Emulating a Cooperative Behavior in a Generic Association Rule Visualization Tool

I. Nsir[1], S. Ben Yahia[1,2], and E. Mephu Nguifo[2]

[1]Départment des Sciences de l'Informatique, Faculté des Sciences de Tunis
Campus Universitaire, 1060 Tunis, Tunisie.
sadok.benyahia@fst.rnu.tn
[2]Centre de Recherche en Informatique de Lens-IUT de Lens
Rue de l'Université SP 16, 62307 Lens Cedex
mephu@cril.univ-artois.fr

Abstract. Traditional framework for mining association rules has pointed out the derivation of many redundant rules. In order to be reliable in a decision making process, such discovered rules have to be both concise and easily understandable for users, and/or an input to visualization tools. In this paper, we present a 3 graphical visualization prototype for handling generic bases of association rules. We discuss also the most adequate graphical visualization technique depending on the intrinsic structure of the generic bases of association rules. An interesting feature of the prototype is that it provides a "contextual" exploration of such rule set. Such additional displayed knowledge, based on the discovery of fuzzy meta-rules, enhances man-machine interaction by emulating a cooperative behavior.

1 Introduction

Modern hardware and database technology has made it possible to store gigabytes of information in databases. However, this rapid digitalization has pointed out an important need for tools and/or techniques to delve and efficiently discover valuable, non-obvious information from large databases. Data mining has been proposed and studied to help users better understand and analyze the information. Much research in data mining from large databases has focused on the discovery of association rules [1–3]. Association rule generation is achieved from a set F of frequent itemsets in an extraction context $\mathcal{D}$, for a minimal support *minsup*. An association rule r is a relation between itemsets of the form $r : X \Rightarrow (Y - X)$, in which X and Y are frequent itemsets, and $X \subset Y$. Itemsets X and $(Y - X)$ are called, respectively, *premise* and *conclusion* of the rule r. The valid association rules are those of which the measure of confidence Conf(r)= $\frac{support(Y)}{support(X)}$[3] is greater than or equal to a minimal threshold of confidence, named *minconf*. If $Conf(r) = 1$ then r is called *exact association rule (ER)*, otherwise it is called *approximative association rule (AR)*.

The problem of the relevance and usefulness of extracted association rules is of primary importance. Indeed, in most real life databases, thousands and even millions of high-confidence rules are generated, among which many are redundant.

[3] The number of transactions of $\mathcal{D}$ containing Y, i.e., support(Y)=$|\{t \in \mathcal{D} \mid Y \subseteq t\}|$.

© V. Snášel, R. Bělohlávek (Eds.): CLA 2004, pp. 34–46, ISBN 80-248-0597-9.
VŠB – Technical University of Ostrava, Dept. of Computer Science, 2004.

Various techniques are used to limit the number of reported rules, starting by basic pruning techniques based on thresholds for both the frequency of the represented pattern and the strength of the dependency between antecedent and conclusion. This pruning can be based on patterns defined by the user (*user-defined templates*), on boolean operators [4–6]. The number of rules can be reduced through pruning based on additional information such as a taxonomy on items [7] or a metric of specific interest [8] (e.g., Pearson's correlation or χ^2-test).

More advanced techniques that produce only a limited number of the entire set of rules rely on closures and Galois connections [9], which are in turn derived from Galois lattice theory and formal concept analysis (FCA) [10]. Finally, works on FCA have yielded a row of results on compact representations of closed set families, also called *bases*, whose impact on association rule reduction is currently under intensive investigation within the community [9].

In this paper, we are interested in the most used kind of of visualization categories in data mining, i.e., use visualization techniques to present the information catched out from the mining process. Visualization tools became more appealing when handling large data sets with complex relationships, since information presented in the form of images is more direct and easily understood by humans. Visualization tools allow users to work in an interactive environment with ease in understanding rules. In a based-tabular view of association rules, all strong rules are represented as in a tabular representation format (rule table), in which each entry corresponds to a rule. All rules can be displayed in different order, such as order by premise, conclusion, support or confidence. This helps users to have a clearer view of the rules and locate a particular rule more easily. The tabular view is advocated for representing a large number of rules with varied length. As a drawback, tabular-based technique draws heavily on "boring" logical inference that a user should perform, and is not suitable for visualizing rules from different aspects. For example, if a user is interested in a comprehensive view of the relationship between rules and items. In this case, the tabular view is not very convenient, since an item repetitively appears in the rule table as long as it is contained by a rule. These facts underline the importance of graphical rule visualization tools, permitting a clearer and more user-friendly view of rules and items. Indeed, visual representation has the capability of shifting load from the user's cognitive system to the perceptual system [11]. However, as pointed out in the dedicated literature, the graphical based techniques (e.g., 3D Histograms or 2D matrix) are actually interesting if only a small size of rules are handled.

In this paper, we are interested in presenting a graphical-based visualization prototype for handling generic bases of association rules. We try to find an answer to the following question : "*Which is the most adequate visualisation technique depending on the intrinsic structure of generic bases of association rules, specially as their sizes is by far lower than the set of all (redundant) association rules*"? As we will show later, 3D histograms-based technique is particularly advocated for visualizing generic bases extracted from sparse contexts. While, 2D matrix-based technique is indicated to visualize generic bases extracted from dense contexts.

An interesting feature of the aforementioned prototype is that it provides a ”contextual” exploration of such rule set. Indeed, additional information are provided to a user selecting a given displayed rule :

1. All the derivable rule are displayed. To derive such rules, we use the set of inference axioms, that we introduced in [12].
2. All the ”connected” rules are displayed to the user in an interactive manner. The contextual interaction is performed through the construction of fuzzy meta-rules, i.e., rules where both premises and conclusions are composed of rules.

Interestingly, such additional displayed knowledge allows improved man-machine interaction by emulating a cooperative behavior.

The remainder of this paper is organized as follows. Section 2 sketches briefly the mathematical background for the extraction of generic bases of association rules. In section 3, we present the process of the extraction of fuzzy association meta-rules. Then we discuss in depth in section 4 the opportunity of the rule set contextual exploration. Section 5 concludes this paper and points out future research directions.

2 Extracting generic bases of association rules

In the following, we recall some key results from the Galois lattice-based paradigm in FCA and its applications to association rules mining. A formal context is a triplet $\mathfrak{K} = (\mathcal{O}, \mathcal{A}, \mathcal{R})$, where $\mathcal{O}$ represents a finite set of objects (or transactions), $\mathcal{A}$ is a finite set of attributes and $\mathcal{R}$ is a binary relation (i.e., $\mathcal{R} \subseteq \mathcal{D} \times \mathcal{T}$). Each couple $(o, a) \in \mathcal{R}$ expresses that the transaction $o \in \mathcal{O}$ contains the attribute $a \in \mathcal{A}$. We define two functions that map sets of objects to sets of attributes and *vice versa*. Thus, for a set $O \subseteq \mathcal{O}$, we define $\phi(O) = \{a \mid \forall o, o \in O \Rightarrow (o, a) \in \mathcal{R}\}$; and for $A \subseteq \mathcal{A}$, $\psi(A) = \{o \mid \forall a, a \in A \Rightarrow (o, a) \in \mathcal{R}\}$. Both functions ϕ and ψ form a Galois connection between the respective power sets $\mathcal{P}(\mathcal{A})$ and $\mathcal{P}(\mathcal{O})$ [13]. Consequently, both compound operators of ϕ and ψ are closure operators, in particular $\omega = \phi \circ \psi$.

Frequent closed itemset: An itemset $A \subseteq \mathcal{A}$ is said to be *closed* if $A = \omega(A)$, and is said to be *frequent* with respect to *minsup* threshold if $support(A) = \frac{|\psi(A)|}{|\mathcal{O}|} \geq$ *minsup*. An itemset $g \subseteq \mathcal{A}$ is *called minimal generator* of a closed itemset A, if and only if $\omega(g) = A$ and $\nexists g' \subseteq g$ such that $\omega(g') = A$.

Iceberg Galois lattice: When only frequent closed itemsets are considered with set inclusion, the resulting structure $(\hat{\mathcal{L}}, \subseteq)$ only preserves the joint operator. In the remaining of the paper, such structure is referred to ”*Iceberg Galois Lattice*”.

With respect to [14] and [15], given an Iceberg Galois lattice, representing precedence relation-based ordered closed itemsets, generic bases of association rules can be derived in a straightforward manner. We assume that, in such structure, each closed itemset is ”decorated” with its associated list of minimal generators. Hence, generic approximative association rules ($\mathcal{GAR}$) represent ”inter-node” implications, assorted with a statistical information, i.e., the confidence, from a sub-closed-itemset to a super-closed-itemset while starting from a given node in an ordered structure. Inversely, generic exact association rules ($\mathcal{GER}$) are ”intra-node” implications extracted from each node in the ordered structure.

For example, let us consider the transaction database given by Table 1 (Left). The set of extracted closed itemsets, with their associated minimal generators, is depicted by Table 1 (Right). Therefore, Table 2 yields, respectively, the set of generic exact association rules and the set of approximative association rules that it may be possible to derive for minsup=1.

Table 1. Left: Transaction database **Right:** Closed itemsets list

TID	items
1	aek
2	aek
3	cg
4	cegk
5	g
6	ek
7	gk

Gen.	extent	intent
a	$\{1, 2\}$	aek
c	$\{3, 4\}$	cg
e	$\{1, 2, 4, 6\}$	ek
g	$\{3, 4, 5, 7\}$	g
k	$\{1, 2, 4, 6, 7\}$	k
ce	$\{4\}$	cegk
ck	$\{4\}$	cegk
eg	$\{4\}$	cegk
gk	$\{4\}$	cegk

Table 2. Generic association rule set **Left:**Exact. **Right:**Approximative

#	”⇒”
R_1	a⇒ek
R_2	c⇒g
R_3	e⇒k
R_4	ce⇒gk
R_5	ck⇒eg
R_6	eg⇒ck
R_7	gk⇒ec

#	”⇒”	Conf.
R_8	g⇒c	0.5
R_9	g⇒cek	0.25
R_{10}	c⇒egk	0.5
R_{11}	k⇒e	0.8
R_{12}	k⇒ae	0.4
R_{13}	e⇒ak	0.5
R_{14}	k⇒ceg	0.5
R_{15}	e⇒cgk	0.25

3 Extracting fuzzy association rules

In the following, we present the basic fuzzy constructs and our proper notations.

3.1 Basic Fuzzy constructs

Fuzzy Sets : Let U be a finite classical set of objects, called *universe of discourse*. A fuzzy set $\widetilde{F}$, in a universe of discourse U, is characterized by a membership function

$\mu_{\widetilde{F}}$: $U \rightarrow [0, 1]$, where $\mu_{\widetilde{F}}(u)$ denotes the degree of membership of u in the fuzzy set $\widetilde{F}$. Hence, the fuzzy set $\widetilde{F}$ is denoted by : $\widetilde{F} = \{\overset{\mu_{\widetilde{F}}(u_1)}{u_1}, \overset{\mu_{\widetilde{F}}(u_2)}{u_2}, \ldots, \overset{\mu_{\widetilde{F}}(u_n)}{u_n}\}$.

In the following, the notion of fuzzy extraction context, made up of objects and attributes, related under a fuzzy type-1 relation, is formalized.

Fuzzy extraction context :A fuzzy extraction context is a triple $\widetilde{\mathcal{D}}$=($\mathcal{O}$,$\widetilde{\mathcal{I}}$,$\widetilde{R}$) describing a finite set $\mathcal{O}$ of objects, a fuzzy finite set $\widetilde{\mathcal{I}}$ of database items (or descriptors) and a fuzzy binary relation $\widetilde{R}$ (i.e.,$\widetilde{R} \subseteq \mathcal{O} \times \widetilde{\mathcal{I}}$). Each couple $(o, \overset{\alpha}{i}) \in \widetilde{R}$, denotes that the object $o \in \mathcal{O}$, has the item $i \in \widetilde{\mathcal{I}}$, at least with a degree α.

Fuzzy Galois connection: Let $\widetilde{\mathcal{D}} = (\mathcal{O}, \widetilde{\mathcal{I}}, \widetilde{R})$ be a fuzzy extraction context. For $X \subseteq \mathcal{O}$ and $\widetilde{Y} \subseteq \widetilde{\mathcal{I}}$, the functions : $\widetilde{\phi} : P(\mathcal{O}) \rightarrow \mathcal{P}(\widetilde{\mathcal{I}})$ and $\widetilde{\psi} : \mathcal{P}(\widetilde{\mathcal{I}}) \rightarrow P(\mathcal{O})$ are defined as follows [16]:

$\widetilde{\phi}(X) = \{\overset{\alpha}{b} | \ \alpha = min\{\mu_{\widetilde{R}}(a, b) \mid a \in X\}\}$,
$\widetilde{\psi}(\widetilde{Y}) = \{a \in O \mid \forall b, b \in \widetilde{B} \Rightarrow \mu_{\widetilde{R}}(a, b) \geq \mu_{\widetilde{Y}}(b)\}$

The fuzzy operator $\widetilde{\phi}$ is applied on a crisp set of objects and determines to which degree each property is satisfied by all objects, according to their respective degrees. Note that $\widetilde{\phi}$, as defined formerly, presents a desired *abstraction* vocation. Indeed, the retrieved fuzzy set is the least generalization, through the "*min*" function, of all fuzzy sets (or descriptions) associated respectively to the input set objects.

Similarly, the fuzzy operator $\widetilde{\psi}$ is applied on a set of properties, represented by a fuzzy set, and determines to which degree each object satisfies all of them. The implicit use of the *Rescher-Gaines* fuzzy implication permits to obtain the desired *interpretation* effect. Hence, the operator "$\geq$" permits to filter only objects fulfilling the constraint that the associated descriptions are more general than the input description.

Remark 1. It is noteworthy that the proposed definition of fuzzy Galois connection -without context transformation- can not be unique. This constatation is due to the existence of multitude of semantic/syntactic parameters to take into account in any extension to the fuzzy context (e.g., fuzzy implication choice). For instance, we mention based-Lukasiewicz-implication propositions of Belohlavek [17] and Pollandt [18], where a fuzzy formal concept is defined by a pair of two fuzzy sets $\widetilde{X}$ and $\widetilde{Y}$, where $\widetilde{X} = \widetilde{\omega}(\widetilde{Y})$ and $\widetilde{Y} = \widetilde{\phi}(\widetilde{X})$. The reader is referred to [16] for a critical discussion on these propositions, based on semantic interpretations attached to membership degrees in a fuzzy set (i.e.,fulfillment, preference,interpretation).

3.2 The FARD algorithm

The FARD [19] algorithm falls in the "test-and-generate" characterization for mining frequent (closed) patterns algorithms. It handles as input an extended transaction database, e.g, a database in which quantities of the purchased items are available. The peculiarity of the FARD algorithm stands in the fact that it is dedicated. Indeed, it tackles the original context without binarizing it, since it takes advantage of the particular properties of the fuzzy descriptions to limit the computation effort.

The algorithm traverses iteratively the search space in a level-wise manner. During each iteration corresponding to a level, a set of candidate patterns is created by joining frequent patterns discovered during the previous iteration, the supports of all candidate patterns are counted and infrequent ones are discarded. Fuzzy association rules are generated in two successive steps:

1. Discovering all frequent fuzzy closed itemsets (FuFCs),
2. From the discovered frequent fuzzy closed itemsets, generate all fuzzy association rules that it is possible to derive.

As input the FARD algorithm takes a fuzzy extraction context $\widetilde{\mathcal{D}}$, *minsup* and *minconf*. In this paper, we consider that the fuzzy extraction context $\widetilde{\mathcal{D}}$ is constituted as follows: $\widetilde{\mathcal{D}} = (\mathcal{O}, \mathcal{RS}, \widehat{\mathcal{R}})$, where $\mathcal{O}$ represents the finite set of objects, $\mathcal{RS}$ is a finite set of $\mathcal{GER}$ and $\mathcal{GAR}$ and $\widehat{\mathcal{R}}$ is a fuzzy binary relation, where $\mu_{\widehat{\mathcal{R}}}(o, R) = conf(R)$, $\forall$ o $\in \mathcal{O}$ and R $\in \{\mathcal{GER} \cup \mathcal{GAR}\}$.

Following the general principle of a level-wise algorithm, the FARD algorithm performs in each iteration the following two steps (assuming that items are sorted in a lexicographic order):

1. **Construction step**: The GEN-CLOSED function, is applied to each generator in CFuFC$_i$, determining its support and possibly its closure (if it is frequent enough). This set is pruned with respect the anti-monotonous constraint, through the $minsup$ threshold.
2. **Pruning step** : The set of generators to be utilized in the next iteration, i.e. CFuFC$_{i+1}$, is computed by applying the GEN-NEXT function to the set CFuFC$_i$.

The algorithm terminates when there are no more generators to process, i.e. CFuFC$_i$.gen-list, is empty. FARD algorithm pseud-code is given in Algorithm 1.1. Note that subroutines pseudo-code is not given due to lack of available space.

Algorithm 1.1: FARD

```
Input: D̃ : fuzzy extraction context, S̃:items user-constraints, minsup
Output: FuFC= ∪_i FuFC_i
begin
    CFuFC_1.gen-list={1-fuzzy itemsets};
    for (i = 1;CFuFC_i.gen-list≠ ∅;i++) do
        CFuFC_i.clos= ∅;
        FuFC_i=GEN-CLOSED(CFuFC_i);
        CFuFC_{i+1}=GEN-NEXT(FuFC_i);
    return FuFC=∪_i FuFC_i
end
```

Example 1. Let us consider the fuzzy extraction context depicted by Table 3. In this context, each $\overset{\alpha}{(R_i, j)}$ couple means that the object j is covered by the rule R$_i$ with a confidence equal to α. From such context, by applying the FARD algorithm, it is possible to extract the fuzzy closed itemsets from which we derive fuzzy generic exact fuzzy association rules, which are depicted by Table 4.

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	R_{11}	R_{12}	R_{13}	R_{14}	R_{15}
1	1		1								.8	.4		.5	
2	1		1								.8	.4		.5	
3			1					.5							
4		1	1	1	1	1	1	.5	.25	.5	.8		.2		.25
5															
6			1								.8				
7															

Table 3. Fuzzy meta-context

#	”⇒”
MR_1	$R_1^1 \Rightarrow R_3^1 R_{11}^{.8} R_{12}^{.4} R_{14}^{.5}$
MR_2	$R_2^1 \Rightarrow R_8^1$
MR_3	$R_3^1 \Rightarrow R_1^1 R_{11}^{.8} R_{12}^{.4} R_{14}^{.5}$
MR_4	$R_4^1 \Rightarrow R_5^1 R_6^1 R_7^1 R_8^1 R_9^{.25} R_{10}^{.5} R_{11}^{.8} R_{13}^{.2} R_{15}^{.25}$
MR_5	$R_5^1 \Rightarrow R_4^1 R_6^1 R_7^1 R_8^1 R_9^{.25} R_{10}^{.5} R_{11}^{.8} R_{13}^{.2} R_{15}^{.25}$
MR_6	$R_6^1 \Rightarrow R_4^1 R_5^1 R_7^1 R_8^1 R_9^{.25} R_{10}^{.5} R_{11}^{.8} R_{13}^{.2} R_{15}^{.25}$
MR_7	$R_7^1 \Rightarrow R_4^1 R_5^1 R_6^1 R_8^1 R_9^{.25} R_{10}^{.5} R_{11}^{.8} R_{13}^{.2} R_{15}^{.25}$
MR_8	$R_8^1 \Rightarrow R_2^1$
MR_9	$R_9^{.25} \Rightarrow R_4^1 R_5^1 R_6^1 R_7^1 R_8^1 R_{10}^{.5} R_{11}^{.8} R_{13}^{.2} R_{15}^{.25}$
MR_{10}	$R_{10}^{.5} \Rightarrow R_4^1 R_5^1 R_6^1 R_7^1 R_8^1 R_9^{.25} R_{11}^{.8} R_{13}^{.2} R_{15}^{.25}$
MR_{11}	$R_{11}^{.8} \Rightarrow R_3^1$
MR_{12}	$R_{12}^{.4} \Rightarrow R_1^1 R_3^1 R_{11}^{.8} R_{14}^{.5}$
MR_{13}	$R_{13}^{.2} \Rightarrow R_4^1 R_5^1 R_6^1 R_7^1 R_8^1 R_9^{.25} R_{10}^{.5} R_{11}^{.8} R_{15}^{.25}$
MR_{14}	$R_{14}^{.5} \Rightarrow R_1^1 R_3^1 R_{11}^{.8} R_{12}^{.4}$
MR_{15}	$R_{15}^{.25} \Rightarrow R_4^1 R_5^1 R_6^1 R_7^1 R_8^1 R_9^{.25} R_{10}^{.5} R_{11}^{.8} R_{13}^{.2}$

Table 4. Exact fuzzy meta-association Rules set

4 Graphical visualization and cooperative exploration

Let us keep in mind that given a generic association rule set, we aim to set up a graphical visualization prototype permitting to enhance the man-machine interaction by providing a contextual ”knowledge”, minimizing a boring large amount of knowledge exploration. To do so, we construct a set of fuzzy meta-rules, where both premise and conclusion rule’s are made up of classical association rules. The role of such fuzzy meta-rules is to highlight ”connections” between association rules without loosing rule’s confidence information.

Here we come to a turning point: Why is it interesting to try to understand why a user and/or a knowledge expert may be interested in a particular rule, and to determine what interesting information or knowledge, not explicitly requested, we could provide him, in addition to the proper answer? Indeed, improving man-machine interaction by emulating a cooperative behavior has been proposed by some researchers through various techniques [20]. In [21], the author states: ”*requests for data can be classified roughly into two kinds : specific requests and goals. A specific request establishes a rigid qualification, and is concerned only with data that matches it precisely. A goal,* *on the other hand, establishes a target qualification and is concerned with data which is* ***close*** *to the target*". For Cuppens and Demolombe [22], "*the basic idea is that when a person asks a question, he is not interested to know the answer just to increase his knowledge, but he has the intention to realize some action, and the answer contains necessary, or useful information to realize this action*". In our context, when such additional knowledge is not supplied, this forces the user to retry a tedious rule exploration repeatedly, until obtaining a satisfactory "matching".

The visualization prototype takes as input an XML file complying to the DTD depicted by Figure 1. This storing format is argued by the fact that XML, the eXtensible Markup Language, has recently emerged as a new standard for data representation and exchange on the Internet. As output, the set of selected rules can be saved in a file with HTML or TXT format.

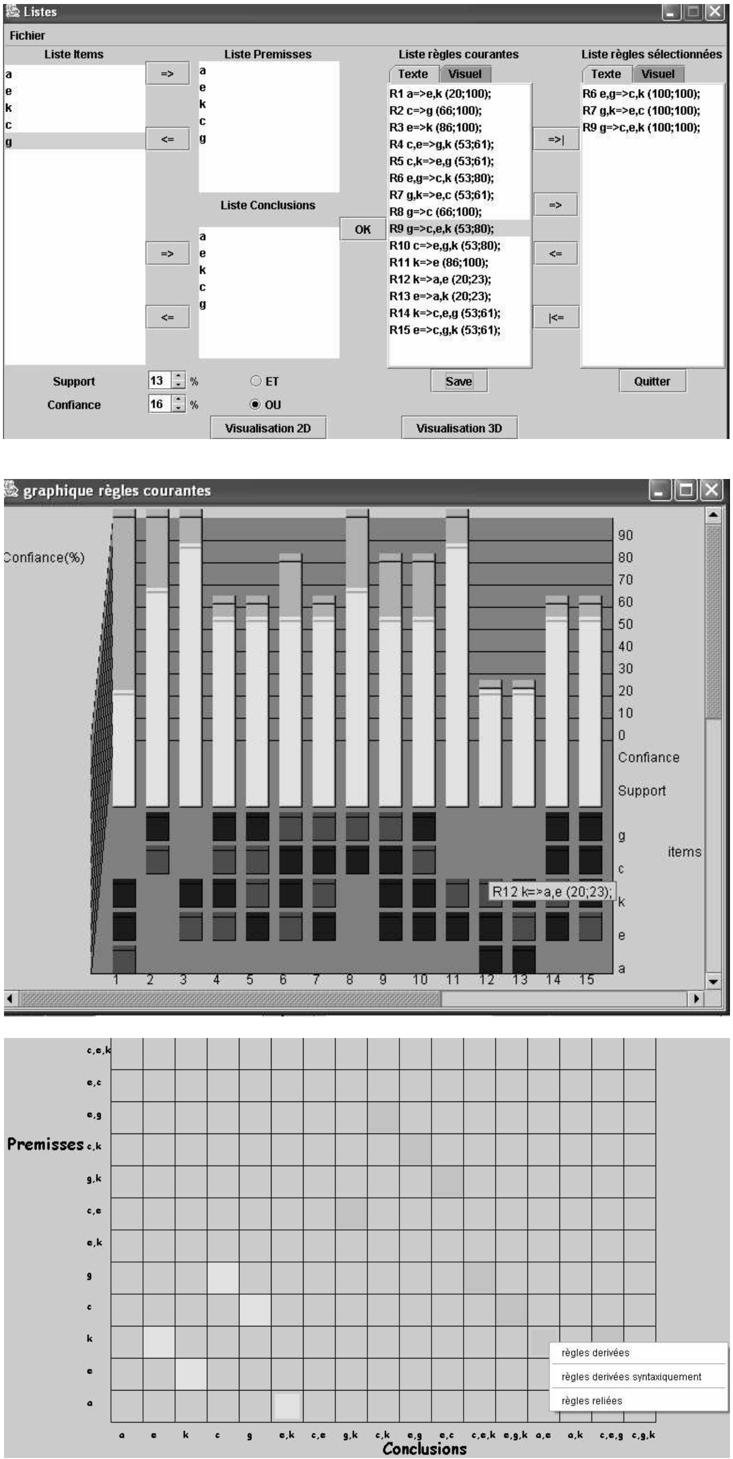
```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE ruleSet[
<!ELEMENT ruleSet (rule+)>
<!ELEMENT rule (premise, conclusion,listObjects?)>
<!ELEMENT premise (item+)>
<!ELEMENT conclusion (item+)>
<!ELEMENT listObjects #PCDATA>
<!ATTLIST rule id ID #REQUIRED>
<!ATTLIST rule conf CDATA #REQUIRED >
<!ATTLIST premise id ID #REQUIRED>
<!ATTLIST premise deg CDATA #REQUIRED >
]>
```

Fig. 1. Associated DTD of accepted input XML file

In the prototype interface, c.f., Figure 2(Up), the user can select items that can both appear in the premise and/or the conclusion part. Once the minsup and minconf thresholds are fixed, the user can generate the desired rules. Both textual and graphical representation are provided. The user can further filter specific rules from the generated ones.

The suitable visualization technique: In the 3D histograms based visualization technique, c.f., the screenshot depicted by Figure 2(Middle), matrix floor rows represent items and columns represent item associations. The red and blue blocks of each column (rule) represent the premise and the conclusion, respectively. Item identities are shown along the right side of the matrix. The associated confidence and support are represented by a scaled histogram. While on 2D matrix-based visualization technique, rule premise items are on one axis, and the rule conclusion items are on the other axis. We use the solution introduced by `MineSet` software [4] allowing multiple items in rule premise

[4] Available at http://www.sgi.com/software/mineset/index.html.


Listes
Fichier
Liste Items
a
e
k
c
g
Liste Premisses
a
e
k
c
g
Liste Conclusions
a
e
k
c
g
OK
Liste règles courantes
Texte
Visuel
R1 a=>e,k (20;100);
R2 c=>g (66;100);
R3 e=>k (86;100);
R4 c,e=>g,k (53;61);
R5 c,k=>e,g (53;61);
R6 e,g=>c,k (53;80);
R7 g,k=>e,c (53;61);
R8 g=>c (66;100);
R9 g=>c,e,k (53;80);
R10 c=>e,g,k (53;80);
R11 k=>e (86;100);
R12 k=>a,e (20;23);
R13 e=>a,k (20;23);
R14 k=>c,e,g (53;61);
R15 e=>c,g,k (53;61);
Liste règles sélectionnées
Texte
Visuel
R6 e,g=>c,k (100;100);
R7 g,k=>e,c (100;100);
R9 g=>c,e,k (100;100);
Support
13 %
Confiance
16 %
ET
OU
Save
Quitter
Visualisation 2D
Visualisation 3D
graphique règles courantes
Confiance(%)
90
80
70
60
50
40
30
20
10
0
Confiance
Support
g
c
k
e
a
items
R12 k=>a,e (20;23);
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Premisses
c,e,k
e,c
e,g
c,k
g,k
c,e
e,k
g
c
k
e
a
règles derivées
règles derivées syntaxiquement
règles reliées
a e k c g e,k c,e g,k c,k e,g e,c c,e,k e,g,k a,e a,k c,e,g c,g,k
Conclusions

or conclusion, by grouping each of the items combinations in rule premise or rule conclusion as one unit. The confidence value is indicated by different colors. The higher the confidence value is, the darker the color. Currently, the user has to indicate which visualization technique he (she) prefers. However, we are trying to set up an automatic indicator which can choose the most adequate visualization technique depending on an assessment of the density of the extraction context. In fact, as indicated by experimental experiments on real-life and synthetic extraction contexts whose parameters are summarized by Table 5, when the extraction context is dense, generic association rules drawn from such context tend to present large conclusions (depending on the number of items). Indeed in Table 6, we tried to assess, for multiple *minsup* values, the difference between the size of the minimal generator and the average size of its associated frequent closed itemsets. the larger the difference is, the larger the generic conclusion's association rule part. From the entries dedicated to dense extraction contexts in Table 6, we remark that this difference is important comparatively to that pointed out in by sparse contexts. Therefore, when we consider sparse extraction contexts, the length of premise and the length of the conclusion of generic association rule tend to be equal. Then given that it is known, 3D histograms visualization technique is adapted to visualize rules with a small number of items in the conclusion part, we can conclude it is better to use them for generic rules extracted from sparse contexts.

Base	Type	$\lvert\mathcal{A}\rvert$	|largest itemset |	$\lvert\mathcal{O}\rvert$	Size
T10I4D100	Sparse	1000	29	100000	4.05MB
T10I10D100	Sparse	1000	77	100000	15MB
ROOMOUT	Dense	120	23	8124	580KB
CONNECT	Dense	130	44	49840	8.82MB
CHESS	Dense	76	37	3196	334KB

Table 5. Extraction contexts parameters

Displaying additional information

– **Displaying derivable rules**: Once a user is interested in the graphical visualization window (c.f., for example Figure 2 Down), then by activating the contextual menu, two options are displayed "Syntactic Rule derivation" and "Rule derivation". If the user selects the first option, then the system provides all syntactically derivable rules [5] in a new window. Indeed, for these derivable rules the only available information is that their support and confidence values are at least equal, respectively, to those of the generic association rule used to derive them. For example in Figure 3, five association rule are presented to the user once he selected the generic association rule: k ⇒ae. On the other hand, if the user selects the "Rule derivation"

[5] The syntactic derivation is based on the *Cover* operator introduced in [23], i.e., Cover(X ⇒Y) = $\{X \cup Z \Rightarrow V \mid Z, V \subseteq Y \wedge Z \cap V = \emptyset \wedge V \neq \emptyset\}$, with |Cover(X ⇒Y)|=$3^m$-$2^m$ where |Y|=m.

Base	support (%)	$\|\mathcal{FCI}\|$	size of minimal generator : Average size of closed itemset
T10I4D100	0.5	1073	1:1- 2:2- 3:3-4:4-5:5
	1	385	1:1- 2:2- 3:3
	2	155	1:1
	5	10	1:1
T10I10D100	5	316	1:1- 2:2
	10	82	1:1
	15	19	1:1
ROOMOUT	10	4897	1 : 3.767 - 2 : 5.840 - 3 : 7.138 - 4 : 8.364 - 5 : 8.998 - 6 : 9.803 - 7 : 11.179
	20	1197	1 : 3.418 - 2 : 4.621 - 3 : 6.358 - 4 : 7.332 - 5 : 8.218 - 6 : 9.169 - 7 : 11
	30	427	1 : 2.5 - 2 : 4.041 - 3 : 5.052 - 4 : 6.209 - 5 : 7.333 - 6 : 8.5
CONNECT	89	9017	1: 2.045 -2 :3.796 - 3: 5.322- 4 :6.7 -5:7.962 -6 :9.151 -7 : 10.272 -8 : 11.341
	90	7475	1 : 1.952 -2: 3.691 -3 : 5.174 -4 : 6.512 -5 : 7.751 -6 : 8.911 -7 : 10.006 -8 : 11.032
	91	6126	1 : 1.952 -2 : 3.596 -3 : 5.017 -4 : 6.323 -5 : 7.539 -6 : 8.671 -7 : 9.732 -8 : 10.71
CHESS	78	7111	1 :1.227 -2 :2.417 -3: 3.543 -4: 4.61 -5: 5.612 -6: 6.573 -7: 7.525 -8: 8.453 -9: 9.319 -10: 10
	80	5083	11 : 1.21 -2: 2.401 -3: 3.537 -4: 4.564 -5: 5.546 -6: 6.518 -7: 7.472 -8: 8.369 -9: 9.181
	85	1885	1 : 1.25 -2: 2.366 -3: 3.406 -4: 4.414 -5: 5.394 -6: 6.322-7: 7.185 -8: 8

Table 6. Experimental results

option, then the system provides only all the derivable rules that have exactly the same support and confidence values as those of the generic association rule used to derive them. As depicted in Figure 3, only one rule, k⇒a, is derivable and has exactly the same support and confidence as k⇒ae.

– **Displaying connected rules**: This option is currently under implementation and we will illustrate it through an example. Suppose that a user may be interested in rule the R_1. Then, when the user clicks on R_1, the visualization interface provides all rules that are connected to R_1 in a displaying area. From the meta-rule MR_1 depicted in Table 4, we can check that the following rules are connected to R_1: $\mathbf{R}_3$: e⇒k (conf.=1), $\mathbf{R}_{11}$: k⇒e (conf.=0.8), $\mathbf{R}_{12}$: k⇒ae (conf.=0.4) and $\mathbf{R}_{14}$: k⇒ceg (conf.=0.5).

5 Conclusion

In this paper, we presented an approach for providing a cooperative exploration of generic bases of association rules. This additional knowledge highlighting connected rules to a user-select rules allows an improvement of man-machine interaction. To do

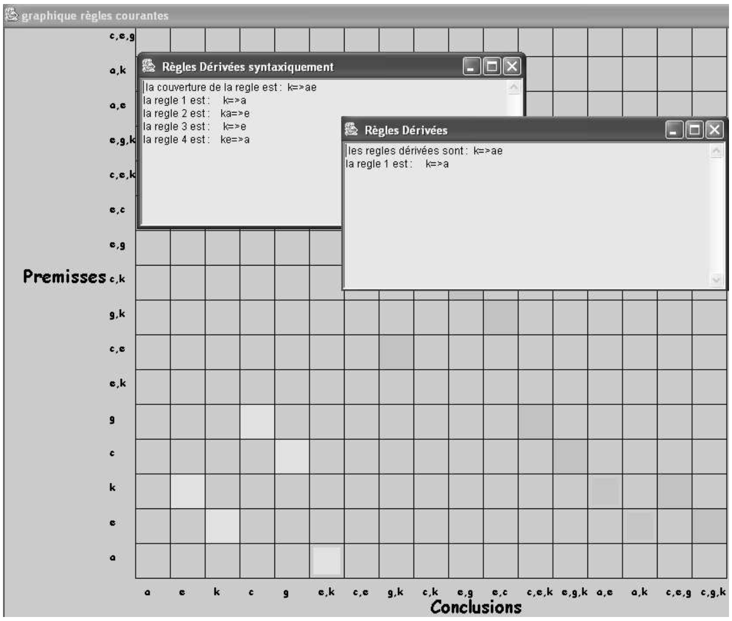


Fig. 3. Displaying derived association rule.

so, we constructed a set of fuzzy meta-rules extracted from the discovered fuzzy closed itemsets.

The visualization prototype is currently under implementation and in the near future, we plan also to include recommended visualization tasks, e.g., *history* and *extract* [24], and to lead extensive experimental results to assess their satisfaction vs the user graphical interface.

Acknowledgments The authors would like to thank Mrs LEGROY Audrey for its helpful efforts in the implementation of visualization prototype.

References

1. Agrawal, R., Skirant, R.: Fast algorithms for mining association rules. In: Proceedings of the 20th International Conference on Very Large Databases. (1994) 478–499
2. Brin, S., Motwani, R., Ullman, J.: Dynamic itemset counting and implication rules. In: Proceedings ACM SIGMOD, International conference on Management of Data ,Tucson, Arizona, USA. (1997) 255–264
3. Manilla, H., Toinoven, H., Verkamo, I.: Efficient algorithms for discovering association rules. In: AAAI Worshop on Knowledge Discovery in Databases. (1994) 181–192
4. Meo, R., Psaila, G., Ceri, S.: A new sql-like operator for mining association rules. In: Proceedings of the VLDB Conference. (1996) 122–133

5. Ng, R.T., Lakshmanan, V.S., han, J., Pang, A.: Exploratory mining and pruning optimizations of constrained association rules. In: Proceedings of the SIGMOD Conference. (1998) 13–24
6. Srikant, R., Vu, Q., Agrawal, R.: Mining association rules with item constraints. In: Proc. of the 3rd Intl. Conference on Knowledge Discovery in Databases and Data Mining, Newport Beach, California. (1997) 67–73
7. Han, J., Fu, Y.: Discovery of multiple-level association rules from large databases. In: Proceedings of the VLDB Conference. (1995) 420–431
8. Brin, S., Motawni, R., Silverstein, C.: Beyond market baskets : Generalizing association rules to correlation. In: Proceedings of the SIGMOD, Tucson, Arizona (USA). (1997) 265–276
9. Bastide, Y., Pasquier, N., Taouil, R., Lakhal, L., Stumme, G.: Mining minimal non-redundant association rules using frequent closed itemsets. In: Proceedings of the International Conference DOOD'2000, Lecture Notes in Computer Sciences, Springer-verlag. (2000) 972–986
10. Ganter, B., Wille, R.: Formal Concept Analysis. Springer-Verlag, Heidelberg (1999)
11. Buono, P., Costabile, M.F., Lisi, F.A.: Supporting data analysis through visualizations. In: Proceedings of Workshop on Visual Data Mining, Freiburg, Germany. (2001) 67–78
12. BenYahia, S., Nguifo, E.M.: Revisiting generic bases of association rules. In: Proceedings of 6th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2004) (Springer-Verlag) to appear, Zaragoza, Spain. (2004)
13. Barbut, M., Monjardet, B.: Ordre et classification. Algèbre et Combinatoire. Hachette, Tome II (1970)
14. Luxenburger, M.: Implication partielles dans un contexte. Mathématiques et Sciences Humaines **29** (1991) 35–55
15. Guigues, J., Duquenne, V.: Familles minimales d'implications informatives résultant d'un tableau de données binaires. Mathématiques et Sciences Humaines (1986) 5–18
16. Jaoua, A., Elloumi, S., BenYahia, S., Alvi, F.: Galois connection in fuzzy binary relations: applications for discovering association rules and decision making. Methodos Publisher (2002)
17. Belohlavek, R.: Fuzzy Galois connections. Math. Logic **45** (1999) 497–504
18. Pollandt, S.: Fuzzy-Begriffe: Formale Begriffsanalyse unscharfer Daten. Springer-verlag (1997)
19. BenYahia, S., Jaoua, A.: Discovering knowledge from fuzzy concept lattice. In: Data mining and Computational Intelligence. Studies in Fuzziness and Soft Computing, Vol. 68. Physica Verlag, Heidelberg (2001) 167–190
20. Frantz, V., Shapiro, J.: Algorithms for automatic construction of query formulations in boolean form. Journal of the American Society for Information Science **1** (1991) 16–26
21. Motro, A.: Extending the relational Databases Model to support Goal queries. In Lerschberg, L., ed.: Expert Database Systems. The Benjamin/Cummings Publishing Company (1987) 129–150
22. Cuppens, F., Demolombe, R.: How to recognize interesting topics to provide cooperative answering. Information Systems **14** (1989) 163–173
23. Kryszkiewicz, M.: Concise representations of association rules. In Hand, D.J., Adams, N., Bolton, R., eds.: Proceedings of Pattern Detection and Discovery, ESF Exploratory Workshop, London, UK. Volume 2447 of Lecture Notes in Computer Science., Springer (2002) 92–109
24. Shneiderman, B.: The eyes have it: A task by data type taxonomy for information visualization. In Press, I.C.S., ed.: Proceedings IEEE Symposium on Visual Languages, Boulder, Colorado, 336–343 (1996)