values) is accepted per column.

The added-value of our characterizing method is illustrated not only on a benchmark data set but also on three real-life data sets. The obtained characterizations are consistent with the available knowledge. This paper extends the preliminary version [21] by further developments on the motivation and the possible applications of the method, the study of some δ-bi-set properties, and further experiments. Indeed, we added the application of our approach to a gene expression data analysis task for which Cocluster provides unstable bi-partitions.

Section 2 formally defines the characterizing framework. Section 3 discusses which type of local pattern can be used. Section 4 is dedicated to our empirical validation of the proposed method. Finally, Section 5 concludes.

2 Bi-cluster characterization using bi-sets

Let us consider a set of objects $\mathcal{O} = \{o_1, \ldots, o_m\}$ and a set of Boolean properties $\mathcal{P} = \{p_1, \ldots, p_n\}$. The Boolean context to be mined is $\mathbf{r} \subseteq \mathcal{O} \times \mathcal{P}$, where $r_{ij} = 1$ if the property p_j is true for object o_i. We assume that a co-clustering algorithm, e.g., [11], provides a bijective mapping between K clusters of objects and K clusters of properties forming K bi-clusters $\{(C_1^o, C_1^p) \ldots (C_K^o, C_K^p)\}$ with $C_k^o \subset \mathcal{O}$ and $C_k^p \subset \mathcal{P}$. A first characterization comes from this mapping.

Our goal is to support each bi-cluster interpretation by collections of bi-sets which are locally pointing out interesting associations between groups of objects and groups of properties. Formally, a bi-set is an element of $2^{\mathcal{O}} \times 2^{\mathcal{P}}$. Therefore, we assume that a collection of N bi-sets $\mathcal{B} = b_1, \ldots, b_N$ has been extracted from the data. First, we associate each of them to one of the K bi-clusters. Each bi-set characterizes the bi-cluster to which it is associated with some degree of accuracy. We can now define a similarity measure between a bi-set (T, G) $(T \subseteq \mathcal{O}, G \subseteq \mathcal{P})$ and a bi-cluster (C_k^o, C_k^p) as follows:

$$sim\left((T, G), (C_k^o, C_k^p)\right) = \frac{|T \cap C_k^o| \cdot |G \cap C_k^p|}{|T \cup C_k^o| \cdot |G \cup C_k^p|}$$

Intuitively, (T, G) and (C_k^o, C_k^p) denote rectangles in the matrix (modulo permutations over the lines and the columns) and we measure the area of the intersection of the two rectangles normalized by the area of their union.

Each bi-set b which is a candidate characterization pattern can now be assigned to the bi-cluster (C_k^o, C_k^p) for which $sim(b, (C_k^o, C_k^p))$ is maximal. Doing so, we get K groups of potentially characterizing bi-sets.

Example 1 *In* **r** *from Tab. 1, a possible bi-partition is*

$$\{(C_1^o, C_1^p), (C_2^o, C_2^p)\} \quad = \quad \{(\{o_1, o_3, o_4\}, \{p_1, p_3, p_4\}), (\{o_2, o_5, o_6, o_7\}, \{p_2, p_5\})\}$$

If we consider the bi-set $b_1 = (\{o_1, o_3, o_5\}, \{p1\})$*, its similarity measures w.r.t.* (C_1^o, C_1^p) *and* (C_2^o, C_2^p) *are:*

$$sim(b_1, (C_1^o, C_1^p)) = \frac{2 \cdot 1}{3 \cdot 1 + 3 \cdot 3 - 2 \cdot 1} = 0.2$$

$$sim(b_1, (C_2^o, C_2^p)) = \frac{1 \cdot 0}{3 \cdot 1 + 4 \cdot 2 - 1 \cdot 0} = 0$$

The bi-set b_1 *is then associated to the first bi-cluster. If we consider now the bi-set* $b_2 = (\{o_5\}, \{p_1, p_2, p_5\})$*, we get:*

$$sim(b_2, (C_1^o, C_1^p)) = \frac{0 \cdot 1}{1 \cdot 3 + 3 \cdot 3 - 0 \cdot 1} = 0$$

$$sim(b_2, (C_2^o, C_2^p)) = \frac{1 \cdot 2}{1 \cdot 3 + 4 \cdot 2 - 1 \cdot 2} = 0.22$$

This bi-set b_2 *is thus associated to the second bi-cluster.*

Finally, we can use an accuracy measure to select the most relevant bi-sets. For that purpose, we propose to measure the exception ratios for the two set components of the bi-sets. Given a bi-set (T, G) and a bi-cluster (C_k^o, C_k^p), it can be computed as follows:

$$\begin{aligned} \epsilon_o(T, C_k^o) &= \frac{|\{o_i \in T |\ o_i \notin C_k^o\}|}{|T|} \\ \epsilon_p(G, C_k^p) &= \frac{|\{p_i \in G |\ p_i \notin C_k^p\}|}{|G|} \end{aligned}$$

Example 2 *In our toy example from Tab. 1, the bi-set* $b_1 = (\{o_1, o_3, o_5\}, \{p1\})$ *contains the object* o_5 *which does not belong to* C_1^o*: we have the exception ratio* $\epsilon_o(\{o_1, o_3, o_5\}, C_1^o) = \frac{1}{3} = 0.33$*. The bi-set* b_2 *contains property* p_1 *which does not belong to* C_2^p*: we have* $\epsilon_p(\{p_1, p_2, p_5\}, C_2^p) = \frac{1}{3} = 0.33$*.*

It is then possible to consider thresholds to select only the bi-sets that have small exception ratios, i.e., $\epsilon_o < \varepsilon_o$ and $\epsilon_p < \varepsilon_p$ where $\varepsilon_o, \varepsilon_p \in [0, 1]$. There are several possible interpretations for these measures. If we are interested in characterizing a cluster of objects (resp. properties), we can look for all the sets of properties (resp. objects) for which the ϵ_o (resp. ϵ_p) values of the related bi-sets are less than a threshold ε_o (resp. ε_p). Alternatively, we can consider the whole bi-cluster and characterize it with all the bi-sets for which the two exception ratios ϵ_o and ϵ_p are less than two thresholds ε_o and ε_p.

3 Choosing a bi-set type for characterization

We now discuss the type of bi-sets which will be post-processed for bi-cluster characterization. It is clear that bi-clusters are, by construction, interesting characterizing bi-sets but they only support a global interpretation. We are interested in strong associations between sets of objects and sets of properties that can locally explain the global behavior. Clearly, formal concepts are candidates.

3.1 Using formal concepts

Definition 1 (formal concept [25]) *If $T \subseteq \mathcal{O}$ and $G \subseteq \mathcal{P}$, assume $\phi(T, \mathbf{r}) = \{p \in \mathcal{P} \mid \forall o \in T,\ (o,p) \in \mathbf{r}\}$ and $\psi(G, \mathbf{r}) = \{o \in \mathcal{O} \mid \forall p \in G,\ (o,p) \in \mathbf{r}\}$. A bi-set (T, G) is a formal concept in $\mathbf{r}$ when $T = \psi(G, \mathbf{r})$ and $G = \phi(T, \mathbf{r})$. By construction, G and T are closed sets, i.e., $G = \phi \circ \psi(G, \mathbf{r})$ and $T = \psi \circ \phi(T, \mathbf{r})$.*

Formal concepts are maximal association of sets of objects and sets of properties: if one adds a property (resp. an object) one might remove at least an object (resp. a property) to get only true values in the encoded Boolean relation.

Example 3 *Eight formal concepts hold in $\mathbf{r}$ from Tab. 1. $(\{o_1, o_3\}, \{p_1, p_3, p_4\})$, $(\{o_1, o_3, o_4\}, \{p_3, p_4\})$, and $(\{o_5, o_6\}, \{p_2, p_5\})$ are among them.*

Efficient algorithms have been developed to extract complete collections of formal concepts which satisfy also user-defined constraints (e.g., minimal size constraint on set components) [24, 3]. Indeed, the popular frequent closed set mining task for a frequency threshold ν fundamentally computes each formal concept (T, G) such that $|T| \geq \nu$.

A major problem with formal concepts is that the Galois connection (ϕ, ψ) is, in some sense, a too strong one: we have to capture every maximal set of objects and its maximal set of associated properties. As a result, the number of formal concepts even in small matrices can be huge. It is indeed common to get several millions of formal concepts even from rather small matrices. A solution is to look for "dense" rectangles in the matrix, i.e., bi-sets with mainly true values but also a bounded (and small) number of false values or exceptions. Some approaches for dense bi-set mining have been recently discussed (see, e.g., [4] for a starting point). We now propose a new type of bi-set which can be efficiently computed and which is an extension of formal concepts towards fault-tolerance.

3.2 Mining δ-bi-sets

We want to compute efficiently smaller collections of bi-sets which still capture strong associations. We recall some definitions about the association rule mining task [1] since it is used for both the definition of the δ-bi-set pattern type and for bi-cluster characterization.

Definition 2 (association rule) *An association rule in* $\mathbf{r}$ *is an expression of the form* $X \Rightarrow Y$*, where* $X, Y \subseteq \mathcal{P}$*,* $Y \neq \emptyset$ *and* $X \cap Y = \emptyset$*. Its absolute frequency is* $|\psi(X \cup Y, \mathbf{r})|$ *and its confidence 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, $X \cup Y$ associated to $\psi(X, \mathbf{r})$ is then a dense bi-set: it contains few false values. We now consider our technique for computing association rules with high confidence, the so-called δ-strong rules [6, 7].

Definition 3 (δ-strong rule) *Given an integer* δ*, a* δ*-strong rule in* $\mathbf{r}$ *is an association rule* $X \Rightarrow Y$ $(X, Y \subset \mathcal{P})$ *s.t.* $|\psi(X, \mathbf{r})| - |\psi(X \cup Y, \mathbf{r})| \leq \delta$*, 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 [6, 7, 10] and their δ-closures.

Definition 4 (δ-free set, δ-closure) *Let* δ *be an integer and* $X \subset \mathcal{P}$*,* X *is a* δ*-free-set in* $\mathbf{r}$ *iff there is no* δ*-strong rule which holds between two of its own proper subsets. The* δ*-closure of* X *in* $\mathbf{r}$*,* $h_\delta(X, \mathbf{r})$*, is the maximal superset* Y *of* X *s.t.* $\forall p \in Y \setminus X$*,* $|\psi(X \cup \{p\})| \geq |\psi(X, \mathbf{r})| - \delta$*.*

In other terms, the frequency of the δ-closure of X in $\mathbf{r}$ is almost the same than the frequency of X when $\delta << |\mathcal{O}|$ and X is frequent. Moreover, $\forall p \in h_\delta(X) \setminus X$, $X \Rightarrow p$ is a δ-strong rule.

Example 4 *In the data from Tab. 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* $\{p_3, p_4\}$*. 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 that $h_0 \equiv \phi \circ \psi$, i.e., the classical closure operator. Looking for a 0-free-set, say X, and its 0-closure, say Y, provides the closed set $X \cup Y$ and thus the formal concept $(\psi(X \cup Y, \mathbf{r}), X \cup Y)$.

Definition 5 (δ-bi-set) *A* δ*-*γ*-bi-set* (T, G) *in* $\mathbf{r}$ *is built on each* δ*-free-set* $X \subset \mathcal{P}$ *with* $T = \psi(X, \mathbf{r})$ *and* $G = h_\gamma(X, \mathbf{r})$*. When* $\delta = \gamma$ *we call them* δ*-bi-sets.*

Example 5 *In the data from Tab. 1, the* 1*-bi-sets derived from the* 1*-free-sets* $\{p_3\}$ *and* $\{p_5\}$ *are* $(\{o_1, o_3, o_4\}, \{p_1, p_3, p_4\})$ *and* $(\{o_2, o_5, o_6, o_7\}, \{p_2, p_5\})$*.*

When $\delta << |T|$, δ-bi-sets are dense bi-sets with a small number of exceptions per column. In order to experiment, we implemented a straightforward extension of AC-Miner [7] which provides the supporting set for each extracted δ-free-set. Let us now discuss some properties of δ-bi-sets. It is clear that 0-bi-sets are formal concepts. However, some important properties of formal concepts do not hold for δ-bi-sets when $\delta > 0$. In particular we lack of a function which

associates the set G to the set T and vice-versa. As a result, we do not have a Galois connection anymore. For example, in Tab. 1, $(\{o_2, o_5, o_6, o_7\}, \{p_2, p_5\})$ (the δ-bi-set generated by the δ-free set $\{p_5\}$), and $(\{o_2, o_5, o_6\}, \{p_2, p_5\})$ (the δ-bi-set generated by the δ-fee set $\{p_2\}$) have the same property set, while the first set of objects includes the second one. Among others, it makes the interpretation process (in terms of characterization) less natural. We now consider how the parameters δ and γ influence the properties of the δ-γ-bi-set collection.

Property 1 *Given a Boolean context* $\mathbf{r}$*, two positive integers* μ *and* δ *such that* $\mu < \delta$*. Let us denote* $Free_\delta(\mathbf{r})$ *the collection of the* δ*-free sets on* $\mathbf{r}$*, and* $Free_\mu(\mathbf{r})$ *the collection of the* μ*-free sets on* $\mathbf{r}$*, we have:* $Free_\delta(\gamma, \mathbf{r}) \subseteq Free_\mu(\gamma, \mathbf{r})$*.*

Proof 1 *X is a δ-free-set iff $\forall Y \subset X \; |\psi(Y, \mathbf{r})| - |\psi(X, \mathbf{r})| > \delta$. Thus $|\psi(Y, \mathbf{r})| - |\psi(X, \mathbf{r})| > \mu$ and X is also a μ-free-set.*

As a consequence, any collection of δ-free sets ($\delta > 0$) is included in the collection of 0-free sets.

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

Table 2: Boolean context $\mathbf{r\prime}$

Example 6 *Let us consider the Boolean data set given in Tab. 2. The set of properties $A = \{p_1, p_3\}$ is 0-free (see Tab. 3), but not 1-free with (see Tab. 4). Its 1-closure is $\{p_1, p_2, p_3\}$. The corresponding δ-bi-set is $b_A = (\{o_4, o_5, o_6\}, \{p_1, p_2, p_3\})$ with one exception on p_2. Since A is not in the collection of 1-free sets, this bi-set can not be built using A, and we have neither another 1-free set which can generate b_A nor any other bi-set covering b_A (see Tab. 4).*

Property 2 *Given a Boolean context* $\mathbf{r}$ *and two positive integers* ρ *and* γ *such that* $\rho \leq \gamma$*. Given a set* $X \subseteq \mathcal{P}$ *we have:* $h_\rho(X) \subseteq h_\gamma(X)$*.*

When the parameter γ increases, then the size of the attribute component of the γ-δ-bi-set increases too.

Property 3 *Given a δ-free set X, $\forall Y \subset X$, then $X \not\subseteq h_\delta(Y, \mathbf{r})$, i.e., X is not included in the δ-closure of any of its own proper subsets.*

X	$h_1(X, \mathbf{r}\prime)$	$\psi(X, \mathbf{r}\prime)$
$\{\emptyset\}$	$\{p_3\}$	$\{o_1, o_2, o_3, o_4, o_5, o_6, o_7\}$
$\{p_1\}$	$\{p_1, p_3\}$	$\{o_3, o_4, o_5, o_6\}$
$\{p_2\}$	$\{p_2, p_3\}$	$\{o_1, o_2, o_4, o_5\}$
$\{p_3\}$	$\{p_3\}$	$\{o_1, o_2, o_4, o_5, o_6, o_7\}$
$\{p_4\}$	$\{p_1, p_2, p_3, p_4\}$	$\{o_1, o_5\}$
$\{p_1, p_2\}$	$\{p_1, p_2, p_3, p_4\}$	$\{o_4, o_5\}$
$\{p_1, p_3\}$	$\{p_1, p_2, p_3\}$	$\{o_4, o_5, o_6\}$
$\{p_1, p_4\}$	$\{p_1, p_2, p_3, p_4\}$	$\{o_5\}$

Table 3: 0-free sets, 1-closures and supporting sets of objects in $\mathbf{r}\prime$

Proof 2 *If $Y \subset X$, and $X \subseteq h_\delta(Y, \mathbf{r})$, then there exists $Z \subset X$, $Z \cap Y = \emptyset$, s.t. $Y \Rightarrow Z$ is a δ-strong rule, i.e., there exists a δ-strong rule which holds between two own proper subsets of X (Y and Z), but this contradicts that X is a δ-free set.*

As a consequence, when $\gamma > \delta$, a set X can belong to the γ-closure of one of its subsets which is not the case when $\gamma \leq \delta$.

X	$h_1(X, \mathbf{r}\prime)$	$\psi(X, \mathbf{r}\prime)$
$\{\emptyset\}$	$\{p_3\}$	$\{o_1, o_2, o_3, o_4, o_5, o_6, o_7\}$
$\{p_1\}$	$\{p_1, p_3\}$	$\{o_3, o_4, o_5, o_6\}$
$\{p_2\}$	$\{p_2, p_3\}$	$\{o_1, o_2, o_4, o_5\}$
$\{p_4\}$	$\{p_1, p_2, p_3, p_4\}$	$\{o_1, o_5\}$
$\{p_1, p_2\}$	$\{p_1, p_2, p_3, p_4\}$	$\{o_4, o_5\}$

Table 4: 1-free sets, 1-closures and supporting sets of objects in $\mathbf{r}\prime$

Example 7 *In the data from Tab. 1, we have eight 1-free sets: $\{\emptyset\}$, $\{p_1\}$, $\{p_2\}$, $\{p_3\}$, $\{p_4\}$, $\{p_5\}$, $\{p_1, p_2\}$, and $\{p_1, p_5\}$. The collection of 0-free sets contains two more sets $\{p_1, p_3\}$ and $\{p_1, p_4\}$ which are contained in the 1-closures of $\{p_1\}$ (i.e., $\{p_1, p_3, p_4\}$ which is also the 1-closure of $\{p_3\}$ and $\{p_4\}$). The supporting set of objects for both $\{p_1, p_3\}$ and $\{p_1, p_4\}$ is $\{o_1, o_3\}$ and it is a subset of the supporting set of objects for $\{p_1\}$ (i.e., $\{o_1, o_3, o_5\}$), and the supporting set of objects for $\{p_3\}$ and $\{p_4\}$ as well. Indeed, the two 0-δ-bi-sets are already included in larger bi-sets obtained from 1-free sets.*

We have considered several settings for computing δ-bi-sets. As the γ-closure of a 0-free set X is equal to the γ-closure of $h_0(X, \mathbf{r})$, by computing 0-δ-bi-sets we get either formal concepts (0-closure of a 0-free set) or their extension towards

fault-tolerance (number of exceptions bounded per column). Computing formal concepts by extracting the free sets and their closure, may become intractable in some data sets, while δ-free set mining for $\delta > 0$ remains quite feasible at the price of missing some associations. On the other hand, using a value of δ greater than γ may result in a further loss in information, even if the size of collection of produced bi-sets could be reduced. Our answer to the previously given question, is that using the same δ value for computing the free sets and their closure is a good trade-off to preserve information and to reduce both the search space and the size of the extracted bi-set collection.

3.3 Formal concepts vs. δ-bi-sets

To study the relevancy of δ-bi-sets w.r.t. formal concepts, we have considered the addition of noise to a synthetical data set. Hereafter, **r** denotes a reference data set from which we generate noisy data sets by adding a given quantity of uniform random noise. Then, we compare the collection of formal concepts which are "built-in" within **r** with various collections of 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 subsets of the reference collection in each of them. Since both set components of each formal concept can be changed when adding noise, we identify those having the largest area in common with the reference ones, and we compute the σ measure which takes into account the common area:

$$\sigma(\mathcal{C}_r, \mathcal{C}_e) = \frac{\rho(\mathcal{C}_r, \mathcal{C}_e) + \rho(\mathcal{C}_e, \mathcal{C}_r)}{2}$$

ρ is defined as follows:

$$\rho(\mathcal{C}_1, \mathcal{C}_2) = \frac{1}{|\mathcal{C}_1|} \sum_{(X_i,Y_i)\in\mathcal{C}_1} \max_{(X_j,Y_j)\in\mathcal{C}_2} \frac{|X_i \cap X_j| \cdot |Y_i \cap Y_j|}{|X_i \cup X_j| \cdot |Y_i \cup Y_j|}$$

$\mathcal{C}_r$ is the collection of formal concepts computed on the reference dataset, $\mathcal{C}_e$ is a collection of patterns in a noised dataset. When $\sigma(\mathcal{C}_r, \mathcal{C}_e) = 1$, all the bi-sets $\in \mathcal{C}_r$ have identical instances in $\mathcal{C}_e$.

In the experiment, **r** has 30 objects and 15 properties and it contains 3 formal concepts of the same size which are pair-wise disjoints: the built-in 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 to **r** increasing quantities of noise (from 1% to 40% of the matrix). Then, 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). Finally, we looked for the occurrence of the 3 formal concepts in each of these extracted collections by using our σ measure. Results are in Fig. 1.

The σ measure decreases when the noise level increases. Interestingly, its values for δ-bi-set collections are always greater or similar to the values for the collections of formal concepts. The collections of δ-bi-sets contain always less

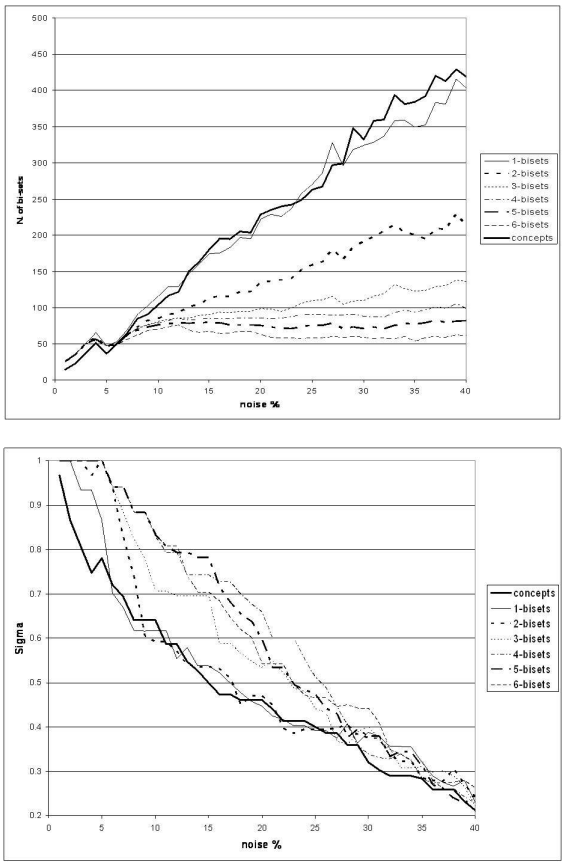

Figure 1: Size of different collections of bi-sets (top) and related values of σ (bottom) depending on noise level

patterns than the collections 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, it enables to reduce significantly the size of the extracted collections and this is important to support the interpretation process.

3.4 Using association rules

Association rules can be derived from extracted bi-sets and used for bi-cluster characterization. For characterization but also classification, heuristics have been studied which select relevant association rules based on their frequency and confidence values [18, 17, 10, 23]. In our case, we propose to use exception

ratios on the extracted bi-sets to provide characterization rules. They have the form $X \Rightarrow k$ where X is a set of properties (resp. objects) and k is a property denoting a cluster of objects (resp. an object denoting a cluster of properties). When considering formal concepts, deriving characterization rules from them is straightforward.

Property 4 *Given a bi-cluster (C_k^o, C_k^p), if (T, G) is a formal concept, then $G \Rightarrow k$ (resp. $T \Rightarrow k$) is a rule with frequency equal to $|T| \cdot (1 - \epsilon_o(T, C_k^o))$ (resp. $|G| \cdot (1 - \epsilon_p(G, C_k^p))$ and confidence equal to $1 - \epsilon_o(T, C_k^o)$ (resp. $1 - \epsilon_p(G, C_k^p)$).*

Table 5: A Boolean context $\mathbf{r}_1$

	p_1	p_2	p_3	p_4	p_5	c_1^o	c_2^o
o_1	1	0	1	1	0	1	0
o_2	0	1	0	0	1	0	1
o_3	1	0	1	1	0	1	0
o_4	0	0	1	1	0	1	0
o_5	1	1	0	0	1	0	1
o_6	0	1	0	0	1	0	1
o_7	0	0	0	0	1	0	1
c_1^p	1	0	1	1	0		
c_2^p	0	1	0	0	1		

Example 8 *Consider the toy example from Tab. 1 with two new columns (resp. rows) to denote the values of the object cluster variable $v_o \in \{c_1^o, c_2^o\}$ (resp. the property cluster variable $v_p \in \{c_1^p, c_2^p\}$). For each object belonging to C_1^o (resp. C_2^o), we have $c_1^o = 1$ and $c_2^o = 0$ (resp. $c_1^o = 0$ and $c_2^o = 1$). We obtain the Boolean data in Tab. 5. Bi-set $b_1 = (T_1, G_1) = (\{o_1, o_3, o_5\}, \{p_1\})$ is a formal concept that can be used to form the association rule $p_1 \Rightarrow c_1^o$. Its relative frequency is $|T_1| \cdot (1 - \epsilon_o(T_1, C_1^o))/|\mathcal{O}| = 3 \cdot (1 - 1/3)/7 = 29\%$, its confidence is $(1 - \epsilon_o(T_1, C_1^o)) = (1 - 1/3) = 67\%$. The formal concept $b_2 = (T_2, G_2) = (\{o_5\}, \{p_1, p_2, p_5\})$ forms the association rule $o_5 \Rightarrow c_2^p$. Its relative frequency is $|G_2| \cdot (1 - \epsilon_p(G_2, C_2^p))/|\mathcal{P}| = 3 \cdot (1 - 1/3)/5 = 40\%$, its confidence is $(1 - \epsilon_p(G_2, C_2^p)) = (1 - 1/3) = 67\%$*

When we use δ-bi-sets instead of formal concepts, Property 4 does not hold because $|\psi(G, \mathbf{r})| < |T|$. However, if we are interested in characterizing a cluster of objects, we can use the following property.

Property 5 *Given a cluster C_k^o, if (T, G) is a δ-bi-set, and $X \subseteq G$ is a δ-free-set then $X \Rightarrow k$ is a rule with frequency equal to $|T| \cdot (1 - \epsilon_o(T, C_k^o))$ and confidence equal to $1 - \epsilon_o(T, C_k^o)$.*

Example 9 *Consider our toy example in Tab. 1, and its extension (Tab. 5). Bi-set $b_\delta = (T_\delta, G_\delta) = (\{o_1, o_3, o_4\}, \{p_1, p_3, p_4\})$ is a 1-bi-set generated by the*

1-free set $X_\delta = \{p_3\}$*. It is associated with* (C_1^o, C_1^p) *since* $sim(b_\delta, (C_1^o, C_1^p)) = 0$*, and* $\epsilon_o(T_\delta, C_1^o) = 0$*. Indeed,* X_δ *forms an association rule* $p_3 \Rightarrow c_1^o$ *with relative frequency* $|T_\delta| \cdot (1-0)/|\mathcal{O}| = 43\%$ *and confidence 100%.*

Such rules are interesting in practice because X is often a rather small set such that its interpretation is easier. However, this approach can not be applied to data sets with large numbers of properties (e.g., for gene expression data sets where we can have thousands of properties). In such cases, we propose to use the ϵ_o and ϵ_p measures. Notice however that a recent work studies δ-free set mining for very large numbers of properties [15].

4 Experimental validation

4.1 Mining a benchmark data set

First, we applied our characterization method to the well-known benchmark voting-records [5]. It contains 435 objects and 48 Boolean attributes (removing class variables). We used COCLUSTER [11] to get 2 bi-clusters:

bi-cluster	$\lvert\boldsymbol{\tau}\rvert$	rep.	dem.	$\lvert\boldsymbol{\gamma}\rvert$
bi-cluster1	193	153	40	16
bi-cluster2	242	15	227	32
total	435	168	267	48

To characterize each bi-cluster, we used D-MINER [3] to extract all formal concepts, and our slight extension of ACMINER to extract two collections δ-bi-sets (δ=1,2). We obtained 227 031 formal concepts, 130 313 1-bi-sets and 66 908 2-bi-sets. The collections have been post-processed by looking for rules with increasing values of the relative minimal frequency (15% up to 40%) and confidence (90% up to 100%). Results for the first bi-cluster are in Fig. 2. Results for the second one look similar. The number of characterizing rules decreases when we increase the frequency and confidence thresholds. When we use δ-bi-sets, we have to process significantly smaller collections. Two examples of characterizing rules which are consistent with the domain knowledge associated to voting-records are now given. The first one (resp. the second) has a 42% relative frequency (resp. 31%) and both have a 100% confidence, i.e., we have $\epsilon_o = 0$.

`el-salvador-aid = yes` $\wedge$ `anti-satellite-test-ban = yes`
$\wedge$ `aid-to-nicaraguan-contras = yes` $\Rightarrow$ `bi-cluster2`

`handicapped-infants = no` $\wedge$ `physician-fee-freeze = yes`
$\wedge$ `el-salvador-aid = yes` $\Rightarrow$ `bi-cluster1`

4.2 Mining a medical data set

We applied the method to the real world medical data set meningitis already used in [23]. It has been gathered from children hospitalized for acute meningitis. The pre-processed Boolean data set is composed of 329 examples described

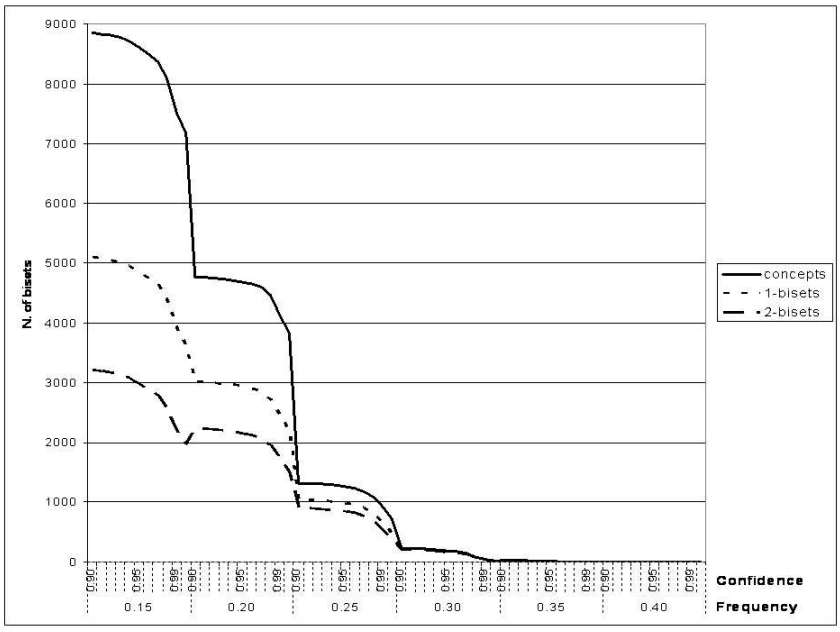

Figure 2: Characterizing patterns for bi-cluster1 in voting-records w.r.t. different values of minimal frequency and confidence.

by 60 Boolean attributes encoding clinical signs (hemodynamic troubles, consciousness troubles, ...), cytochemical analysis of the cerebrospinal fluid (C.S.F proteins, C.S.F glucose, ...), and blood analysis (sedimentation rate, white blood cell count, ...). In meningitis, the majority of the cases are known to be viral infections whereas about one quarter are are known to be caused by bacteria. Furthermore, medical knowledge is available which can be used to assess characterization relevancy. Using COCLUSTER, we got two bi-clusters:

bi-cluster	$\lvert\boldsymbol{\tau}\rvert$	bact.	vir.	$\lvert\boldsymbol{\gamma}\rvert$
bi-cluster1	100	81	19	21
bi-cluster2	229	3	226	39
total	329	84	245	60

The first bi-cluster contains a majority of bacterial cases while the second one contains almost only viral cases. We selected characterization rules based on a collection of formal concepts and 2 collections of δ-bi-sets (δ=1,2). We obtained the results in Fig. 3. Here again, using δ-bi-sets leads to smaller collections of candidate characterization patterns. The number of characterization rules for the first bi-cluster is always very low and it does not significantly change when using δ-bi-sets instead of formal concepts. If we select the rules with a minimal body, a 10% frequency threshold, a 98% confidence threshold, and for which the

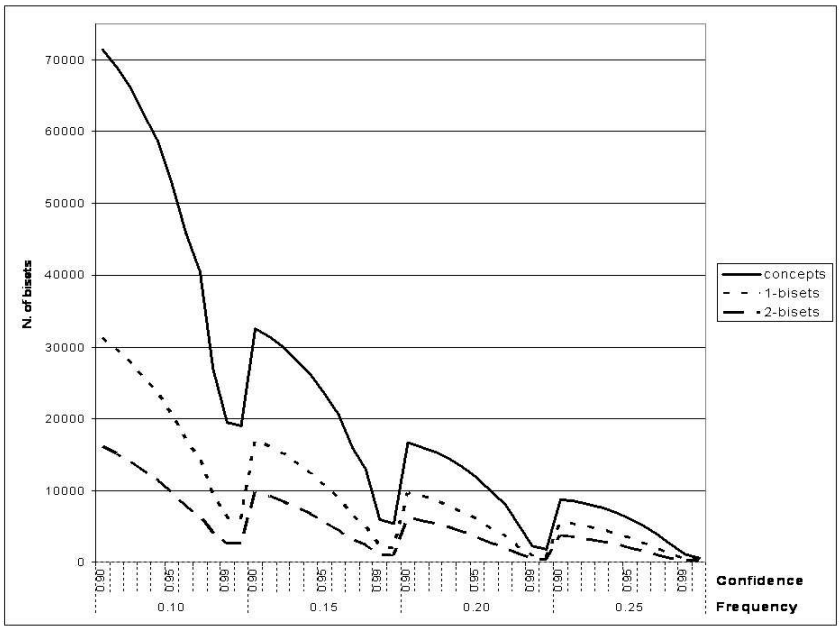

Figure 3: Characterizing patterns for the bi-cluster2 in meningitis w.r.t. different values of minimal frequency and confidence.

property exception ratio ϵ_p is zero, we obtain only 9 rules which are consistent with the medical knowledge (see [23] for details). Examples of rules are:

`presence of bacteria in C.S.F. analysis = yes` $\Rightarrow$ `bi-cluster1`
`polynuclear percent` $> 80 \wedge$ `C.S.F. proteins` $> 0.8 \Rightarrow$ `bi-cluster1`
`C.S.F. proteins` $> 0.8 \wedge$ `C.S.F. glucose` $< 1.5 \Rightarrow$ `bi-cluster1`

4.3 Mining Boolean gene expression data

An other experiment concerns the analysis of plasmodium, a public gene expression data set concerning Plamodium falciparum (i.e., a causative agent of human malaria) described in [9]. It records the expression profile of 3 719 genes in 46 biological samples. Each sample corresponds to a time point of the developmental cycle. It is divided into 3 phases: the ring, the trophozoite and the schizont stages. The numerical expression data have been preprocessed by using one of the property encoding methods described in [22]. We used Cocluster to get the following bi-clusters:

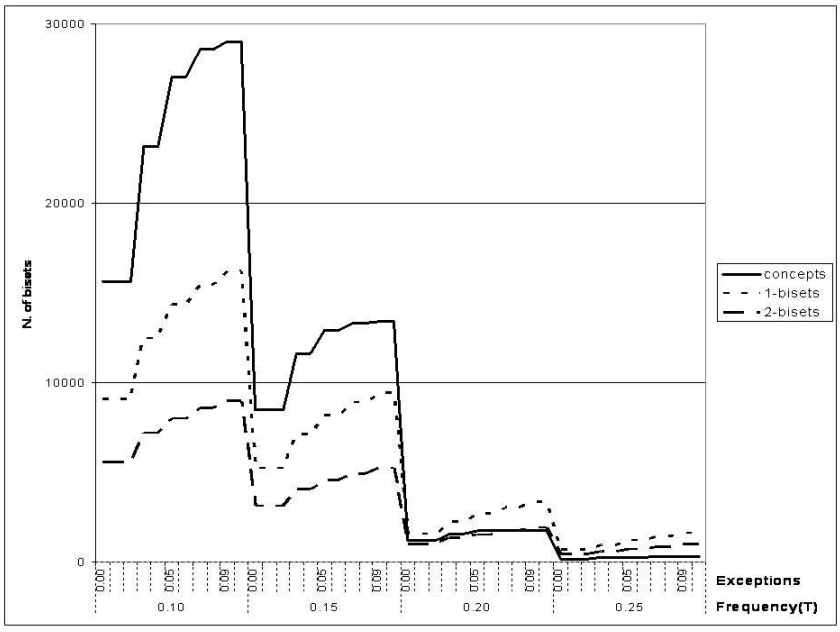

Figure 4: Characterizing bi-sets for bi-cluster1 in plasmodium w.r.t. different values of minimal size and maximal exception ratio.

bi-cluster	$\|\boldsymbol{\tau}\|$	ring	troph	schiz.	$\|\boldsymbol{\gamma}\|$
bi-cluster1	20	15	5	0	558
bi-cluster2	16	0	5	11	1699
bi-cluster3	10	6	0	4	1462
total	46	21	10	15	3719

We extracted collections of bi-sets to characterize clusters of samples by means of sets of genes. Here, the number of properties was too large and we extracted the δ-bi-sets on the transposed matrix. It means that the frequency and the confidence measures can not be used since they are computed on samples while we are looking for patterns on genes. Therefore, to evaluate a bi-set (T, G), we have considered $|T|$, $|G|$, ϵ_o, and ϵ_p. Results for a minimal size from 10% up to 25% of $|\mathcal{O}|$ and for maximal values of ϵ_o from 0% up to 10% are in Fig. 4. Considering bi-cluster1, we analyzed the characterizing 2-bi-sets when the minimal size for their sets of objects was 25% of $|\mathcal{O}|$ and for a maximal exception ratio $\varepsilon_o = 0$. Among the 442 bi-sets characterizing bi-cluster1, only 4 of them concern genes that belong to the same bi-cluster. In each of them, we found at least one gene belonging to the cytoplasmic translation machinery group which is known to be active in the ring stage (see [9] for details), i.e., the main developmental phase corresponding to bi-cluster1.

4.4 Characterization of unstable bi-partitions

In some application, clustering results are quite ambiguous. Algorithms generally return local optimum solutions for the considered objective function. Usually, such local optima are close to the global one, and the computed bi-partitions are quite similar after many randomly initialized executions of the algorithm. However, in some cases, the local optima may give rise to very different bi-partitions. How does our technique behave in this particular situation? How does characterization changes between two different bi-partitions? Does bi-partition quality influence the relevancy of the characterizing patterns? To answer such questions, we have analyzed the data set described in [2]. It concerns the expression profiles of 3 433 genes during 10 time points of adult drosophila melanogaster life cycle. The expression levels are measured for both males and females, i.e., the data involve 20 biological situations. We applied again a discretization method from [22] for gene expression property encoding. We then executed 100 randomly initialized instances of Cocluster (to find 2 bi-clusters), and compared the results by considering both the Goodman-Kruskal's τ coefficient [14] and the loss in mutual information [11]. Notice that this later is the objective function which Cocluster wants to minimize. Both coefficient are evaluated in a contingency table $\mathbf{p}$. Let p_{ij} be the frequency of relations between an object of a cluster C_i^o and a property of a cluster C_j^p, and $p_{i.} = \sum_j p_{ij}$ and $p_{.j} = \sum_i p_{ij}$. The Goodman-Kruskal's τ coefficient, which evaluates the proportional reduction in error given by the knowledge of C^o on the prediction of C^p and vice versa, is defined as follows:

$$\tau = \frac{\frac{1}{2}\sum_i \sum_j \left(p_{ij} - p_{i.}p_{.j}\right)^2 \frac{p_{i.}+p_{.j}}{p_{i.}p_{.j}}}{1 - \frac{1}{2}\sum_i p_{i.}^2 - \frac{1}{2}\sum_j p_{.j}^2}$$

The mutual information, which compute the amount of information C^o contains about C^p, is:

$$I(C^o; C^p) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{p_{i.}p_{.j}}$$

Then, given two different bi-partitions (C^o, C^p) and $(\hat{C}^o, \hat{C}^p)$, the loss in mutual information is given by:

$$I(C^o; C^p) - I(\hat{C}^o; \hat{C}^p)$$

When computing such coefficients on the 100 bi-partitions returned by Cocluster, we found that results were significantly unstable (see Tab. 6). It seems that there are two optimum points for which the two measures are distant. For 56 runs, we got a high τ coefficient (the mean is about 0.5605), for the other 44 ones the τ coefficient was sensibly smaller (about 0.1156). If we consider each group of results separately, the standard deviation is significantly smaller. It means that these two results are two local optima for the Cocluster heuristics.

From a semantical point of view, the first group of solution reflects the male and female repartition of the individuals, while in the second group each cluster contains both male and female individuals. Indeed, it seems that the

		τ		$I - \hat{I}$	
bi-partition	**instances**	**mean**	**std.dev**	**mean**	**std.dev**
males vs. females	56	0.5605	0.0381	1.6615	0.0390
mixed	44	0.1156	0.0166	2.0256	0.0258
overall	100	0.3648	0.2240	1.8217	0.1847

Table 6: Clustering results on adult drosophila individuals.

bi-partition	$\lvert\mathcal{B}_1\rvert/\lvert\mathcal{B}_2\rvert$	$\mathcal{P}$**-freq**	$\mathcal{P}$**-conf**	$\mathcal{O}$**-freq**	$\mathcal{O}$**-conf**
males vs. females	0.87	0.6%	91%	6%	78%
mixed	0.97	0.4%	73%	3%	47%

Table 7: Characterization interestingness measures on adult drosophila data.

the first co-clustering is more relevant w.r.t. the biological knowledge. We use then our characterization technique to post-process the collection of all formal concepts contained in the matrix. Obviously the characterization changes, but we want also to evaluate this change in co-clustering interpretation. To do that, we computed the means of all our interestingness measures (frequency, confidence), one instance for each group of solutions. The two instances have been chosen by considering those with the minimum deviation from the mean. The interestingness measures were computed on all the 5 936 formal concepts, without setting any frequency or confidence constraint. Results are in Tab. 7.

In the first bi-partition, the average frequency and confidence of the characterizing rules are higher than in the second one. This is true for rules computed on both objects and properties clusters. This means that local patterns (formal concepts) reflects more the first bi-partition than the second one. In other words, the consistency of the first global model is validated by the local associations within the matrix. The fact that both Goodman-Kruskal and mutual information loss measures are better in the first group of solutions, is a further mean to link global and local consistency. It means that, characterizing a global bi-partition by means of local patterns makes sense, and could be a new way to assess bi-partition quality.

5 Conclusion

We presented a new (bi-)cluster characterization method based on extracted local patterns, more precisely formal concepts and δ-bi-sets. It is now possible to use quite efficient constraint-based mining techniques for computing various local patterns. While a bi-partition provides a global and somehow expected characterization, selected collections of characterizing bi-sets point out local as-

sociation which might lead to more unexpected but yet relevant information. Global and local patterns are both useful during a knowledge discovery process, and it is important to support these intrinsically interactive processes. Our approach suggest the use of characterizing queries, i.e., queries in which analysts can used the proposed accuracy measures to select relevant characterizing patterns. Examples of typical characterizing queries are as follows:

- Select all the bi-sets which characterize bi-cluster (C^o, C^p) with a maximum exception ratio of ε for both objects and properties;
- Select all the association rules with minimal body characterizing bi-cluster (C^o, C^p) with a minimal frequency f, a minimal confidence c, and a maximal exception ratio ε for the set of properties;
- Select all the association rules with minimal body characterizing bi-cluster (C^o, C^p) with a minimal frequency f, a minimal confidence c, and a minimal exception ratio ε for the set of properties.

The two first types of queries are obviously useful for bi-cluster characterization. The third one concerns knowledge discovery thanks to unexpectedness. Indeed, it might return patterns that are exceptions, i.e., they concern objects belonging to bi-cluster (C^o, C^p) that are characterized by some properties from other bi-clusters. If a global pattern like a bi-partition captures some important structures in the data, it seems also interesting to look at the collections of local associations which are somehow far from it. Assume that the popular association R which points out frequent transactions with beers and diapers among male customers is somehow valid. A co-clustering on a complete basket data set may group beer together with male customers within one bi-cluster, and diapers with female customers in a second bi-cluster. In such a case, a query which would select frequent and high-confidence association rules with a high exception ratio on properties ($\epsilon > \varepsilon$) would support the discovery of the "unexpected" association R. Characterizing queries might be studied further. They are interesting examples of queries which have to process both the data and multiple types of patterns holding in the data, i.e., interesting objects for the study of the promising inductive database framework [20, 8]. An other perspective is also to better understand the convergent techniques developed for (conceptual) clustering, subgroup discovery [13], and association rule discovery.

Acknowledgements. The authors thank P. Francois and B. Crémilleux who provided the data set meningitis. They also thank C. Rigotti and J. Besson for exciting discussions. This research is partially funded by ACI MD 46 Bingo (French government funding) and by EU contract IQ FP6-516169 (FET arm of the IST programme).

References

[1] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proceedings of ACM SIGMOD'93*, pages 207–216, Washington (USA), 1993.

[2] M.N. Arbeitman, E.E. Furlong, F. Imam, E. Johnson, B.H. Null, B.S. Baker, M.A. Krasnow, M.P. Scott, R.W. Davis, and K.P. White. Gene expression during the life cycle of drosophila melanogaster. *Science*, 297:2270–2275, september 2002.

[3] J. Besson, C. Robardet, and J-F. Boulicaut. Constraint-based mining of formal concepts in transactional data. In *Proceedings PaKDD'04*, volume 3056 of *LNAI*, pages 615–624, Sydney (Australia), May 2004. Springer-Verlag.

[4] J. Besson, C. Robardet, and J-F. Boulicaut. Mining formal concepts with a bounded number of exceptions from transactional data. In *Proceedings KDID'04*, volume 3377 of *LNCS*, pages 33–45, Pisa (I), 2004. Springer-Verlag.

[5] C.L. Blake and C.J. Merz. UCI repository of machine learning databases, 1998.

[6] J-F. Boulicaut, A. Bykowski, and C. Rigotti. Approximation of frequency queries by mean of free-sets. In *Proceedings PKDD'00*, volume 1910 of *LNAI*, pages 75–85, Lyon (F), September 2000. Springer-Verlag.

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

[8] J-F. Boulicaut, L. De Raedt, and H. Mannila (Editors). *Constraint-based mining and inductive databases*. Springer-Verlag LNAI 3848, 2006. 405 pages.

[9] Z. Bozdech, M. Llinás, B. Lee Pulliam, E.D. Wong, J. Zhu, and J.L. DeRisi. The transcriptome of the intraerythrocytic developmental cycle of plasmodium falciparum. *PLoS Biology*, 1(1):1–16, October 2003.

[10] B. Crémilleux and J-F. Boulicaut. Simplest rules characterizing classes generated by delta-free sets. In *Proceedings ES 2002*, pages 33–46, Cambridge (UK), dec 2002. Springer-Verlag.

[11] I. S. Dhillon, S. Mallela, and D. S. Modha. Information-theoretic coclustering. In *Proceedings ACM SIGKDD 2003*, pages 89–98, Washington (USA), 2003.

[12] D. H. Fisher. Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, 2:139–172, 1987.

[13] D. Gamberger and N. Lavrac. Expert-guided subgroup discovery: Methodology and application. *Journal on Artificial Intelligence Research*, 17:501–527, 2002.

[14] L. A. Goodman and W. H. Kruskal. Measures of association for cross classification. *Journal of the American Statistical Association*, 49:732–764, 1954.

[15] C. Hébert and B. Crémilleux. Mining frequent delta-free patterns in large databases. In *Proceedings DS'05*, volume 3735 of *LNCS*, pages 124–136, Singapore, October 2005. Springer-Verlag.

[16] A.K. Jain and R.C. Dubes. *Algorithms for clustering data*. Prentice Hall, Englewood cliffs, New Jersey, 1988.

[17] W. Li, J. Han, and J. Pei. CMAR: Accurate and efficient classification based on multiple class-association rules. In *Proceedings IEEE ICDM'01*, pages 369–376, San Jose (USA), November 2001.

[18] B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rule mining. In *Proceedings KDD'98*, pages 80–86, New York (USA), August 1998. AAAI Press.

[19] S. C. Madeira and A. L. Oliveira. Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 1(1):24–45, 2004.

[20] R. Meo, P.-L. Lanzi, and M. Klemettinen (Editors). *Database Support for Data Mining Applications: Discovering Knowledge with Inductive Queries*. Springer-Verlag LNCS 2682, 2004. 385 pages.

[21] R. Pensa and J-F. Boulicaut. From local pattern mining to relevant bi-cluster characterization. In *Proceedings IDA'05*, volume 3646 of *LNCS*, pages 293–304, Madrid (E), 2005. Springer-Verlag.

[22] R. G. Pensa, C. Leschi, J. Besson, and J-F. Boulicaut. Assessment of discretization techniques for relevant pattern discovery from gene expression data. In *Proceedings ACM BIOKDD'04*, pages 24–30, Seattle, USA, August 2004.

[23] C. Robardet, B. Crémilleux, and J.-F. Boulicaut. Characterization of unsupervized clusters by means of the simplest association rules: an application for child's meningitis. In *Proceedings IDAMAP'02 co-located with ECAI'02*, pages 61–66, Lyon (F), July 2002.

[24] G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, and L. Lakhal. Computing iceberg concept lattices with TITANIC. *Data & Knowledge Engineering*, 42:189–222, 2002.

[25] R. Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. In I. Rival, editor, *Ordered sets*, pages 445–470. Reidel, 1982.