



**HAL**  
open science

# Extracting Compact and Information Lossless Sets of Fuzzy Association Rules

Sarra Ayouni, Sadok Ben Yahia, Anne Laurent

► **To cite this version:**

Sarra Ayouni, Sadok Ben Yahia, Anne Laurent. Extracting Compact and Information Lossless Sets of Fuzzy Association Rules. *Fuzzy Sets and Systems*, 2011, 183 (1), pp.1-25. 10.1016/j.fss.2011.06.019 . lirmm-00604420

**HAL Id: lirmm-00604420**

**<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00604420v1>**

Submitted on 31 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Extracting Compact and Information Lossless Sets of Fuzzy Association Rules <sup>☆</sup>

S. AYOUNI<sup>1,2</sup> and S. BEN YAHIA<sup>1</sup> and A. LAURENT<sup>2</sup>

<sup>1</sup>*Faculty of Sciences of Tunis, 1060, Tunis, Tunisia*

*sarra.ayouni@fst.rnu.tn*

*sadok.benyahia@fst.rnu.tn*

<sup>2</sup>*LIRMM – CNRS, 161 rue Ada, Montpellier, France*

*ayouni@lirmm.fr*

*laurent@lirmm.fr*

---

## Abstract

Applying classical association rule extraction framework on fuzzy data sets leads to an unmanageably highly sized association rule sets – compounded with an information loss due to the discretization operation – that often constitutes a hamper towards an efficient exploitation of the mined knowledge. To overcome such a drawback, we advocate the extraction and the exploitation of compact and informative generic basis of fuzzy association rules. This generic basis constitutes a compact nucleus of fuzzy association rules, from which it is possible to informatively derive all the remaining rules. In order to ensure a sound and complete derivation process, we introduce an axiomatic system allowing the complete derivation of all the redundant rules. The results obtained from experiments carried out on benchmark datasets, are very encouraging. They highlight a very important reduction of the number of the extracted fuzzy association rules without information loss.

*Key words:* Fuzzy Sets, Fuzzy Galois Connection, Fuzzy Generic Association Rules, Axiomatic System.

---

<sup>☆</sup>This article is a revised and extended version of a paper presented at the 2007 IEEE International Conference on Fuzzy Systems FUZZ-IEEE'2007, London, UK 2007.

## 1. Introduction

Association rule mining, introduced by [1], has been applied over market basket data in order to identify groups of products frequently bought together. This knowledge may be of valuable help for shop-keepers to make decisions about what to put on sale, how to place merchandize on shelves, to maximize a cross-selling effect etc. Such an association rule can tell, for example, customers that buy bread and milk will also buy butter. This kind of rules used to be called binary association rules (*i.e.*, which involve binary attributes). Mining binary association rules has been studied for several years and has become a well established technique.

However, when tackling numerical contexts using the binary data mining background, the reported results are closely dependent on the discretization method, *i.e.*, how the original context is translated to a binary one. Usually, the corresponding entry is set to crossed whenever the corresponding item fulfills a given property with respect to a given threshold. Clearly, such translation is far from being information lossless. Indeed, the main drawback is that discretization is not able to describe the "actual" situation. For this reason, it is of paramount importance to handle an extended extraction context, *i.e.*, without binarizing the original context. In this respect, introducing soft computing techniques seems to be a promising solution. In fact, the use of soft computing techniques mainly based on fuzzy sets in connection with association rules grasped the interest of numerous authors. By allowing "soft" rather than "harsh" boundaries of intervals, fuzzy sets can avoid some undesirable threshold effects. Furthermore, fuzzy association rules are very appealing from a knowledge representational point of view: fuzzy set theory features an interesting capability to bridge the gap between quantitative patterns and qualitative knowledge structures expressible in terms of natural language. Thus, association rules discovered in a database might be presented in a linguistic and, hence, comprehensible and user-friendly way. Thus, the literature witnesses a determined effort to introduce the soft computing techniques, e.g., [2, 8, 14, 15, 16, 28, 30], to cite but a few. Unfortunately, these approaches were introduced regardless the effectiveness of the mined knowledge. In fact, they paid little attention to the amount of association rules that may be drawn. Hence, at the end of the process, the user is obliged to face an overwhelming quantity of association rules among which a large number is redundant, what badly affects the quality of their interpretability. Beyond, the quantitative aspect and as pointed out by [32], the existing mining approaches did not differentiate the data items in terms of the interestingness users have on them. Thus, avoiding the extraction of a huge amount of knowledge is of primary importance as it guarantees extra

value knowledge usefulness and reliability. This fact is reinforced while handling highly dense data.

To tackle the above mentioned limitations, we introduce a novel approach for extracting fuzzy association rules, with the following features:

1. **Highlight user's item interestingness** : laying within a user-driven approach, the user has the possibility to highlight the importance of the products, items or attributes, *e.g.*, the *total income* attribute is more interesting than the *height* of a person in a household. Indeed, according to Hackman and Oldham [23], the larger the user's skills involvement during the mining process, the stronger sense of independence and responsibility.
2. **The extraction of compact and information lossless fuzzy association rules**: in fact, generic bases of association rules – backboneed on the concepts of minimal generator and closed itemset, –constituted so far irreducible compact nuclei of association rules. In addition, we provide an axiomatic system, that we show that it is sound and complete, to ensure the derivation mechanism of all redundant fuzzy association rules. The experiments show important rates of compactness.

The remainder of the paper is organized as follows. Section 2 scrutinizes the related work that focussed on fuzzy association rule mining. Section 3 introduces the generic basis of fuzzy association rules. The derivation mechanism by means of an axiomatic system, shown to be sound and complete, is also given. Section 4 introduces the GEN-IFF algorithm that mines fuzzy closed itemsets. In Section 5, we report the encouraging compactness rates obtained by applying our approach on SAGE data as well as on benchmark datasets. Section 6 presents concluding remarks and sketches our future perspectives.

## 2. Related work

The problem of mining association rules in large relational tables containing quantitative attributes was ignited by Agrawal *et al.* [30]. An example of such an association might be "10% of married people between age 50 and 60 have 2 cars". The authors proposed to divide the attribute domain into discret intervals and to combine adjacent ones if necessary. Then, each original attribute is replaced by a set of (*attribute, interval*) pairs. Thus, the quantitative problem is converted into a binary one. This discretizing method certainly generates information loss. Besides, defining such intervals may not be concise and intuitive for human experts. For example, the partition into intervals method might classify a person

as "young", if his age is less than 40, and as "adult" if his age is greater than 40. However, this obviously corresponds to a very subjective human perception to "young" and "adult". Moreover, another drawback consists in the number of extracted association rules that will evidently be huge, since every attribute  $att_i$  of the basis will be replaced by a certain number of pairs  $(att_i, interval_{i_j})$  (i.e., each attribute  $att_i$  of the basis is associated to  $k$  intervals  $interval_{i_1}, \dots, interval_{i_j}, \dots, interval_{i_k}$ ). To avoid the pitfall of quantitative attribute discretization problems, several authors opted for introducing fuzzy logic in the association rule mining technique. In fact, the fuzzy logic was shown to be an advocable mechanism for handling uncertain and imprecise data [17]. Thus, many fuzzy association rule mining algorithms have been proposed in literature.

Au and Chan [2] introduced a novel technique called FARM. This technique employs linguistic terms and their corresponding fuzzy sets whom have to be previously predefined by human experts. In order to identify interesting fuzzy association rules (FARs), FARM uses *adjusted difference analysis* which does not require any user supplied thresholds (i.e., a rule is interesting when its adjusted difference is greater than 1.96 which is the 95 percentiles of the normal distribution). Furthermore, FARM uses the *weight of evidence* measure [2] as a confidence metric for FARs.

In the proposal of Kuok et al. [28], to each attribute is assigned to a set of fuzzy sets. These fuzzy sets and the corresponding membership functions are provided by human experts. The interestingness of a rule is assessed by means of two measures called, respectively *significance* and *certainty* factor. Indeed, if a rule is interesting, then it should present enough significance as well as a high certainty factor. The significance factor is a generalization of support and computed by summing all the votes<sup>(1)</sup> of each record with respect to the specified itemset and then dividing it by the total number of records. Two different methods were proposed to compute the certainty factor of a fuzzy rule. The first one uses the significance factor, thus the certainty factor is a generalization of the confidence. Whereas, the second method is based on correlation measure.

In He et al. [24], the authors proposed a novel adaptive algorithm for mining FARs. The algorithm aims at building an effective *Decision Support System* for binary classification problems in the biomedical domain. In fact, it combines data clustering techniques with fuzzy interval partitions on input features. Doing so,

---

<sup>1</sup>The vote of a record is computed as the product of membership grade of each item in that record.

high-level data abstraction is extracted and the quantitative data can be efficiently transformed into fuzzy discrete transactions. On the latter, the traditional Apriori algorithm is applied to mine association rules that can be used for classification and decision support.

The proposal of Delgado et al. [16] for mining FARs is based on "*fuzzy transactions*". The main characteristic of this proposal is to model fuzzy transactions with crisp items. The support of an itemset and the support and the confidence of a fuzzy association rule is assessed by the evaluation of *quantified sentences*. The authors propose an alternative measure to confidence which is *certainty factor*.

In Ben Yahia and Jaoua,[8], a different definition of FARs was proposed. It is based on fuzzy degrees, which are respectively associated to items. It does not necessarily employ linguistic terms. An item  $i$  having a degree  $\alpha$  is present in each transaction  $\tau$  when  $\tau(i) \geq \alpha$ . A fuzzy *itemset*  $\tilde{I}$  is a set of items associated to their fuzzy degrees. The definitions of the support and the confidence of a FAR are similar to those of the crisp case. An efficient algorithm for mining FARs, based on the pruning of the "fuzzy concept lattice" was proposed in [8]. Recently, Helm proposed an implementation of a mining FARs algorithm [25]. The author adopted the FP-GROWTH algorithm to deal with fuzzy data.

Recently and as an extension of their former approach [11], Weng and Chen [12] introduced a new Apriori-like algorithm, called UDM, for extracting fuzzy association rules using the theory of possibility to represent uncertain data. The theory of possibility was of use to represent uncertain data. An additional statistical metric, called *deviation*, was introduced towards a further pruning and a withdrawal of uncertain fuzzy association rules. Experiments showed that the proposed algorithm "increases linearly with respect to the database size". Unfortunately, the latter approach is another one that is introduced regardless the effectiveness of the mined knowledge. In fact, the authors paid little attention to the amount of association rules that may be drawn. Hence, at the end of the process, the user faces an overwhelming quantity of association rules among which a large number is redundant, what badly affects the quality. Thus the effort is padded out in optimizing approaches that will generate "useless" knowledge for end-users.

In table 1, we report a comparative survey of the different approaches mentioned above. For this purpose, we consider as criteria of comparison the following features:

- **Transformation of original attributes:** In all propositions illustrated above a transformation of original context attributes was performed before beginning the FARs extraction process.

- **FARs measure assessment:** Every approach has identified its own measure to assess FARs.
- **Number of FARs:** For the same context the number of generated fuzzy association rules differs from one approach to another. In fact, this number varies, respectively with the number of intervals and the number of fuzzy sets assigned to each attribute according, respectively to Agrawal *et al.* [30] and Kuok *et al.* [28]. In the propositions of Chan and Au [13] and Delgado *et al.* [16], the number of generated FARs varies with the number of linguistic terms associated to each attribute. Finally, the number of generated FARs varies with the number of formal concepts generated with the FARD algorithm [8].

	<b>Transformation of the original attributes</b>	<b>FARs measure assessment</b>	<b>Number of FARs</b>
<b>Srikant and Agrawal [30] (Discretization)</b>	(attribute-intervals)	Support, confidence	varies with the number of intervals
<b>Kuok <i>et al.</i> [28]</b>	(attribute-fuzzy sets)	Significance, certainty factors	varies with the number of fuzzy sets
<b>Chan and Au [13] (FARM)</b>	(attribute-linguistic terms)	Weight of evidence, Adjusted difference	varies with the number of linguistic terms
<b>Delgado <i>et al.</i> [16]</b>	(attribute-linguistic labels)	Fuzzy quantifiers evaluation	varies with linguistic terms
<b>Ben Yahia and Jaoua [8] (FARD)</b>	-	Support, confidence	varies with the number of fuzzy formal concepts

Table 1: Comparative survey of the different approaches.

**Remark 1.** *In order to enable the evaluation of a fuzzy association rule, several measures were proposed in the literature. According to [18], the standard approach is to replace set-theoretic operations, namely Cartesian product and cardinality, by corresponding fuzzy set-theoretic operations. Modeling the Cartesian*

product by a *t*-norm<sup>2</sup>. and the cardinality of a fuzzy set by the sum of the values of its membership function

It is noteworthy that all above cited propositions were not interested in reducing the huge number of FARs that may be drawn even from small contexts. They were only limited to define the adopted approach to generate FARs. In this paper, we thus put the focus on FARs, as previously defined in [8]. The main thrust is to solve the problem of the huge amount of rules discovered from databases containing fuzzy attributes. As quite expectable, this amount of knowledge seems to be overwhelming and beyond human capabilities to analyze. Thus, in the remainder, we propose a new algorithm for extracting generic FARs from frequent fuzzy closed itemsets *w.r.t* the user's preferences.

### 3. Generic basis of FARs

To cope with the problem of the overwhelming number of binary association rules that can often be extracted from even reasonably sized real-life databases, several solutions towards an information lossless reduction were proposed. These reduced association rules are commonly referred to as "generic basis". A solution consists in using the battery of results of the FCA (*Formal Concept Analysis*) [21] to define generic basis of association rules. Several studies focused on extracting such generic basis [3, 7, 26, 27].

To the best of our knowledge, no previous study in the literature paid attention to define a generic basis of FARs. However, the extension of Galois connection from the point of view of fuzzy logic grasped the interest (see *e.g.* [6] for further information and references).

#### 3.1. Characterization of the search space

In the following section, we present the fundamental properties of the search space induced by a fuzzy extraction context. Along all this section,  $J$  stands for

---

<sup>2</sup>A triangular norm *t*-norm is a function  $\top : [0, 1] \times [0, 1] \rightarrow [0, 1]$  verifying,  $\forall x$  and  $y \in [0, 1]$ , these following properties:

- $\top$  is commutative  $\top(x, y) = \top(y, x)$ ,
- $\top$  is associative  $\top(x, \top(y, z)) = \top(\top(x, y), z)$ ,
- $\top$  is increasing  $\top(x, y) \leq \top(z, t)$  if  $x \leq z$  and  $y \leq t$ ,
- $\top(x, 1) = x$ .



any set of indices.

**Definition 1. User's constraint**

A user's constraint is a fuzzy subset  $\tilde{C}$  in a universe of discourse  $\mathcal{I}$  (i.e., finite set of attributes) characterized by the membership function  $\mu_{\tilde{C}}: \mathcal{I} \rightarrow [0, 1]$ . The fuzzy subset  $\tilde{C}$  is denoted by:

$$\tilde{C} = \{i_1^{\mu_{\tilde{C}}(i_1)}, i_2^{\mu_{\tilde{C}}(i_2)}, \dots, i_n^{\mu_{\tilde{C}}(i_n)}\}$$

**Definition 2. Fuzzy extraction context under constraint**

A fuzzy formal context under constraint is a quadruplet  $\mathcal{K} = (\mathcal{O}, \tilde{\mathcal{I}}, \tilde{R}, \tilde{C})$  describing a finite set  $\mathcal{O}$  of objects (or transactions), a fuzzy finite set  $\tilde{\mathcal{I}}$  of attributes (or items), a fuzzy binary relation  $\tilde{R}$  (i.e.,  $\tilde{R} \subseteq \mathcal{O} \times \tilde{\mathcal{I}}$ ) and a fuzzy finite set of attributes  $\tilde{C}$  (i.e., the user's constraint).

Each pair  $(o, i^\alpha) \in \tilde{R}$ , means that the attribute (item)  $i$  belonging to  $\tilde{\mathcal{I}}$  is contained in the object  $O$  belonging to  $\mathcal{O}$  with the degree  $\alpha$ . The value  $\alpha$  has to be greater than or equal to a minimal threshold given by  $\mu_{\tilde{C}}(i)$ .

In order to obtain a fuzzy extraction context, it is possible to fuzzify a quantitative database by normalizing each item quantity. A fuzzy extraction context can also be an indexing relation between documents and terms, in which membership values are obtained by considering only the local frequency of the term  $t_i$  in the document  $d_j$ , normalized with respect to the maximal frequency of a term  $t_k$  belonging to the same document  $d_j$ .

Within the framework of basket (binary) association rules, all the products are treated uniformly, and all the rules are mined based on the counts of the itemsets. However, in the social science research, the analysts may be interested in highlighting the importance of the products, items or attributes in the minable knowledge. For example, *total income* attribute is more interesting than the *height* of a person in a household.

In the remainder, we tackle this issue by considering the case where items are given weights to reflect their importance for the users. Hence, an extracted generic FAR has to fulfill the user's constraint, i.e., this constraint is a materialization of the importance assignment to attributes. Therefore, each item in a fuzzy association rule must have a membership degree at least equal to the corresponding degree indicated by the underlying constraint.

$\tilde{R}$	B	C	E	M
1	0.5	1.0	0.7	0.5
2	0.6	0.7	1.0	0.5
3	1.0	0.9	1.0	0.1
4	1.0	0.9	0.9	0.1
$\tilde{C}$	<b>0.0</b>	<b>0.2</b>	<b>0.7</b>	<b>0.1</b>

Figure 1: Fuzzy formal context under constraint  $\tilde{C}$

**Example 1.** An example of a fuzzy formal context under constraint  $\mathcal{K} = (\mathcal{O}, \tilde{\mathcal{I}}, \tilde{R}, \tilde{C})$  is sketched by Figure 1. According to the constraint  $\tilde{C}$ , we are interested in extracting FARs such that the items "B", "C", "E" and "M" must appear with degrees, respectively, greater than or equal to 0.0, 0.2, 0.7 and 0.1.

### 3.1.1. Mapping fuzzy to crisp

In the following, we show how fuzzy context can be mapped into a crisp one. It is important to note that the fuzzy context introduces a peculiarity, with respect to the crisp case, that the extraction context is not fixed "a priori". Indeed, for fuzzy attribute description we have to mine the fuzzy context and to extract different values associated to each attribute. With respect to a generalization order, these irreducible elements form a chain made up of elements corresponding to the set of distinct values associated to each attribute.

**Example 2.** From the fuzzy extraction context given by Figure 1, the chains associated to each attribute is depicted in Figure 2 (Up).

Therefore, under a generalization order view, a fuzzy context can be mapped into a crisp one as follows: Let  $A$  be an attribute and  $\text{Val}_A$  its associated distinct values. Then, the set  $A$  is mapped into set of distinct  $|\text{Val}_A|$  binary attributes  $A_i$ , such that  $A_i \leq A_j$   $i, j \in \text{Val}_A$  and  $i < j$ . The set of distinct "crisp" associated values of the context extracted is given by Figure 2 (Down).

Let  $\tilde{R}$  be a relation over a schema  $\mathcal{R} = \{a_1, a_2, \dots, a_n\}$ .  $\prod a_i$  stands for the projection of  $\tilde{R}$  over  $a_i$ .

The search space includes all valid combinations built up by considering the value sets of  $\mathcal{R}$ , and is defined as follows :  $\mathcal{S}(\tilde{R}) = (\times_{a_i \in \mathcal{R}} \prod a_i)$  where  $\times$  stands for the cartesian product.

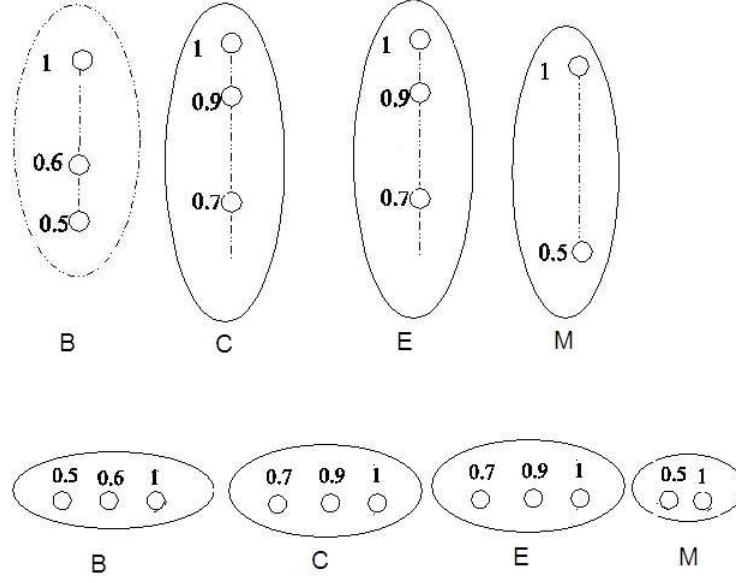


Figure 2: (**Up**)Attribute chains from the extraction context (**Down**) Crisp attributes from the extraction context

**Example 3.** If we consider a projection over the attributes  $\{B, C, E\}$  from the relation scheme associated to the context given by Figure 1, then Figure 3 depicts its associated search space.

The search space of  $\tilde{R}$  is structured under a generalization order, denoted  $\leq_g$ , between its descriptions.

**Definition 3.** Let  $\tilde{d}_1$  and  $\tilde{d}_2$  be two fuzzy descriptions of the search space  $\mathcal{S}(\tilde{R})$  :

$$\tilde{d}_1 \leq_g \tilde{d}_2 \Leftrightarrow \forall a_i \in \mathcal{R} \tilde{d}_1[a_i] \supseteq \tilde{d}_2[a_i]$$

The covering relation  $\prec_g$  associated to  $\leq_g$  is defined as follows:  $\forall \tilde{d}_1, \tilde{d}_2 \in \mathcal{S}(\mathcal{R}), \tilde{d}_1 \prec_g \tilde{d}_2 \Leftrightarrow \tilde{d}_1 <_g \tilde{d}_2$  and  $\forall \tilde{d}_3 \in \mathcal{S}(\mathcal{R}), \tilde{d}_1 \leq_g \tilde{d}_3 \leq_g \tilde{d}_2 \Rightarrow \tilde{d}_1 = \tilde{d}_2$ .

**Proposition 1.** Let  $Tr$  be the set of fuzzy descriptions of  $\tilde{R}$  ordered by the relation  $\leq_g$ .  $(Tr, \leq_g)$  is a complete and distributive lattice, with least element  $\langle 0, \dots, 0 \rangle$ , and greatest one  $\langle 1, \dots, 1 \rangle$ . The supremum ( $\bigvee$ ) and the in-

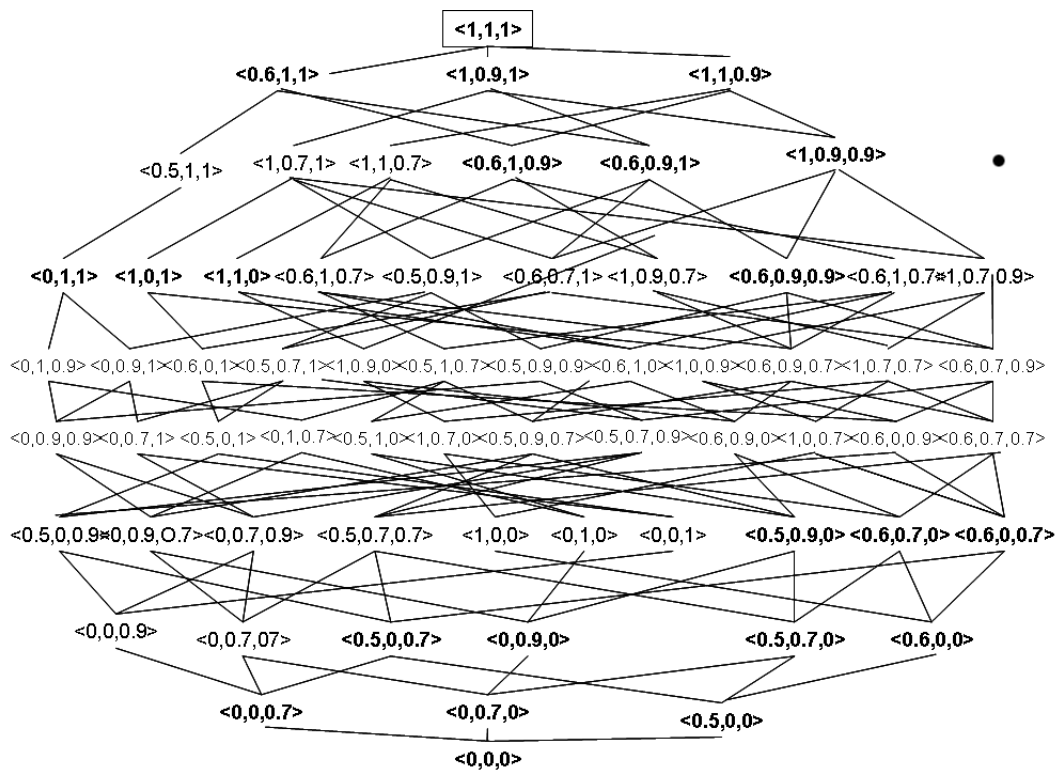


Figure 3: Search space associated to  $\tilde{R}$

imum ( $\bigwedge$ ) of any set of fuzzy descriptions of  $Tr$  are, respectively, as follows:

$$\bigvee_{j \in J} \tilde{X}_j = \bigcap_{j \in J} \tilde{X}_j \quad (1)$$

$$\bigwedge_{j \in J} \tilde{X}_j = \bigcup_{j \in J} \tilde{X}_j \quad (2)$$

**Proof 1.** First, let us show that any set of fuzzy rectangles of  $\tilde{R}$  has a least upper bound and a greatest lower one, which are both fuzzy rectangles of  $\tilde{R}$ . Then, we show that  $\bigvee$  is distributive relatively to  $\bigwedge$  and conversely.

- *Least upper Bound :*

$$\forall j \in J, (\tilde{X}_j) \leq_g (\tilde{Y})$$

$$\Leftrightarrow \forall j \in J, \tilde{Y} \subseteq \tilde{X}_j$$

$$\Leftrightarrow \tilde{Y} \subseteq \bigcap_{j \in J} \tilde{X}_j$$

$$\Leftrightarrow (\bigcap_{j \in J} \tilde{X}_j) \leq_g (\tilde{Y}).$$

Hence, any set of fuzzy description of  $Tr$  has a least upper bound, which is:  $\bigvee_{j \in J} (\tilde{X}_j) = (\bigcap_{j \in J} \tilde{X}_j)$ .

- *Greatest lower Bound :*

Let  $(\tilde{X}_j)$  be a fuzzy description of  $Tr$  :

$$\forall j \in J, (\tilde{Y}) \leq_g (\tilde{X}_j)$$

$$\Leftrightarrow \forall j \in J, \tilde{X}_j \subseteq \tilde{Y}$$

$$\Leftrightarrow \bigcup_{j \in J} \tilde{X}_j \subseteq \tilde{Y}$$

$$\Leftrightarrow (\tilde{Y}) \leq_g (\bigcup_{j \in J} \tilde{X}_j).$$

Hence, any set of fuzzy descriptions of  $Tr$  has a greatest lower bound, which is :  $\bigwedge_{j \in J} (\tilde{X}_j) = (\bigcup_{j \in J} \tilde{X}_j)$ .

- we could easily show that  $\forall X_1, X_2, X_3 \in Tr$   $X_1 \wedge (X_2 \vee X_3) = (X_1 \vee X_2) \wedge (X_1 \vee X_3)$ . Hence,  $\bigvee$  is distributive relatively to  $\bigwedge$  and conversely by using the corresponding properties and the duality of the operators  $\cup$  and  $\cap$ .

- According to the definition of  $\leq_g$ , we could easily show that  $\langle 0, \dots, 0 \rangle$  is the least element of  $(Tr, \leq_g)$  and that  $\langle 1, \dots, 1 \rangle$  is the greatest one.

### 3.1.2. Fuzzy Galois connection under constraint

**Definition 4.** Let  $\mathcal{K}_{\tilde{C}} = (\mathcal{O}, \tilde{\mathcal{I}}, \tilde{R}, \tilde{C})$  be a fuzzy formal context under constraint. For  $O \subseteq \mathcal{O}$  and  $\tilde{I} \subseteq \tilde{\mathcal{I}}$ , we define:

$$\begin{aligned} \tilde{f}_{\tilde{C}} &: P(\mathcal{O}) \rightarrow P(\tilde{\mathcal{I}}) \\ \tilde{f}_{\tilde{C}}(O) &= \{d^\alpha \mid \forall o \in O, \alpha = \min \mu_{\tilde{R}}(o, d) \wedge \alpha \geq \mu_{\tilde{C}}(d), d \in \tilde{\mathcal{I}}\} \end{aligned}$$

$$\begin{aligned} \tilde{g}_{\tilde{C}} &: P(\tilde{\mathcal{I}}) \rightarrow P(\mathcal{O}) \\ \tilde{g}_{\tilde{C}}(\tilde{I}) &= \{o \in \mathcal{O} \mid \forall d \in \tilde{I}, [(\mu_{\tilde{I}}(d) \xrightarrow{I_{RG}} \mu_{\tilde{R}}(o, d)) = 1 \wedge (\mu_{\tilde{C}}(d) \xrightarrow{I_{RG}} \mu_{\tilde{R}}(o, d)) = 1]\} \end{aligned}$$

In the above definition  $I_{RG}$  denotes the *Rescher-Gaines* fuzzy implication, which is an *R-implication* [5, 19]. This implies that:

$$\tilde{g}_{\tilde{C}}(\tilde{I}) = \{o \in \mathcal{O} \mid \forall d \in \tilde{I}, \mu_{\tilde{I}}(d) \leq \mu_{\tilde{R}}(o, d) \wedge \mu_{\tilde{C}}(d) \leq \mu_{\tilde{R}}(o, d)\}.$$

$\tilde{f}_{\tilde{C}}$  and  $\tilde{g}_{\tilde{C}}$  are defined respectively over the power sets of  $\mathcal{O}$  and  $\tilde{\mathcal{I}}$  (i.e.,  $P(\mathcal{O})$  and  $P(\tilde{\mathcal{I}})$ ).

The fuzzy operator  $\tilde{f}_{\tilde{C}}$  is applied on a crisp set of objects and determines to which degree each property is satisfied by all the objects, according to their respective degrees. Note that  $\tilde{f}_{\tilde{C}}$ , as defined formerly, presents the 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 fulfilling the constraint  $\tilde{C}$ .

On the other hand, the  $\tilde{g}_{\tilde{C}}(\tilde{I})$ , applied on a fuzzy set of attributes, permits to obtain the desired *interpretation* effect by filtering objects presenting more general descriptions than that of  $\tilde{I}$  as well as the user's constraint  $\tilde{C}$ .

**Example 4.** Let us consider the constrained fuzzy context illustrated in Figure 1. Thus, we have :

$$\tilde{f}_{\tilde{C}}(\{2, 3, 4\}) = B^{0.6} C^{0.7} E^{0.9} M^{0.1} \text{ and } \tilde{g}_{\tilde{C}}(B^{0.6} C^{0.9}) = \{3, 4\}.$$

**Proposition 2.** The composite operators  $\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}$  and  $\tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}$  are called *fuzzy constrained Galois closure operator*. Indeed, considering both functions ( $\tilde{f}_{\tilde{C}}$  and  $\tilde{g}_{\tilde{C}}$ ) as previously defined, the following properties hold  $\forall \tilde{I}, \tilde{I}_i, \tilde{I}_j \in \tilde{\mathcal{I}}$ :

$$(A1) \quad O_i \subseteq O_j \Rightarrow \tilde{f}_{\tilde{C}}(O_i) \supseteq \tilde{f}_{\tilde{C}}(O_j)$$

$$(B1) \quad \tilde{I}_i \subseteq \tilde{I}_j \Rightarrow \tilde{g}_{\tilde{C}}(\tilde{I}_i) \supseteq \tilde{g}_{\tilde{C}}(\tilde{I}_j)$$

$$(A2) \quad O \subseteq \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O)$$

$$(B2) \quad \tilde{I} \subseteq \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I})$$

$$(A3) \quad \tilde{f}_{\tilde{C}}(O) = \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O)$$

$$(B3) \quad \tilde{g}_{\tilde{C}}(\tilde{I}) = \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I})$$

$$(A4) \quad O_i \subseteq O_j \Rightarrow \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O_i) \subseteq \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O_j)$$

$$(B4) \quad \tilde{I}_i \subseteq \tilde{I}_j \Rightarrow \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I}_i) \subseteq \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I}_j)$$

$$(A5) \quad \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(\tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O)) = \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O)$$

$$(B5) \quad \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I})) = \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I})$$

**Proof 2.** In a dual manner, these properties are valid for the fuzzy Galois closure operators  $\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}$  and  $\tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}$ .

(A1) We have :

$$\begin{aligned} \tilde{f}_{\tilde{C}}(O_i) &= \{d^{\alpha_i} \mid \forall o \in O, \alpha_i = \min \mu_{\tilde{R}}(o, d) \wedge \alpha_i \geq \mu_{\tilde{C}}(d), d \in \tilde{I}\} \text{ and} \\ \tilde{f}_{\tilde{C}}(O_j) &= \{d^{\alpha_j} \mid \forall o \in O, \alpha_j = \min \mu_{\tilde{R}}(o, d) \wedge \alpha_j \geq \mu_{\tilde{C}}(d), d \in \tilde{I}\} \\ \text{If } O_i \subseteq O_j &\Rightarrow \alpha_i \geq \alpha_j. \text{ Thus, } \tilde{f}_{\tilde{C}}(O_i) \supseteq \tilde{f}_{\tilde{C}}(O_j). \end{aligned}$$

**(B1)** We have :

$$\begin{aligned}\tilde{g}_{\tilde{C}}(\tilde{I}_i) &= \{o_i \in O \mid \forall d \in \tilde{I}_i, \mu_{\tilde{I}_i}(d) \leq \mu_{\tilde{R}}(o_i, d) \wedge \mu_{\tilde{C}}(d) \leq \mu_{\tilde{R}}(o_i, d)\} \text{ and} \\ \tilde{g}_{\tilde{C}}(\tilde{I}_j) &= \{o_j \in O \mid \forall d \in \tilde{I}_j, \mu_{\tilde{I}_j}(d) \leq \mu_{\tilde{R}}(o_j, d) \wedge \mu_{\tilde{C}}(d) \leq \mu_{\tilde{R}}(o_j, d)\} \\ \text{If } \tilde{I}_i \subseteq \tilde{I}_j &\Rightarrow \mu_{\tilde{I}_i}(d) \leq \mu_{\tilde{I}_j}(d) \quad (1)\end{aligned}$$

$$\text{If } o_j \in \tilde{g}_{\tilde{C}}(\tilde{I}_j) \Rightarrow \begin{cases} \mu_{\tilde{I}_j}(d) \leq \mu_{\tilde{R}}(o_j, d) \\ \mu_{\tilde{C}}(d) \leq \mu_{\tilde{R}}(o_j, d) \end{cases} \quad (2) \text{ According to (1) and (2), we}$$

have  $\mu_{\tilde{R}}(o_j, d) \geq \mu_{\tilde{I}_j}(d) \geq \mu_{\tilde{I}_i}(d)$  and  $o_j$  satisfies the constraint  $\tilde{C}$ . Hence,  $o_j \in \tilde{g}_{\tilde{C}}(\tilde{I}_i)$ . Therefore,  $\tilde{g}_{\tilde{C}}(\tilde{I}_i) \supseteq \tilde{g}_{\tilde{C}}(\tilde{I}_j)$ .

**(A2)** We have :

$$\begin{aligned}\tilde{f}_{\tilde{C}}(O) &= \{d^\alpha \mid \forall o, o \in O, \alpha = \min \mu_{\tilde{R}}(o, d) \wedge \alpha \geq \mu_{\tilde{C}}(d), d \in \tilde{I}\}, \\ \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O) &= \{o \in O \mid \forall d \in \tilde{f}_{\tilde{C}}(O), \mu_{\tilde{R}}(o, d) \geq \min \mu_{\tilde{R}}(o, d) \wedge \mu_{\tilde{C}}(d) \leq \mu_{\tilde{R}}(o, d)\}. \text{ Obviously, if } o \in O \Rightarrow o \in \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O) \Rightarrow O \subseteq \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O).\end{aligned}$$

**(B2)** Let  $d \in \tilde{I}$  with  $\mu_{\tilde{I}}(d)$ , We have :

$$\begin{aligned}\tilde{g}_{\tilde{C}}(\tilde{I}) &= \{o \in O \mid \forall d \in \tilde{I}, \mu_{\tilde{I}}(d) \leq \mu_{\tilde{R}}(o, d) \wedge \mu_{\tilde{C}}(d) \leq \mu_{\tilde{R}}(o, d)\} \quad (3) \\ \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I}) &= \{d^\alpha \mid \forall o, o \in \tilde{g}_{\tilde{C}}(\tilde{I}), \alpha = \min \mu_{\tilde{R}}(o, d) \wedge \alpha \geq \mu_{\tilde{C}}(d), d \in \tilde{I}\} \quad (4)\end{aligned}$$

From (3) and (4),  $\alpha \geq \mu_{\tilde{I}}(d) \Rightarrow \tilde{I} \subseteq \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I})$ .

**(A3)** Let  $\tilde{I} = \tilde{f}_{\tilde{C}}(O)$ , according to **(B2)**, we have:

$$\tilde{f}_{\tilde{C}}(O) \subseteq \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O) \quad (5)$$

And from **(A1)**  $O \subseteq \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O)$ , we have:

$$\tilde{f}_{\tilde{C}}(O) \supseteq \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O) \quad (6)$$

According to (5) and (6), we get  $\tilde{f}_{\tilde{C}}(O) = \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O)$ .

**(B3)** Let  $O = \tilde{g}_{\tilde{C}}(\tilde{I})$ , according to **(A2)**, we have :

$$\tilde{g}_{\tilde{C}}(\tilde{I}) \subseteq \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I}) \quad (7)$$

We have:  $\tilde{I} \subseteq \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I})$ , from **(B1)**, we have :

$$\tilde{g}_{\tilde{C}}(\tilde{I}) \supseteq \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I}) \quad (8)$$

According to (7) and (8), we get  $\tilde{g}_{\tilde{C}}(\tilde{I}) = \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I})$ .



(A4) We have from (A1) :

If  $O_i \subseteq O_j$  then  $\tilde{f}_{\tilde{C}}(O_i) \supseteq \tilde{f}_{\tilde{C}}(O_j)$ . Therefore,  $\tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O_i) \subseteq \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O_j)$ .

(B4) We have from (B1) :

If  $\tilde{I}_i \subseteq \tilde{I}_j$  then  $\tilde{g}_{\tilde{C}}(\tilde{I}_i) \supseteq \tilde{g}_{\tilde{C}}(\tilde{I}_j)$ . Therefore,  $\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I}_i) \subseteq \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I}_j)$ .

(A5) We have from (A3) :

$\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{f}_{\tilde{C}}(O)) = \tilde{f}_{\tilde{C}}(O)$ , hence  $\tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(\tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O)) = \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(O)$ .

(B5) We have from (B3) :

$$\begin{aligned} \tilde{g}_{\tilde{C}} \circ \tilde{f}_{\tilde{C}}(\tilde{g}_{\tilde{C}}(\tilde{I})) &= \tilde{g}_{\tilde{C}}(\tilde{I}) \\ \Rightarrow \tilde{g}_{\tilde{C}}(\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I})) &= \tilde{g}_{\tilde{C}}(\tilde{I}) \\ \Rightarrow \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I})) &= \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I}). \end{aligned}$$

**Remark 2.** It is noteworthy that in case we omit the user's constraint (i.e., membership degrees of the items in  $\tilde{C}$  are fixed to zero) or transforming beforehand the extraction context by assigning the value 0 to each  $(o_j, a_i)$  whenever  $o_j[a_i] < \mu_{\tilde{C}}(i)$ , we recover the same definitions of a fuzzy formal context and fuzzy Galois connection operators as they were previously defined by Ben Yahia and Jaoua [8].

**Definition 5. Fuzzy formal concept** The pair  $(O, \tilde{I})$ , such that  $O \in \mathcal{O}$  and  $\tilde{I} \in \tilde{\mathcal{I}}$ , is called a fuzzy formal concept if and only if  $\tilde{f}_{\tilde{C}}(O) = \tilde{I}$  and  $\tilde{g}_{\tilde{C}}(\tilde{I}) = O$ .  $O$  is called the extension and  $\tilde{I}$  the intension of the fuzzy formal concept.  $\tilde{g}_{\tilde{C}}(\tilde{I})$  is called the domain of  $\tilde{I}$ . A set of fuzzy formal concepts  $\mathcal{FC}_{\tilde{K}}$ , extracted from a fuzzy formal constrained context  $\mathcal{K}_{\tilde{C}}$  and ordered using the set inclusion relation, form a complete fuzzy lattice  $\mathcal{L}_{\mathcal{FC}_{\tilde{K}}} = (\mathcal{FC}_{\tilde{K}}, \subseteq)$ , called fuzzy Galois lattice.

**Definition 6. Fuzzy closed itemset** Let us consider the fuzzy context under constraint  $\mathcal{K}_{\tilde{C}} = (\mathcal{O}, \tilde{\mathcal{I}}, \tilde{\mathcal{R}}, \tilde{\mathcal{C}})$ , a fuzzy subset  $\tilde{I} \subseteq \tilde{\mathcal{I}}$ ,  $\tilde{I}$  is called fuzzy closed itemset if and only if it is equal to its closure, i.e.,  $\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{I}) = \tilde{I}$ ,  $\tilde{I}$  is said to be frequent with respect to the minimum threshold if  $\text{support}(\tilde{I}) = \frac{|\tilde{g}_{\tilde{C}}(\tilde{I})|}{|\tilde{\mathcal{I}}|} \geq \text{minsup}$ .

**Definition 7. Fuzzy Iceberg Concept lattice** When only fuzzy frequent closed itemsets are considered with the set inclusion relation  $\subseteq$ , the resulting structure only preserves the join operator. This structure forms an upper semi-lattice and it is designated by "Fuzzy Iceberg Concept lattice".

**Definition 8. Fuzzy minimal generator** A fuzzy itemset  $\tilde{c}$  is said to be fuzzy minimal generator of a fuzzy closed itemset  $\tilde{I}$ , if and only if  $\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{c}) = \tilde{I}$  and  $\nexists \tilde{c}_1 \subset \tilde{c}$  such that  $\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{c}_1) = \tilde{I}$ . The set  $\mathcal{GMF}_{\tilde{I}}$  of the fuzzy minimal generators of a fuzzy closed itemset  $\tilde{I}$  is defined as follows:  
 $\mathcal{GMF}_{\tilde{I}} = \{\tilde{c} \subseteq \tilde{I} \mid \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{c}) = \tilde{I} \wedge \nexists \tilde{c}_1 \subset \tilde{c} \text{ s.t. } \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(\tilde{c}_1) = \tilde{I}\}.$

**Definition 9. Fuzzy equivalence class** The fuzzy constrained Galois closure operators  $(\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}})$  induces an equivalence relation on the power set of  $\tilde{I}$ , i.e., the set of parts are split into disjoint subsets, called fuzzy equivalence classes. In each class, all elements have the same value of support since they share the same closure. Within a given fuzzy equivalence class, minimal generators are the smallest incomparable elements (w.r.t set inclusion relation), while the fuzzy closed itemset is the largest element.

**Example 5.** Let us consider the constrained fuzzy context given by Figure 1. For a  $\text{minsup} = \frac{1}{4}$ , the set of fuzzy minimal generators as well as their domains and respective fuzzy closures are given in table 6. The associated fuzzy Iceberg concept lattice is depicted by Figure 4.

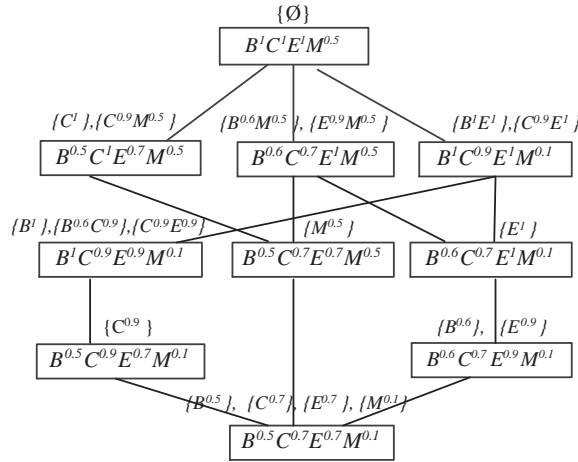


Figure 4: Fuzzy Iceberg Concept lattice associated to the constrained fuzzy context  $\mathcal{K}_{\tilde{C}}$  depicted in Figure 1, for  $\text{minsup} = \frac{1}{4}$ .

In order to extract fuzzy closed itemsets and their corresponding fuzzy minimal generators (i.e., that will be of use during the FARs extraction process), we introduce an algorithm, called GEN-IFF, which is detailed below.

Table 2: List of fuzzy minimal generators and their corresponding fuzzy closed itemsets extracted from the constrained fuzzy context  $\mathcal{K}_{\tilde{C}}$ , illustrated by Figure 1 for  $minsup = \frac{1}{4}$ .

Fuzz.min.gen	Domain	closure	Support
$B^1$	{3, 4}	$B^1 C^{0.9} E^{0.9} M^{0.1}$	$\frac{1}{2}$
$B^{0.6}$	{2, 3, 4}	$B^{0.6} C^{0.7} E^{0.9} M^{0.1}$	$\frac{3}{4}$
$B^{0.5}$	{1, 2, 3, 4}	$B^{0.5} C^{0.7} E^{0.7} M^{0.1}$	1
$C^1$	{1}	$B^{0.5} C^1 E^{0.7} M^{0.5}$	$\frac{1}{4}$
$C^{0.9}$	{1, 3, 4}	$B^{0.5} C^{0.9} E^{0.7} M^{0.1}$	$\frac{3}{4}$
$C^{0.7}$	{1, 2, 3, 4}	$B^{0.5} C^{0.7} E^{0.7} M^{0.1}$	1
$E^1$	{2, 3}	$B^{0.6} C^{0.7} E^{1.0} M^{0.1}$	$\frac{3}{4}$
$E^{0.9}$	{2, 3, 4}	$B^{0.6} C^{0.7} E^{0.9} M^{0.1}$	$\frac{3}{4}$
$E^{0.7}$	{1, 2, 3, 4}	$B^{0.5} C^{0.7} E^{0.7} M^{0.1}$	1
$M^{0.5}$	{1, 2}	$B^{0.5} C^{0.7} E^{0.7} M^{0.5}$	$\frac{1}{2}$
$M^{0.1}$	{1, 2, 3, 4}	$B^{0.5} C^{0.7} E^{0.7} M^{0.1}$	1
$B^{0.6} C^{0.9}$	{3, 4}	$B^1 C^{0.9} E^{0.9} M^{0.1}$	$\frac{1}{2}$
$B^1 E^1$	{3}	$B^1 C^{0.9} E^1 M^{0.1}$	$\frac{1}{4}$
$B^{0.6} M^{0.5}$	{2}	$B^{0.6} C^{0.7} E^1 M^{0.5}$	$\frac{1}{4}$
$C^{0.9} E^1$	{3}	$B^1 C^{0.9} E^1 M^{0.1}$	$\frac{1}{4}$
$C^{0.9} E^{0.9}$	{3, 4}	$B^1 C^{0.9} E^{0.9} M^{0.1}$	$\frac{1}{2}$
$C^{0.9} M^{0.5}$	{1}	$B^{0.5} C^1 E^{0.7} M^{0.5}$	$\frac{1}{4}$
$E^{0.9} M^{0.5}$	{2}	$B^{0.6} C^{0.7} E^1 M^{0.5}$	$\frac{1}{4}$

#### 4. GEN-IFF: Fuzzy closed itemsets extraction algorithm

GEN-IFF is an iterative algorithm taking as input a constrained fuzzy formal context and the *minsup* value. As a result, the algorithm outputs the set of all fuzzy closed itemsets and their corresponding fuzzy minimal generators. The algorithm proceeds in two steps as described below **Finding all frequent fuzzy minimal generators fulfilling the *minsup* threshold and the user's constraint:** The GEN-IFF algorithm starts by determining the set of candidate fuzzy minimal generators denoted by  $\mathcal{GMFC}_k$  and their respective domains. It adopts a levelwise search starting by finding the set of all candidate 1-sized fuzzy minimal generators  $\mathcal{GMFC}_1$ . This set will be pruned with respect to both the *minsup* threshold and the user's constraint. This pruning step is carried out after the application of the GET-DOMAIN function on  $\mathcal{GMFC}_1$  set (line 3). The GET-DOMAIN function consists in applying the  $\tilde{g}_{\tilde{C}}$  operator on the  $\mathcal{GMFC}_1$  set. The domain cardinality of each 1-fuzzy minimal generator is compared to *minsup* value. If it is greater than or equal to *minsup*, then the fuzzy minimal generator is potentially frequent. If the 1-fuzzy minimal generator fulfils the user's constraint then it will be inserted in the  $\mathcal{GMFF}_1$  set (line 8). In order to find the  $\mathcal{GMFC}_k$  set, the subroutine GEN-NEXT is invoked (line 11). GEN-NEXT considers frequent fuzzy minimal generators of the previous iteration as an input and returns frequent fuzzy generators with one more item. Indeed, it joins frequent  $k$ -fuzzy generators together, forming candidate  $(k + 1)$ -fuzzy generators (*i.e.*, the join step). After that, a pruning step will remove any fuzzy generator whose fuzzy subsets have not been part of the discovered sets during the previous iterations. The following step *i.e.*, the domain's computing step, will compute the domain of all frequent  $(k+1)$ -fuzzy minimal generators by calling the GET-DOMAIN2 function. This function computes the domain of a  $(k + 1)$ -fuzzy generator which is the result of a two  $k$ -fuzzy generators combination. This will be achieved without the need to scan the original context any more. This step is based on proposition 3.

After that, the candidate fuzzy generators set is pruned with respect to the *minsup* threshold as well as to the structural property of a fuzzy minimal generator, *i.e.*, for any  $(k + 1)$ -fuzzy generator, we have to make sure that there is no frequent  $k$ -fuzzy generator, having the same domain. If such a case exists, this  $(k + 1)$ -fuzzy generator will not be inserted in the  $\mathcal{GMFF}_{(k+1)}$  set.

**Proposition 3.**  $\tilde{g}_{\tilde{C}}(\tilde{I}_1 \cup \tilde{I}_2) = \tilde{g}_{\tilde{C}}(\tilde{I}_1) \cap \tilde{g}_{\tilde{C}}(\tilde{I}_2), \forall \tilde{I}_1, \tilde{I}_2 \in \tilde{I}$ .

**Proof 3.** Let  $\tilde{I}_3 = \tilde{I}_1 \cup \tilde{I}_2$ , we have:

$$\begin{aligned}
\tilde{g}_{\tilde{C}}(\tilde{I}_3) &= \{g | \forall d \in \tilde{I}_3, \implies \mu_{\tilde{R}}(g, d) \geq \mu_{\tilde{I}_3}(d) \wedge \mu_{\tilde{C}}(d) \leq \mu_{\tilde{R}}(g, d)\} \\
&= \{g | \forall d \in \tilde{I}_3, \implies \mu_{\tilde{R}}(g, d) \geq \mu_{\tilde{I}_1 \cup \tilde{I}_2}(d) \wedge \mu_{\tilde{C}}(d) \leq \mu_{\tilde{R}}(g, d)\} \\
&= \{g | \forall d \in \tilde{I}_3, \implies \mu_{\tilde{R}}(g, d) \geq \max(\mu_{\tilde{I}_1}(d), \mu_{\tilde{I}_2}(d)) \wedge \mu_{\tilde{C}}(d) \leq \mu_{\tilde{R}}(g, d)\} \\
&= \{g | \forall d \in \tilde{I}_3, \implies \mu_{\tilde{R}}(g, d) \geq \mu_{\tilde{I}_1}(d) \text{ and } \mu_{\tilde{R}}(g, d) \geq \mu_{\tilde{I}_2}(d) \wedge \mu_{\tilde{C}}(d) \leq \\
&\mu_{\tilde{R}}(g, d)\} \\
&= \{g | \forall d \in \tilde{I}_3, \implies \mu_{\tilde{R}}(g, d) \geq \mu_{\tilde{I}_1}(d) \wedge \mu_{\tilde{C}}(d) \leq \mu_{\tilde{R}}(g, d)\} \text{ and } \{g | \forall d \in \\
&\tilde{I}_3, \implies \mu_{\tilde{R}}(g, d) \geq \mu_{\tilde{I}_2}(d) \wedge \mu_{\tilde{C}}(d) \leq \mu_{\tilde{R}}(g, d)\} \\
&= \tilde{g}_{\tilde{C}}(\tilde{I}_1) \cap \tilde{g}_{\tilde{C}}(\tilde{I}_2)
\end{aligned}$$

**Generating fuzzy closed itemsets:** This step consists in applying the  $\tilde{f}_{\tilde{C}}$  operator on the  $\mathcal{GMFF}$  set (line 12).

The pseudo-code and the notations used in the GEN-IFF algorithm are, respectively presented in algorithm 1 and table 3. The GEN-NEXT procedure is given by algorithm 2.

$\mathcal{K}_{\tilde{C}}$	: Constrained fuzzy formal context.
$\mathcal{GMFC}_k$	: Set of $k$ -fuzzy minimal generators.
$\mathcal{GMFF}_k$	: Set of $k$ -frequent fuzzy minimal generators ( <i>i.e.</i> , having a support greater than or equal to <i>minsup</i> and fulfilling the constraint $\tilde{C}$ ).
$\mathcal{GMFF}$	: Set of all frequent fuzzy minimal generators.
$\mathcal{IF}\tilde{\mathcal{F}}$	: Set of all fuzzy closed itemsets.
<i>minsup</i>	: Minimal support.

Table 3: Notations used in the GEN-IFF algorithm

**Input:**

- $\mathcal{K}_{\tilde{C}}$  : Fuzzy formal context under constraint
- $minsup$

**Output:**  $\mathcal{IF}\tilde{\mathcal{F}}$  : Fuzzy closed itemsets

**1 Begin**

2  $\mathcal{GMFC}_1 = \{1\text{-fuzzy itemsets}\};$

3  $\text{GET-DOMAIN}(\mathcal{GMFC}_1);$

4  $\mathcal{GMFF}_0 = \{\emptyset\}$

5  $\mathcal{GMFF} = \{\emptyset\}$

6 **foreach** ( $\tilde{g} \in \mathcal{GMFC}_1$ ) **do**

7     **if** ( $|\tilde{g}.dom| \geq minsup$ ) **and** ( $\tilde{g}.dom \subseteq \tilde{C}.dom$ ) **then**

8          $\mathcal{GMFF}_1 = \mathcal{GMFF}_1 \cup \tilde{g};$

9 **for** ( $k=1; \mathcal{GMFF}_k \neq \emptyset; k++$ ) **do**

10      $\mathcal{GMFF} = \mathcal{GMFF} \cup \mathcal{GMFF}_k;$

11      $\mathcal{GMFF}_{(k+1)} = \text{GEN-NEXT}(\mathcal{GMFF}_k);$

12  $\mathcal{IF}\tilde{\mathcal{F}} = \text{GEN-CLOSURE}(\mathcal{GMFF});$

13 **return**  $\mathcal{IF}\tilde{\mathcal{F}};$

**14 End**

**Algorithm 1:** GEN-IFF

1 **Procedure:** GEN-NEXT

**Input:**  $\mathcal{GMFF}_k$

**Output:**  $\mathcal{GMFF}_{(k+1)}$

2 **Begin**

```
3   /* Join step. */
4   INSERT INTO  $\mathcal{GMFC}_{(k+1)}$ 
5   SELECT  $\tilde{g}_1.item_1^{\alpha_1}, \tilde{g}_1.item_2^{\alpha_2}, \dots, \tilde{g}_1.item_i^{\alpha_i}, \tilde{g}_2.item_1^{\alpha_1}, \tilde{g}_2.item_2^{\alpha_2}, \dots,$ 
6    $\tilde{g}_2.item_i^{\alpha_i}$ 
7   FROM  $\mathcal{GMFF}_k \tilde{g}_1, \mathcal{GMFF}_k \tilde{g}_2$ 
8   WHERE  $\tilde{g}_1 \neq \tilde{g}_2,$ 
9          $\tilde{g}_1.item_1^{\alpha_1} = \tilde{g}_2.item_1^{\alpha_1},$ 
10         $\tilde{g}_1.item_2^{\alpha_2} = \tilde{g}_2.item_2^{\alpha_2},$ 
11        ...
12         $\tilde{g}_1.item_{i-1}^{\alpha_{i-1}} = \tilde{g}_2.item_{i-1}^{\alpha_{i-1}},$ 
13         $\tilde{g}_1.item_i^{\alpha_i} \neq \tilde{g}_2.item_i^{\alpha_i}$ 
14
15   /* Pruning Step. */
16   foreach ( $\tilde{g} \in \mathcal{GMFC}_{(k+1)}$ ) do
17     foreach ( $\tilde{g}_1$  s.t.  $|\tilde{g}_1| = k$  and  $\tilde{g}_1 \subset \tilde{g}$ ) do
18       if  $\tilde{g}_1 \notin \mathcal{GMFF}_k$  then
19          $\mathcal{GMFC}_{(k+1)} = \mathcal{GMFC}_{(k+1)} - \tilde{g}_1$ 
20         exit;
21
22   /* Domain computing step. */
23   foreach  $\tilde{g} \in \mathcal{GMFC}_{(k+1)}$  do
24     GET-DOMAIN2( $\tilde{g}$ );
25     if ( $|\tilde{g}.dom| \geq \text{minsup}$ ) and ( $\nexists \tilde{g}' \in \mathcal{GMFF}_k$  s.t. ( $\tilde{g}' \subset \tilde{g}$ ) and
26       ( $\tilde{g}.dom = \tilde{g}'.dom$ )) then
27        $\mathcal{GMFF}_{(k+1)} = \mathcal{GMFF}_{(k+1)} \cup \tilde{g};$ 
28
29   return  $\mathcal{GMFF}_{(k+1)}$ 
30 End
```

**Algorithm 2:** GEN-NEXT

**Example 6.** The GEN-IFF algorithm starts by determining the set of fuzzy candidate minimal 1-generators  $\mathcal{GMFC}_1 = \{B^1, B^{0,6}, B^{0,5}, C^1, C^{0,9}, C^{0,7}, E^1, E^{0,9}, E^{0,7}, M^{0,5}, M^{0,1}\}$ , as well as their associated domains. For  $\text{minsup} = \frac{1}{4}$  and setting values in  $\tilde{C}$  to 0, we find that  $\mathcal{GMFC}_1$  is equal to  $\mathcal{GMFF}_1$  (cf. Figure 5.a). Then, the GEN-NEXT procedure is invoked to obtain the set of fuzzy candidate minimal 2-generators  $\mathcal{GMFC}_2$  (cf. Figure 5.b) as well as their associated domains. The obtained set is pruned with respect to the property that has to fulfill any fuzzy minimal generator. Hence, the generator  $B^1 C^{0,9}$  is withdrawn, since it exists another generator  $B^{0,6} C^{0,9}$  such that  $B^{0,6} C^{0,9} \subset B^1 C^{0,9}$  and the domain of  $B^1 C^{0,9}$  is equal to that of  $B^{0,6} C^{0,9}$ , i.e., both generators share the same closure. The  $\mathcal{GMFF}_2$  set, shown in Figure 5.c, is obtained after the pruning of the  $\mathcal{GMFC}_2$  set. The process comes to an end at the third iteration, since the respective domains of all the fuzzy candidate fuzzy 3-generators are shown to be equal to the empty set (cf. Figure 5.d). Once the set of all frequent fuzzy minimal generators is outputted, the GEN-CLOSURE is invoked in order to compute the associated closures, by applying the  $\tilde{f}_{\tilde{C}}$  operator on the associated domain of the retained frequent fuzzy minimal generators. The obtained output is sketched by Figure 6.

In the sake of better performances for the GEN-IFF algorithm, we suggest storing the constrained fuzzy context (i.e., the input data of the algorithm) in a condensed structure that can be loaded in main memory. In what follows, we describe the construction of this data structure that we baptized "FUZZYTREE".

#### 4.1. The FUZZYTREE data structure: A condensed representation of the original constrained fuzzy context

##### 4.1.1. Presentation

The GEN-IFF takes as an input data the constrained fuzzy context that will be explored to identify fuzzy minimal generators with their respective domains (i.e., the application of the fuzzy operator  $\tilde{g}_{\tilde{C}}$ ). One bottleneck encountered by GEN-IFF is the computation of 1-sized fuzzy items domains. In fact, let us suppose that every 1-sized fuzzy minimal generator has  $\alpha$  degrees over the  $|\mathcal{O}|$  transactions of the constrained fuzzy context. So,  $(\alpha \times |\mathcal{O}|)$  scans are needed for computing the associated domain of each 1-sized fuzzy minimal generator. Thus, the total number of I/O operations is estimated to be equal to  $(\alpha \times |\mathcal{O}| \times |\mathcal{I}|)$ . In order to make GEN-IFF more efficient, original constrained fuzzy context can be loaded in a compact lossless representation. This is achieved by organizing the data in a suitable data structure, called "FUZZYTREE". From this data structure, it is straight-



Fuzz.min.gen	Domain	Support
$B^1$	{t3 t4}	$\frac{1}{2}$
$B^{0.6}$	{t2 t3 t4}	$\frac{1}{4}$
$B^{0.5}$	{t1 t2 t3 t4}	1
$C^1$	{t1}	$\frac{1}{4}$
$C^{0.9}$	{t1 t4}	1
$C^{0.7}$	{t1 t2 t3 t4}	1
$E^1$	{t2 t3}	$\frac{3}{4}$
$E^{0.9}$	{t2 t3 t4}	$\frac{1}{4}$
$E^{0.7}$	{t1 t2 t3 t4}	1
$M^{0.5}$	{t1 t2}	$\frac{1}{2}$
$M^{0.1}$	{t1 t2 t3 t4}	1

(a)

Fuzz.min.gen	Domain	Support
$B^1 C^{0.9}$	{t4}	1
$B^1 C^{0.7}$	{t3 t4}	1
$B^1 E^1$	{t3}	1
$B^1 E^{0.9}$	{t3 t4}	1
$B^1 E^{0.7}$	{t3 t4}	1
$B^1 M^{0.1}$	{t3 t4}	1
$B^{0.6} C^{0.9}$	{t4}	1
$B^{0.6} C^{0.7}$	{t2 t3 t4}	1
$B^{0.6} E^1$	{t2 t3}	1
$B^{0.6} E^{0.9}$	{t2 t3 t4}	1
$B^{0.6} E^{0.7}$	{t2 t3 t4}	1
$B^{0.6} M^{0.5}$	{t2}	1
$B^{0.6} M^{0.1}$	{t2 t3 t4}	1
$B^{0.5} C^1$	{t1}	1
$B^{0.5} C^{0.9}$	{t1 t4}	1
$B^{0.5} C^{0.7}$	{t1 t2 t3 t4}	1
$B^{0.5} E^1$	{t2 t3}	1
$B^{0.5} E^{0.9}$	{t2 t3 t4}	1
$B^{0.5} E^{0.7}$	{t1 t2 t3 t4}	1
$B^{0.5} M^{0.5}$	{t1 t2}	1
$B^{0.5} M^{0.1}$	{t1 t2 t3 t4}	1
$C^1 E^{0.7}$	{t1}	1
$C^1 M^{0.5}$	{t1}	1
$C^1 M^{0.1}$	{t1}	1
$C^{0.9} E^{0.9}$	{t4}	1
$C^{0.9} E^{0.7}$	{t1 t4}	1
$C^{0.9} M^{0.5}$	{t1}	1
$C^{0.9} M^{0.1}$	{t1 t4}	1
$C^{0.7} E^1$	{t2 t3}	1
$C^{0.7} E^{0.9}$	{t2 t3 t4}	1
$C^{0.7} E^{0.7}$	{t1 t2 t3 t4}	1
$C^{0.7} M^{0.5}$	{t1 t2}	1
$C^{0.7} M^{0.1}$	{t1 t2 t3 t4}	1
$E^1 M^{0.5}$	{t2}	1
$E^1 M^{0.1}$	{t2 t3}	1
$E^{0.9} M^{0.5}$	{t2}	1
$E^{0.9} M^{0.1}$	{t2 t3 t4}	1
$E^{0.7} M^{0.5}$	{t1 t2}	1
$E^{0.7} M^{0.1}$	{t1 t2 t3 t4}	1

(b)

Fuzz.min.gen	Domain	Support
$B^1 E^1$	{t3}	$\frac{1}{4}$
$B^{0.6} C^{0.9}$	{t4}	$\frac{1}{4}$
$B^{0.6} M^{0.5}$	{t2}	$\frac{1}{4}$
$C^{0.9} E^{0.9}$	{t4}	$\frac{1}{4}$
$C^{0.9} M^{0.5}$	{t1}	$\frac{1}{4}$
$E^{0.9} M^{0.5}$	{t2}	$\frac{1}{4}$

(c)

Fuzz.min.gen	Domain	Support
$B^{0.6} C^{0.9} M^{0.5}$	$\emptyset$	-
$C^{0.9} E^{0.9} M^{0.5}$	$\emptyset$	-

(d)

Figure 5: (a): Frequent fuzzy closed 1-itemsets, (b): Candidate fuzzy minimal 2-generators, (c): Frequent fuzzy closed 2-itemsets, (d): Candidate fuzzy minimal 3-generators.

Fuzz.min.gen	Domain	Closure	Support
$B^1$	$\{t_3 t_4\}$	$B^1 C^{0.7} E^{0.9} M^{0.1}$	$\frac{1}{5}$
$B^{0.6}$	$\{t_2 t_3 t_4\}$	$B^{0.6} C^{0.7} E^{0.9} M^{0.1}$	$\frac{3}{4}$
$B^{0.5}$	$\{t_1 t_2 t_3 t_4\}$	$B^{0.5} C^{0.7} E^{0.7} M^{0.1}$	1
$C^1$	$\{t_1\}$	$B^{0.5} C^1 E^{0.7} M^{0.5}$	$\frac{1}{4}$
$C^{0.9}$	$\{t_1 t_4\}$	$B^{0.5} C^{0.9} E^{0.7} M^{0.1}$	$\frac{1}{4}$
$C^{0.7}$	$\{t_1 t_2 t_3 t_4\}$	$B^{0.5} C^{0.7} E^{0.7} M^{0.1}$	1
$E^1$	$\{t_2 t_3\}$	$B^{0.6} C^{0.7} E^{1.0} M^{0.1}$	$\frac{3}{4}$
$E^{0.9}$	$\{t_2 t_3 t_4\}$	$B^{0.6} C^{0.7} E^{0.9} M^{0.1}$	$\frac{3}{4}$
$E^{0.7}$	$\{t_1 t_2 t_3 t_4\}$	$B^{0.5} C^{0.7} E^{0.7} M^{0.1}$	1
$M^{0.5}$	$\{t_1 t_2\}$	$B^{0.5} C^{0.7} E^1 M^{0.5}$	$\frac{1}{2}$
$M^{0.1}$	$\{t_1 t_2 t_3 t_4\}$	$B^{0.5} C^{0.7} E^{0.7} M^{0.1}$	1
$B^1 E^1$	$\{t_3\}$	$B^1 C^{0.7} E^1 M^{0.1}$	$\frac{1}{4}$
$B^{0.6} C^{0.9}$	$\{t_4\}$	$B^1 C^{0.9} E^{0.9} M^{0.1}$	$\frac{1}{4}$
$B^{0.6} M^{0.5}$	$\{t_2\}$	$B^{0.6} C^{0.7} E^1 M^{0.5}$	$\frac{1}{4}$
$C^{0.9} E^{0.9}$	$\{t_4\}$	$B^1 C^{0.9} E^{0.9} M^{0.1}$	$\frac{1}{4}$
$C^{0.9} M^{0.5}$	$\{t_1\}$	$B^{0.5} C^1 E^{0.7} M^{0.5}$	$\frac{1}{4}$
$E^{0.9} M^{0.5}$	$\{t_2\}$	$B^{0.6} C^{0.7} E^1 M^{0.5}$	$\frac{1}{4}$

Figure 6: Fuzzy minimal generators and closed itemsets extracted from the fuzzy context  $\mathcal{K}$  for a  $minsup = \frac{1}{4}$ .

forward to get out the domain of fuzzy minimal generators. The “FUZZYTREE” is a two-level tree. The first level contains nodes representing the identifiers of items and a link to upper level. The second level contains nodes storing the degree of the fuzzy item (*i.e.*, which is presented in the upper level), the set of objects corresponding to that degree and a link to the following *sibling* node. The construction process of the FUZZYTREE corresponding to the constrained fuzzy context of the Figure 1 is shown below.

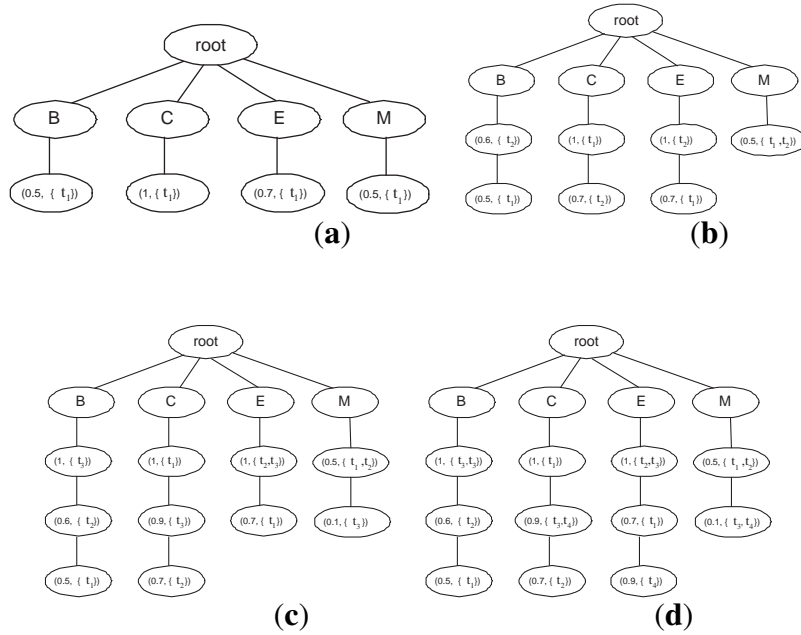


Figure 7: Construction process of the FUZZYTREE corresponding to the constrained fuzzy context of the Figure 1. (a): First iteration, (b): Second iteration, (c): Third iteration, (d): Fourth iteration.

Initially, the FUZZYTREE contains only one node (*i.e.*, the root). The handling of the first transaction generates four nodes of level 1 (B, C, D, E) and their corresponding elder sons  $(0.5, t_1)$ ,  $(0.9, t_1)$ ,  $(0.7, t_1)$  and  $(0.5, t_1)$ <sup>3</sup> (Figure 7.a). The second transaction  $t_2$ , generates three new nodes as follows:

Degrees are read and inserted according to a decreasing order. So, the degree of the item "B" in the second transaction is equal to 0.6 which is greater than that

<sup>3</sup>The  $i^{th}$  transaction is denoted here by  $t_i$ .

of the node  $(0.5, t_1)$ . Consequently, we insert a new node  $(0.6, t_2)$  before  $(0.5, t_1)$ . For the node "C", it has a degree equal to 0.7 which is lower than the degree 1 of the first node. Therefore, we insert the node  $(0.7, t_2)$  as a son of the node  $(1, t_1)$ . For the item "M", its membership degree in the second transaction is equal to 0.5, which is equal to that of the node  $(0.5, t_1)$ . In that case, the transaction  $t_2$  is inserted in the domain of the node  $(0.5, t_1)$ . So, this node becomes  $(0.5, t_1; t_2)$  (c.f., Figure 7.b). The process continues until scanning the whole constrained fuzzy context. So, the domain of a fuzzy itemset can be found in a straightforward manner by scanning the data structure FUZZYTREE. Indeed, the domain of a 1-sized fuzzy generator  $g$ , having at least the degree  $\alpha$ , is determined as follows:

- To cross FUZZYTREE from the root to the node  $g$  of the first level;
- To go downward in the tree structure until the second level node having the degree  $\alpha$ , while concatenating all transactions of previously visited nodes.

**Example 7.** *The domain of the candidate fuzzy generator  $C^{0.7}$ , is equal to  $\{t_1, t_2, t_3, t_4\}$ .*

#### 4.1.2. FUZZYTREE construction algorithm complexity issues

Let us consider a fuzzy formal context under constraint  $\mathcal{K}_{\tilde{C}} = (\mathcal{O}, \tilde{\mathcal{I}}, \tilde{\mathcal{R}}, \tilde{\mathcal{C}})$ . In order to represent it in the FUZZYTREE structure, we suppose that  $|\mathcal{O}| = m$  (the number of transactions) and  $|\tilde{\mathcal{I}}| = n$  (the number of items). The insertion of a first level node costs  $O(1)$ . To insert a second level node in the FUZZYTREE, we have first to find the position of its corresponding first level node which costs  $O(m)$ . Then, we have to go down in the tree structure, and find the position of this second level node by comparing its degree with the degrees of all visited nodes. In the worst case, we perform  $p$  comparisons, where  $p$  is the total number of degrees that can be taken by an item. So, the theoretical complexity of the FUZZYTREE construction algorithm is bounded by:

$$O(n \times m \times p) \simeq O(n \times m)$$

#### 4.2. Fuzzy generic basis of FARs

In this paper, we are interested in extracting fuzzy association rules of the form  $\tilde{r}: \tilde{I}_1 \Rightarrow \tilde{I}_2$ , where  $\tilde{I}_1, \tilde{I}_2 \subseteq \tilde{I} = \{a_1^{\alpha_1}, a_2^{\alpha_2}, \dots, a_p^{\alpha_p}, a_q^{\alpha_q}, \dots, a_n^{\alpha_n}\}$ , and  $\tilde{I}_1 = \{a_1^{\alpha_1}, a_2^{\alpha_2}, \dots, a_p^{\alpha_p}\}$  and  $\tilde{I}_2 = \{a_q^{\alpha_q}, \dots, a_n^{\alpha_n}\}$ .  $\tilde{I}_1$  and  $\tilde{I}_2$  are called, respectively, the *premise part* and *conclusion part* of the fuzzy rule  $\tilde{r}$ . The value  $\alpha_i, i = 1, \dots, n$ , is called

*local weight* of the item  $a_i$ . This value indicates the relative degree of importance of each item contributing to the conclusion of the rule, and plays an important role in many real life problems [33]. For example, in medical diagnostic systems, it is common to assign a local weight to each symptom in order to show the relative importance (weight) of each symptom leading to the conclusion (a disease).

A fuzzy association rule is *valid* whenever its confidence is greater than or equal to a minimal threshold of confidence *minconf*. The confidence and the support of a rule are defined as follows<sup>4</sup>:

$$Conf(R) = \frac{|\tilde{g}_{\tilde{C}}(\tilde{I}_1 \cup \tilde{I}_2)|}{|\tilde{g}_{\tilde{C}}(\tilde{I}_1)|}$$

$$Supp(R) = \frac{|\tilde{g}_{\tilde{C}}(\tilde{I}_1 \cup \tilde{I}_2)|}{|O|}$$

**Remark 3.** *Classical (or crisp) association rules can be defined as a special case of fuzzy association rules. Indeed, when  $\tilde{I} = \{a_1^{\alpha_1}, a_2^{\alpha_2}, \dots, a_p^{\alpha_p}, a_q^{\alpha_q}, \dots, a_n^{\alpha_n}\}$  and  $\alpha_i = 1, i = 1, \dots, n$ , then a fuzzy association rule is reduced to a classical one.*

It is noteworthy that extracting fuzzy association rules is far from being a trivial task, compared to the boolean case, mostly because of the huge size of the fuzzy extraction context and consequently the number of potentially interesting fuzzy rules that can be drawn from such a fuzzy dataset. In this respect, reducing such a set of fuzzy rules is a critical issue. Beyond basic statistical techniques used to prune the number of rules (i.e., support and confidence), more advanced techniques allowing at producing only limited number of rules. This techniques rely on closures and Galois connections [4, 31, 34], which are in turn derived from Galois lattice theory and formal concept analysis (FCA) [21]. In order to select without loss of information a generic subset of all fuzzy association rules, we define three fuzzy generic basis from which remaining (redundant) FARs are generated.

**Definition 10. Generic basis for exact FARs**

*Let  $\mathcal{FC}_{\tilde{C}}$  be the set of fuzzy frequent closed itemsets extracted from a fuzzy context under constraint, and, for each fuzzy frequent closed itemset  $\tilde{I}$ , let us denote  $\mathcal{FG}_{\tilde{I}}$  the set of its fuzzy minimal generators. Thus, the generic basis for exact FARs denoted by  $\mathcal{GBEF}$ , is defined as follows :*

---

<sup>4</sup>The interested reader is referred to [20], in which techniques were investigated to identify and evaluate associations in a relational database that are expressible by fuzzy if-then rules.

$$\mathcal{GBEF} = \{R: \tilde{g} \Rightarrow (\tilde{I} - \tilde{g}) \mid \tilde{I} \in \mathcal{FC}_{\mathcal{K}} \wedge \tilde{g} \in \mathcal{FG}_{\tilde{I}} \wedge \tilde{g} \neq \tilde{I}^{(5)}\}$$

**Example 8.** The generic basis for exact FARs extracted from the constrained fuzzy context  $\mathcal{K}_{\tilde{C}}$ , shown in Figure 1, is depicted by table 4. From this context, 18 exact FARs are extracted.

Table 4: The generic basis  $\mathcal{GBEF}$  extracted from the constrained fuzzy context  $\mathcal{K}_{\tilde{C}}$ , for  $\text{minsup}=\frac{1}{4}$  and  $\text{minconf}=\frac{1}{4}$ .

Exact FARs	Support	Confidence
$B^{0.5} \Rightarrow C^{0.7} E^{0.7} M^{0.1}$	1	1
$C^{0.7} \Rightarrow B^{0.5} E^{0.7} M^{0.1}$	1	1
$E^{0.7} \Rightarrow B^{0.5} C^{0.7} M^{0.1}$	1	1
$M^{0.1} \Rightarrow B^{0.5} C^{0.7} E^{0.7}$	1	1
$C^{0.9} \Rightarrow B^{0.5} E^{0.7} M^{0.1}$	1	1
$B^{0.6} \Rightarrow C^{0.7} E^{0.9} M^{0.1}$	$\frac{3}{4}$	1
$E^{0.9} \Rightarrow B^{0.6} C^{0.7} M^{0.1}$	$\frac{3}{4}$	1
$B^1 \Rightarrow C^{0.9} E^{0.9} M^{0.1}$	$\frac{1}{2}$	1
$M^{0.5} \Rightarrow B^{0.5} C^{0.7} E^{0.7}$	$\frac{1}{2}$	1
$E^1 \Rightarrow B^{0.6} C^{0.7} M^{0.1}$	$\frac{1}{2}$	1
$C^1 \Rightarrow B^{0.5} E^{0.7} M^{0.5}$	$\frac{1}{2}$	1
$C^{0.9} M^{0.5} \Rightarrow B^{0.5} C^1 E^{0.7}$	$\frac{1}{4}$	1
$B^{0.6} M^{0.5} \Rightarrow C^{0.7} E^1$	$\frac{1}{4}$	1
$E^{0.9} M^{0.5} \Rightarrow B^{0.6} C^{0.7} E^1$	$\frac{1}{4}$	1
$B^{0.6} C^{0.9} \Rightarrow B^1 E^{0.9} M^{0.1}$	$\frac{1}{4}$	1
$B^1 E^1 \Rightarrow C^{0.9} M^{0.1}$	$\frac{1}{4}$	1
$C^{0.9} E^{0.9} \Rightarrow B^1 M^{0.1}$	$\frac{1}{4}$	1
$C^{0.9} E^1 \Rightarrow B^1 M^{0.1}$	$\frac{1}{4}$	1

**Definition 11. Generic basis for approximate FARs**

The generic basis for approximative FARs, denoted by  $\mathcal{GBAF}$ , is defined as follows :

<sup>5</sup>This condition ensures discarding non-informative rules of the form  $\tilde{g} \rightarrow \emptyset$ .

$$\mathcal{GBAF} = \{ R: \tilde{g} \Rightarrow (\tilde{I}_1 - \tilde{g}) \mid \tilde{I}, \tilde{I}_1 \in \mathcal{FC}_{\tilde{\mathcal{K}}}, \tilde{g} \in \mathcal{FG}_{\tilde{I}} \wedge \tilde{I} \subset \tilde{I}_1 \wedge \text{Conf}(R) \geq \text{minconf} \}$$

A transitive reduction of generic basis for approximate FARs can be defined as:

**Definition 12. Generic basis for transitive FARs**

The generic basis for transitive FARs, denoted by  $\mathcal{RTF}$ , is defined as follows :

$$\mathcal{RTF} = \{ R: \tilde{g} \Rightarrow (\tilde{I}_1 - \tilde{g}) \mid \tilde{I}, \tilde{I}_1 \in \mathcal{FC}_{\tilde{\mathcal{K}}} \wedge \tilde{g} \in \mathcal{FG}_{\tilde{I}} \wedge \tilde{I} \subset \tilde{I}_1 \wedge \nexists \tilde{I}_2 \text{ s.t. } \tilde{I} \subset \tilde{I}_2 \subset \tilde{I}_1 \wedge \text{Conf}(R) \geq \text{minconf} \}$$

**Example 9.** From the context shown in Figure 1, 47 approximate generic FARs are extracted vs 22 transitive generic FARs. The generic basis for approximative FARs ( $\mathcal{GBAF}$ ) extracted from the constrained fuzzy context  $\mathcal{K}_{\tilde{C}}$  is shown in table 5, while, the generic basis for transitive FARs ( $\mathcal{RTF}$ ) is shown in table 6.

**Remark 4.** Given a fuzzy Iceberg concept lattice in which each fuzzy closed itemset is “decorated” with its associated list of fuzzy minimal generators, the generic exact FARs can be derived in a straightforward manner. In fact, generic exact fuzzy rules represent “intra-node” implications, between a fuzzy minimal generator and the corresponding fuzzy closed itemset. Generic approximate (transitive) fuzzy rules represent “inter-node” implications, assorted with the confidence measure, between two comparable equivalence classes, i.e., from a sub fuzzy closed itemset to a super (immediate) fuzzy closed itemset, when starting from a given node in the partially ordered structure.

#### 4.3. Redundant FARs derivation

A generic basis extraction approach should be performed **without information loss**. Therefore, extracting generic basis process has to fulfill the following requirements [7] :

- **“Derivability”**: An inference mechanism should be provided (e.g., an axiomatic system). The axiomatic system has to be valid (i.e., should forbid derivation of non valid rules) and complete (i.e., should enable derivation of all valid rules).
- **“Informativeness”**: The generic basis of association rules allows to exactly retrieve the support and confidence of the derived (redundant) association rules.

Table 5: The generic basis ( $\mathcal{GBAF}$ ) corresponding to the constrained fuzzy context  $\mathcal{K}_{\bar{C}}$  for  $min\text{-}sup=\frac{1}{4}$  and  $min\text{-}conf=\frac{1}{4}$ .

App. FARs	Support	Confidence
$B^{0.5} \Rightarrow B^{0.6} C^{0.7} E^{0.9} M^{0.1}$	3	3
$B^{0.5} \Rightarrow C^{0.9} E^{0.7} M^{0.1}$	3	3
$B^{0.5} \Rightarrow C^{0.7} E^{0.7} M^{0.5}$	4	4
$B^{0.5} \Rightarrow B^1 C^{0.9} E^{0.9} M^{0.1}$	1	1
$B^{0.5} \Rightarrow B^{0.6} C^{0.7} E^1 M^{0.1}$	2	2
$B^{0.5} \Rightarrow C^1 E^{0.7} M^{0.5}$	1	1
$B^{0.5} \Rightarrow B^{0.6} C^{0.7} E^1 M^{0.5}$	4	4
$B^{0.5} \Rightarrow B^1 C^{0.9} E^1 M^{0.1}$	4	4
$C^{0.7} \Rightarrow B^{0.6} E^{0.9} M^{0.1}$	3	3
$C^{0.7} \Rightarrow B^{0.5} C^{0.9} E^{0.7} M^{0.1}$	3	3
$C^{0.7} \Rightarrow B^{0.5} E^{0.7} M^{0.5}$	4	4
$C^{0.7} \Rightarrow B^{0.6} E^1 M^{0.1}$	2	2
$C^{0.7} \Rightarrow B^1 C^{0.9} E^{0.9} M^{0.1}$	1	1
$C^{0.7} \Rightarrow B^{0.5} C^1 E^{0.7} M^{0.5}$	2	2
$C^{0.7} \Rightarrow B^{0.6} E^1 M^{0.5}$	4	4
$C^{0.7} \Rightarrow B^1 C^{0.9} E^1 M^{0.1}$	4	4
$E^{0.7} \Rightarrow B^{0.5} C^{0.9} M^{0.1}$	3	3
$E^{0.7} \Rightarrow B^{0.6} C^{0.7} E^{0.9} M^{0.1}$	3	3
$E^{0.7} \Rightarrow B^1 C^{0.9} E^{0.9} M^{0.1}$	1	1
$E^{0.7} \Rightarrow B^{0.5} C^{0.7} M^{0.5}$	2	2
$E^{0.7} \Rightarrow B^{0.6} C^{0.7} E^1 M^{0.1}$	2	2
$E^{0.7} \Rightarrow B^{0.5} C^1 M^{0.5}$	1	1
$E^{0.7} \Rightarrow B^{0.6} C^{0.7} E^1 M^{0.5}$	4	4
$E^{0.7} \Rightarrow B^1 C^{0.9} E^1 M^{0.1}$	4	4
$M^{0.1} \Rightarrow B^{0.5} C^{0.9} E^{0.7}$	3	3
$M^{0.1} \Rightarrow B^{0.6} C^{0.7} E^{0.9}$	3	3
$M^{0.1} \Rightarrow B^1 C^{0.9} E^{0.9}$	4	4
$M^{0.1} \Rightarrow B^{0.5} C^{0.7} E^{0.7} M^{0.5}$	1	1
$M^{0.1} \Rightarrow B^{0.6} C^{0.7} E^1$	1	1
$M^{0.1} \Rightarrow B^{0.5} C^1 E^{0.7} M^{0.5}$	4	4
$M^{0.1} \Rightarrow B^{0.6} C^{0.7} E^1 M^{0.5}$	4	4
$M^{0.1} \Rightarrow B^1 C^{0.9} E^1$	4	4
$B^{0.6} \Rightarrow C^{0.7} E^1 M^{0.1}$	1	1
$B^{0.6} \Rightarrow B^1 C^{0.9} E^1 M^{0.1}$	2	2
$B^{0.6} \Rightarrow C^{0.7} E^1 M^{0.5}$	4	4
$E^{0.9} \Rightarrow B^{0.6} C^{0.7} E^1 M^{0.1}$	1	1
$E^{0.9} \Rightarrow B^1 C^{0.9} E^1 M^{0.1}$	4	4
$E^{0.9} \Rightarrow B^{0.6} C^{0.7} E^1 M^{0.5}$	1	1
$C^{0.9} \Rightarrow B^1 C^{0.9} E^{0.9} M^{0.1}$	1	1
$C^{0.9} \Rightarrow B^1 C^{0.9} E^1 M^{0.1}$	2	2
$B^1 \Rightarrow C^{0.9} E^1 M^{0.1}$	4	4
$B^{0.6} C^{0.9} \Rightarrow B^1 E^1 M^{0.1}$	4	4
$C^{0.9} E^{0.9} \Rightarrow B^1 E^1 M^{0.1}$	4	4
$M^{0.5} \Rightarrow B^{0.5} C^1 E^{0.7}$	4	4
$M^{0.5} \Rightarrow B^{0.6} C^{0.7} E^1$	1	1
$E^1 \Rightarrow B^{0.6} C^{0.7} M^{0.5}$	4	4
$E^1 \Rightarrow B^1 C^{0.9} M^{0.1}$	4	4



Table 6: The generic basis ( $\mathcal{RTF}$ ) corresponding to the constrained fuzzy context  $\mathcal{K}_{\bar{C}}$ , for  $min\text{-}sup=\frac{1}{4}$  and  $min\text{-}conf=\frac{1}{4}$ .

Trans. FARs	Support	Confidence
$B^{0.5} \Rightarrow B^{0.6} C^{0.7} E^{0.9} M^{0.1}$	3	3
$B^{0.5} \Rightarrow C^{0.9} E^{0.7} M^{0.1}$	4	4
$B^{0.5} \Rightarrow C^{0.7} E^{0.7} M^{0.5}$	3	3
$C^{0.7} \Rightarrow B^{0.6} E^{0.9} M^{0.1}$	4	4
$C^{0.7} \Rightarrow B^{0.6} C^{0.9} E^{0.7} M^{0.1}$	1	1
$C^{0.7} \Rightarrow B^{0.5} E^{0.7} M^{0.5}$	2	2
$E^{0.7} \Rightarrow B^{0.6} C^{0.7} E^{0.9} M^{0.1}$	3	3
$E^{0.7} \Rightarrow B^{0.5} C^{0.9} M^{0.1}$	4	4
$E^{0.7} \Rightarrow B^{0.5} C^{0.7} M^{0.5}$	3	3
$M^{0.1} \Rightarrow B^{0.6} C^{0.7} E^{0.9}$	4	4
$M^{0.1} \Rightarrow B^{0.5} C^{0.9} E^{0.7}$	3	3
$M^{0.1} \Rightarrow B^{0.5} C^{0.7} E^{0.7} M^{0.5}$	4	4
$B^{0.6} \Rightarrow C^{0.7} E^1 M^{0.1}$	1	1
$E^{0.9} \Rightarrow B^{0.6} C^{0.7} E^1 M^{0.1}$	2	2
$C^{0.9} \Rightarrow B^1 E^{0.9} M^{0.1}$	2	2
$B^1 \Rightarrow C^{0.9} E^1 M^{0.1}$	1	1
$B^{0.6} C^{0.9} \Rightarrow B^1 E^1 M^{0.1}$	4	4
$C^{0.9} E^{0.9} \Rightarrow B^1 E^1 M^{0.1}$	4	4
$M^{0.5} \Rightarrow B^{0.5} C^1 E^{0.7}$	1	1
$M^{0.5} \Rightarrow B^{0.6} C^{0.7} E^1$	4	4
$E^1 \Rightarrow B^{0.6} C^{0.7} M^{0.5}$	1	1
$E^1 \Rightarrow B^1 C^{0.9} M^{0.1}$	4	4

In the following, we start by defining the notion of a redundant FAR and then, we prove that the derivation of the pair  $(\mathcal{GBEF}, \mathcal{GBAF})$  is a lossless information approach.

**Definition 13.** Let  $\mathcal{FAR}$  be the set of all FARs derived from a constrained fuzzy context  $\mathcal{K}_{\tilde{C}}$ . A fuzzy rule  $R: \tilde{I}_1 \xrightarrow{s,c} \tilde{I}_2 \in \mathcal{FAR}$  is said to be redundant to (or derivable from)  $R' : \tilde{I}_1' \xrightarrow{s,c} \tilde{I}_2'$ , when  $R$  fulfills:

1.  $Support(R) = Support(R') = s \wedge Confidence(R) = Confidence(R') = c$ .
2.  $\tilde{I}_1' \subseteq \tilde{I}_1 \wedge \tilde{I}_2 \subseteq \tilde{I}_2'$ .

In the remainder, we introduce rule inference mechanisms, by means of an axiomatic system.

**Proposition 4.** Let us consider the pair  $(\mathcal{GBEF}, \mathcal{GBAF})$  and the set of all valid fuzzy rules extracted from a constrained fuzzy context  $\mathcal{K}_{\tilde{C}}$  denoted by  $\mathcal{FAR}$ . The following axiomatic system is valid and complete:

**A1. Reflexivity:**

If  $R : X \xrightarrow{s,c} Y \in (\mathcal{GBEF}, \mathcal{GBAF})$  then  $R : X \xrightarrow{s,c} Y \in \mathcal{FAR}$ .

**A2. Left augmentation:**

- If  $R : X \xrightarrow{s,c} Y \in \mathcal{GBEF}$  and  $Z \subset Y$ , then  $R' : XZ \xrightarrow{s,c} (Y - Z) \in \mathcal{FAR}$ ,  $\forall Z \subset Y$  (i.e.,  $R'$  is valid).
- If  $R : X \xrightarrow{s,c} Y \in \mathcal{GBAF}$  then  $R' : XZ \xrightarrow{s,c} (Y - Z) \in \mathcal{FAR}$ , such that  $support(XZ) = support(X)$  and  $Z \subset Y$ .

**A3. Right decomposition:**

- If  $R : X \xrightarrow{s,c} Y \in \mathcal{GBEF}$  then  $R' : X \xrightarrow{s,c} Z \in \mathcal{FAR}$ ,  $\forall Z \subset Y$ .
- If  $R : X \xrightarrow{s,c} Y \in \mathcal{GBAF}$  then  $R' : X \xrightarrow{s,c} Z \in \mathcal{FAR}$ , such that  $support(XZ) = support(XY)$  and  $Z \subset Y$ .

**Proposition 5.** The rule inference mechanism (i.e., the axiomatic system), as defined above is **valid**.

**Proof 4.** In order to prove the axiomatic system's validity, we have to prove that all the FARs derived from the pair  $(\mathcal{GBEF}, \mathcal{GBAF})$ , are **valid** (i.e., their support and confidence are, respectively, greater than or equal to minsup and minconf).

**A1. Reflexivity :**

This can be deduced from the definition of both generic bases  $\mathcal{GBEF}$  and  $\mathcal{GBAF}$ .

**A2. Left augmentation :**

- If  $R : X \xrightarrow{s,c} Y \in \mathcal{GBEF}$  then  $\text{confidence}(R) = \frac{\text{support}(XY)}{\text{support}(X)} = 1 = c$ . So,  $\text{support}(XY) = \text{support}(X)$ .

Let  $R' : XZ \xrightarrow{s',c'} (Y - Z)$ , be a fuzzy association rule such that  $Z \subset Y$ . Since,  $X \subset XZ$ , then we have  $\text{support}(X) \geq \text{support}(XZ)$ . On the one hand, we have  $\text{confidence}(R') = c' = \frac{\text{support}(XY)}{\text{support}(XZ)} \geq c$ . On the other hand, the confidence of a fuzzy association rule is a statistic metric ranging in the interval  $[0, 1]$ . We have  $\text{confidence}(R) = c = 1$  and  $\text{confidence}(R') = c' \geq c$ , then,  $\text{confidence}(R') = 1$  (i.e.,  $R'$  is a valid fuzzy association rule).

- If  $R : X \xrightarrow{s,c} Y \in \mathcal{GBAF}$  then  $\text{confidence}(R) = \frac{\text{support}(XY)}{\text{support}(X)} = c$ .

Let  $R' : XZ \xrightarrow{s',c'} (Y - Z)$ , be a fuzzy association rule such that  $Z \subset Y$ . We have,  $Z \subset Y$  then  $\text{support}(R') = s' = |\tilde{g}_C(XZ \cup (Y - Z))| = \text{support}(XY) = \text{support}(R) = s$ . On the other hand, we have  $\text{support}(XZ) = \text{support}(X)$ , then  $\text{confidence}(R') = \frac{\text{support}(XY)}{\text{support}(XZ)} = c' = \frac{\text{support}(XY)}{\text{support}(X)} = \text{confidence}(R) = c$ .

**A3. Right decomposition :**

- If  $R : X \xrightarrow{s,c} Y \in \mathcal{GBEF}$  then  $\text{confidence}(R) = \frac{\text{support}(XY)}{\text{support}(X)} = 1$ . So,  $\text{support}(XY) = \text{support}(X)$ . Let  $R' : X \Rightarrow Z$ , be a fuzzy association rule such that  $Z \subset Y$ . So,  $\text{confidence}(R') = \frac{\text{support}(XZ)}{\text{support}(X)}$ . We have,  $Z \subset Y$  then  $XZ \subset XY$ . Hence, we have  $\text{support}(XZ) \geq \text{support}(XY)$  and  $\text{confidence}(R') = \frac{\text{support}(XZ)}{\text{support}(X)} \geq \text{confidence}(R) = \frac{\text{support}(XY)}{\text{support}(X)}$ . However, the confidence is a statistic measure belonging to the interval  $[0, 1]$ . We can deduce then, that  $\text{confidence}(R') = \frac{\text{support}(XZ)}{\text{support}(X)} = 1$  and  $\text{support}(R') = \text{support}(R) = s$ .

- If  $R : X \xrightarrow{s,c} Y \in \mathcal{GBAF}$  then  $\text{confidence}(R) = \frac{\text{support}(XY)}{\text{support}(X)} = c$ . Let  $R' : X \xrightarrow{s',c'} Z$ , be a fuzzy association rule such that  $Z \subset Y$  and  $\text{support}(XZ) = \text{support}(XY)$ .  
From these last conditions, we can deduce that  $\text{support}(R' : X \Rightarrow Z) = \text{support}(XZ) = \text{support}(XY) =$  and  $\text{confidence}(R') = \frac{\text{support}(XZ)}{\text{support}(X)} = c' = \frac{\text{support}(XY)}{\text{support}(X)} = \text{confidence}(R) = c$ .

**Proposition 6.** *The rule inference mechanism (i.e., the axiomatic system), as defined above is **complete**.*

**Proof 5.** *Proving the completeness of the axiomatic system comes back to show that, when it is applied to the pair  $(\mathcal{GBEF}, \mathcal{GBAF})$  it allows the derivation of **all** valid fuzzy associative rules that can be extracted from a fuzzy formal context.*

Let  $R : X \xrightarrow{s,c} Y - X$  be a valid fuzzy association rule between two fuzzy itemsets  $X$  and  $Y$ :

- If  $R$  is an **exact** fuzzy association rule (i.e.,  $\text{confidence}(R) = c = 1$ ), we obviously have  $X \subset Y$  and  $\text{confidence}(R) = 1$ , we have  $\text{support}(X) = \text{support}(Y)$ . Therefore, we can conclude that  $\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(X) = \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(Y) = \tilde{I}$ . The fuzzy itemset  $\tilde{I}$  is a fuzzy closed itemset and there exists a fuzzy association rule  $R' : \tilde{g} \Rightarrow (\tilde{I} - \tilde{g}) \in \mathcal{GBEF}$  such that  $\tilde{g}$  is a fuzzy minimal generator of  $\tilde{I}$  for which we have :  
 $\tilde{g} \subset X$  and  $\tilde{g} \subset Y$ , (application of the augmentation axiom);  
 $\tilde{g} = X$  and  $\tilde{g} \subset Y$ , (application of the decomposition axiom). We will prove that the fuzzy rule  $R$  and its support can be deduced from the fuzzy rule  $R'$ . Since  $\tilde{g} \subseteq X$  and  $\tilde{g} \subset Y \subseteq \tilde{I}$ , the fuzzy association rule  $R$  can be derived from  $R'$  (by applying the augmentation or the decomposition axiom). From  $\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(X) = \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(Y) = \tilde{I}$ , we can deduce that  $\text{support}(R) = s = \text{support}(Y) = \text{support}(\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(Y)) = \text{support}(R')$ .
- If  $R$  is an **approximative** fuzzy association rule (i.e.,  $\text{confidence}(R) = s \leq 1$ ) then we have  $X \subset Y$ . Since  $\text{confidence}(R) \leq 1$ , we have necessarily,  $\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(X) \subset \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(Y)$ . For both fuzzy itemsets  $X$  and  $Y$ , there exist two fuzzy minimal generators,  $\tilde{g}_1$  and  $\tilde{g}_2$  having, respectively,  $\tilde{I}_1$  and  $\tilde{I}_2$  as fuzzy closures such that:

$$\begin{aligned} \tilde{g}_1 &\subset X \subset \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(X) = \tilde{I}_1; \\ \tilde{g}_2 &\subset Y \subseteq \tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(Y) = \tilde{I}_2. \end{aligned}$$

Since  $X \subset Y$ , we have  $X \subseteq \tilde{I}_1 \subset Y \subseteq \tilde{I}_2$  and the fuzzy rule  $R' : \tilde{g}_1 \xrightarrow{s', c'} (\tilde{I}_2 - \tilde{g}_1) \in \mathcal{GBAF}$ . We will prove now that the fuzzy rule  $R$  (with its support  $s$  and confidence  $c$ ) can be deduced from the fuzzy rule  $R'$  (with its support  $s'$  and confidence  $c'$ ).

Since  $\tilde{g}_1 \subset X \subseteq \tilde{I}_1 \subset Y \subseteq \tilde{I}_2$  then the premise and the conclusion parts of the fuzzy rule  $R$  can be derived from  $R'$ . Besides, we have  $\tilde{f}_{\tilde{C}} \circ \tilde{g}_{\tilde{C}}(Y) = \tilde{I}_2$  then  $\text{support}(R) = s = \text{support}(Y) = \text{support}(\tilde{I}_2) = \text{support}(R') = s'$ . Since  $X \subseteq \tilde{I}_1$ , we have  $\text{support}(X) = \text{support}(\tilde{I}_1)$  and we can deduce that :

$$\text{confidence}(R) = c = \frac{\text{support}(Y)}{\text{support}(X)} = \frac{\text{support}(\tilde{I}_2)}{\text{support}(\tilde{I}_1)} = \text{confidence}(R') = c'.$$

**Example 10.** The cardinality of the redundant exact FARs set extracted from the  $\mathcal{GBEF}$  basis given by table 4 is equal to 179 FARs which are enumerated below.

- 18 exact fuzzy rules obtained by applying the “reflexivity” axiom;
- 69 exact fuzzy rules obtained by applying the “left augmentation” axiom;
- 92 exact fuzzy rules obtained by applying the “right decomposition” axiom.

## 5. Experiments

We implemented our algorithm in the C++ language using gcc version 3.3.1 under Linux Fedora core 5 platform. Experiments were conducted on a Centrino Dual Core with a 1.66 GHz and 2 GB of main memory. The evaluation of the proposed approach is split in two parts. Firstly, the focus is put on the scalability issue of the GEN-IFF algorithm. Secondly, the compactness ratios obtained by the different introduced generic bases of fuzzy association rules are assessed. We begin by presenting the dataset’s characteristics used during this evaluation.

### 5.1. Test dataset description

To rate the behavior of our algorithm presented in the previous section, we ran experiments on two different types of datasets. The first type originates from biological domain, while the second one presents benchmark datasets of the data mining field. The first dataset is the publicly available SAGE (Serial Analysis of Gene Expression)<sup>(6)</sup> data produced from human cells. One of the crucial task

---

<sup>6</sup><http://www.ncbi.nlm.nih.gov/SAGE/index.cgi>.

in genomic research is the analysis of SAGE expression to identify a priori interesting sets of genes, e.g., sets of genes that are frequently co-regulated. Such *matrices* provide expression values for given biological situations (the lines) and given genes (columns) [29]. The SAGE dataset contains 12000 genes (in columns) and only 56 patients (in lines). This type of context is different from the traditional ones since the number of lines is generally very important compared to that of attributes (i.e., usually of use in the data mining field). Mining the whole basis with its 12000 genes was an intractable task. This is the reason why we have chosen to work with some randomly chosen excerpts of the base.

The *Mushroom* dataset contains characteristics of various mushroom species. The *Chess* dataset is derived steps of associated game. These datasets were originally taken from the UC Irvine Machine Learning Database Repository<sup>(7)</sup>. Typically, these real datasets are dense, i.e., they produce long frequent itemsets even for very high values of support. Please note that both latter datasets were "fuzzified", by assigning a random value, ranging in the unit interval, to each item appearing in each dataset rows. Characteristics of the considered datasets are summarized in Table 7.

<b>Dataset</b>	<b>Type</b>	<b>#items</b>	<b>Avg. transaction size</b>	<b># transactions</b>
SAGE	dense	12, 000	–	56
Mushroom	dense	119	23	8, 124
Chess	dense	75	37	3, 196

Table 7: Benchmark dataset characteristics.

## 5.2. Scalability issue of the GEN-IFF algorithm

We begin this subsection by providing an idea on the usefulness of the introduced FUZZYTREE data structure. To illustrate it, we have chosen the SAGE dataset since it is the most dense one from the considered datasets.

At a glance, Table 8, reporting the computation time and the memory space needed to build the FUZZYTREE, sheds light on the compactness of the proposed data structure as well as on the scalability of its construction algorithm.

During the scalability assessment experiments, we took the liberty to omit considering the user's constraint in order to test our algorithm in the worst case.

---

<sup>7</sup><http://www.ics.uci.edu/~mlearn/MLRepository.html>.

Excerpts of the SAGE base	Space Memory(MB)	Computation time (sec.)
Version 1 : 2000 genes	1.5	2
Version 2 : 4000 genes	2.3	4
Version 3 : 6000 genes	3.2	8
Version 4 : 8000 genes	4.8	21
Version 5 : 10000 genes	5.6	35
Version 6 : 12000 genes	6.2	45

Table 8: Variation of the memory space and the computation time vs. the number of genes

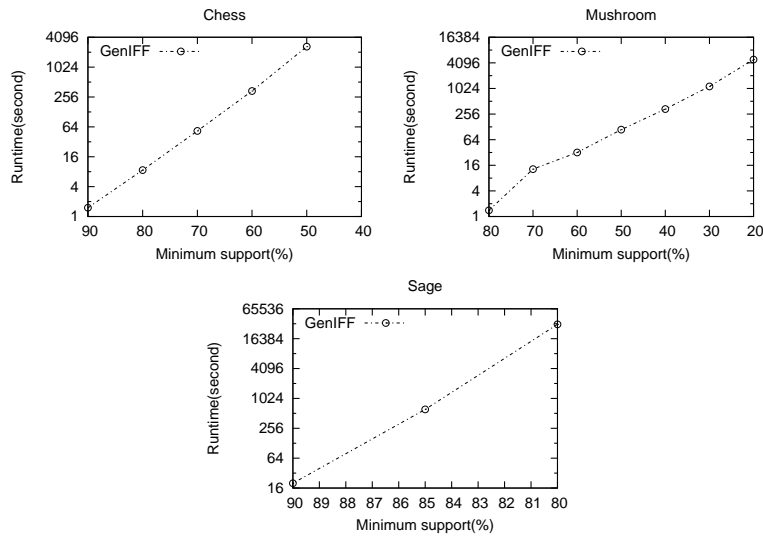


Figure 8: Evolution of the computation time for discovering fuzzy closed itemsets and fuzzy minimal generators vs the variation of the *minsup* value for the for the considered datasets.

Table 9 reports statistics on the number of fuzzy closed itemsets and their corresponding fuzzy minimal generators that can be extracted from the considered datasets. From the reported values in Table 9, we can highlight that:

- We remark that even from a small excerpt (that represents the  $\frac{1}{80}$  of the totality of the dataset), we extracted a huge number of fuzzy minimal generators and their corresponding fuzzy closed itemsets (that reaches millions for a *minsup* value lower than 0.5), when compared to those extracted from the other datasets. This result confirms that our SAGE dataset is a very dense one. The overwhelming number of fuzzy minimal generators extracted from the SAGE dataset has a direct influence on the performances of the GEN-IFF

<i>Dataset</i>	<i>minsup</i>	Fuzzy minimal generators (Step 1)	Fuzzy closed itemsets (Step 2)	Runtime (Sec)	
				<i>Step 1</i>	<i>Step 2</i>
SAGE (an excerpt of of 150 genes)	0.80	496 509	247 521	31 474	291
	0.85	56 502	30 539	577	37
	0.90	10 135	5 824	20	0
	0.95	484	231	1	0
CHESS	0.50	21 142	19 764	2 527	126
	0.60	2 630	1 876	321	17
	0.70	406	365	56	3
	0.80	121	48	8	1
MUSHROOM	0.20	28 934	26 815	4681	228
	0.30	3 344	3 098	1085	61
	0.40	836	695	332	11
	0.50	320	253	109	6
	0.60	187	64	36	2

Table 9: Variation of the number of fuzzy closed itemsets and their fuzzy minimal generators vs the *minsup* value variation for the considered datasets.

algorithm. Indeed, the computation time required for the SAGE dataset is by far greater than those dedicated respectively to the Chess and Mushroom datasets.

- As expectable, the number of fuzzy minimal generators as well as that of fuzzy closed itemsets increases as far as the *minsup* value decreases.
- The computation time, corresponding to the extraction of fuzzy minimal generators, exponentially increases as far as the *minsup* value decreases. Whereas, the computation time, corresponding to the generation of fuzzy closed itemsets, remains reasonable even for low *minsup* values, specially for the Chess and Mushroom datasets.
- For all the considered datasets, the computation time for extracting fuzzy minimal generators is by far greater than that dedicated to computing closed fuzzy itemsets. This is due to time consumed by the construction of the FUZZYTREE.



### 5.3. Compactness ratio's assessment

In the following, we put the focus on the variation of the reported generic fuzzy rule number of  $\mathcal{GBEF}$  basis. The compactness rate is measured as follows:

$$\text{comp} = 1 - \left( \frac{\#basis}{\#FAR} \right).$$

Table 10: Variation of generic FARs vs that of gene number ( $minsup=95\%$ ).

Genes	Fuzz.Closed itemsets	Tran. FARs $\mathcal{RTF}$	Exact FARs $\mathcal{GBEF}$	Redundant exact FARs	Compactness ratio
100	115	420	268	52 098	0.0051
200	418	3 135	1 003	394 880	0.0025
300	1 589	33 973	4 627	2 743 210	0.0016
400	2 649	105 169	10 572	8 379 892	0.0012
500	3 049	189 534	16 990	16 863 554	0.0010
600	4 084	448 869	33 562	40 019 440	0.0008
700	5 391	1 018 551	64 282	89 500 100	0.0007
800	6 373	1 512 339	87 563	139 423 724	0.0006
900	7 029	1 854 399	102 858	184 348 722	0.0005

From the statistics reported in table 10, we can remark that:

- Even for a high  $minsup$  value (*i.e.*, 0.95), the number of generic exact and transitive fuzzy rules is very important ranging from 688 rules (for an excerpt of 100 genes) to 1 957 257 (for an excerpt of 900 genes).
- The number of generic fuzzy rules is by far smaller than that of the total number of FARs. In fact, the total number of association rules that can be extracted from any context have been assessed to be equal to  $2^{2 \times l}$ , where  $l$  is the length of the longest frequent itemset [35]. So, if we consider an excerpt of 200 items (genes), the number of FARs that can be extracted from such a context will be roughly equal to  $2^{400}$  (a number that we cannot even read). Thus, fuzzy generic rules ( $\mathcal{GBEF}$  and  $\mathcal{RTF}$ ) drastically reduce this number (reaching 115 741 generic fuzzy rules), without loss of information.

Table 12 reports the compactness provided by the pair ( $\mathcal{RTF}$ ,  $\mathcal{GBEF}$ ) generic bases vs. the total number of all redundant fuzzy association rules that we extracted from respectively the Chess and Mushroom datasets. From the reported statistics in Table 12, we can point out that:

Table 11: Evolution of the generic fuzzy association rules number vs.  $minconf$  value variation for the Chess and Mushroom datasets.

Dataset ( $minsup$ %)	$minconf$ (%)	Exact. FARs $ \mathcal{GBEF} $	Tran. FARs $ \mathcal{RTF} $	$ \mathcal{GBEF}  +  \mathcal{RTF} $
CHESS ( <b>70</b> )	30	406	756	1 162
	70	406	756	1 162
	80	406	727	1 133
	90	406	394	800
	100	406	84	490
MUSHROOM ( <b>40</b> )	30	836	2 752	3 588
	70	836	2 335	3 171
	80	836	1 949	2 785
	90	836	1 381	2 217
	100	836	856	1 692

- The reduced number of generic exact fuzzy association rules is due to the small difference between the number of fuzzy minimal generators and that of the fuzzy closed itemsets. Even though the Chess and Mushroom datasets are considered as dense ones, this behavior is typical of sparse datasets. This fact can be due to the process of fuzzification of these datasets.
- The obtained compactness ratios are very interesting. The ratio flagged by the Mushroom dataset shows that it is more dense than the Chess dataset.

Table 12: Compactness rate of ( $\mathcal{RTF}$ ,  $\mathcal{GBEF}$ ) for CHESS and MUSHROOM datasets.

Dataset	$minsup$ %	$minconf$ %	$ \mathcal{RTF}  +  \mathcal{GBEF} $	Redundant FARs	Compactness ratio
CHESS	70	100	490	58 484	0.0083
MUSHROOM	50	100	1 692	189 564	0.0089

## 6. Conclusion

In this paper, we presented a novel approach for the extraction of generic bases of fuzzy association rules. Beyond an interesting scalability feature, the proposed approaches aims at :

1. Reinforcing the user intervention within the knowledge mining process by giving him the possibility to interact within in order to highlight the importance of items or attributes.
2. The extraction of compact and information lossless fuzzy association rules. We also provided an expansion mechanism via a sound and complete axiomatic system, to enable the derivation of all redundant fuzzy association rules whenever of need. The experiments shed light on important rates of compactness.

Other avenues for future work mainly address the following issues:

1. Handling type-2 fuzzy contexts: To the best of our knowledge, no previous work was led for defining closure operator for type-2 fuzzy contexts. In addition, it is worth carrying out an in depth study of the applicability of other definitions of generic bases of association rules, e.g., the Guigues-Duquenne generic basis is known to present the maximal compactness rate for exact association rules [7].
2. A thorough study of the approximation by factorization of fuzzy concept lattices: This study was initiated by Bělohlavek *et al.* [10] and presents the smart idea to get out more comprehensible but less accurate approximations of the original concept lattice by means of factorization of the original fuzzy context. In this respect, we have to put the focus on the scalability issue of such an approach as well as the scrutiny of the trade off between accuracy of the extracted knowledge and approximation.
3. The design of alternative interestingness measures: During the last decade, the designing of quality measures has become an important challenge in data mining [22]. As stated in the well known "No Free Lunch Theorems", interestingness measures have many different qualities or flaws, since there is no optimal outstanding measure. In this respect, we are trying to introduce alternative interestingness measures, e.g. that presented in [9], to that based on the confidence measure. Even though it is very popular, the latter quality measure was shown to present many flaws specially from a statistical point of view.

## References

- [1] Agrawal, R., Imielinski, T., Swami, A., May 1993. Mining associations between sets of items in massive databases. In: Proceedings of the Intl. Conference on Management of Data.

- [2] Au, W., Chan, K., November 1999. FARM: A data mining system for discovering fuzzy association rules. In: Proceedings of FUZZ-IEEE'99, Seoul, South Korea. Vol. III. pp. 22–25.
- [3] Bastide, Y., Pasquier, N., Taouil, R., Lakhal, L., Stumme, G., July 2000. Mining minimal non-redundant association rules using frequent closed itemsets. In: Proceedings of the International Conference DOOD'2000, LNAI, volume 1861, Springer-Verlag, London, UK. pp. 972–986.
- [4] Bastide, Y., Pasquier, N., Taouil, R., Lakhal, L., Stumme, G., July 2000. Mining minimal non-redundant association rules using frequent closed itemsets. In: Proceedings of the International Conference DOOD'2000, LNCS, Springer-Verlag. pp. 972–986.
- [5] Bellman, R., Kalaba, L., Zadeh, L., 1966. Abstraction and pattern classification. *Journal of Mathematical Analysis and Applications* 13, 1–7.
- [6] Belohlavěk, R., Vychodil, V., 2005. What is a fuzzy concept lattice? In: Proceedings of the Int. Conf. on Concept Lattice and their Applications, Oloumouc, Czech Republic. pp. 34–45.
- [7] Ben Yahia, S., Gasmı, G., Mephu Nguifo, E., 2008. A new generic basis of factual and implicative association rules. Accepted to appear in *Intelligent Data Analysis*.
- [8] BenYahia, S., Jaoua, A., March 2001. Discovering knowledge from fuzzy concept lattice. In: A. Kandel, M. L., Bunke, H. (Eds.), *Data mining and Computational Intelligence, Studies in Fuzziness and Soft Computing*. Vol. 68. pp. 167–190. Physica-Verlag.
- [9] Berzal, F., Blanco, I., Sanchez, D., Vila, M., 2002. Measuring the accuracy and interest of association rules: A new framework. *Intelligent Data Analysis* 6 (3), 221–235.
- [10] Bělohlavěk, R., Outrata, J., Vychodil, V., April 2008. Fast factorization by similarity of fuzzy concept lattices with heges. *the International Journal of Foundations of Computer Sciences (IJFCS)* 19 (2), 255–270.
- [11] C-H.Weng, Chen, Y.-L., 2008. Mining fuzzy association rules from imprecise ordinal data. *Fuzzy Sets and Systems* (159), 460–474.
- [12] C-H.Weng, Chen, Y.-L., june 2009. Mining fuzzy association rules from uncertain data. *Knowledge Information Systems*, \*\*\*\*\*\_\*\*\*.

- [13] Chan, K., Au, W., November 1997. Mining fuzzy association rules. In: Proceedings of the Sixth Int'l Conf. on Information and Knowledge Management (CIKM'97), Las Vegas, Nevada, USA. pp. 209–215.
- [14] Chen, G., Wei, Q., Kerre, E. E., 2000. Fuzzy data mining: Discovery of fuzzy generalized association rules. In: G. Bordogna and G. Pasi, editors, Recent Issues on Fuzzy Databases. Springer-Verlag.
- [15] Chen, Y.-L., Weng, C.-H., 2008. Mining association rules from imprecise ordinal data. *Fuzzy Sets and Systems* (159), 460–474.
- [16] Delgado, M., Marin, N., Martin-Bautista, M., Sanchez, D., Vila, M., 2003. Mining fuzzy association rules: An overview. In: Proceedings of BISC International Workshop on Soft Computing for Internet and Bioinformatics.
- [17] Delgado, M., Sanchez, D., Vila, M., N.Marin, 2003. Fuzzy association rules: General model and applications. *IEEE Transactions on Fuzzy Systems* 11, 214–225.
- [18] Dubois, D., Hüllermeier, E., Prade, H., 2006. A systematic approach to the assessment of fuzzy association rules. *Data Mining and Knowledge Discovery* 13.
- [19] Dubois, D., Prade, H., 1991. Fuzzy sets in approximate reasoning, part II: Logical approaches. *Fuzzy Sets and Systems Journal* 40, 203–244.
- [20] Dubois, D., Prade, H., Sudkamp, T., April 2005. On the representation, measurement, and discovery of fuzzy associations. *IEEE Transactions on Fuzzy Systems* 13 (2), 250–262.
- [21] Ganter, B., Wille, R., 1999. *Formal Concept Analysis*. Springer-Verlag.
- [22] Guillet, F., Hamilton, H., 2007. Quality measures in data mining. Vol. 43 of *Studies in Computational Intelligence*. Springer-Verlag.
- [23] Hackman, J., Oldham, G., 1975. Development of the job diagnostic survey. *Journal of Applied Psychology* 60, 159–170.
- [24] He, Y., Tang, Y., Zhang, Y., Sunderraman, R., 2006. Adaptive fuzzy association rule mining for effective decision support in biomedical applications. *Int. J. Data Mining and Bioinformatics* 1 (1), 107–123.
- [25] Helm, B. L., August 2007. Fuzzy association rules an implementation in R. Master thesis, Vienna University of Economics and Business Administration, Vienna.

- [26] Kryszkiewicz, M., 1998. Representative association rules and minimum condition maximum consequence association rules. In: Proceedings of Second European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD), 1998, LNCS, volume 1510, Springer-Verlag, Nantes, France. pp. 361–369.
- [27] Kryszkiewicz, M., 2001. Concise representation of frequent patterns based on disjunction-free generators. In: Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM), San Jose, California, USA. pp. 305–312.
- [28] Kuok, C., Fu, A., Wong, M., 1998. Mining fuzzy association rules in databases. SIGMOD Record 27 (1), 41–46.
- [29] Rioult, F., Robardet, C., Blachon, S., Crémilleux, B., Gandrillon, O., Boulicaut, J., 2003. Mining concepts from large sage gene expression matrices. In: Proceedings of the Int. Conf. on Knowledge Discovery In Databases. pp. 107–118.
- [30] Srikant, R., Agrawal, R., June 1996. Mining quantitative association rules in large relational tables. In: Proceedings of the ACM-SIGMOD Int. Conf., Montreal, Canada. pp. 1–12.
- [31] Stumme, G., Taouil, R., Bastide, Y., Pasquier, N., Lakhal, L., September 2001. Intelligent structuring and reducing of association rules with formal concept analysis. In: Proc. KI'2001 conference, LNAI 2174, Springer-Verlag. pp. 335–350.
- [32] Tseng, S.-M., June 2001. Mining association rules with interestingness constraints in large databases. International Journal of Fuzzy Systems 3 (2), 415–421.
- [33] Yeung, D., Tsang, E., 1997. Weighted fuzzy production rules. Fuzzy Sets and Systems Journal 88, 299–313.
- [34] Zaki, M., August 2000. Generating non-redundant association rules. In: Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, MA. pp. 34–43.
- [35] Zaki, M. J., 2004. Mining non-redundant association rules. In: Data Mining and Knowledge Discovery. Vol. 9. pp. 223–248.