Fuzzy Association Rules: General Model and Applications

Miguel Delgado, Nicolás Marín, Daniel Sánchez, and María-Amparo Vila

***Abstract*—The theory of fuzzy sets has been recognized as a suitable tool to model several kinds of patterns that can hold in data. In this paper, we are concerned with the development of a general model to discover association rules among items in a (crisp) set of fuzzy transactions. This general model can be particularized in several ways; each particular instance corresponds to a certain kind of pattern and/or repository of data. We describe some applications of this scheme, paying special attention to the discovery of fuzzy association rules in relational databases.**

***Index Terms*—Association rules, data mining, fuzzy transactions, quantified sentences.**

I. Introduction

KNOWLEDGE discovery, whose objective is to obtain useful knowledge from data stored in large repositories, is recognized as a basic necessity in many areas, specially those related to business. Since data represent a certain real-world domain, patterns that hold in data show us interesting relations that can be used to improve our understanding of that domain. Data mining is the step in the knowledge discovery process that attempts to discover novel and meaningful patterns in data.

The theory of fuzzy sets can certainly help data mining to reach this goal [1]. It is widely recognized that many real world relations are intrinsically fuzzy. For instance, fuzzy clustering generally provides a more suitable partition of a set of objects than crisp clustering do. Moreover, fuzzy sets are an optimal tool to model imprecise terms and relations as commonly employed by humans in communication and understanding. As a consequence, the theory of fuzzy sets is an excellent basis to provide knowledge expressed in a meaningful way.

One of the best studied models for data mining is that of *association rules* [2]. This model assumes that the basic object of our interest is an *item*, and that data appear in the form of sets of items called *transactions*. Association rules are "implications" that relate the presence of items in transactions. The classical example are the rules extracted from the content of market baskets. Items are things we can buy in a market, and transactions are market baskets containing several items. Association rules relate the presence of items in the same basket, for example, "every basket that contains bread contains butter," usually noted bread $\Rightarrow$ butter.

Transactions are the basic structure of data, from which association rules are obtained. However, as we mentioned above, in many cases real-world relations, and hence data patterns, are fuzzy rather than crisp. Even otherwise, it could happen that we were interested in mapping crisp data to fuzzy data (e.g., to diminish the granularity and/or to improve the semantic content of the patterns).

In this paper, we introduce the concept of fuzzy transaction as a fuzzy subset of items. In addition we present a general model to discover association rules in fuzzy transactions. We call them fuzzy association rules. One of the main problems is how to measure the support and accuracy of fuzzy rules. This problem is related to fuzzy cardinality [3]–[5] and evaluation of quantified sentences [6], [7]. We show how this general model can be particularized for different applications by mapping the abstract concepts of item and fuzzy transaction to different types of objects and structures.

The paper is organized as follows. Section II contains the general model for fuzzy association rules. Section III is devoted to explain applications of the model, in particular fuzzy association rules in relational databases. Section IV is an overview of related approaches to the discovery of fuzzy rules. Section V contains our conclusions and references to our future research.

II. General Model

A. Association Rules

Let I be a set of items and T a set of transactions with items in I, both assumed to be finite. An association rule is an expression of the form $A \Rightarrow C$, where $A, C \subseteq I$, $A, C \neq \emptyset$, and $A \cap C = \emptyset$. The rule $A \Rightarrow C$ means "every transaction of T that contains A contains C."

The usual measures to assess association rules are support and confidence, both based on the concept of support of an *itemset* (i.e., a subset of items). The support of an itemset $I_0 \subseteq I$ is

$$\mathrm{supp}(I_0, T) = \frac{|\{\tau \in T \mid I_0 \subseteq \tau\}|}{|T|} \tag{1}$$

i.e., the probability that a transaction of T contains I_0. The support of the association rule $A \Rightarrow C$ in T is

$$\mathrm{Supp}(A \Rightarrow C, T) = \mathrm{supp}(A \cup C) \tag{2}$$

and its confidence is

$$\mathrm{Conf}(A \Rightarrow C, T) = \frac{\mathrm{supp}(A \cup C)}{\mathrm{supp}(A)} = \frac{\mathrm{Supp}(A \Rightarrow C)}{\mathrm{supp}(A)}. \tag{3}$$

It is usual to assume that T is known, so the previous examples are noted as $\mathrm{supp}(I_0)$, $\mathrm{Supp}(A \Rightarrow C)$, and $\mathrm{Conf}(A \Rightarrow C)$, respectively. Notice that supp is the notation for items, and Supp for rules.

Manuscript received February 7, 2001; revised January 23, 2003.

The authors are with the Department of Computer Science and Artificial Intelligence, University of Granada, Granada 18071, Spain (e-mail: daniel@decsai.ugr.es).

Digital Object Identifier 10.1109/TFUZZ.2003.809896

TABLE I
SET T_6 OF FUZZY TRANSACTIONS

	i_1	i_2	i_3	i_4
$\tilde{\tau}_1$	0	0.6	0.7	0.9
$\tilde{\tau}_2$	0	1	0	1
$\tilde{\tau}_3$	1	0.5	0.75	1
$\tilde{\tau}_4$	1	0	0.1	1
$\tilde{\tau}_5$	0.5	1	0	1
$\tilde{\tau}_6$	1	0	0.75	1

Support is the percentage of transactions where the rule holds. Confidence is the conditional probability of C with respect to A or, in other words, the relative cardinality of C with respect to A. The techniques employed to mine for association rules attempt to discover rules whose support and confidence are greater than two user-defined thresholds called $\mathrm{minsupp}$ and $\mathrm{minconf}$, respectively. Such rules are called *strong rules*.

B. Fuzzy Transactions and Fuzzy Association Rules

Definition 1: A fuzzy transaction is nonempty fuzzy subset $\tilde{\tau} \subseteq I$.

For every $i \in I$, we note $\tilde{\tau}(i)$ the membership degree of i in a fuzzy transaction $\tilde{\tau}$. We note $\tilde{\tau}(I_0)$ the degree of inclusion of an itemset $I_0 \subseteq I$ in a fuzzy transaction $\tilde{\tau}$, defined as

$$\tilde{\tau}(I_0) = \min_{i \in I_0} \tilde{\tau}(i).$$

According to Definition 1, a transaction is a special case of fuzzy transaction. We represent a set of fuzzy transactions by means of a table. Columns and rows are labeled with identifiers of items and transactions, respectively. The cell for item i_k and transaction $\tilde{\tau}_j$ contains a [0,1]-value: the membership degree of i_k in $\tilde{\tau}_j$, $\tilde{\tau}_j(i_k)$.

Example 1: Let $I = \{i_1, i_2, i_3, i_4\}$ be a set of items. Table I shows six transactions defined on I.

Here, $\tilde{\tau}_1 = 0.6/i_2 + 0.7/i_3 + 0.9/i_4$, $\tilde{\tau}_2 = 1/i_2 + 1/i_4$, and so on. In particular, $\tilde{\tau}_2$ is a crisp transaction, $\tilde{\tau}_2 = \{i_2, i_4\}$.

Some inclusion degrees are $\tilde{\tau}_1(\{i_3, i_4\}) = 0.7$, $\tilde{\tau}_1(\{i_2, i_3, i_4\}) = 0.6$, $\tilde{\tau}_4(\{i_1, i_4\}) = 1$.

We call *T-set* a set of ordinary transactions, and *FT-set* a set of fuzzy transactions. Example 1 shows the FT-set $\{\tilde{\tau}_1, \ldots, \tilde{\tau}_6\}$ that contains six transactions. Let us remark that a FT-set is a crisp set.

Definition 2: Let I be a set of items, T a FT-set, and $A, C \subseteq I$ two crisp subsets, with $A, C \neq \emptyset$, and $A \cap C = \emptyset$. A fuzzy association rule $A \Rightarrow C$ holds in T iff

$$\tilde{\tau}(A) \leq \tilde{\tau}(C) \quad \forall \tilde{\tau} \in T$$

i.e., the inclusion degree of C is greater than that of A for every fuzzy transaction $\tilde{\tau}$.

This definition preserves the meaning of association rules, because if we assume $A \subseteq \tilde{\tau}$ in some sense, we must assume $C \subseteq \tilde{\tau}$ given that $\tilde{\tau}(A) \leq \tilde{\tau}(C)$. Since a transaction is a special case of fuzzy transaction, an association rule is a special case of fuzzy association rule.

C. Support and Confidence of Fuzzy Association Rules

1) Generalizing the Support/Confidence Framework: We employ a semantic approach based on the evaluation of quantified sentences [6]. A quantified sentence is an expression of the form "Q of F are G," where F and G are two fuzzy subsets of a finite set X, and Q is a relative fuzzy quantifier. Relative quantifiers are linguistic labels for fuzzy percentages that can be represented by means of fuzzy sets on [0,1], such as "most," "almost all," or "many."

A family of relative quantifiers, called coherent quantifiers [8], is specially relevant for us. Coherent quantifiers are those that verify the following properties:

- $Q(0) = 0$ and $Q(1) = 1$;
- if $x < y$, then $Q(x) \leq Q(y)$ (monotonicity).

An example is "many young people are tall," where $Q =$ many, and F and G are possibility distributions induced in the set $X =$ people by the imprecise terms "young" and "tall," respectively. A special case of quantified sentence appears when $F = X$, as in "most of the terms in the profile are relevant." The evaluation of a quantified sentence yields a [0,1]-value, that assesses the accomplishment degree of the sentence.

Definition 3: Let $I_0 \subseteq I$. The support of I_0 in T is the evaluation of the quantified sentence

$$Q \text{ of } T \text{ are } \widetilde{\Gamma}_{I_0}$$

where $\widetilde{\Gamma}_{I_0}$ is a fuzzy set on T defined as

$$\widetilde{\Gamma}_{I_0}(\tilde{\tau}) = \tilde{\tau}(I_0).$$

Definition 4: The support of the fuzzy association rule $A \Rightarrow C$ in the set of fuzzy transactions T is $\mathrm{supp}(A \cup C)$, i.e., the evaluation of the quantified sentence

$$Q \text{ of } T \text{ are } \widetilde{\Gamma}_{A \cup C} = Q \text{ of } T \text{ are } \left(\widetilde{\Gamma}_A \cap \widetilde{\Gamma}_C\right).$$

Definition 5: The confidence of the fuzzy association rule $A \Rightarrow C$ in the set of fuzzy transactions T is the evaluation of the quantified sentence

$$Q \text{ of } \widetilde{\Gamma}_A \text{ are } \widetilde{\Gamma}_C.$$

Let us remark that these definitions establish families of support and confidence measures, depending on the evaluation method and the quantifier of our choice. We have chosen to evaluate the sentences by means of method GD[7], that has been shown to verify good properties with better performance than others. The evaluation of "Q of F are G" by means of GD is defined as

$$GD_Q\left(\frac{G}{F}\right) = \sum_{\alpha_i \in \Delta(G/F)} (\alpha_i - \alpha_{i+1})\, Q\left(\frac{|(G \cap F)_{\alpha_i}|}{|F_{\alpha_i}|}\right) \tag{4}$$

where $\Delta(G/F) = \Lambda(G \cap F) \cup \Lambda(F)$, $\Lambda(F)$ being the level set of F, and $\Delta(G/F) = \{\alpha_1, \ldots, \alpha_p\}$ with $\alpha_i > \alpha_{i+1}$ for every

$i \in \{1, \ldots, p\}$. The set F is assumed to be normalized. If not, F is normalized and the normalization factor is applied to $G \cap F$.

2) Choosing a Quantifier: The choice of the quantifier allows us to change the semantics of the values in a linguistic way. This flexibility is very useful when using this general model to fit different types of patterns, as we shall see in the applications section. The evaluation of a quantified sentence "Q of F are G" by means of method GD can be interpreted as

- the evidence that the percentage of objects in F that are also in G (relative cardinality of G with respect to F) is Q[9];
- a quantifier-guided aggregation [10], [11] of the relative cardinalities of G with respect to F for each α cut of the same level of both sets.

Hence, $\mathrm{Supp}(A \Rightarrow C)$ can be interpreted as the evidence that the percentage of transactions in $\widetilde{\Gamma}_{A\cup C}$ is Q, and $\mathrm{Conf}(A \Rightarrow C)$ can be seen as the evidence that the percentage of transactions in $\widetilde{\Gamma}_A$ that are also in $\widetilde{\Gamma}_C$ is Q. In both cases, the quantifier is a linguistic parameter that determines the final semantics of the measures.

Many evaluation methods and quantifiers can be chosen to characterize and assess support and confidence of fuzzy association rules, provided that the following four intuitive properties of the measures for ordinary association rules hold.

1) If $\widetilde{\Gamma}_A \subseteq \widetilde{\Gamma}_C$, then $\mathrm{Conf}(A \Rightarrow C) = 1$.
2) If $\widetilde{\Gamma}_A \cap \widetilde{\Gamma}_C = \emptyset$, then $\mathrm{Supp}(A \Rightarrow C) = 0$ and $\mathrm{Conf}(A \Rightarrow C) = 0$.
3) If $\widetilde{\Gamma}_A \subseteq \widetilde{\Gamma}_{A'}$ (particularly when $A' \subseteq A$), then $\mathrm{Conf}(A' \Rightarrow C) \le \mathrm{Conf}(A \Rightarrow C)$.
4) If $\widetilde{\Gamma}_C \subseteq \widetilde{\Gamma}_{C'}$ (particularly when $C' \subseteq C$), then $\mathrm{Conf}(A \Rightarrow C) \le \mathrm{Conf}(A \Rightarrow C')$.

According to the properties of the evaluation method GD[7], it is easy to show that any coherent quantifier yields support and confidence measures that verify the four aforementioned properties.

We have chosen the quantifier Q_M defined by $Q_M(x) = x$, since it is coherent and the measures obtained by using it in Definitions 3–5 are the ordinary measures in the crisp case, as the following propositions show.

Proposition 1: Let $I_0 \subset I$ such that $\widetilde{\Gamma}_{I_0}$ is crisp. Then, $\mathrm{supp}(I_0)$ measured by GD with Q_M is the ordinary support of an itemset.

Proof: GD verifies that if F and G are crisp then the evaluation of "Q of F are G" is [7]

$$GD_Q\left(\frac{G}{F}\right) = Q\left(\frac{|F \cap G|}{|F|}\right).$$

Hence

$$\mathrm{supp}(I_0) = GD_{Q_M}\left(\frac{\widetilde{\Gamma}_{I_0}}{T}\right) = \frac{\left|\widetilde{\Gamma}_{I_0}\right|}{|T|}.$$

■

Proposition 2: Let $A \Rightarrow C$ be an ordinary association rule on T. Then, $\mathrm{Supp}(A \Rightarrow C)$, measured by GD with Q_M, is the ordinary support of the rule.

TABLE II
SUPPORT IN T_6 OF FOUR ITEMSETS

Itemset	Support
$\{i_1\}$	0.583
$\{i_4\}$	0.983
$\{i_2, i_3\}$	0.183
$\{i_1, i_3, i_4\}$	0.266

TABLE III
SUPPORT AND CONFIDENCE IN T_6 OF THREE FUZZY RULES

Rule	Support	Confidence
$\{i_2\} \Rightarrow \{i_3\}$	0.183	0.283
$\{i_1, i_3\} \Rightarrow \{i_4\}$	0.266	1
$\{i_1, i_4\} \Rightarrow \{i_3\}$	0.266	0.441

Proof: From the properties of GD

$$\begin{aligned}\mathrm{Supp}(A \Rightarrow C) =& GD_{Q_M}\left(\frac{\widetilde{\Gamma}_{A\cup C}}{T}\right)\\ =& \frac{\left|\widetilde{\Gamma}_{A\cup C}\right|}{|T|}\\ =& \mathrm{supp}(A \cup C).\end{aligned}$$

■

Proposition 3: Let $A \Rightarrow C$ be an ordinary association rule on T. Then, $\mathrm{Conf}(A \Rightarrow C)$, measured by GD with Q_M, is the ordinary confidence of the rule.

Proof: From the properties of GD

$$\begin{aligned}\mathrm{Conf}(A \Rightarrow C) =& GD_{Q_M}\left(\frac{\widetilde{\Gamma}_C}{\widetilde{\Gamma}_A}\right)\\ =& \frac{\left|\widetilde{\Gamma}_A \cap \widetilde{\Gamma}_C\right|}{\left|\widetilde{\Gamma}_A\right|}\\ =& \frac{\left|\widetilde{\Gamma}_{A\cup C}\right|}{\left|\widetilde{\Gamma}_A\right|}\\ =& \frac{\mathrm{Supp}(A \Rightarrow C)}{\mathrm{supp}(A)}.\end{aligned}$$

■

Hence, a possible interpretation of the values of the measures for crisp association rules is the evidence that the support (respectively, confidence) of the rule is Q_M. Unless a specific reference to the quantifier is given, from now on we shall consider support and confidence based on Q_M and GD. The study of the support/confidence framework with other quantifiers is left to future research.

Example 2: Let us illustrate the concepts introduced in this subsection. According to Definition 3, the support of several itemsets in Table I is shown in Table II.

Table III shows the support and confidence of several fuzzy association rules in T_6.

We remark that

$$\mathrm{Conf}(\{i_1, i_3\} \Rightarrow \{i_4\}) = GD_{Q_M}\left(\frac{\widetilde{\Gamma}_{\{i_4\}}}{\widetilde{\Gamma}_{\{i_1,i_3\}}}\right) = 1$$

since $\widetilde{\Gamma}_{\{i_1,i_3\}} \subseteq \widetilde{\Gamma}_{\{i_4\}}$.

D. New Framework to Measure Accuracy and Importance

Several authors have pointed out some drawbacks of the support/confidence framework to assess association rules [12], [13], [9], [14].

To avoid some of these drawbacks and to ensure that the discovered rules are interesting and accurate, a new approach was proposed in [9] and [14]. It employs certainty factors [15] and the new concept of very strong rule.

1) Certainty Factors:

Definition 6: We call certainty factor of a fuzzy association rule $A \Rightarrow C$ to the value

$$CF(A \Rightarrow C) = \frac{\mathrm{Conf}(A \Rightarrow C) - \mathrm{supp}(C)}{1 - \mathrm{supp}(C)}$$

if $\mathrm{Conf}(A \Rightarrow C) > \mathrm{supp}(C)$, and

$$CF(A \Rightarrow C) = \frac{\mathrm{Conf}(A \Rightarrow C) - \mathrm{supp}(C)}{\mathrm{supp}(C)}$$

if $\mathrm{Conf}(A \Rightarrow C) \leq \mathrm{supp}(C)$, assuming by agreement that if $\mathrm{supp}(C) = 1$ then $CF(A \Rightarrow C) = 1$ and if $\mathrm{supp}(C) = 0$ then $CF(A \Rightarrow C) = -1$.

The certainty factor takes values in $[-1, 1]$. It is positive when the dependence between A and C is positive, 0 when there is independence, and a negative value when the dependence is negative. The following proposition is an interesting property shown in [16].

Proposition 4: $\mathrm{Conf}(A \rightarrow C) = 1$ if and only if $CF(A \Rightarrow C) = 1$

This property guarantees that the certainty factor of a fuzzy association rule achieves its maximum possible value, 1, if and only if the rule is totally accurate.

From now on, we shall use certainty factors to measure the accuracy of a fuzzy association rule. We say that a fuzzy association rule is strong when its certainty factor and support are greater than two user-defined thresholds minCF and $\mathrm{minsupp}$, respectively.

2) Very Strong Rules:

Definition 7: A fuzzy association rule $A \Rightarrow C$ is very strong if both $A \Rightarrow C$ and $\neg C \Rightarrow \neg A$ are strong.

The itemsets $\neg A$ and $\neg C$, whose meaning is "absence of A" (respectively, C) in a transaction, are defined in the usual way: $\widetilde{\Gamma}_{\neg A}(\tilde{\tau}) = 1 - \widetilde{\Gamma}_A(\tilde{\tau})$ and $\widetilde{\Gamma}_{\neg C}(\tilde{\tau}) = 1 - \widetilde{\Gamma}_C(\tilde{\tau})$. The logical basis of this definition is that the rules $A \Rightarrow C$ and $\neg C \Rightarrow \neg A$ represent the same knowledge. Therefore, if both rules are strong we can be more certain about the presence of that knowledge in a set of transactions.

Experiments described in [9] and [14] have shown that by using certainty factors and very strong rules we avoid to report a large amount of false, or at least doubtful, rules. In some experiments, the number of rules was diminished by a factor of 20 and even more. This is important not only because the discovered rules are reliable, but also because the set of rules is smaller and more manageable.

E. Algorithms

Many papers have been devoted to develop algorithms to mine ordinary association rules. The early efficient algorithms like AIS [2], Apriori and AprioriTid [17], SETM [18], OCD [19], and DHP [20] were continued with more recent developments like DIC [12], CARMA [21], TBAR [22], and FP-Growth [23]. See [24] for a recent survey of the problem. Most of the existing algorithms work in two steps.

Step P.1. To find the itemsets whose support is greater than minsupp (the so-called *frequent itemsets*). In this step it is usual to consider transactions one by one, updating the support of the itemsets each time a transaction is considered. Algorithm A shows this basic procedure. This step is the most expensive from the computational point of view.

Step P.2. To obtain rules with accuracy greater than an user-defined threshold, from the frequent itemsets obtained in step P.1. Specifically, if the itemsets A and $A \cup C$ are frequent, we can obtain the rule $A \Rightarrow C$. The support of that rule will be high enough, since it is equal to the support of the itemset $A \cup C$. However, we must verify the accuracy of the rule, in order to determine whether it is strong.

In the algorithm of Appendix A, the items are processed in the order given by its size. First 1-itemsets, next 2-itemsets and so on. The variable l stores the actual size. The set L_l stores the l itemsets that are being analyzed and, at the end, it stores the frequent l itemsets. The procedure CreateLevel(i, L) generates a set of i-itemsets such that every proper subset with $i-1$ items is frequent (i.e., is in L_{i-1}) together with the associated counters. Since every proper subset of a frequent itemset is also a frequent itemset, we avoid analyzing itemsets that don't verify this property, hence, saving space and time.

The basic procedure described in the algorithm of Appendix A (i.e., the procedure of step P.1.) may be adapted to the case of fuzzy transactions. For that purpose, we apply the following changes.

- We store the difference between the cardinality of every α cut of $\widetilde{\Gamma}_{I_0}$ and the cardinality of the corresponding strong α cut, $\alpha \in [0, 1]$, for all the considered itemsets I_0. Specifically

$$\left|\left(\widetilde{\Gamma}_{I_0}\right)_\alpha\right| - \left|\left(\widetilde{\Gamma}_{I_0}\right)_{\alpha+}\right|$$

where

$$\left(\widetilde{\Gamma}_{I_0}\right)_\alpha = \left\{\tilde{\tau} \in T \mid \widetilde{\Gamma}_{I_0}(\tilde{\tau}) \geq \alpha\right\}$$
$$\left(\widetilde{\Gamma}_{I_0}\right)_{\alpha+} = \left\{\tilde{\tau} \in T \mid \widetilde{\Gamma}_{I_0}(\tilde{\tau}) > \alpha\right\}.$$

We use a fixed number of k equidistant α cuts, (specifically, $k = 100$). This information allows us to obtain the fuzzy cardinality of the representation of the items. It is stored in an array V_{I_0} that can be easily obtained from a FT-set in a similar way that the support of an itemset in a T-set: each time a transaction $\tilde{\tau}$ is considered, we add 1 to $V_{I_0}\left(\widetilde{\Gamma}_{I_0}(\tilde{\tau})\right)$ for every itemset I_0, see the algorithm in Appendix C

- We obtain the support and confidence from the information about the cardinality (the certainty factor is obtained directly from them according to definition 6). We use GD to evaluate the corresponding quantified sentences, by means of the algorithm in Appendix B, proposed in [25]. In this algorithm, k is a constant: the number of equidistant levels we are using. As we mentioned before, we shall employ Q_M and $k = 100$ in general. We have empirically shown that $k = 100$ is a good value. However, the algorithm is valid for any coherent quantifier Q and any value k, although it seems reasonable to request $k \geq 10$ in order to obtain a good representation of the cardinality of the itemsets. We shall deal with how to establish the minimum suitable value of k in future research.

The algorithm in Appendix C is a modification of the algorithm in Appendix A to find frequent itemsets in an FT-set. The function $R(x, k)$ maps the real value x to the nearest value in the set of k levels we are using.

One important objective for us is to keep (in the worst case) the complexity of the existing algorithms when they are modified in order to find fuzzy rules. This is accomplished, since algorithm B has complexity $O(1)$ and, hence, algorithm A has the same complexity than algorithm C. Space complexity remains the same as well. In both cases, the hidden constant is increased in a factor that depends on k, as this value affects the size of the arrays V.

To adapt step P.2. of the general procedure is rather straightforward. We obtain the confidence by using algorithm B to evaluate the corresponding sentence. We only modify this step in the sense that we obtain the certainty factor of the rules from the confidence and the support of the consequent, both available. Finally, it is easy to decide whether the rule is strong, because support and certainty factor of the rule are available in this step.

There are several possible solutions to deal with very strong rules. They are described in [16], and they can be easily adapted (as well as the basic algorithm) to find strong rules.

Let us remark that most of the existing association rule mining algorithms can be adapted in order to discover fuzzy association rules, by using algorithm B. In this work we are not interested in obtaining a very efficient algorithm, but in the semantics of the rules, so we have adapted a very basic algorithm (algorithm A). To adapt other algorithms is left to future research. In any case, the complexity of the adapted algorithms will be at worst the same as the complexity of the original ones.

III. Applications

"Item" and "transaction" are abstract concepts that are usually associated to "an object" and "a subset of objects." By particularizing them, association rules can provide different kinds of patterns. In this section we shall describe briefly how this simple idea yields different applications.

A. *Fuzzy Association Rules in Relational Databases*

Nowadays, most of data available all over the world are stored in relational databases. As such, the development of models to find patterns in relational databases is a must. Roughly speaking, data in relational databases are stored in tables, where each row is the description of an object and each column is one characteristic/attribute of the object. For each object (row) t, $t[X]$ stands for the value of attribute (column) X.

Algorithms to mine for association rules have been applied to represent patterns in relational databases by defining items as pairs $\langle$attribute, value$\rangle$ and transactions as tuples. The following formalization is described in [26]. Let $RE = \{X_1, \ldots, X_m\}$ be a set of attributes. We denote I^{RE} to the set of items associated to RE, i.e.,

$$I^{RE} = \{\langle X_j, x\rangle \text{ such that } x \in \mathrm{Dom}(X_j), j \in \{1, \ldots, m\}\}.$$

Every instance r of RE is associated to a T-set, denoted T^r, with items in I^{RE}. Each tuple $t \in r$ is associated to an unique transaction $\tau^t \in T^r$ in the following way:

$$\tau^t = \{\langle X_j, t\,[At_j]\rangle \mid j \in \{1, \ldots, m\}\}\,.$$

A special feature of transactions in relational databases is that no pair of items in one transaction share the same attribute, because of the first normal form constraint. Any other itemset must also verify this restriction.

One frequent problem in relational databases is the discovery of association rules involving quantitative attributes. Such rules are called *quantitative association rules* [27], and the problem is that its support and its semantic content use to be poor [26]. Moreover, the mining task can be more expensive [27], [28]. One commonly used solution is to split the domain of the quantitative attributes into intervals, and to take the set of clusters as the new domain of the attribute. Several approaches based on this idea have been proposed, either performing the clustering during the mining process [27], [29] or before it [30], [31]. This solution has two drawbacks: it is difficult for clusters to fit a meaningful (for users) concept [9], and the importance and accuracy of rules can be sensitive to small variations of boundaries [27], [32].

To avoid these drawbacks, a soft alternative is to define a set of meaningful linguistic labels represented by fuzzy sets on the domain of the quantitative attributes, and to use them as a new domain. Now, the meaning of the values in the new domain is clear, and the rules are not sensitive to small changes of the boundaries because they are fuzzy. With this solution, we have fuzzy transactions and rules.

The following formulation is a summary of that in [26]: let $\mathrm{Lab}(X_j) = \{L_1^{X_j}, \ldots, L_{c_j}^{X_j}\}$ be a set of linguistic labels for attribute X_j. We shall use the labels to name the corresponding fuzzy set, i.e.,

$$L_k^{X_j} : \mathrm{Dom}(X_j) \to [0, 1].$$

Let $L = \bigcup_{j \in \{1, \ldots, m\}} \mathrm{Lab}(X_j)$. Then, the set of items with labels in L associated to RE is

$$I_L^{RE} = \left\{ \langle X_j, L_k^{X_j}\rangle \mid X_j \in RE, \quad \begin{matrix} k \in \{1, \ldots, c_j\} \\ j \in \{1, \ldots, m\} \end{matrix} \right\}.$$

TABLE IV
AGE AND HOUR OF BIRTH OF SIX PEOPLE

	Age	Hour
t_1	60	20:15
t_2	80	23:45
t_3	22	15:30
t_4	55	01:00
t_5	3	19:30
t_6	18	06:51

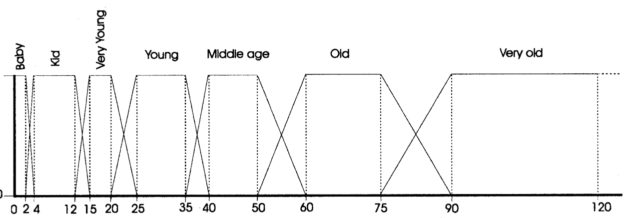

Fig. 1. Representation of some linguistic labels for "Age."

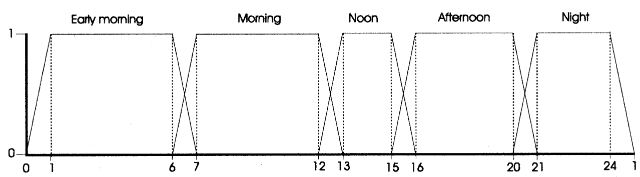

Fig. 2. Representation of some linguistic labels for "Hour."

Every instance r of RE is associated to a FT-set, denoted T_L^r, with items in I_L^{RE}. Each tuple $t \in r$ is associated to a unique fuzzy transaction $\tilde{\tau}_L^t \in T_L^r$

$$\tilde{\tau}_L^t : I_L^{RE} \rightarrow [0, 1]$$

such that

$$\tilde{\tau}_L^t \left(\left\langle X_j, L_k^{X_j} \right\rangle \right) = L_k^{X_j} \left(t[X_j] \right).$$

In this case, a fuzzy transaction can contain more than one item corresponding to different labels of the same attribute, because it is possible for a single value in the table to match more than one label to a certain degree. However, itemsets keep being restricted to contain at most one item per attribute, because otherwise fuzzy rules would not make sense (e.g., "if Height $=$ tall and Height $=$ medium then … ").

The following example is from [26].

Example 3: Let r be the relation of Table IV, containing the age and hour of birth of six people. The relation r is an instance of $ER = \{\text{Age}, \text{Hour}\}$.

We shall use for age the set of labels

$$\text{Lab(Age)} = \{\text{Baby}, \text{Kid}, \text{Very young}, \text{Young}, \text{Middle age}, \text{Old}, \text{Very old}\}$$

of Fig. 1. Fig. 2 shows a possible definition of the set of labels for hour

$$\text{Lab(Hour)} = \{\text{Early morning}, \text{Morning}, \text{Noon}, \text{Afternoon}, \text{Night}\}.$$

TABLE V
FUZZY TRANSACTIONS WITH ITEMS IN I_L^{ER} FOR THE RELATION OF TABLE IV

	$\tilde{\tau}_L^{t_1}$	$\tilde{\tau}_L^{t_2}$	$\tilde{\tau}_L^{t_3}$	$\tilde{\tau}_L^{t_4}$	$\tilde{\tau}_L^{t_5}$	$\tilde{\tau}_L^{t_6}$
$\langle Age, Baby \rangle$	0	0	0	0	**0.5**	0
$\langle Age, Kid \rangle$	0	0	0	0	**0.5**	0
$\langle Age, Very\ young \rangle$	0	0	**0.6**	0	0	**1**
$\langle Age, Young \rangle$	0	0	**0.4**	0	0	0
$\langle Age, Middle\ age \rangle$	0	0	0	**0.5**	0	0
$\langle Age, Old \rangle$	**1**	**0.67**	0	**0.5**	0	0
$\langle Age, Very\ old \rangle$	0	**0.33**	0	0	0	0
$\langle Hour, Early\ morning \rangle$	0	0	0	**1**	0	**0.85**
$\langle Hour, Morning \rangle$	0	0	0	0	0	**0.15**
$\langle Hour, Noon \rangle$	0	0	**0.5**	0	0	0
$\langle Hour, Afternoon \rangle$	**0.75**	0	**0.5**	0	**1**	0
$\langle Hour, Night \rangle$	**0.25**	**1**	0	0	0	0

Then

$$L = \text{Lab(Age)} \cup \text{Lab(Hour)}$$

and

$$\begin{aligned} I_L^{ER} = \{ & \langle \text{Age, Baby} \rangle, \langle \text{Age, Kid} \rangle, \langle \text{Age, Very young} \rangle, \\ & \langle \text{Age, Young} \rangle, \langle \text{Age, Middle age} \rangle, \\ & \langle \text{Age, Very Old} \rangle, \langle \text{Age, Old} \rangle, \\ & \langle \text{Hour, Early morning} \rangle, \langle \text{Hour, Morning} \rangle, \\ & \langle \text{Hour, Noon} \rangle, \langle \text{Hour, Afternoon} \rangle, \langle \text{Hour, Night} \rangle \}. \end{aligned}$$

The FT-set T^r on I_L^{ER} is

$$T_L^r = \{\tilde{\tau}_L^{t_1}, \tilde{\tau}_L^{t_2}, \tilde{\tau}_L^{t_3}, \tilde{\tau}_L^{t_4}, \tilde{\tau}_L^{t_5}, \tilde{\tau}_L^{t_6}\}.$$

The columns of Table V define the fuzzy transactions of T_L^r as fuzzy subsets of I_L^{ER} (we have interchanged rows and columns from the usual representation of fuzzy transactions, for the sake of space). For instance

$$\begin{aligned} \tilde{\tau}_L^{t_1} &= \left\{ \frac{1}{\langle \text{Age, Old} \rangle} + \frac{0.75}{\langle \text{Hour, Afternoon} \rangle} + \frac{0.25}{\langle \text{Hour, Night} \rangle} \right\} \\ \tilde{\tau}_L^{t_3} &= \left\{ \frac{0.6}{\langle \text{Age, Very young} \rangle} + \frac{0.4}{\langle \text{Age, Young} \rangle} + \frac{0.5}{\langle \text{Hour, Noon} \rangle} + \frac{0.5}{\langle \text{Hour, Afternoon} \rangle} \right\}. \end{aligned}$$

In Table V, the row for item i_L contains the fuzzy set $\tilde{\Gamma}^r_{\{i_L\}}$. For instance

$$\begin{aligned} \tilde{\Gamma}^r_{\{\langle \text{Age,Old} \rangle\}} &= \left\{ \frac{1}{\tilde{\tau}_L^{t_1}} + \frac{0.67}{\tilde{\tau}_L^{t_2}} + \frac{0.5}{\tilde{\tau}_L^{t_4}} \right\} \\ \tilde{\Gamma}^r_{\{\langle \text{Hour,Night} \rangle\}} &= \left\{ \frac{0.25}{\tilde{\tau}_L^{t_1}} + \frac{1}{\tilde{\tau}_L^{t_2}} \right\}. \end{aligned}$$

Descriptions of itemsets with more than one fuzzy item are, for instance

$$\begin{aligned} \tilde{\Gamma}^r_{\{\langle \text{Age,Old} \rangle, \langle \text{Hour,Night} \rangle\}} &= \left\{ \frac{0.25}{\tilde{\tau}_L^{t_1}} + \frac{0.67}{\tilde{\tau}_L^{t_2}} \right\} \\ \tilde{\Gamma}^r_{\{\langle \text{Age,Kid} \rangle, \langle \text{Hour,Afternoon} \rangle\}} &= \left\{ \frac{0.5}{\tilde{\tau}_L^{t_5}} \right\}. \end{aligned}$$

TABLE VI
SOME ATTRIBUTES FROM THE CENSUS DATABASE

Name	Description
AMARITL	Marital status
AHGA	Education
ACLSWKR	Class of worker
AAGE	Age
ARACE	Race
ASEX	Sex
PENATVTY	Country of birth

TABLE VII
NUMBER OF RULES OBTAINED USING SEVERAL ACCURACY MEASURES AND THRESHOLDS, $\mathrm{minsupp} = 0.05$

α	**0.1**	**0.5**	**0.9**
minConf=α	182	82	18
minCF=α	62	14	8
minCF=α and VSR	60	14	8

Some rules involving fuzzy items in I_L^{ER} are

$$\langle \text{Age, Old} \rangle \Rightarrow \langle \text{Hour, Afternoon} \rangle$$
$$\langle \text{Hour, Afternoon} \rangle \Rightarrow \langle \text{Age, Baby} \rangle.$$

We have used the general model presented in this paper to find fuzzy association rules in relational databases. Algorithms, implementations and experimental results are detailed in [26], [9].

We have considered the situation when the database is crisp and we employ a set of linguistic labels defined by fuzzy sets on the domains. However we cannot forget the possibility that data is fuzzy. Several fuzzy relational database models have been developed and implemented. In this situation, models like those in Section II are needed.

Experiments: We have applied the model to discover fuzzy association rules in the CENSUS database. The database we have worked with was extracted by Lane and Kohavi using the Data Extraction System from the census bureau database.[1] Specifically, we have worked with a test database containing 99 762 instances, obtained from the original database by using MineSet's MIndUtil mineset-to-mlc utility.

The database contains 40 attributes, but we have employed only those in Table VI. Let us remark that in relational databases, the usual interpretation is that items take the form [attribute = value].

We have employed the set of fuzzy labels in Fig. 1 to diminish the granularity of AAGE. Instead of items like $\langle \text{AAGE} = 7 \rangle$ we have worked with items of the form $\langle \text{AAGE} = L \rangle$ with $L \in \text{Lab(Age)}$, like for example $\langle \text{AAGE} = \text{Kid} \rangle$.

A comparison between the number of rules obtained by using confidence, certainty factors, and very strong rules with different accuracy thresholds is shown in Tables VII and VIII. Only rules with a single item in the antecedent and consequent have been considered. These tables have been obtained by using $\mathrm{minsupp} = 0.05$ and $\mathrm{minsupp} = 0.15$, respectively. As it is shown in [14], misleading rules are discarded since certainty factors are able to detect statistical independence and negative dependence, and hence less rules (though more reliable) are obtained.

[1] http://www.census.gov/ftp/pub/DES/www/welcome.html

TABLE VIII
NUMBER OF RULES OBTAINED USING SEVERAL ACCURACY MEASURES AND THRESHOLDS, $\mathrm{minsupp} = 0.15$

α	**0.1**	**0.5**	**0.9**
minConf=α	82	45	10
minCF=α	28	9	5
minCF=α and VSR	18	7	5

TABLE IX
FUZZY ASSOCIATION RULES THAT RELATE AAGE AND PENATVTY IN THE CENSUS DATABASE

Rule	Supp	Conf	CF
$\langle AAGE, Baby \rangle \Rightarrow \langle PENATVTY, USA \rangle$	0.052	0.98	0.85
$\langle AAGE, Kid \rangle \Rightarrow \langle PENATVTY, USA \rangle$	0.157	0.94	0.56
$\langle AAGE, Very\ young \rangle \Rightarrow \langle PENATVTY, USA \rangle$	0.108	0.89	0.07
$\langle AAGE, Young \rangle \Rightarrow \langle PENATVTY, USA \rangle$	0.194	0.83	-0.05
$\langle AAGE, Middle\ age \rangle \Rightarrow \langle PENATVTY, USA \rangle$	0.194	0.85	-0.03
$\langle AAGE, Old \rangle \Rightarrow \langle PENATVTY, USA \rangle$	0.160	0.89	0.05

If we focus on very strong rules, we discard those rules whose consequent has very high support (which are misleading rules). The best results are achieved when $\mathrm{minsupp} = 0.15$. This fact can be appreciated in Table VIII, since the reduction in the number of rules between the last two rows is more important than in Table VIII.

Table IX shows some rules that relate AAGE and PENATVTY with support greater than 0.05. According to the confidence criterion, all the rules in Table IX are very interesting if we consider $\mathrm{minconf} \leq 0.83$ (a rather high value). However, not all the rules are intuitive. For example, the rule for $\langle \text{AAGE} = \text{Old} \rangle$ tell us that if we know that a person is Old, we should believe that she was born in the U.S. with confidence 0.89. The certainty factor give us a more appropriate value of 0.05, meaning there is almost independence between both facts.

In the case of $\langle \text{AAGE} = \text{Baby} \rangle$, both confidence and certainty factor are high, but this is more reasonable because most of the babies we can find in the U.S. were born in the U.S. In fact, as age increases, we tend to believe that age and country of birth are more independent, as can be appreciated by looking at the certainty factors of the rules in Table IX.

The problem with confidence here is that the support of the item $\langle \text{PENATVTY, U.S.} \rangle$ is 0.88, and confidence does not take this into account. As a consequence, most items seem to be a good predictor of being born in the U.S., in particular those related to age. Certainty factors do take into account the support of the consequent, so they provide more accurate information. As we can see in this example, confidence with $\mathrm{minconf} = 0.83$ yields six rules, while certainty factors provide only one with the same threshold (two if we use $\mathrm{minCF} = 0.56$). Something similar happens when we find $\langle \text{ARACE, White} \rangle$ in the consequent, since its support is 0.83.

Other reasonable rules are shown in Table X.

B. Fuzzy and Approximate Functional Dependencies

In the previous section, we have explained the usual concept of association rule in a relational database. But by using a suit-

TABLE X
SOME FUZZY ASSOCIATION RULES THAT HOLD IN THE CENSUS DATABASE

Rule	Supp	Conf	CF
$\langle AAGE, Baby\rangle \Rightarrow \langle AMARITL, NeverMarried\rangle$	0.05	0.99	0.99
$\langle AAGE, Baby\rangle \Rightarrow \langle AHGA, Children\rangle$	0.16	1	1
$\langle AAGE, Baby\ young\rangle \Rightarrow \langle ACLSWKR, Notinuniverse\rangle$	0.05	1	1
$\langle AAGE, Kid\rangle \Rightarrow \langle AHGA, Children\rangle$	0.16	1	1
$\langle AAGE, Kid\ young\rangle \Rightarrow \langle ACLSWKR, Notinuniverse\rangle$	0.16	1	1

able definition of items and transactions, fuzzy association rules can be employed to define and to mine other kind of structures. Some of them are related to the concept of functional dependence.

Let RE be a set of attributes and r an instance of RE. A functional dependence $X \to Y, X, Y \subset RE$, holds in r if the value $t[X]$ determines $t[Y]$ for every tuple $t \in r$. The dependence holds in RE if it holds in every instance of RE. Formally, a functional dependence that holds in a relation r is a rule of the form

$$\forall t, s \in r \text{ if } t[X] = s[X], \qquad \text{then } t[Y] = s[Y]. \tag{5}$$

To disclose such knowledge is very interesting but, like in the case of association rules, it is difficult to find perfect dependencies, mainly because of the usual existence of exceptions. To cope with this, two main groups of smoothed dependencies have been proposed: fuzzy functional dependencies and approximate dependencies. The former introduce some fuzzy components into (5) (e.g., the equality can be replaced by a similarity relation) while the latter establishes the functional dependencies with exceptions (i.e., with some uncertainty). Approximate dependencies can be interpreted as a relaxation of the universal quantifier in rule (5). A detailed study of different definitions of fuzzy and approximate dependencies can be found in [33], [9], and [34].

We have used association rules to represent approximate dependencies. For this purpose, transactions and items are associated to pairs of tuples and attributes respectively. We consider that the item associated to the attribute X, I_X, is in the transaction τ_{ts} associated to the pair of tuples $\langle t, s\rangle$ when $t[X] = s[X]$. The set of transactions associated to an instance r of RE is denoted T_r, and contains $|r|^2$ transactions. We define an approximate dependence in r to be an association rule in T_r[9], [35], and [34].

The support and certainty factor of an association rule $I_X \Rightarrow I_W$ in T_r measure the importance and accuracy of the approximate dependence $X \to W$. The main drawback of this approach is the computational complexity of the process, since $|T_r| = |r|^2$ and the complexity of algorithm C is linear on the number of transactions. We have solved the problem by analyzing several transactions at a time. The algorithm, detailed in [9] and [34], stores the support of every item of the form $\langle X, x\rangle$ with $x \in \mathrm{Dom}(X)$ in order to obtain the support of I_X. Its complexity is linear on the number of tuples in r. An additional feature is that it finds dependencies and the associated models at the same time. We have shown that our definitions and algorithms provide a reasonable and manageable set of dependencies [9], [34]. The following example is from [34].

TABLE XI
TABLE r_3 WITH DATA ABOUT THREE STUDENTS

ID	Year	Course	Lastname
1	1991	3	Smith
2	1991	4	Smith
3	1991	4	Smith

TABLE XII
SET T_{r_3} OF TRANSACTIONS FOR r_3

Pair	it_{ID}	it_{Year}	it_{Course}	$it_{Lastname}$
$\langle 1,1\rangle$	1	1	1	1
$\langle 1,2\rangle$	0	1	0	1
$\langle 1,3\rangle$	0	1	0	1
$\langle 2,1\rangle$	0	1	0	1
$\langle 2,2\rangle$	1	1	1	1
$\langle 2,3\rangle$	0	1	1	1
$\langle 3,1\rangle$	0	1	0	1
$\langle 3,2\rangle$	0	1	1	1
$\langle 3,3\rangle$	1	1	1	1

TABLE XIII
SOME ASSOCIATION RULES IN T_{r_3} THAT DEFINE APPROXIMATE DEPENDENCIES IN r_3

Ass. Rule	Confidence	Support	App. dependence
$\{it_{ID}\} \Rightarrow \{it_{Year}\}$	1	1/3	$ID \to Year$
$\{it_{Year}\} \Rightarrow \{it_{Course}\}$	5/9	5/9	$Year \to Course$
$\{it_{Year}, it_{Course}\} \Rightarrow \{it_{ID}\}$	3/5	1/3	$Year, Course \to ID$

Example 4: To illustrate our definition of AD, we shall use the relation r_3 of Table XI. It is an instance of $RE = \{ID, \text{Year}, \text{Course}, \text{Lastname}\}$.

Table XII shows the T-set T_{r_3} and Table XIII contains some association rules that hold in T_{r_3}. They define approximate dependencies that hold in r_3. Confidence and support of the association rules in Table XIII measure the accuracy and support of the corresponding dependencies.

Fuzzy association rules are necessary in this context when quantitative attributes are involved. Our algorithms provide not only an approximate dependency, but also a model that consists of a set of association rules (in the usual sense in relational databases) relating values of the antecedent with values of the consequent of the dependencies. The support and certainty factor of the dependencies have been shown to be related to the same measures of the rules in the model [34]. However, when attributes are quantitative, this model suffers from the same problem discussed in the previous subsection. To cope with this, we propose to use a set of linguistic labels. A set of labels $\mathrm{Lab}(X_j)$ induces a fuzzy similarity relation $S_{\mathrm{Lab}(X_j)}$ in the domain of X in the following way:

$$S_{\mathrm{Lab}(X_j)}(x_1, x_2) = \max_{L_k^{X_j} \in \mathrm{Lab}(X_j)} \min\left(L_k^{X_j}(x_1), L_k^{X_j}(x_2)\right)$$

for all $x_1, x_2 \in \mathrm{Dom}(X_j)$, assuming that for every $x \in \mathrm{Dom}(X_j)$ there is one $L_k^{X_j} \in \mathrm{Lab}(X_j)$ such that $L_k^{X_j}(x) = 1$.

Then, the item I_X is in the transaction τ_{ts} with degree $S_{\mathrm{Lab}(X_j)}(t[X], s[X])$. Now, the transactions for the table r are fuzzy, and we denote $T_{S_L}^r$ this FT-set. In this new situation, we can find approximate dependencies in r by looking for fuzzy association rules in $T^r_{S_L}$. The model of such approximate dependencies will be a set of association rules in T^r_L. These dependencies can be used to summarize data in a relation.

This is only one of the possible relaxations of functional dependencies into fuzzy functional dependencies [33]. We have shown that most of the existing relaxations can be defined by replacing the equality and the universal quantifier in rule 5 by a similarity relation S and a fuzzy quantifier Q respectively [36].

For instance, let S_Y be a resemblance relation [37] and $\varphi \in (0, 1]$ such that

$$S(y_1, y_2) = \begin{cases} 1, & S_Y(y_1, y_2) \geq \varphi \\ 0, & \text{otherwise.} \end{cases}$$

Also, let $Q = \forall$, with

$$\forall(x) = \begin{cases} 1 & x = 1 \\ 0, & \text{otherwise.} \end{cases}$$

The fuzzy functional dependency $X \to W$ defined by [38]

$$\forall t, s \in r, \qquad \text{if } t[X] = s[X] \text{ then } S_Y(t[Y], s[Y]) \geq \varphi$$

can be modeled in r by an association rule in T^r_S. Here, T^r_S stands for the FT-set of fuzzy similarities given by S between pairs of tuples of r.

We have also faced a more general problem: the integration of fuzzy and approximate dependencies in what we have called *fuzzy quantified dependencies* [36] (i.e., fuzzy functional dependencies with exceptions). Let us remark that our semantic approach, based on the evaluation of quantified sentences, allows to assess rules in a more flexible way. Hence, to deal with certain kinds of patterns is possible, as we have seen before.

C. Gradual Rules

Gradual rules are expressions of the form "The more X is L^X_i, the more Y is L^Y_j", like "the more *Age* is *Young*, the more *Height* is *Tall*." The semantics of that rules are "the greater the membership degree of the value of X in L^X_i, the greater the membership degree of the value of Y in L^Y_j" [39], [40]. There are several formal specifications of this idea. In [33], the authors propose

$$\forall t \in r \quad L^X_i(t[X]) \leq L^Y_j(t[Y]). \tag{6}$$

In this case, the items are pairs $\langle \text{Attribute}, \text{Label} \rangle$ and the transactions are associated to tuples. The item $\langle X, L^X_i \rangle$ is in the transaction τ_t associated to the tuple t when $L^X_i(t[X]) \leq L^Y_j(t[Y])$, and the set of transactions, denoted $T^r_{G_1(L)}$, is a T-set. This way, ordinary association rules in $T^r_{G_1(L)}$ are gradual rules in r. A more general expression of this kind is

$$\forall t \in r \quad L^X_i(t[X]) \to_* L^Y_j(t[Y]) \tag{7}$$

where $\to_*$ is a fuzzy implication. Expression (6) is a particular case where Rescher–Gaines implication ($\alpha \to_{R-G} \beta = 1$ when $\alpha \leq \beta$ and 0 otherwise) is employed. One interesting possibility is to use the FT-set T^r_L described in Section III-A and the quantifier $Q = \forall$. From the properties of GD, the evaluation of "$\forall$ of F are G" provides an inclusion degree of F in G that can be interpreted as a kind of implication.

In our opinion, the meaning of the previous rules is closer to "the membership degree of the value of Y to L^Y_j is greater than the membership degree of the value of X to L^X_i". But expression (7) is not the only possible general semantics for a gradual rule. Another possibility is $\forall t, s \in r$

$$\text{if } L^X_i(t[X]) \leq L^X_i(s[X]) \text{ then } L^Y_j(t[Y]) \leq L^Y_j(s[Y]) \tag{8}$$

where it is not assumed that the degrees in Y are greater than those of X. Items keep being pairs $\langle Attribute, Label \rangle$ but transactions are associated to pairs of tuples. The item $\langle X, L^X_i \rangle$ is in the transaction τ_{ts} when $L^X_i(t[X]) \leq L^X_i(s[X])$, and the set of transactions, denoted $T^r_{G_2(L)}$, is a T-set. Now, $\left| T^r_{G_2(L)} \right| = |r|^2$. This alternative can be extended with fuzzy implications in a similar way that (7) extends (6).

IV. Some Relevant Related Works

Most of the papers in this field are devoted to the specific task of mining association rules involving quantitative attributes in relational databases. In the following, we describe some of them for the sake of offering here a somewhat complete overview of the topic.

- To our knowledge, [41] is the first paper relating fuzzy sets and association rules. In this work, fuzzy sets are introduced to diminish the granularity of quantitative attributes in the sense detailed in Section III-A. The model uses a membership threshold to change fuzzy transactions into crisp ones before looking for ordinary association rules in the set of crisp transactions. Items keep being pairs $\langle attribute, label \rangle$.
- In [42]–[44], a set of predefined linguistic labels is employed. The importance and accuracy of fuzzy association rules are obtained by means of two measures called *adjusted difference* and *weight of evidence*. A rule is said to be important when its adjusted difference is greater than 1.96 (i.e., the 95 percentile of the normal distribution). This avoids the need for a user-supplied importance threshold, but has the drawback that the adjusted difference is symmetric, i.e., if a rule $A \Rightarrow C$ is found to be interesting, then $C \Rightarrow A$ will be too. The weight of evidence is a measure of information gain that is provided to the user as an estimation of how interesting a rule is.
- In [32], the usefulness of itemsets and rules is measured by means of a *significance factor*, defined as a generalization of support based on sigma-counts (to count the percentage of transactions where the item is) and the product (for the intersection in the case of k itemsets with $k > 1$). The accuracy is based on a kind of *certainty factor* (with different formulation and semantics of our measure). In fact, two different formulations of the certainty factor are proposed in this work: the first one is based on the significance factor, in the same way that confidence is based on support. This provides a generalization of the ordinary support/confidence framework for association rules. The second proposal is based on correlation and it is not a generalization of confidence.
- The methodology in [45] finds the fuzzy sets that represent suitable linguistic labels for data (in the sense that

they allow to obtain rules with good support/accuracy) by using fuzzy clustering techniques. This way, the user does not need to define them, and that can be an advantage in certain cases. However, it could happen that the fuzzy sets so obtained are hard to fit to meaningful labels. Another methodology that follows this line is proposed in [46].

- In [47], only one item per attribute is considered: the pair $\langle$attribute,label$\rangle$ with greater support among those items based on the same attribute. The model is the usual generalization of support and confidence based on sigma-counts. In [48], an extension of the equi-depth (EDP) algorithm [27] for mining fuzzy association rules involving quantitative attributes is presented. The approach combines the obtained partitions with predefined linguistic labels.
- In [49], fuzzy taxonomies are used instead of simpler sets of labels. This allows to find rules at different granularity levels. The model is based on a generalization of support and confidence by means of sigma-counts, and the algorithms are extensions of classical Srikant and Agrawal's algorithms [2], [50].
- In [51] the concept of "ordinal fuzzy set" is introduced as an alternative interpretation of the membership degrees of values to labels. This carries out an alternative interpretation of fuzzy rules. Reference [52] studies fuzzy association rules with weighted items, i.e., an importance degree is given to each item. Weighted support and confidence are defined. Also, numerical values of attributes are mapped into linguistic terms by using Kohonen's self-organized maps. A generalization of the Apriori algorithm is proposed to discover fuzzy rules between weighted items.
- The definition of fuzzy association rule introduced in [53] is different from most of the existing in the literature. Fuzzy degrees are associated to items, and their meaning is the relative importance of items in rules. The model is different from [52], because linguistic labels are not considered. An item i with associated degree α is said to be in a fuzzy transaction $\tilde{\tau}$ when $\tilde{\tau}(i) \geq \alpha$. This seems to be a generalization of the model in [41] which uses the degree associated to an item as the threshold to turn fuzzy transactions into crisp ones, instead of using the same thresholds for all the items. In summary, the support of a "fuzzy itemset" $\tilde{I}$ (a set of items with associated degrees) is the percentage of fuzzy transactions $\tilde{\tau}$ such that $\tilde{I} \subseteq \tilde{\tau}$. Ordinary support and confidence are employed. A very interesting algorithm is proposed, which has the valuable feature that performs only one pass over the database in the mining process.

V. Conclusion

Mining fuzzy association rules (i.e., association rules in fuzzy transactions) is a useful technique to find patterns in data in the presence of imprecisione, either because data are fuzzy in nature or because we must improve their semantics. The proposed model has been tested on some of the applications described in this paper, specifically to discover fuzzy association rules in relational databases that contain quantitative data.

The model can be employed in mining distinct types of patterns, from ordinary association rules to fuzzy and approximate functional dependencies and gradual rules. They will be used in multimedia data mining and web mining. In the first case we shall mine transactional data about images given by artificial vision models. With respect to web mining, we are now facing the problem of mining user profiles characterized by fuzzy subsets of items [54].

Other technical issues we will study in the future, such as the analysis of measures given by quantifiers others than Q_M, have been pointed out in previous sections.

Appendix I
Basic Algorithm to Find Frequent Itemsets in a T-Set

Input: a set I of items and a T-set T based on I.
Output: a set of frequent itemsets F.
1. { Initialization }
(a) Create a counter $c_{\{i\}}$ for every $i \in I$
(b) $L_1 \leftarrow \{\{i\} \mid i \in I\}$
(c) $F = \emptyset$
(d) $l \leftarrow 1$
2. Repeat until $l > |I|$ or $L_l = \emptyset$
(a) For every $\tau \in T$
 i. For every $I_* \in L_l$
  A. If $I_* \subseteq \tau$ then $c_{I_*} \leftarrow c_{I_*} + 1$
(b) For every $I_* \in L_l$
 i. If $c_{I_*} < \mathrm{minsupp} \times |T|$
  A. $L_l \leftarrow L_l \backslash \{I_*\}$
  B. Free the memory used by c_{I_*}
(c) {Variables updating}
 i. $F = F \cup L_l$
 ii. $L_{l+1} \leftarrow CreateLevel(l+1, L_l)$
 iii. $l \leftarrow l+1$
3. Return(F)

Appendix II
Algorithm to Obtain $GD_Q(C/A)$ From V_A and $V_{A \cup C} j \leftarrow k$

1. $j \leftarrow k$
$GD \leftarrow 0$
$nf(A)^* \leftarrow k$
$acum_A \leftarrow 0$
$acum_{A \cup C} \leftarrow 0$
2. {Calculate $nf(A)^* = nf(A) \times k$ }
{This is the normalization factor}
While $(nf(A)^* > 0)$ and $(V_A(nf(A)^*) = 0)$
(a) $nf(A)^* \leftarrow nf(A)^* - 1$
3. If $(nf(A)^* = 0)$ then return("Error"); End
4. While $j > 0$
(a) $acum_{A \cup C} \leftarrow acum_{A \cup C} + V_{A \cup C}(j)$
(b) $acum_A \leftarrow acum_A + V_A(j)$
(b) If $(j \leq nf(A)^*)$
i. $GD \leftarrow GD + Q\left(acum_{A \cup C}/acum_A\right)$
(d) $j \leftarrow j - 1$
5. {Normalization }
$GD \leftarrow GD/nf(A)^*$
6. return(GD); End

APPENDIX III
BASIC ALGORITHM TO FIND FREQUENT ITEMSETS IN AN FT-SET

Input: a set I of items and an FT-set T based on I.
Output: a set of frequent itemsets F.
1. { Initialization }
(a) Create an array $V_{\{i\}}$ of size $k + 1$ for every $i \in I$
(b) $L_1 \leftarrow \{\{i\} \mid i \in I\}$
(c) $F = \emptyset$
(d) $l \leftarrow 1$
2. Repeat until $l > |I|$ or $L_l = \emptyset$
(a) For every $\tilde{\tau} \in T$
 i. For every $I_* \in L_l$
  A. $V_{I_*}\left(R\left(\widetilde{\Gamma}_{I_*}(\tilde{\tau}), k\right)\right) \leftarrow V_{I_*}\left(R\left(\widetilde{\Gamma}_{I_*}(\tilde{\tau}), k\right)\right) + 1$
(b) For every $I_* \in L_l$
 i. Use algorithm B to calculate $GD_Q\left(\widetilde{\Gamma}_{I_*}/T\right)$
 ii. If $GD_Q\left(\widetilde{\Gamma}_{I_*}/T\right) < \mathrm{minsupp}$
  A. $L_l \leftarrow L_l \backslash \{I_*\}$
  B. Free the memory used by V_{I_*}
(c) {Updating}
 i. $F = F \cup L_l$
 ii. $L_{l+1} \leftarrow CreateLevel(l+1, L_l)$
 iii. $l \leftarrow l + 1$
3. Return(F)

REFERENCES

[1] W. Pedrycz, "Fuzzy set technology in knowledge discovery," *Fuzzy Sets Syst.*, vol. 98, pp. 279–290, 1998.
[2] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," in *Proc. ACM SIGMOD Conf.*, 1993, pp. 207–216.
[3] A. De Luca and S. Termini, "Entropy and energy measures of a fuzzy set," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds. Amsterdam, The Netherlands: North-Holland, 1979, vol. 20, pp. 321–338.
[4] M. Wygralak, *Vaguely Defined Objects. Representations, Fuzzy Sets and Nonclassical Cardinality Theory*. Boston, MA: Kluwer, 1996.
[5] M. Delgado, M. J. Martín-Bautista, D. Sánchez, and M. A. Vila, "A probabilistic definition of a nonconvex fuzzy cardinality," *Fuzzy Sets Syst.*, vol. 126, no. 2, pp. 41–54, 2002.
[6] L. A. Zadeh, "A computational approach to fuzzy quantifiers in natural languages," *Comput. Math. Applicat.*, vol. 9, no. 1, pp. 149–184, 1983.
[7] M. Delgado, D. Sánchez, and M. A. Vila, "Fuzzy cardinality based evaluation of quantified sentences," *Int. J. Approx. Reason.*, vol. 23, pp. 23–66, 2000.
[8] J. C. Cubero, J. M. Medina, O. Pons, and M. A. Vila, "The generalized selection: An alternative way for the quotient operations in fuzzy relational databases," in *Fuzzy Logic and Soft Computing*, B. Bouchon-Meunier, R. Yager, and L. A. Zadeh, Eds. Singapore: World Scientific, 1995.
[9] D. Sánchez, "Adquisición de relaciones entre atributos en bases de datos relacionales," Ph.D. dissertation (in Spanish), Dept. Comput. Sci. Artificial Intell., Univ. Granada, Granada, Spain, 1999.
[10] J. Kacprzyk, "Fuzzy logic with linguistic quantifiers: A tool for better modeling of human evidence aggregation processes?," in *Fuzzy Sets in Psychology*, T. Zétényi, Ed. Amsterdam, The Netherlands: North-Holland, 1988, pp. 233–263.
[11] R. R. Yager, "Quantifier guided aggregation using OWA operators," *Int. J. Intell. Syst.*, vol. 11, pp. 49–73, 1996.
[12] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, "Dynamic itemset counting and implication rules for market basket data," *SIGMOD Record*, vol. 26, no. 2, pp. 255–264, 1997.
[13] C. Silverstein, S. Brin, and R. Motwani, "Beyond market baskets: Generalizing association rules to dependence rules," *Data Mining Knowl. Disc.*, vol. 2, pp. 39–68, 1998.
[14] F. Berzal, I. Blanco, D. Sánchez, and M. A. Vila, "Measuring the accuracy and interest of association rules: A new framework," presented at the An Extension of Advances in Intelligent Data Analysis: 4th Int. Symp., IDA'01. Intelligent Data Analysis, Cascais, Portugal, 2002.
[15] E. Shortliffe and B. Buchanan, "A model of inexact reasoning in medicine," *Math. Biosci.*, vol. 23, pp. 351–379, 1975.
[16] F. Berzal, M. Delgado, D. Sánchez, and M. A. Vila, "Measuring the accuracy and importance of association rules," Dept. Comput. Sci. Artificial Intell., Univ. Granada, Granada, Spain, CCIA-00–01-16, 2000.
[17] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th VLDB Conf.*, Sep. 1994, pp. 478–499.
[18] M. Houtsma and A. Swami, "Set-oriented mining for association rules in relational databases," in *Proc. 11th Int. Conf. Data Engineering*, 1995, pp. 25–33.
[19] H. Mannila, H. Toivonen, and I. Verkamo, "Efficient algorithms for discovering association rules," in *Proc. AAAI Workshop Knowledge Discovery Databases*, 1994, pp. 181–192.
[20] J.-S. Park, M.-S. Chen, and P. S. Yu, "An effective hash based algorithm for mining association rules," *SIGMOD Record*, vol. 24, no. 2, pp. 175–186, 1995.
[21] C. Hidber, "Online association rule mining," in *Proc. ACM SIGMOD Int. Conf. Management Data*, 1999, pp. 145–156.
[22] F. Berzal, J. C. Cubero, N. Marín, and J. M. Serrano, "TBAR: An efficient method for association rule mining in relational databases," *Data Knowledge Eng.*, vol. 31, no. 1, pp. 47–64, 2001.
[23] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proc. ACM SIGMOD Int. Conf. Management Data*, Dallas, TX, 2000, pp. 1–12.
[24] J. Hipp, U. Güntzer, and G. Nakhaeizadeh, "Algorithms for association rule mining – A general survey and comparison," *SIGKDD Explor.*, vol. 2, no. 1, pp. 58–64, 2000.
[25] M. Delgado, D. Sánchez, J. M. Serrano, and M. A. Vila, "A survey of methods to evaluate quantified sentences," *Math. Soft Comput.*, vol. VII, no. 2–3, pp. 149–158, 2000.
[26] M. Delgado, D. Sánchez, and M. A. Vila, "Acquisition of fuzzy association rules from medical data," in Medicine. ser. Studies in Fuzziness and Soft Computing, S. Barro and R. Marín, Eds. Heidelberg, Germany: Physica-Verlag, 2002, vol. 83, pp. 286–310, to be published.
[27] R. Srikant and R. Agrawal, "Mining quantitative association rules in large relational tables," in *Proc. ACM SIGMOD Int. Conf. Management Data*, 1996, pp. 1–12.
[28] J. Wijsen and R. Meersman, "On the complexity of mining quantitative association rules," *Data Mining Knowled. Disc.*, vol. 2, pp. 263–281, 1998.
[29] R. J. Miller and Y. Yang, "Association rules over interval data," in *Proc. ACM—SIGMOD Int. Conf. Management Data*, 1997, pp. 452–461.
[30] S.-J. Yen and A. L. P. Chen, "The analysis of relationships in databases for rule derivation," *J. Intell. Inform. Syst.*, vol. 7, pp. 235–259, 1996.
[31] Z. Zhang, Y. Lu, and B. Zhang, "An effective partitioning-combining algorithm for discovering quantitative association rules," in *KDD: Techniques and Applications*, H. Lu, H. Motoda, and H. Liu, Eds. Singapore: World Scientific, 1997, pp. 241–251.
[32] C.-M. Kuok, A. Fu, and M. H. Wong, "Mining fuzzy association rules in databases," *SIGMOD Record*, vol. 27, no. 1, pp. 41–46, 1998.
[33] P. Bosc and L. Lietard, "Functional dependencies revisited under graduality and imprecision," in *Proc. Annu. Meeting NAFIPS*, 1997, pp. 57–62.
[34] I. Blanco, M. J. Martín-Bautista, D. Sánchez, and M. A. Vila, "On the support of dependencies in relational databases: Strong approximate dependencies," *Data Mining Knowled. Disc.*, 2003, submitted for publication.
[35] M. Delgado, M. J. Martín-Bautista, D. Sánchez, and M. A. Vila, "Mining strong approximate dependencies from relational databases," presented at the IPMU'2000, Madrid, Spain, 2000.
[36] M. Delgado, D. Sánchez, and M. A. Vila, "Fuzzy quantified dependencies in relational databases," presented at the EUFIT'99, Aachen, Germany, 1999.
[37] J. C. Cubero, O. Pons, and M. A. Vila, "Weak and strong resemblance in fuzzy functional dependencies," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 1994, pp. 162–166.
[38] J. C. Cubero and M. A. Vila, "A new definition of fuzzy functional dependency in fuzzy relational databases," *Int. J. Intell. Syst.*, vol. 9, no. 5, pp. 441–448, 1994.

[39] D. Dubois and H. Prade, "Fuzzy rules in knowledge-based systems. modeling gradedness, uncertainty and preference," in *An Introduction to Fuzzy Logic Applications in Intelligent Systems*, R. R. Yager and L. A. Zadeh, Eds. Dordrecht, The Netherlands: Kluwer, 1992, pp. 45–68.

[40] B. Bouchon-Meunier, G. Dubois, L. L. Godó, and H. Prade, *Fuzzy Sets and Possibility Theory in Approximate and Plausible Reasoning*, D. Dubois and H. Prade, Eds. Norwell, MA: Kluwer, 1999, Handbooks of Fuzzy Sets, ch. 1, pp. 15–190.

[41] J. H. Lee and H. L. Kwang, "An extension of association rules using fuzzy sets," presented at the IFSA'97, Prague, Czech Republic, 1997.

[42] W. H. Au and K. C. C. Chan, "Mining fuzzy association rules," in *Proc. 6th Int. Conf. Information Knowledge Management*, Las Vegas, NV, 1997, pp. 209–215.

[43] ——, "An effective algorithm for discovering fuzzy rules in relational databases," in *Proc. IEEE Int. Conf. Fuzzy Systems*, vol. II, 1998, pp. 1314–1319.

[44] ——, "FARM: A data mining system for discovering fuzzy association rules," in *Proc. FUZZ-IEEE'99*, vol. 3, 1999, pp. 22–25.

[45] 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. Intelligent Data Engineering Learning (IDEAL'98)*, Hong Kong, 1998, pp. 263–268.

[46] M. Vazirgiannis, "A classification and relationship extraction scheme for relational databases based on fuzzy logic," in *Proc. Research Development Knowledge Discovery Data Mining*, Melbourne, Australia, 1998, pp. 414–416.

[47] T. P. Hong, C. S. Kuo, and S. C. Chi, "Mining association rules from quantitative data," *Intell. Data Anal.*, vol. 3, pp. 363–376, 1999.

[48] W. Zhang, "Mining fuzzy quantitative association rules," in *Proc. 11th Int. Conf. Tools Artificial Intelligence*, Chicago, IL, 1999, pp. 99–102.

[49] G. Chen, Q. Wei, and E. Kerre, "Fuzzy data mining: Discovery of fuzzy generalized association rules," in *Recent Issues on Fuzzy Databases*, G. Bordogna and G. Pasi, Eds. Heidelberg, Germany: Physica-Verlag, 2000, Studies in Fuzziness and Soft Computing Series.

[50] R. Srikant and R. Agrawal, "Mining generalized association rules," in *Proc 21st Int. Conf. Very Large Data Bases*, Sept. 1995, pp. 407–419.

[51] J. W. T. Lee, "An ordinal framework for data mining of fuzzy rules," in *FUZZ IEEE 2000*, San Antonio, TX, 2000, pp. 399–404.

[52] J. Shu-Yue, E. Tsang, D. Yenng, and S. Daming, "Mining fuzzy association rules with weighted items," in *Proc. IEEE Int. Conf. Systems, Man, Cybernetics*, Nashville, TN, 2000, pp. 1906–1911.

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

[54] M. J. Martín-Bautista, "Modelos de computación flexible para la recuperación de información," Ph.D. dissertation (in Spanish), Dept. Comput. Sci. Artificial Intell., Univ. Granada, Granada, Spain, 2000.

[55] F. Berzal, I. Blanco, D. Sánchez, and M. A. Vila, "A new framework to assess association rules," in *Proc. Advances Intelligent Data Analysis: 4th Int. Symp.*, Lecture Notes in Computer Science 2189, F. Hoffmann, Ed., 2001, pp. 95–104.

Miguel Delgado was born in Granada, Spain, in May, 1951. He received the M.S. degree in mathematics, the Dipl. in statistics, the Ph.D. degree in science, and the O.R. Dipl. in science of education, all from the University of Granada, Granada, Spain, in 1973, 1974, 1975, and 1989, respectively.

Since 1989, he has been a Full Professor of computer science and artificial intelligence at the University of Granada. From 1996 to 2001, he was Vicerector of the same university. His teaching experience includes the topics of decision theory, mathematical programming, theory of algorithms, systems theory, operations research, information theory, knowledge engineering, and artificial intelligence. He has been the Principal Investigator as well as Member of the teams of more than ten research projects. He has published two books and more than 80 papers, 50 of them in international journals. He has attended and presented communications or invited lectures in more than 30 national or international conferences and workshops. He has been and is currently a Member of different national and international program committees. Additionally, he has been the Advisor of more than 15 Ph.D. degree dissertations on topics related to decision making and optimization in fuzzy environment, knowledge representation, knowledge engineering, neural nets, machine learning, and data mining, and he has been an Invited Lecturer at several European universities (Trento, Budapest, Wroclaw, etc.) and scientific conferences. His main areas of interest are approximate reasoning, optimization problems, neural nets, learning models, decision support systems, and data and text mining.

Nicolás Marín was born in Granada, Andalusia, Spain, in 1975. He received the Ph.D. degree in computer science from the University of Granada, Granada, Spain, in 2001.

He currently works as a Lecturer in the Department of Computer Science and Artificial Intelligence of the University of Granada. He is Member of the Intelligent Databases and Information Systems Research Group of the Andalusian Goverment and is Member of the team of several projects. His research interest is mainly focused on the fields of fuzzy databases, knowledge discovery and data mining, fuzzy sets theory, and soft computing.

Daniel Sánchez was born in Almería, Spain, in 1972. He received the M.S. and Ph.D. degrees in computer science, both from the University of Granada, Granada, Spain, in 1995 and 1999, respectively.

Since 2001 he has been an Associate Professor in the Department of Computer Science and Artificial Intelligence of the University of Granada. He has participated and is currently a Member of the teams of several projects, and he has published more than 30 papers in international journals and conferences. His current main research interests are in the fields of knowledge discovery and data mining, relational databases, information retrieval, fuzzy sets theory, and soft computing.

María-Amparo Vila received the M.S. and Ph.D. degrees in mathematics, both from the University of Granada, Granada, Spain, in 1973 and 1978, respectively.

While atthe University of Granada, she was Assistant Professor in the Department of Statistics until 1982, Associate Professor in the same department until 1986, and Associate Professor in the Department of Computer Science and Artificial Intelligence until 1992. Since 1992, she has been a Professor in the same department. Since 1997, she has also been Head of the Department and the IdBIS Research Group. Her research activity is centered around the application of soft computing techniques to different areas of computer science and artificial intelligence, such as theoretical aspects of fuzzy sets, decision and optimization processes in fuzzy environments, fuzzy databases including relational, logical and object-oriented data models, and information retrieval. Currently, she is interested in the application of soft computing techniques to data, text, and web mining. She has been responsible for ten research projects and advisor of seven Ph.D. degree dissertations. She has published more than 50 papers in prestigious international journals, more than 60 contributions to international conferences, and many book chapters.