Managing the Absence of Items in Fuzzy Association Mining

C. Molina[1] D. Sanchez[2] J.M. Serrano[1] M.A. Vila[2]

1.Department of Computer Science, University of Jaen
Jaen, Spain
2.Department of Computer Science and Artificial Intelligence, University of Granada
Granada, Spain
Email: carlosmo@ujaen.es, daniel@decsai.ugr.es, jschica@ujaen.es, vila@decsai.ugr.es

Abstract— *One of the most well-known and extended data mining techniques is that of association rule mining, a helpful tool to discover relations between items present in sets of transactions. Nevertheless, in some other scenarios, another interesting issue is that of considering not only the possible relations involving presence of items, but the absence of them. The problem gets more complex when it is necessary to represent also imprecision and/or uncertainty in the information. In this paper, we introduce a methodology to obtain fuzzy association rules involving absent items. Additionally, our proposal is based on restriction level sets, a recent representation of imprecision that extends that of fuzzy sets, and introduces some new operators, covering some misleading results obtained from usual fuzzy operators as, for example, negation. In our methodology, we define new measures of interest and accuracy for fuzzy association rules as RL-numbers, as well as we propose a new way of summarizing the resulting set of fuzzy association rules, distributed in restriction levels.*

Keywords— Absence of items, fuzzy association rules, restriction levels

1 Introduction

As the amount of information stored in databases grows continuously day after day, it is desirable the development of tools, not only to properly manage all this knowledge, but to get a profit from it. Over the past two decades, considerable efforts have been devoted to the study and development of techniques in data mining and knowledge discovery in databases. One of the most well-known and extended data mining techniques is that of association rule mining, a helpful tool to discover implicit, non-trivial and potentially useful relations between items in sets of transactions. The most known example if that of market baskets, but association rules can be also applied on relational databases. Since the original association rule definition was proposed in [1], many studies and related methodologies have been devoted to extend this tool in order to manage different scenarios and knowledge representations.

One interesting issue, approached in works like [2], and [3], is that of considering both present and absent items. In these cases, it is not only interesting to detect relations between items included in transactions, but also if the absence of given items in the transaction can be related to the presence of some others. One possible application can be the discovery of conflicts or complements between items, as well as the establishment of constraints in the sets of items involved in the data mining procedure. Some other authors have shown interest in this problem, and additional information and efficient algorithms can be found in [4],[5],[6], and [7].

On the other hand, another interesting issue is that of applying the theory of fuzzy sets [8] in the modeling of how items are included in transactions, leading to the definition of fuzzy association rules. In the literature, some interesting approaches are discussed in [9], [10], [11], [12], and [13], among others.

But, as far as we know, none of the cited works manage properly the absence of items in the fuzzy case. In particular, consider an item i that belongs to a fuzzy transaction with a membership degree of 0.5. According to the usual fuzzy extension of the complement operator, the complementary item $\neg i$ should belong to the same transaction with a membership degree of $1 - 0.5 = 0.5$, leading us to counterintuitive results. Restriction level sets [14], a recently introduced representation of fuzzy quantities that extends the representation by means of fuzzy sets, manage to solve scenarios like these.

The representation via restriction level sets can be seen as equivalent to that of fuzzy sets seen as collections of alpha cuts, and hence it offers a alternative approach to the representation and management of fuzzy quantities, differing from fuzzy sets in the sense that operations are defined on levels. However, some operations as negation result in a representation that may not correspond to the alpha cuts of a fuzzy set.

Actually, although a parallel definition of a logic model for association rules based also on restriction level sets can be found in [15], our proposal, in this paper, is a methodology for mining fuzzy association rules via restriction level sets, instead of fuzzy sets, extending the definition introduced in [16], and, in addition, considering both present and absent items.

The paper is organized as follows. Sections 2 and 3 are devoted to introduce the reader in the followed representations and previous methodologies. Next, we introduce our proposal in section 4, where we define new measures for fuzzy association rules, based on RL-numbers, and we describe some derived properties from these. Next, some aspects with respect to the implementation and our first experimentation results are discussed in section 5. Finally, we present our concluding remarks as well as propose some interesting open issues in the conclusions sections.

2 Data mining tools

2.1 *Association rules*

Given a set I (“set of items”) and a set of transactions T (also called T-set), each transaction being a subset of I, association

ISBN: 978-989-95079-6-8

rules [1] are "implications" of the form $A \Rightarrow C$ that relate the presence of itemsets (sets of items) A and C in transactions of T, assuming $A, C \subseteq I$, $A \cap C = \emptyset$ and $A, C \neq \emptyset$.

In the case of relational databases, it is usual to consider that items are pairs $\langle attribute, value \rangle$, and transactions are tuples in a table. For example, the item $\langle X, x_0 \rangle$ is in the transaction associated to a tuple t iff $t[X] = x_0$.

The ordinary measures proposed in [1] to assess association rules are *confidence* (the conditional probability $p(C|A)$) and *support* (the joint probability $p(A \cup C)$). An alternative framework was proposed in [17]. In this framework, accuracy is measured by means of Shortliffe and Buchanan's certainty factors [18], in the following way: the certainty factor of the rule $A \Rightarrow C$ is

$$CF(A \Rightarrow C) = \frac{(Conf(A \Rightarrow C)) - S(C)}{1 - S(C)} \quad (1)$$

if $Conf(A \Rightarrow C) > S(C)$, and

$$CF(A \Rightarrow C) = \frac{(Conf(A \Rightarrow C)) - S(C)}{S(C)} \quad (2)$$

if $Conf(A \Rightarrow C) < S(C)$, and 0 otherwise.

Certainty factors take values in $[-1, 1]$, indicating the extent to which our belief that the consequent is true varies when the antecedent is also true. It ranges from 1, meaning maximum increment (i.e., when A is true then C is true) to -1, meaning maximum decrement.

2.2 Fuzzy association rules

In [16], the model for association rules is extended in order to manage fuzzy values in databases. The approach is based on the definition of fuzzy transactions as fuzzy subsets of items. Let $I = \{i_1, \ldots, i_m\}$ be a set of items and T' be a set of fuzzy transactions, where each fuzzy transaction is a fuzzy subset of I. Let $\tilde{\tau} \in T'$ be a fuzzy transaction, we note $\tilde{\tau}(i_k)$ the membership degree of i_k in $\tilde{\tau}$. A fuzzy association rule is an implication of the form $A \Rightarrow C$ such that $A, C \subset I$ and $A \cap C = \emptyset$.

It is immediate that the set of transactions where a given item appears is a fuzzy set. We call it *representation* of the item. For item i_k in T' we have the following fuzzy subset of T':

$$\tilde{\Gamma}_{i_k} = \sum_{\tilde{\tau} \in T'} \tilde{\tau}(i_k)/\tilde{\tau} \quad (3)$$

This representation can be extended to itemsets as follows: let $I_0 \subset I$ be an itemset, its representation is the following subset of T':

$$\tilde{\Gamma}_{I_0} = \bigcap_{i \in I_0} \tilde{\Gamma}_i = min_{i \in I_0} \tilde{\Gamma}_i \quad (4)$$

In order to measure the interest and accuracy of a fuzzy association rule, we must use approximate reasoning tools, because of the imprecision that affects fuzzy transactions and, consequently, the representation of itemsets. In [16], a semantic approach is proposed based on the evaluation of quantified sentences (see [19]). Let Q be a fuzzy coherent quantifier:

- The support of an itemset $\tilde{\Gamma}_{I_0}$ is equal to the result of evaluating the quantified sentence Q of T' are $\tilde{\Gamma}_{I_0}$.
- The support of the fuzzy association rule $A \Rightarrow C$ in the FT-set T', $Supp(A \Rightarrow C)$, is the evaluation of the quantified sentence Q of T are $\tilde{\Gamma}_{A \cup C}$ = Q of T are $(\tilde{\Gamma}_A \cap \tilde{\Gamma}_C)$.
- The confidence of the fuzzy association rule $A \Rightarrow C$ in the FT-set T', $Conf(A \Rightarrow C)$, is the evaluation of the quantified sentence Q of $\tilde{\Gamma}_A$ are $\tilde{\Gamma}_C$.

As seen in [16], the proposed method is a generalization of the ordinary association rule assessment framework in the crisp case.

3 Restriction-level representation

3.1 Representation

The RL-representation of an imprecise property is a collection of crisp sets, each crisp set corresponding to a crisp realization of the property under a *restriction* rule. We distinguish between *atomic* and *derived* properties. Atomic properties are those that cannot be defined in terms of other properties in our problem. Derived properties are defined by logical operations on other properties.

In [14] we consider that atomic imprecise properties are represented by fuzzy sets, and hence restrictions are of the form $degree \geq \alpha$ with $\alpha \in (0, 1]$, and restriction levels are associated to values $\alpha \in (0, 1]$. In the same case, the crisp realization of an atomic imprecise property represented by a fuzzy set A in the restriction level α corresponds to the α-cut A_α.

For every property we assume that there is a finite set of restriction levels $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ verifying that $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$, $m \geq 1$. We call such sets *RL-sets*. The consideration that a RL-set is finite is not a practical limitation since humans are able to distinguish a limited number of restriction or precision levels and, in practice, the limit in precision and storage of computers allows us to work with a finite number of degrees (and consequently, of levels) only. In practice, the RL-set for an atomic property represented by a fuzzy set A on an universe X is

$$\Lambda_A = \{A(x) \mid x \in support(A)\} \cup \{1\} \quad (5)$$

The RL-set employed to represent a derived property is obtained as the union of the RL-sets of the atomic properties in terms of which the property is defined. Finally, the RL-representation of an imprecise property on X is defined in [14] as follows:

Definition 1 *A RL-representation is a pair (Λ, ρ) where Λ is a RL-set and ρ is a function*

$$\rho : \Lambda \rightarrow \mathcal{P}(X) \quad (6)$$

The function ρ indicates the crisp realization that represents the imprecise property for each restriction level. As an example, the RL-representation for an atomic imprecise property defined by a fuzzy set A is the pair (Λ_A, ρ_A), where Λ_A is obtained using equation (5), and $\rho_A(\alpha) = A_\alpha \; \forall \alpha \in \Lambda_A$.

Given an imprecise property P represented by (Λ_P, ρ_P), we define the set of crisp representatives of P, Ω_P, as

$$\Omega_P = \{\rho_P(\alpha) \mid \alpha \in \Lambda_P\} \quad (7)$$

ISBN: 978-989-95079-6-8

For an atomic property A, the set of crisp representatives Ω_A is the set of significant α-cuts of A, as we have seen. However, notice that in definition 1 there is no restriction about the possible crisp representatives for non-atomic properties. In particular, as a consequence of operations, they don't need to be nested, so the final RL-representation of a derived property is not always equivalent to the α-cut representation of fuzzy sets.

In order to define properties by operations, it is convenient to extend the function ρ to any RL $\alpha \in (0,1]$. Let (Λ, ρ) be a RL-representation with $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ verifying $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$. Let $\alpha \in (0,1]$ and $\alpha_i, \alpha_{i+1} \in \Lambda$ such that $\alpha_i \geq \alpha > \alpha_{i+1}$. Then

$$\rho(\alpha) = \rho(\alpha_i) \tag{8}$$

Finally, let us remark that a RL-representation (Λ, ρ) on X is a crisp set $A \subseteq X$ iff $\forall \alpha \in \Lambda$, $\rho(\alpha) = A$.

3.2 Interpretation in terms of evidence

Given a RL-representation (Λ_A, ρ_A) for an atomic property A, the values of Λ_A can be interpreted as values of possibility of a possibility measure defined $\forall \rho_A(\alpha_i) \in \Omega_A$ as

$$Pos(\rho_A(\alpha_i)) = \alpha_i. \tag{9}$$

Following this interpretation we define a basic probability assignment in the usual way:

Definition 2 *Let (Λ, ρ) be a RL-representation with crisp representatives Ω. The associated probability distribution $m : \Omega \to [0,1]$ is*

$$m(Y) = \sum_{\alpha_i \mid Y = \rho(\alpha_i)} \alpha_i - \alpha_{i+1}. \tag{10}$$

The basic probability assignment m_F gives us information about how representative of the property F is each crisp set in Ω_F. For each $Y \in \Omega_F$, the value $m_F(Y)$ represents the proportion to which the available evidence supports the claim that the property F is represented by Y. From this point of view, a RL-representation can be seen as a basic probability assignment in the sense of the theory of evidence, *plus a structure indicating dependencies between the possible representations of different properties*.

3.3 RL-numbers

On the basis of RL-representations and operations, RL-numbers were introduced in [20] as a representation of imprecise quantities.

Definition 3 *A RL-real number is a pair $(\Lambda, \mathcal{R})$ where Λ is a RL-set and $\mathcal{R} : (0,1] \to \mathbb{R}$.*

We shall note $\mathbb{R}_{RL}$ the set of RL-real numbers. The RL-real number R_x is the representation of a (precise) real number x iff $\forall \alpha \in \Lambda_{R_x}$, $\mathcal{R}_{R_x}(\alpha) = x$. We shall denote such RL-real number as R_x or, equivalently, x, since in the crisp case, the set Λ_{R_x} is unimportant. Operations are extended as follows:

Definition 4 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $R_1 \ldots R_n$ be RL-real numbers. Then $f(R_1, \ldots, R_n)$ is a RL-real number with*

$$\Lambda_{f(R_1,\ldots,R_n)} = \bigcup_{1 \leq i \leq n} \Lambda_{R_i} \tag{11}$$

and, $\forall \alpha \in \Lambda_{f(R_1,\ldots,R_n)}$

$$\mathcal{R}_{f(R_1,\ldots,R_n)}(\alpha) = f(\mathcal{R}_{R_1}(\alpha), \ldots, \mathcal{R}_{R_n}(\alpha)) \tag{12}$$

This approach offers two main advantages:

- RL-numbers are representations of fuzzy quantities that can be easily obtained by extending usual crisp measurements to fuzzy sets.
- Arithmetic and logical operations on RL-numbers are straightforward and unique extensions of the operations on crisp numbers, verifying the following:
  - They verify all the usual properties of crisp arithmetic and logical operations.
  - The imprecision does not necessarily increase through operations, and can even diminish. The maximum imprecision is related to the number of restriction levels employed.

4 Fuzzy association rules via restriction levels

As we noted before, restriction level sets can be viewed as sets of alpha cuts in fuzzy sets. Actually, if the restriction level set represents an atomic property, the representation is equal to that of fuzzy sets. Following this idea, we extend the definition of fuzzy association rules introduced in [16], using restriction level sets instead. Moreover, since each restriction level corresponds to a crisp set, and crisp operators hold for each level, we can define a fuzzy association rule via restriction levels as an aggregation of all the related (crisp) association rules extracted at each restriction level.

In our approach, we represent items via restriction levels as follows. Let $I = \{i_1, \ldots, i_m\}$ be a set of items and T' be a set of fuzzy transactions, where each fuzzy transaction is a restriction level set of I. Let $\tilde{\tau} \in T'$ be a fuzzy transaction, we note $\tilde{\tau}(i_k)$ the membership degree of i_k in $\tilde{\tau}$. Let $supp_\alpha(i_k)$ be the support of i_k (i.e., the probability that $\tilde{\tau}(i_k) > \alpha, \forall \tilde{\tau} \in T'$) at the restriction level α.

On the other hand, let $\neg i_k$ be the negation (absence) of item i_k, and let $supp_\alpha(\neg i_k) = 1 - supp_\alpha(i_k)$ be the support of $\neg i_k$ at the restriction level α. Note how, in the case of negated items, the subsequent restriction level set may not correspond necessarily to a fuzzy set.

Let $\bar{I} = \{\neg i_1, \ldots, \neg i_m\}$ be the set of negated items from I, and $A \subset \{I \cup \bar{I}\}, C \in \{I \cup \bar{I}\}, A \cap C = \emptyset$, the antecedent and the consequent of the fuzzy association rule $R : A \Rightarrow C$. If we see this fuzzy association rule R under its RL-representation as the pair (Λ_R, ρ_R), we define the set of crisp representations of R, Ω_R, as

$$\Omega_R = \{\rho_R(\alpha) \mid \alpha \in \Lambda_R\} \tag{13}$$

where $\rho_R)(\alpha) = (R)_\alpha$, that is, ρ represents the association rules at each restriction level. In other words, this fuzzy association rule can be viewed as the aggregation of all (crisp)

ISBN: 978-989-95079-6-8

association rules of the form $A \Rightarrow C$ that can be obtained at each restriction level α.

Let $Supp_\alpha(A \Rightarrow C)$, $Conf_\alpha(A \Rightarrow C)$, and $CF_\alpha(A \Rightarrow C)$ be, respectively, the support, confidence and certainty factor of the crisp association rule $Supp_\alpha(A \Rightarrow C)$ at restriction level α.

Following the definitions in section 2.2, we must define the measures of interest and accuracy of a fuzzy association rule in terms of evaluation of quantified sentences. Hence, we define the support, confidence and certainty factor of the fuzzy association rule $A \Rightarrow C$ as RL-numbers, according to the RL-evaluation process proposed in [21].

Definition 5 *Let m_R be the basic probability assignment for association rule $R : A \Rightarrow C$ at each restriction level. The support of the fuzzy association rule $A \Rightarrow C$, RL-Supp$(A \Rightarrow C)$, is given by*

$$RLSupp(A \Rightarrow C) = \sum_{(A \Rightarrow C)_\alpha \in \Omega_R} m_R((A \Rightarrow C)_\alpha) \cdot Supp_\alpha(A \Rightarrow C) \quad (14)$$

Definition 6 *The confidence of the fuzzy association rule $R : A \Rightarrow C$, RL-Conf$(A \Rightarrow C)$, is given by*

$$RLConf(A \Rightarrow C) = \sum_{(A \Rightarrow C)_\alpha \in \Omega_R} m_R((A \Rightarrow C)_\alpha) \cdot Conf_\alpha(A \Rightarrow C) \quad (15)$$

Definition 7 *The certainty factor of the fuzzy association rule $R : A \Rightarrow C$, RL-CF$(A \Rightarrow C)$, is given by*

$$RLCF(A \Rightarrow C) = \sum_{(A \Rightarrow C)_\alpha \in \Omega_R} m_R((A \Rightarrow C)_\alpha) \cdot CF_\alpha(A \Rightarrow C) \quad (16)$$

These measures, RL-Supp$(A \Rightarrow C)$, RL-Conf$(A \Rightarrow C)$ and RL-CF$(A \Rightarrow C)$, can be viewed as the summarization of, respectively, the support, confidence, and certainty factor of every crisp association rule $A \Rightarrow C$ obtained at each restriction level. In fact, if the considered association rule has no negated items involved, the following propositions (as generalizations of the one proposed in [21]) hold.

Proposition 1 *The RL-support of the fuzzy association rule $A \Rightarrow C$, RL-Supp$(A \Rightarrow C)$, is equal to the support of the rule as computed in [16], using the method GD proposed in [22].*

Proposition 2 *The RL-confidence of the fuzzy association rule $A \Rightarrow C$, RL-Conf$(A \Rightarrow C)$, is equal to the confidence of the rule as computed in [16], using the method GD proposed in [22].*

Let us remark that these same propositions for RL-Supp and RL-Conf are not applicable for RL-CF since, again according to [16], the certainty factor for a fuzzy association rule was still computed in the same way as for a crisp one. Thus, we are defining a new measure of accuracy for fuzzy association rules, via restriction levels. The derived properties for this new measure will be the object of study in a separate paper.

Finally, let us notice that, due to the followed RL-representation, and opposite to the fuzzy sets case, it is not possible to find itemsets containing both items i and $\neg i$, which not only is desirable but also sounds logical.

5 Experiments

5.1 Implementation aspects

In our implementation, we start from the same Apriori-based algorithm proposed in [16], with the exception that we must compute the measures of interest and accuracy in terms of evidence in restriction levels, following definitions 5, 6, and 7.

Additionally, we must take into account that now we consider also absent/negated items, which result in a larger number of itemsets to be computed. I.e., if the original Apriori algorithm [1] takes into account up to 2^n itemsets, in our proposal we must consider all the combinations of items and negated items (except those impossible cases containing both i and $\neg i$), resulting in a total number of $(n+1) \cdot 2^n \approx n \cdot 2^n$ itemsets. In each iteration, we must count the occurrences of items in the set of transactions at each restriction level, but also the absences of them. This results in that when considering combinations of items and negated items, we must correctly compute, for each restriction level, the correct support for each itemset. I.e., $supp(\{\neg i_1, \neg i_2\}) = 1 - (supp(i_1) + supp(i_2) - supp(\{i_1, i_2\}))$, $supp(\{i_1, \neg i_2\}) = supp(i_1) - supp(\{i_1, i_2\})$, and so on. One important remark is that we need to keep all the generated itemsets, in order to have access to their support values when necessary. This fact increases the amount of necessary memory, and delays the threshold pruning (by minimum support, etc.) until the second stage of the algorithm (association rule extraction).

According to this, the final efficiency of the algorithm is approximately that of classic Apriori algorithm, but multiplied by $(n+1)$ (due to negated items consideration) and by k (being k the number of restriction levels considered). One of our pending tasks will be the study and development of a more efficient algorithm. In this sense, one choice is that of parallelize the overall process. Since each one of the restriction levels is a crisp set, we can reduce the problem to that of concurrently extract crisp association rules at each level, and then aggregate the obtained measures. We will address this aspect in a future paper.

5.2 Experimentation results

In order to test the resulting algorithm, we first performed our experiments over artificially generated set of transactions. We randomly generated a set of transactions, involving 100 items (and their negations), considered $k = 10$ restriction levels, and obtained a total number of 39600 fuzzy association rules. For space saving purposes, we restrict the number of items in a rule to 2, 1 in the antecendent and 1 in the consequent. In forthcoming papers, we will afford the study of more complex rules.

Each fuzzy association rule has a related set of (crisp) association rules, present in each restriction level. Table 2 shows an example subset of the obtained rules, after establishing a threshold of minimum support and minimum certainty factor of 0.5. Note that, after establishing these thresholds, a rule

ISBN: 978-989-95079-6-8

Table 1: Example of extracted fuzzy association rules

$id\#$	Rule	Evidence	RL-Supp	RL-CF
#39283	$\neg i_{90} \Rightarrow \neg i_{92}$	0.4	0.31	0.31
#39295	$\neg i_{90} \Rightarrow \neg i_{95}$	0.4	0.30	0.25
#39347	$\neg i_{91} \Rightarrow \neg i_{92}$	0.4	0.31	0.31
#39359	$\neg i_{91} \Rightarrow \neg i_{95}$	0.1	0.05	0.05
#39463	$\neg i_{93} \Rightarrow \neg i_{95}$	0.5	0.38	0.36
#39503	$\neg i_{94} \Rightarrow \neg i_{95}$	0.5	0.38	0.36
#39519	$\neg i_{94} \Rightarrow \neg i_{99}$	0.6	0.45	0.38
#39591	$\neg i_{97} \Rightarrow \neg i_{99}$	0.5	0.38	0.37
#39599	$\neg i_{98} \Rightarrow \neg i_{99}$	0.5	0.38	0.37

may not appear in all restriction levels. For sake of simplicity and space saving matters, only the rule number is shown. See Table 1 for a more complete description of some of the rules.

Applying definitions 5 and 7, we can summarize the support and certainty factor of the rules in the restriction levels, and compute the measures for the fuzzy association rules. Table 1 shows some of the obtained fuzzy association rules. The rule numbers correspond to the referred ones in previous Table 2.

In addition, another interesting issue of our proposed methodology is that, in terms of the restriction levels evidence, we can summarize also the obtained set of fuzzy association rules. Actually, we can reduce the number of rules present at each restriction level, by discarding those with low support and/or low certainty factor. After this pruning, we can compute the associated probability distribution (see definition 10) for each rule, and then interpret the resulting set of association rules in terms of evidence. For example, the association rules shown in Table 2 could be summarize according to their evidence in the following expression,

$$\begin{aligned} Ruleset = \{\#39519\}/0.6 + \\ \{\#39463, \#39503, \#39591, \#39599\}/0.5 + \\ \{\#39283, \#39295, \#39347\}/0.4 \end{aligned} \tag{17}$$

Again, for space saving purposes, we only refer to the rules by its rule number. Let us remark that this expression is not actually a fuzzy set, but still can be very helpful in order to interpret the set of results, as we relate a relevance degree, the basic probability assignment, to each rule or set of rules.

Finally, we applied our proposed methodology on real data, obtained from a database containing soil information (see [23] for a more complete description). We reduce the set of data to those attributes modeled as fuzzy quantities ($Averagerainfall$, $Altitude$, $Depth$, PH, and attributes describing the soil chemical composition). A set of linguistic labels $\{Low, Medium, High\}$ was defined for every attribute. Table 3 shows some of the obtained fuzzy association rules, all having $Evidence = 1.0$, restraining the thresholds of minimum support to 0.5 and minimum certainty factor to 0.7, for space saving purposes.

Let us notice that considering absent items allows us to improve the semantics of the rules. I.e., considering item $\neg Avg.Rainfall = Low$ allows us to represent both $Avg.Rainfall = Medium$ and $Avg.Rainfall = High$ concepts. This aspect will be addressed in more detail in a forthcoming paper.

Table 3: Example of fuzzy association rules on real data

Rule	RL-Supp	RL-CF
$PH = High \Rightarrow \neg Avg.Rain. = Low$	0.616	0.97
$PH = Low \Rightarrow \neg Avg.Rain. = Medium$	0.616	0.97
$PH = High \Rightarrow \neg Avg.Rain. = Medium$	0.616	0.97
$PH = High \Rightarrow \neg Avg.Rain. = High$	0.616	0.97
$PH = Medium \Rightarrow \neg Avg.Rain. = Low$	0.616	0.97
$PH = Low \Rightarrow \neg Avg.Rain. = Low$	0.616	0.97
$PH = Medium \Rightarrow \neg Avg.Rain. = Medium$	0.616	0.97
$PH = Low \Rightarrow \neg Avg.Rain. = High$	0.616	0.97
$PH = Medium \Rightarrow \neg Avg.Rain. = High$	0.616	0.97
$\neg Alt. = Low \Rightarrow \neg Avg.Rain. = High$	0.810	0.83
$\neg Alt. = Medium \Rightarrow \neg Avg.Rain. = High$	0.810	0.83
$\neg Alt. = High \Rightarrow \neg Avg.Rain. = Low$	0.810	0.83
$\neg Alt. = High \Rightarrow \neg Avg.Rain. = High$	0.810	0.83
$\neg Alt. = Medium \Rightarrow \neg Avg.Rain. = Low$	0.810	0.83
$\neg Alt. = High \Rightarrow \neg Avg.Rain. = Medium$	0.810	0.83
$\neg Alt. = Low \Rightarrow \neg Avg.Rain. = Medium$	0.810	0.83
$\neg Alt. = Medium \Rightarrow \neg Avg.Rain. = Medium$	0.810	0.83
$\neg Alt. = Low \Rightarrow \neg Avg.Rain. = Low$	0.810	0.83

6 Concluding remarks

We have introduced a new methodology to represent fuzzy association rules in terms of RL-sets, as an alternative representation to that of fuzzy sets. This methodology offers some new advantages as, for example, it allows us to properly consider issues as fuzzy association rules involving absent items. As far as we know, ours is a novel methodology, combining both presence and absence of items with imprecision and uncertainty management. Considering both presence and absence of items can lead us to discern possible conflicts between items or detect when a set of items complements another.

In addition, it also offers a method of summarizing the set of obtained rules, representing them in a format similar to that of fuzzy sets, and thus allowing us to help the user to discern the relevance of a given rule or set of rules, along with the usual measures of interest and accuracy. With respect to these measures, we have seen how the resulting measures values are the same to those obtained in the fuzzy set approach. In this sense, our approach is still coherent with respect to previous ones. Nevertheless, in the general case, and specially in the case of the measure of certainty factor, new challenges appears abroad, and the derived properties for these new definitions for measuring the accuracy in rules will be approached in a future paper.

Finally, we have extended an existing Apriori-based algorithm in order to consider a fuzzy association rule as the aggregation (RL-representation) of the related association rules

ISBN: 978-989-95079-6-8

Table 2: Example of (crisp) association rules distributed through restriction levels

RL \ $Rule$	#39283	#39295	#39347	#39359	#39463	#39503	#39519	#39591	#39599
1.0	×		×		×	×	×	×	×
0.9	×	×	×		×	×	×	×	×
0.8	×	×	×		×	×	×	×	×
0.7	×	×	×		×	×	×	×	×
0.6		×			×	×	×	×	×
0.5							×		
0.4				×					

at each restriction level. As the overall performance is affected by the higher number of itemsets to be taken into account, one pending task will be the optimization of this algorithm in order to improve its efficiency.

Acknowledgment

This work was developed as part of the research project TIN2006-07262.

References

[1] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. Of the 1993 ACM SIGMOD Conference*, pages 207–216, 1993.

[2] Craig Silverstein, Sergey Brin, and Rajeev Motwani. Beyond market baskets: Generalizing association rules to dependence rules. *Data Min. Knowl. Discov*, 2(1):39–68, 1998.

[3] Maria-Luiza Antonie and Osmar R. Zaane. Mining positive and negative association rules: an approach for confined rules. In *In European PKDD Conf*, pages 27–38, 2004.

[4] Xindong Wu, Chengqi Zhang, and Shichao Zhang. Efficient mining of both positive and negative association rules. *ACM Transactions on Information Systems*, 22(3):381–405, July 2004.

[5] Wei-Guang Teng, Ming-Jyh Hsieh, and Ming-Syan Chen. On the mining of substitution rules for statistically dependent items. In *ICDM*, pages 442–449. IEEE Computer Society, 2002.

[6] Ashok Savasere, Edward Omiecinski, and Shamkant Navathe. Mining for strong negative associations in a large database of customer transactions. pages 494–502. IEEE Computer Society, 1998.

[7] Xiaohui Yuan, Bill P. Buckles, Zhaoshan Yuan, and Jian Zhang. Mining negative association rules. *Computers and Communications, IEEE Symposium on*, 0:623, 2002.

[8] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[9] W.H. Au and K.C.C. Chan. Mining fuzzy association rules. In *Proc. Of 6th Int. Conf. On Information and Knowledge Management. Las Vegas, NV, USA*, pages 209–215, 1997.

[10] A.W.C. Fu, M.H. Wong, S.C. Sze, W.C. Wong, W.L. Wong, and W.K. Yu. Finding fuzzy sets for the mining of fuzzy association rules for numerical attributes. In *Proc. Int. Symp. On Intelligent Data Engineering and Learning (IDEAL'98), Hong Kong*, pages 263–268, 1998.

[11] S. Ben-Yahia and A. Jaoua. A top-down approach for mining fuzzy association rules. In *Proc. 8th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'2000*, pages 952–959, 2000.

[12] Shu-Yue-J, Tsang-E, Yenng-D, and Daming-Shi. Mining fuzzy association rules with weighted items. In *Proc. IEEE Int. Conf. On Systems, Man and Cybernetics*, 2000.

[13] D. Dubois, H. Prade, and T. Sudkamp. A discussion of indices for the evaluation of fuzzy associations in relational databases. In T. Bilgic, B. De Baets, and O. Kaynak, editors, *Fuzzy Sets and Systems - Proceedings IFSA'03, LNAI 2715*, pages 111–118. Springer, 2003.

[14] D. Sánchez, M. Delgado, and M.A. Vila. A restriction level approach to the representation of imprecise properties. In *IPMU 2008*, pages 153–159, 2008.

[15] M. Delgado, M.D. Ruiz, and D. Sánchez. A restriction level approach for the representation and evaluation of fuzzy association rules. Submitted to IFSA'2009, 2009.

[16] M. Delgado, N. Marín, D. Sánchez, and M.A. Vila. fuzzy association rules: General model and applications. *IEEE Transactions on Fuzzy Systems 11(2)*, pages 214–225, 2003.

[17] F. Berzal, I. Blanco, D. Sánchez, and M.A. Vila. Measuring the accuracy and interest of association rules: A new framework. *Intelligent Data Analysis*, 6:221–235, 2002.

[18] E. Shortliffe and B. Buchanan. A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23:351–379, 1975.

[19] L.A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computing and Mathematics with Applications*, 9(1):149–184, 1983.

[20] D. Sánchez, M. Delgado, and M.A. Vila. RL-numbers: An alternative to fuzzy numbers for the representation of imprecise quantities. In *FUZZ-IEEE 2008*, pages 2058–2065, 2008.

[21] D. Sánchez, Delgado M., and M.A. Vila. Fuzzy quantification using restriction levels. Submitted to WILF'2009, 2009.

[22] M. Delgado, D. Sánchez, and M.A. Vila. Fuzzy cardinality based evaluation of quantified sentences. *International Journal of Approximate Reasoning*, 23:23–66, 2000.

[23] J. Calero, G. Delgado, M. Sánchez-Marañón, D. Sanchez, MA Vila, and J. Serrano. An experience in management of imprecise soil databases by means of fuzzy association rules and fuzzy approximate dependencies. In *Proceedings of the 6th international conference on enterprise information systems, ICEIS*, volume 4, pages 138–146. Springer, 2004.

ISBN: 978-989-95079-6-8