

FARM: A Data Mining System for Discovering Fuzzy Association Rules

Wai-Ho Au Keith C.C. Chan

Department of Computing, The Hong Kong Polytechnic University

Hung Hom, Kowloon, Hong Kong

E-mail: {cswchau, cskcchan}@comp.polyu.edu.hk

Abstract

In this paper, we introduce a novel technique, called FARM, for mining fuzzy association rules. FARM employs linguistic terms to represent the revealed regularities and exceptions. The linguistic representation is especially useful when those rules discovered are presented to human experts for examination because of the affinity with the human knowledge representations. The definition of linguistic terms is based on fuzzy set theory and hence we call the rules having these terms fuzzy association rules. The use of fuzzy technique makes FARM resilient to noises such as inaccuracies in physical measurements of real-life entities and missing values in the databases. Furthermore, FARM utilizes adjusted difference analysis which has the advantage that it does not require any user-supplied thresholds which are often hard to determine. In addition to this interestingness measure, FARM has another unique feature that the conclusions of a fuzzy association rule can contain linguistic terms. Our technique also provides a mechanism to allow quantitative values be inferred from fuzzy association rules. Unlike other data mining techniques that can only discover association rules between different discretized values, FARM is able to reveal interesting relationships between different quantitative values. Our experimental results showed that FARM is capable of discovering meaningful and useful fuzzy association rules in an effective manner from a real-life database.

Keywords: data mining, knowledge discovery in databases, fuzzy association rules, linguistic terms, interestingness measure.

1. Introduction

An important topic in data mining research is concerned with the discovery of *association rules* [1]. An association rule describes an interesting relationship among different attributes and we refer to such relationship as an *association* in this paper. Existing algorithms (e.g. [11]) involve discretizing the domains of quantitative attributes into intervals so as to discover quantitative association rules. These intervals may not be concise and meaningful enough for human experts to easily obtain nontrivial knowledge from those rules discovered. Instead of using intervals, we introduce a novel technique, called FARM (Fuzzy Association Rule Miner), which employs *linguistic terms* to represent the revealed regularities and exceptions. The

linguistic representation makes those rules discovered to be much natural for human experts to understand. The definition of linguistic terms is based on fuzzy set theory and hence we call the rules having these terms *fuzzy association rules*.

Unlike other data mining algorithms (e.g. [1, 11]) which utilize user-supplied thresholds to identify interesting associations, FARM employs an objective interestingness measure, called *adjusted difference* [2-5], to mine fuzzy association rules. The use of this technique has the advantage that it does not require any user-supplied thresholds which are often hard to determine. Furthermore, FARM also has the advantage that it allows us to discover both *positive* and *negative* association rules. A positive association rule tells us that a record having certain attribute value (or linguistic term) will also have another attribute value (or linguistic term) whereas a negative association rule tells us that a record having certain attribute value (or linguistic term) will not have another attribute value (or linguistic term).

Many data mining algorithms (e.g. [1, 11]) require the conclusions of rules to be crisp and hence quantitative values cannot be inferred from those rules. To be more effective, FARM is able to deal with class boundaries that are fuzzy and to associate qualitative attribute values with quantitative values. This provides a mechanism for FARM to allow quantitative values be inferred. Such mechanism equipped our technique with a unique feature that interesting associations between different quantitative attribute values can be revealed whereas other data mining techniques can only discover association rules between different discretized intervals.

2. Related Work

An example of an association rule is "90% of transactions that contain bread also contain butter; 3% of all transactions contain both of these items." The 90% is referred to as the *confidence* and the 3%, the *support*, of the rule. Let $I = \{i_1, i_2, \dots, i_m\}$ be a set of binary attributes called *items* and T be a set of *transactions*. Each transaction $t \in T$ is represented as a binary vector with $t[k] = 1$ if t contains item i_k and $t[k] = 0$, otherwise, for $k = 1, 2, \dots, m$. An association rule is defined as an implication of the form $X \Rightarrow Y$ where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$. The rule $X \Rightarrow Y$ holds in T with support defined as $\Pr(X \cup Y)$ and confidence defined as $\Pr(Y | X)$. An association rule is interesting if and only if its support and confidence is greater than some user-

supplied threshold. Association rules of such type are often referred to as *boolean association rules*. Since boolean association rules are defined over binary data, they are rather restrictive in many different areas and hence a lot of recent effort has been put into the mining of *quantitative association rules* [11].

Quantitative association rules are defined over quantitative and categorical attributes. In [11], the values of categorical attributes are mapped to a set of consecutive integers and the values of quantitative attributes are first discretized into intervals using *equi-depth partitioning*, if necessary, and then mapped to consecutive integers to preserve the order of the values/intervals. And as a result, both categorical and quantitative attributes can be handled in a uniform fashion as a set of <attribute, integer value> pairs. With the mappings defined in [11], a quantitative association rule is mapped to a set of boolean association rules. After the mappings, the algorithms for mining boolean association rules (e.g. [1]) is then applied to the transformed data set.

For the mining algorithms such as that described in [1, 11] to determine if an association rule is interesting, its support and confidence have to be greater than some user-supplied thresholds. A weakness of such approach is that many users do not have any idea what the thresholds should be. If it is set too high, a user may miss some useful rules but if it is set too low, the user may be overwhelmed by many irrelevant ones.

Furthermore, the intervals in quantitative association rules may not be concise and meaningful enough for human experts to obtain nontrivial knowledge. *Fuzzy linguistic summaries* introduced in [12] express knowledge in linguistic representation which is natural for people to comprehend. An example of linguistic summaries is the statement "about half of people are middle aged." In addition to fuzzy linguistic summaries, an interactive top-down summary discovery process which utilizes *fuzzy is-a hierarchies* as domain knowledge has been described in [9]. This technique aims at discovering a set of *generalized tuples* such as <technical writer, documentation>. In contrast to association rules which involve implications between different attributes, fuzzy linguistic summaries and the generalized tuples only provide summarization on different attributes. The idea of implication has not been taken into consideration and hence these techniques are not developed for the task of rule discovery.

The applicability of fuzzy modeling techniques to data mining has also been discussed in [8]. Given a series of fuzzy sets, $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_c$, *conditional (context-sensitive) Fuzzy C-Means (FCM)* method is able to reveal the rule-based models within a family of patterns by considering their vicinity in a feature space along with the similarity of the values assumed by a certain conditional variable (context) [8]. Nevertheless, the conditional FCM method can only manipulate quantitative attributes and it is for this reason that this technique is inadequate to deal with

databases which consist of both quantitative and categorical attributes.

3. FARM for Mining Fuzzy Association Rules

A first-order fuzzy association rule can be defined as a fuzzy association rule involving one linguistic term in its antecedent; a second-order fuzzy association rule can be defined as a fuzzy association rule involving two linguistic terms in its antecedent and a third-order fuzzy association rule can be defined as a fuzzy association rule involving three linguistic terms in its antecedent, etc. To perform search more effectively, we propose to use an evolutionary algorithm. FARM is, therefore, capable of mining fuzzy association rules in large databases by an iterative process.

3.1. Linguistic Terms

Given a set of records, \mathcal{D} , each of which consists of a set of attributes $\mathcal{J} = \{I_1, I_2, \dots, I_n\}$, where I_1, I_2, \dots, I_n can be quantitative or categorical. For any record, $d \in \mathcal{D}$, $d[I_v]$ denotes the value i_v in d for attribute $I_v \in \mathcal{J}$. For any quantitative attribute, $I_v \in \mathcal{J}$, let $\text{dom}(I_v) = [l_v, u_v] \subseteq \mathcal{R}$ denote the domain of the attribute. A set of linguistic terms can be defined over the domain of each quantitative attribute. Let \mathcal{L}_{vr} , $r = 1, 2, \dots, s_v$, be linguistic terms associated with some quantitative attribute, $I_v \in \mathcal{J}$. \mathcal{L}_{vr} is represented by a fuzzy set, L_{vr} , defined on $\text{dom}(I_v)$ whose membership function is $\mu_{L_{vr}}$ such that

$$\mu_{L_{vr}} : \text{dom}(I_v) \rightarrow [0,1]$$

The fuzzy sets L_{vr} , $r = 1, 2, \dots, s_v$, are then defined as

$$L_{vr} = \begin{cases} \sum_{\text{dom}(I_v)} \frac{\mu_{L_{vr}}(i_v)}{i_v} & \text{if } I_v \text{ is discrete} \\ \int_{\text{dom}(I_v)} \frac{\mu_{L_{vr}}(i_v)}{i_v} & \text{if } I_v \text{ is continuous} \end{cases}$$

for all $i_v \in \text{dom}(I_v)$. The degree of compatibility of $i_v \in \text{dom}(I_v)$ with some linguistic term \mathcal{L}_{vr} is given by $\mu_{L_{vr}}(i_v)$.

For any categorical attribute, $I_v \in \mathcal{J}$, let $\text{dom}(I_v) = \{i_{v1}, i_{v2}, \dots, i_{vm_v}\}$ denote the domain of I_v . We can define a set of linguistic terms, \mathcal{L}_{vr} , $r = 1, 2, \dots, m_v$, for $I_v \in \mathcal{J}$ where \mathcal{L}_{vr} is represented by a fuzzy set, L_{vr} , such that

$$L_{vr} = \frac{1}{i_{vr}}$$

Using the above technique, we can represent the original attributes, \mathcal{J} , using a set of linguistic terms, $\mathcal{L} = \{\mathcal{L}_{vr} \mid v = 1, 2, \dots, n, r = 1, 2, \dots, s_v\}$ where $s_v = m_v$ for categorical attributes. Each linguistic term is represented by a fuzzy set and hence we have a set of fuzzy sets, $\mathcal{L} = \{L_{vr} \mid v = 1, 2, \dots, n, r = 1, 2, \dots, s_v\}$. Given a record, $d \in \mathcal{D}$, and a linguistic term, $\mathcal{L}_{vr} \in \mathcal{L}$, which is represented by a fuzzy set, $L_{vr} \in \mathcal{L}$, the degree of membership of the values in d

with respect to L_{vr} is given by $\mu_{L_{vr}}(d[I_v])$. The degree, $\lambda_{L_{vr}}(d)$, to which d is characterized by L_{vr} is defined as

$$\lambda_{L_{vr}}(d) = \mu_{L_{vr}}(d[I_v])$$

If $\lambda_{L_{vr}}(d) = 1$, d is completely characterized by the term L_{vr} . If $\lambda_{L_{vr}}(d) = 0$, d is undoubtedly not characterized by the term L_{vr} . If $0 < \lambda_{L_{vr}}(d) < 1$, d is partially characterized by the term L_{vr} . If $d[I_v]$ is unknown, $\lambda_{L_{vr}}(d) = 0.5$ which indicates that there is no information available concerning whether d is characterized by the term L_{vr} or not.

In fact, d can also be characterized by more than one linguistic terms. Let φ be a subset of integers such that $\varphi = \{v_1, v_2, \dots, v_m\}$ where $v_1, v_2, \dots, v_m \in \{1, 2, \dots, n\}$, $v_1 \neq v_2 \neq \dots \neq v_m$ and $|\varphi| = m \geq 1$. We further suppose that \mathcal{I}_φ be a subset of \mathcal{J} such that $\mathcal{I}_\varphi = \{I_v \mid v \in \varphi\}$. Given any \mathcal{I}_φ it is associated with a set of linguistic terms, $L_{\varphi r}$, $r = 1, 2, \dots, s_\varphi$ where $s_\varphi = \prod_{v \in \varphi} s_v$. Each $L_{\varphi r}$ is defined by a set of linguistic terms, $L_{v_1 r_1}, L_{v_2 r_2}, \dots, L_{v_m r_m} \in \mathcal{L}$. The degree, $\lambda_{L_{\varphi r}}(d)$, to which d is characterized by the term $L_{\varphi r}$ is defined as

$$\lambda_{L_{\varphi r}}(d) = \min(\mu_{L_{v_1 r_1}}(d[I_{v_1}]), \mu_{L_{v_2 r_2}}(d[I_{v_2}]), \dots, \mu_{L_{v_m r_m}}(d[I_{v_m}]))$$

3.2. Identification of Interesting Associations

In order to decide whether the association between a linguistic term, $L_{\varphi k}$, and another linguistic term, L_{pq} , is interesting, we employ the *adjusted difference* [2-5] which is defined as

$$d_{L_{pq} L_{\varphi k}} = \frac{z_{L_{pq} L_{\varphi k}}}{\sqrt{\gamma_{L_{pq} L_{\varphi k}}}} \quad (1)$$

where $z_{L_{pq} L_{\varphi k}}$ is the *standardized difference* [2-5] given by

$$z_{L_{pq} L_{\varphi k}} = \frac{\deg_{L_{pq} L_{\varphi k}} - e_{L_{pq} L_{\varphi k}}}{\sqrt{e_{L_{pq} L_{\varphi k}}}} \quad (2)$$

$e_{L_{pq} L_{\varphi k}}$ is the sum of degrees to which records are expected to be characterized by L_{pq} and $L_{\varphi k}$ and is calculated by

$$e_{L_{pq} L_{\varphi k}} = \frac{\sum_{i=1}^{s_\varphi} \deg_{L_{pq} L_{\varphi i}} \sum_{i=1}^{s_p} \deg_{L_{pi} L_{\varphi k}}}{\sum_{u=1}^{s_p} \sum_{i=1}^{s_\varphi} \deg_{L_{pu} L_{\varphi i}}} \quad (3)$$

and $\gamma_{L_{pq} L_{\varphi k}}$ is the *maximum likelihood estimate* [2-5] of the variance of $z_{L_{pq} L_{\varphi k}}$ and is given by

$$\gamma_{L_{pq} L_{\varphi k}} = \left(1 - \frac{\sum_{i=1}^{s_\varphi} \deg_{L_{pq} L_{\varphi i}}}{\sum_{u=1}^{s_p} \sum_{i=1}^{s_\varphi} \deg_{L_{pu} L_{\varphi i}}} \right) \left(1 - \frac{\sum_{i=1}^{s_p} \deg_{L_{pi} L_{\varphi k}}}{\sum_{u=1}^{s_p} \sum_{i=1}^{s_\varphi} \deg_{L_{pu} L_{\varphi i}}} \right) \quad (4)$$

If $|d_{L_{pq} L_{\varphi k}}| > 1.96$ (the 95 percentiles of the normal distribution), we can conclude that the association between $L_{\varphi k}$ and L_{pq} is interesting. If $d_{L_{pq} L_{\varphi k}} > +1.96$, the presence of $L_{\varphi k}$ implies the presence of L_{pq} . In other words, it is more *likely* for a record having both $L_{\varphi k}$ and L_{pq} . We say that $L_{\varphi k}$ is *positively* associated with L_{pq} . If $d_{L_{pq} L_{\varphi k}} < -1.96$, the absence of $L_{\varphi k}$ implies the presence of L_{pq} . In other words, it is more *unlikely* for a record having $L_{\varphi k}$ and L_{pq} at the same time. We say that $L_{\varphi k}$ is *negatively* associated with L_{pq} .

3.3. Formation of Association Rules

If the association between $L_{\varphi k}$ and L_{pq} is found to be interesting, there is some evidence for or against a record having L_{pq} given it has $L_{\varphi k}$. Based on an information theoretic concept known as *mutual information*, a confidence measure, called *weight of evidence*, to represent the uncertainty of the fuzzy association rules is defined in [2-5] as

$$w_{L_{pq} L_{\varphi k}} = \log \frac{\Pr(L_{\varphi k} | L_{pq})}{\Pr(L_{\varphi k} | \bigcup_{i \neq q} L_{pi})} \quad (5)$$

The weight of evidence is positive if $L_{\varphi k}$ is positively associated with L_{pq} whereas the weight of evidence is negative if $L_{\varphi k}$ is negatively associated with L_{pq} . Based on this confidence measure, a fuzzy association rule is in the form of $L_{\varphi k} \Rightarrow L_{pq} [w_{L_{pq} L_{\varphi k}}]$.

Since $L_{\varphi k}$ is defined by a set of linguistic terms, $L_{v_1 r_1}, L_{v_2 r_2}, \dots, L_{v_m r_m} \in \mathcal{L}$, we have a high-order fuzzy association rule, $L_{v_1 r_1} \wedge L_{v_2 r_2} \wedge \dots \wedge L_{v_m r_m} \Rightarrow L_{pq} [w_{L_{pq} L_{\varphi k}}]$ where $v_1, v_2, \dots, v_m \in \varphi$.

3.4. The FARM in Details

First of all, a set of first-order fuzzy association rules is found using adjusted difference analysis. After these rules are discovered, they are stored in R_1 . R_1 is then used to generate second-order rules which are, in turn, stored in R_2 . R_2 is then used to generate third-order rules that are stored in R_3 and so on for 4th and higher order. The details are given in Fig. 1. The *terminate* function in Fig. 1 implements the following termination criteria: (i) terminate when the best and the worst performing chromosome differs by less than 0.1%; (ii) terminate when the total number of generations specified by the user is reached; and (iii) terminate when no more interesting rules can be identified.

For the evolutionary process, FARM encodes a complete set of fuzzy association rules in a single chromosome in such a way that each gene encodes a single rule. Specifically, given the following m -th order rule,

$\mathcal{L}_{v_1 r_1} \wedge \mathcal{L}_{v_2 r_2} \wedge \dots \wedge \mathcal{L}_{v_m r_m} \Rightarrow \mathcal{L}_{pq}[\mathcal{W}_{\mathcal{L}_{pq} \mathcal{L}_{pk}}]$, it is encoded in FARM by the allele given in Fig. 2.

```

1)  $R_1 = \{\text{first-order fuzzy association rules}\};$ 
2) for ( $m = 2; |R_{m-1}| \geq \text{minrules}; m++$ ) do
3) begin
4)    $t = 0;$ 
5)    $\text{Population}[t] = \text{initialize}(R_{m-1});$ 
6)    $\text{fitness}(\text{Population}[t]);$ 
7)   while not  $\text{terminate}(\text{Population}[t])$  do
8)     begin
9)        $t = t + 1;$ 
10)       $\text{Population}[t] = \text{replace}(\text{Population}[t-1]);$ 
11)       $\text{fitness}(\text{Population}[t]);$ 
12)    end
13)     $R_m = \text{decode}(\text{the fittest individual in } \text{Population}[t]);$ 
14)  end
15)  $\mathcal{R} = \bigcup_m R_m$ 

```

Fig. 1. Algorithm FARM.

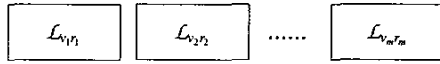


Fig. 2. An allele representing a m -th order fuzzy association rule.

3.5. Initialization of Populations

FARM generates different sets of m -th order rules by randomly combining the $(m-1)$ -th order rules discovered in the previous iteration. The details of the initialization process are given in the *initialize* function in Fig. 3. The $\text{chrom}_i.\text{allele}_j$ in Fig. 3 denotes the j -th allele of the i -th chromosome. The $\text{rand}_m(R)$ function returns a m -th order allele constructed by randomly combining m elements in R . In our experiments, *popsize* was set to 30 and the number of alleles in each chromosome was set to $\text{nalleles} = |R_{m-1}|$.

```

1) population  $\text{initialize}(R_{m-1})$ 
2) begin
3)    $R = \{\text{all conjuncts in the antecedent of all } r \in R_{m-1}\};$ 
4)   for ( $i = 1; i \leq \text{popsize}; i++$ ) do
5)     begin
6)       for ( $j = 1; j \leq \text{nalleles}; j++$ ) do
7)          $\text{chrom}_i.\text{allele}_j = \text{rand}_m(R);$ 
8)       end
9)     return  $\bigcup_i \text{chrom}_i;$ 
10) end

```

Fig. 3. The *initialize* function.

3.6. The Genetic Operators

The genetic operators used by FARM are implemented in the *reproduce* function shown in Fig. 4. The *select* function uses the *roulette wheel selection* scheme [7] to select two different chromosomes, chrom_1 and chrom_2 , from the current population. These two chromosomes are then passed as arguments to the *crossover* function.

The *crossover*($\text{chrom}_1, \text{chrom}_2$) function uses the *two-point crossover* operator [7]. The crossover points are randomly chosen so that they can either occur between two

rules or within one. An example of it is graphically depicted in Fig. 5.

```

1) population  $\text{reproduce}(\text{Population}[t-1])$ 
2) begin
3)    $\text{chrom}_1 = \text{select}(\text{Population}[t-1]);$ 
4)    $\text{chrom}_2 = \text{select}(\text{Population}[t-1]);$ 
5)    $\text{nchrom}_1, \text{nchrom}_2 = \text{crossover}(\text{chrom}_1, \text{chrom}_2);$ 
6)    $\text{mutation}(\text{nchrom}_1);$ 
7)    $\text{mutation}(\text{nchrom}_2);$ 
8)    $\text{Population} = \text{steady-state}(\text{Population}[t-1], \text{nchrom}_1, \text{nchrom}_2);$ 
9)   return  $\text{Population};$ 
10) end

```

Fig. 4. The *reproduce* function.

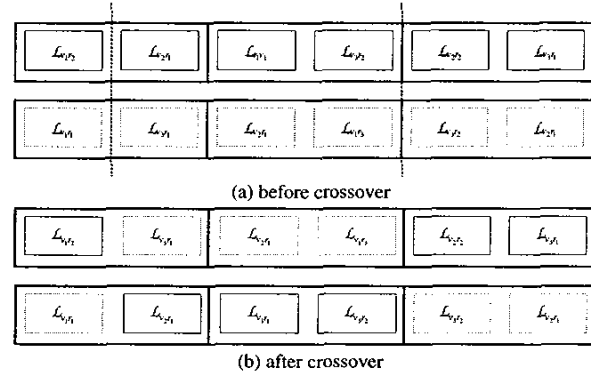


Fig. 5. An example of a two-point crossover (the thick borders indicate the rule boundaries.)

The *mutation* function, which is different from the traditional mutation operator [7], is given in Fig. 6. The *random* function returns a real number between 0 and 1 and the constant *pmutation* contains the mutation rate.

```

1)  $\text{mutation}(\text{nchrom})$ 
2) begin
3)    $R = \{\text{all conjuncts in the antecedent of all } r \in R_{m-1}\};$ 
4)   for ( $j = 1; j \leq \text{nalleles}; j++$ ) do
5)     begin
6)       if  $\text{random} < \text{pmutation}$  then
7)          $\text{nchrom}.\text{allele}_j = \text{rand}_m(R);$ 
8)       end
9)   end

```

Fig. 6. The *mutation* function.

The *steady-state*($\text{Population}[t-1], \text{nchrom}_1, \text{nchrom}_2$) function in *reproduce* produces a new population, $\text{Population}[t]$, by removing the two least-fit chromosomes in $\text{Population}[t-1]$ and replacing them with nchrom_1 and nchrom_2 while keeping the rest of the other chromosomes intact.

3.7. Selection and The Fitness Function

To determine the fitness of a chromosome that encodes a set of m -th order rules, FARM uses a performance measure defined in terms of the probability that the value of an attribute of a record can be correctly predicted based on the rules in $R = R_1 \cup R_2 \cup \dots \cup R_{m-1} \cup \{\text{rules encoded in}$

the chromosome being evaluated}. The details of predicting the value of an attribute will be given in Section 4. If the prediction is correct, we can increment an accuracy count associated with the chromosome whose fitness is being evaluated by one. By putting each of the N records in the database to the test, we define the fitness measure of each chromosome to be: (value of the accuracy count) $\div N$.

4. Inferring Previously Unknown Values Using Fuzzy Association Rules

Given a record, $t \in \text{dom}(I_1) \times \dots \times \text{dom}(I_p) \times \dots \times \text{dom}(I_n)$, let t be characterized by n attribute values, $\alpha_1, \dots, \alpha_p, \dots, \alpha_n$, where α_p is the value to be predicted. Let $L_p, p = 1, 2, \dots, s_p$, be the linguistic terms corresponding to the class attribute, I_p . We further let l_p be a linguistic term with domain $\text{dom}(I_p) = \{L_{p1}, L_{p2}, \dots, L_{ps_p}\}$. The value of α_p is given by the value of l_p . To predict the correct value of l_p , FARM searches the association rules with $L_{pq} \in \text{dom}(I_p)$ as consequents. For any combination of attribute values, α_φ , $p \notin \varphi$, of t , it is characterized by a linguistic term, $L_{\varphi k}$, to a degree of compatibility, $\lambda_{L_{\varphi k}}(t)$, for each $k \in \{1, 2, \dots, s_\varphi\}$.

Given those rules implying the assignment of L_{pq} , $L_{\varphi k} \Rightarrow L_{pq} [w_{L_{pq} L_{\varphi k}}]$, for all $k \in \zeta \subseteq \{1, 2, \dots, s_\varphi\}$, the evidence for or against such assignment is given by

$$w_{L_{pq} \alpha_\varphi} = \sum_{k \in \zeta} w_{L_{pq} L_{\varphi k}} \cdot \lambda_{L_{\varphi k}}(t) \quad (6)$$

Suppose that, of the $n - 1$ attribute values excluding α_p , only some combinations of them, $\alpha_{[1]}, \dots, \alpha_{[j]}, \dots, \alpha_{[\beta]}$ with $\alpha_{[j]} = \{\alpha_i \mid i \in \{1, 2, \dots, n\} - \{p\}\}$, are found to match one or more rules, then the overall weight of evidence for or against the value of l_p to be assigned to L_{pq} is given by

$$w_q = \sum_{j=1}^{\beta} w_{L_{pq} \alpha_{[j]}} \quad (7)$$

In case that I_p is categorical, l_p is assigned to L_{pc} if

$$w_c > w_g, g = 1, 2, \dots, s'_p \text{ and } g \neq c \quad (8)$$

where s'_p ($\leq s_p$) is the number of linguistic terms implied by the rules. α_p is therefore assigned to $i_{pc} \in \text{dom}(I_p)$.

If I_p is quantitative, a novel method is used to assign an appropriate value to α_p . Given the linguistic terms, $L_{p1}, L_{p2}, \dots, L_{ps_p}$, and their overall weights of evidence, w_1, w_2, \dots, w_{s_p} , let $\mu'_{L_{pu}}(i_p)$ be the weighted degree of membership of $i_p \in \text{dom}(I_p)$ to the fuzzy set L_{pu} , $u \in \{1, 2, \dots, s_p\}$. $\mu'_{L_{pu}}(i_p)$ is given by

$$\mu'_{L_{pu}}(i_p) = w_u \cdot \mu_{L_{pu}}(i_p) \quad (9)$$

where $i_p \in \text{dom}(I_p)$ and $u = 1, 2, \dots, s_p$. The defuzzified value, $F^{-1}(\bigcup_{u=1}^{s_p} L_{pu})$, which provide an appropriate value for α_p is then defined by

$$F^{-1}(\bigcup_{u=1}^{s_p} L_{pu}) = \frac{\int_{\text{dom}(I_p)} \mu'_{L_{p1}} \cup \mu'_{L_{p2}} \cup \dots \cup \mu'_{L_{ps_p}}(i_p) i_p di_p}{\int_{\text{dom}(I_p)} \mu'_{L_{p1}} \cup \mu'_{L_{p2}} \cup \dots \cup \mu'_{L_{ps_p}}(i_p) di_p} \quad (10)$$

where $\mu'_{X \cup Y}(i) = \max(\mu'_X(i), \mu'_Y(i))$ for any fuzzy sets X and Y . For quantitative predictions, we use *root-mean-squared error* as a performance measure. Given a set of test records, D , let n be number of records in D . For any record, $r \in D$, let $[l, u] \subset \mathfrak{R}$ denote the domain of the class attribute. We further let t_r be the target value of the class attribute in r and o_r be the value predicted by FARM. The root-mean-squared error, *rms*, is defined as

$$\text{rms} = \sqrt{\frac{1}{n} \sum_{r \in D} (\frac{t_r - l}{u - l} - \frac{o_r - l}{u - l})^2} \quad (11)$$

5. Experimental Results

In order to evaluate the effectiveness of FARM, we applied it to a real-life database concerning with the agriculture business in mainland China. The database contains data obtained from a survey about the economic situations of 280 villages in 1992. The scope includes production process, organizational structure, and working capital, etc. This database consists of 280 records and each record is characterized by 167 attributes. There are 99.6% of records having missing values in at least one of the attributes. As an illustration, let us consider attribute *Capital-of-transportation* in detail. We define the linguistic terms *Very-low*, *Low*, *Moderate*, *High*, *Very-high*, and *Extremely-high* for *Capital-of-transportation* (Fig. 7).

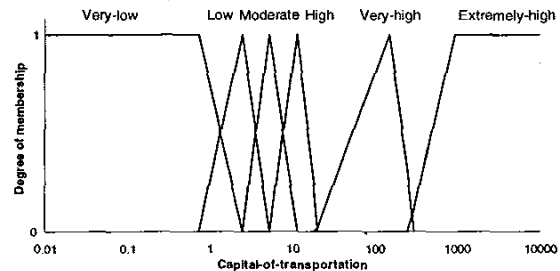


Fig. 7. Definition of linguistic terms for attribute *Capital-of-transportation*.

Using the linguistic terms described in Section 3.1, we applied FARM to the database. Among the fuzzy association rules discovered by FARM, the following rules are particularly interesting.

Investment-in-construction = *Very-high*
 \Rightarrow *Capital-of-transportation* = *High* [1.53]
Capital-of-agriculture = *Very-high*
 \Rightarrow *Capital-of-transportation* = *High* [1.17]
Investment-in-construction = *Very-high*
 \wedge *Capital-of-agriculture* = *High*
 \Rightarrow *Capital-of-transportation* = *High* [3.51]

The first one says that very high investment in construction tends to introduce high capital of transportation

whereas the second one says that very high capital of agriculture is associated with high capital of transportation. Based on these two rules, it is reasonable that both of these two factors is related to high capital of transportation. It is exactly what the third rule says. The fact that the weight of evidence of the third rule is substantially greater than the weight of evidence of the first and the second rules means that the capital of transportation is more likely to be high when both the investment in construction and the capital of agriculture are very high at the same time.

Let us consider the following fuzzy association rule as well.

Income-of-industry = Very-low
 \wedge *Capital-of-commerce-and-food-&-beverage = Very-low*
 \wedge *Capital-of-commerce-and-food-&-beverage-(self) = Very-low*
 \Rightarrow *Capital-of-transportation = Moderate* [-4.80]

This rule states that there will not even be moderate capital of transportation when the income of industry and the overall and the self used capital of commerce and food & beverage are very low. This rule is an example of negative association rules. It should be noted that existing algorithms (e.g. [1, 11]) are unable to discover negative associations. The ability of FARM to mine negative association rules is another unique feature of it.

Among the attributes in the database, one of them, called *Capital-of-transportation*, is identified as the class attribute for further experimentation. Of the 280 records, 220 are randomly selected for training and the others are used for testing. The attribute *Capital-of-transportation* is quantitative and the predictions on this attribute are therefore of quantitative values. Unfortunately, many data mining techniques (e.g. [1, 11]) which can only classify a record into appropriate category are unable to give quantitative values as predictions. As a result, they are not readily applicable to this task. On the contrary, by representing attribute *Capital-of-transportation* with a set of linguistic terms, FARM is able to produce predictions which are of quantitative values. The root mean squared error calculated by (11) over the class attribute *Capital-of-transportation* for the database is equal to 4.5%. In other words, the predictions produced by our algorithm deviated from the target values by 4.5% in average.

6. Conclusions

We presented a novel algorithm, called FARM, which employs linguistic terms to represent the revealed regularities and exceptions in this paper. The definition of linguistic terms is based on fuzzy set theory and hence we call the rules having these terms fuzzy association rules. Unlike other algorithms which discover association rules based on the use of some user supplied threshold such as minimum support and minimum confidence, FARM employs adjusted difference analysis to identify interesting associations among attributes without using any user supplied thresholds. FARM also has unique features that it

is able to discover both positive and negative associations and it uses a confidence measure, called weight of evidence, to represent the uncertainty associated with the fuzzy association rules. Furthermore, FARM has another unique feature that quantitative values can be inferred from fuzzy association rules. This allows interesting associations between different quantitative values to be revealed.

7. References

- [1] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," in *Proc. of the ACM SIGMOD Int'l Conf.*, Washington D.C., May 1993, pp. 207-216.
- [2] W.-H. Au and K.C.C. Chan, "An Effective Algorithm for Discovering Fuzzy Rules in Relational Databases," in *Proc. of the 1998 IEEE Int'l Conf. on Fuzzy Systems*, Anchorage, Alaska, May 1998, pp. 1314-1319.
- [3] K.C.C. Chan and W.-H. Au, "An Effective Algorithm for Mining Interesting Quantitative Association Rules," in *Proc. of the 12th ACM Symp. on Applied Computing*, San Jose, CA, Feb. 1997, pp. 88-90.
- [4] K.C.C. Chan and W.-H. Au, "Mining Fuzzy Association Rules," in *Proc. of the 6th ACM Int'l Conf. on Information and Knowledge Management*, Las Vegas, Nevada, Nov. 1997, pp. 209-215.
- [5] K.C.C. Chan and A.K.C. Wong, "A Statistical Technique for Extracting Classificatory Knowledge from Databases," in [10], pp. 107-123.
- [6] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (Eds.), *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, 1996.
- [7] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.
- [8] K. Hirota and W. Pedrycz, "Linguistic Data Mining and Fuzzy Modelling," in *Proc. of the 5th IEEE Int'l Conf. on Fuzzy Systems*, New Orleans, LA, Sep. 1996, pp. 1488-1492.
- [9] D.H. Lee and M.H. Kim, "Database Summarization Using Fuzzy ISA Hierarchies," *IEEE Trans. on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 27, no. 4, pp. 671-680, Aug. 1997.
- [10] G. Piatetsky-Shapiro and W.J. Frawley (Eds.), *Knowledge Discovery in Databases*, AAAI/MIT Press, 1991.
- [11] R. Srikant and R. Agrawal, "Mining Quantitative Association Rules in Large Relational Tables," in *Proc. of the ACM SIGMOD Int'l Conf.*, Monreal, Canada, June 1996, pp. 1-12.
- [12] R.R. Yager, "On Linguistic Summaries of Data," in [10], pp. 347-363.