2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Optimized Fuzzy Association Rule Mining for Quantitative Data

Hui Zheng, Jing He, Guangyan Huang, Yanchun Zhang

Abstract— **With the advance of computing and electronic technology, quantitative data, for example, continuous data (i.e., sequences of floating point numbers), become vital and have wide applications, such as for analysis of sensor data streams and financial data streams. However, existing association rule mining generally discover association rules from discrete variables, such as boolean data ('0' and '1') and categorical data ('sunny', 'cloudy', 'rainy', etc.) but very few deal with quantitative data. In this paper, a novel optimized fuzzy association rule mining (OFARM) method is proposed to mine association rules from quantitative data. The advantages of the proposed algorithm are in three folds: 1) propose a novel method to add the smoothness and flexibility of membership function for fuzzy sets; 2) optimize the fuzzy sets and their partition points with multiple objective functions after categorizing the quantitative data; and 3) design a two-level iteration to filter frequent-item-sets and fuzzy association-rules. The new method is verified by three different data sets, and the results have demonstrated the effectiveness and potentials of the developed scheme.**

Index Terms—**Quantitative Association Rule; Fuzzy sets; Optimized Partition Points; Objective Function;**

I. Introduction

With the advance of computing and electronic technology, how to analyse quantitative data, for example, continuous data (i.e. sequences of floating point numbers) has become a crucial issue to solve, such as for analysis of sensor data streams and financial data streams. However, classical methods for association rule mining concern only non-quantitative variables such as binary and categorical data objects. Binary variables are also called boolean variables, whose values are either 0 (false) or 1 (true). Categorical variables are often labelled with category names, such as "sunny", "cloudy", "rainy". It is also very often to represent categorical values with integers, which can be considered as groups of binary values. In contrast, the values of quantitative variables are usually represented by floating point numbers, they are so different to binary and categorical variables that conventional association rule approaches are not suitable for quantitative variables [1]. Therefore, several methods have been proposed to convert the quantitative attributes to categorical data objects, so these classical methods can be used.

Frawley et al. proposes a mining approach, which partitions quantitative attributes into intervals [3]. Then, some approaches try to discover interval conditions on quantitative attributes using association rule with clustering [4], rule templates [5], specific interest measures [6] and genetic algorithms [7].

Most studies focus on how to group quantitative data into different sets. In all of the above methods, the quantitative data are transformed into category data and the association rule therefore can be achieved. But as direct discretization methods, they reduce the precision of data objects and if the pre-transformed data are close to the partition points, sharp boundaries problems will emerge when pre-transformed data transform into fuzzy sets. Real-world applications, however, usually need to keep this advantages of quantitative values as well as cutting down the sharp boundaries in partitioning process.

Fuzzy association rule is a suitable method consisting more information for quantitative data. For instance, in Lee et al.'s paper [8], fuzzy sets are first introduced as an extension of association rules, which keep the precision of quantitative data with fuzzy sets and diminish the sharp boundaries while dividing the intervals to change fuzzy transactions into crisp ones. Then Delgado et al. [9] proposes a general model to discover association rules, using the definition of certainty factors and very strong rules to get the proper fuzzy association rules. Different from these, dozens of researchers have presented numerous methods to improve fuzzy association rule mining: Dubois et al. [10] develops an assessment approach partitioning the data into two groups using a given rule: those against the rule (the counterexamples) and those that are irrelevant; De Cock et al. [11] introduce new quality measures identifying the set of positive as well as the set of negative examples; some other papers [12][13] apply extra measures (such as clustering, classifying) to modify fuzzy association rule methods.

Just like the fuzzy association rule mining methods mentioned in these papers, in the fuzzy context, one can extend the boolean values $0, 1$ (indicating absence and presence) to the interval $[0, 1]$. Whether a tuple contains an item is characterized by the membership. Consider blood pressure test as an example. Suppose a patient took a blood pressure test, and a doctor tries to determine whether the patient's blood pressure is high or not. Note that blood pressure is measured quantitatively, i.e., what the doctor measures is a real number, not a boolean or binary value. For example with classification of blood pressure for adults, there are normally two criteria: systolic and diastolic pressure. If the systolic pressure falls into the interval of 120-139 mmHG and the diastolic pressure falls into 80-89 mmHG, the patient will be diagnosed as pre hypertension. But the hard cut might not always apply to all adult patients. Imagine if the patient only

Hui Zheng are with University of Chinese Academy of Sciences, Beijing, China. Jing He, Guangyan Huang and Yanchun Zhang are with Centre for Applied Informatics, College of Engineering and Science, Victoria University, Melbourne, Australia (email: zhenghui12b@mails.ucas.ac.cn {Jing.He,Guangyan.Huang,Yanchun.Zhang}@vu.edu.au).

This work was supported by National Natural Science Foundation of China under Project 61332013 and Australia Research Council under Project LP100200682.

meets one criterion (i.e. diastolic) for hypotension not two, the hard quantitative interval will lose the power to determine if the adults with or without the hypertension. The doctor can estimate the membership value [8] of the blood pressure to the high-blood-pressure fuzzy set. Actually, what the doctor thinks about is that how much the blood pressure belongs to "high", and this membership value lies between 0 and 1, where 0 indicates that the blood pressure is not high, and 1 indicates that the blood pressure is high. However, if the blood pressure given as time series data (continuous data), methods mentioned above are difficult to extend.

Classical fuzzy association rule mining methods are, however, still absence of the optimization to the partition points used for generating the rules (while converting the quantitative data into fuzzy sets, the partition points are used to divide every two neighboring fuzzy sets). These partition points are essential parts of forming fuzzy sets in almost every Fuzzy Association Rule Mining models (abbreviated as FARM in the rest of this paper). So when we originally divide systolic pressure into three fuzzy sets $(-\infty, 110], (110, 150], (150, +\infty)$ with two partition points $110, 150$ (which is not accurate and it is inconsistent with general knowledge $120, 139$), some optimization methods are needed to improve the partition points. However, the common FARM methods take little notice of it. Different from these, the measures to filter association-rules and to smooth the membership function [9][10] are still needed to improve in common FARM methods.

According to the current restrains in FARM, a generic Optimized Fuzzy Association Rule Mining (OFARM) method is proposed in this paper, which is easy to extend for continuous data. It for the first time optimizes the partition points for fuzzy sets, where a multiple objective function is used. Then, a flexible measure is presented to smooth membership function and further lessen the sharp phenomenons. Ultimately, we use a two-level iterative to generate association-rules. To achieve these, main steps of our OFARM are shown as the follows:

1) Use the membership function and original partition points to convert the quantitative data; 2) generate original frequent itemsets with Apriori Algorithm [14]; 3) optimize the partition points in a defined multiple objective function and generate the final frequent itemsets; 4) employ certainty factor [9] and confidence [14] to generate frequent itemsets and filter fuzzy association rules.

The rest of this paper is organized as follows. Section II presents the preliminaries of our algorithm. Section III describes the OFARM algorithms including processes about the multiple objective optimization, membership function and two-level iteration. The experimental results are presented in Section IV to validate the effectiveness of our OFARM algorithm. Finally, Section V concludes the paper.

II. PRELIMINARIES

In this section, we present the preliminary knowledge of fuzzy association rules, which contains the explanations of some user-defined thresholds and the optimized method used in this paper to mine the real-interesting rules.

A. The Support and Confidence of Fuzzy Association Rule

Generally, with the membership function introduced in the fuzzy association rule [12][13], the support of an item-set can be counted as follows. Let the database be D and an item-set $A = A_1, A_2, \cdots, A_l \subset I$, where I is the set comprised of all items in D as in the crisp association rule. The support of a transaction $t \in D$ to the item-set A can be defined as

$$Supp(A,t) = \mu_A(t) = \mu_{\cap_{i=1}^{l} A_i}(t).$$

where $\mu_{A_i}(t)$ denotes the membership value of the item A_i in the transaction t and $\mu_{\cap_{i=1}^{l} A_i}(t)$ denotes the fuzzy logic $\cap$ of these membership values. Taking this fuzzy logic $\cap$ as the well-known product, the support of A from the database D is thus defined as

$$Supp(A) = \sum_{t\in D} Supp(A,t) = \sum_{t\in D} \mu_A(t) = \sum_{t\in D}\prod_{i=1}^{l} \mu_{A_i}(t).$$

There are also other possibilities for the fuzzy logic $\cap$ as long as they are a triangular norm (t-norm for short). For simplicity, however, we use the product to calculate the item-set support because, given a transaction $t = t_1, t_2, \cdots, t_m$ and the set of all items $I = I_1, I_2, \cdots, I_k$, the support of I from the transaction t will be always 1 by the normalization assumption. Therefore, the support of an association rule $A \rightarrow B$ is defined as

$$Supp(A \rightarrow B) = \frac{Supp(A \cup B)}{|D|} = \frac{\sum_{t\in D}\prod_{x\in A\cup B}\mu_x(t)}{|D|}, \quad (1)$$

and the confidence of the rule therefore is

$$Conf(A \rightarrow B) = \frac{Supp(A \cup B)}{Supp(A)} = \frac{\sum_{t\in D}\prod_{x\in A\cup B}\mu_x(t)}{\sum_{t\in D}\prod_{x\in A}\mu_x(t)}. \quad (2)$$

B. The Certainty Factor of Fuzzy Association Rule

The support-confidence framework [16] introduced above is regarded as the classical theory of association rule mining models. It is not perfect, however, and many have pointed out its defects. To complement this framework, new measures for evaluation of association rules were proposed, including conviction [17], interest [18] and certainty factor [9] [15][16]. Among these, certainty factor is regarded as one of the most effective measures. It is defined as below:

$$CF(A \rightarrow B) = \begin{cases} \frac{Conf(A\rightarrow B) - Supp(B)}{1 - Supp(B)}, & Conf(A \rightarrow B) >= Supp(B), \\ \frac{Conf(A\rightarrow B) - Supp(B)}{Supp(B)}, & Conf(A \rightarrow B) < Supp(B). \end{cases} \quad (3)$$

Obviously,

$$-1 \leq CF(A \rightarrow B) \leq 1,$$

and when $CF(A \rightarrow B) > 0$, the antecedent of the rule is positively correlated to the consequence; when $CF(A \rightarrow B) < 0$, it is negatively correlated; when $CF(A \rightarrow B) = 0$, it is independent to the consequence.

C. Gradient-based Optimization Methods

Gradient-based optimization methods [19] are based on the observation that the value of the function grows fastest along the gradient of the function. Among these methods, one of the simplest is the steepest descent method. This method is also called gradient descent, and it is a first-order optimization algorithm. In this method, it takes proportional steps to the negative of the gradient of the objective function, so that it approaches a local minimum of the function. Specifically, suppose that the multivariate objective function $F(\mathbf{x})$ is defined and differentiable in a neighborhood of the point P_0, then the direction, along which $F(\mathbf{x})$ decreases fastest, is the negative direction of the gradient $\nabla F(P_0)$. The method therefore goes through the following iteration

$$\mathbf{x}_n = \mathbf{x}_{n-1} - \gamma_n \nabla F(\mathbf{x}_{n-1}) \tag{4}$$

to find the local minimum, where the sequence $\{\mathbf{x}_n\}$ is constructed by the method, and it converges to the local minimum if the parameters $\gamma_1, \gamma_2, \ldots$ are chosen carefully. A common approach is to do an iterative line search for each γ_n by halving the value of such γ_n repeatedly until the new point $\mathbf{x}_n$ makes $F(\mathbf{x}_n) < F(\mathbf{x}_{n-1})$.

For our optimization algorithm, we extend the idea of steepest descend method by relaxing the constraint on the search direction. It is unnecessary to search strictly along the gradient, especially when we search for improvement of multiple objective functions at the same time. Our algorithm is described in next section.

III. Optimized Fuzzy Association Rule

The Optimized Fuzzy Association Rule Mining (OFAR-M) we proposed is differ itself from the common fuzzy association rule in taking an additional two-level iteration to optimize the mining-output after generating the initial frequent-item-sets and association-rules, as illustrated by Figure 1. The inner level of the iteration aims at finding optimized membership functions of the fuzzy sets with several objective functions with respect to the parameters of the fuzzy-set membership functions (actually, the parameters are the partition points in our algorithm), whereas the outer level of the iteration finds updates frequent item-sets and association rules with the fuzzy-sets improved by the inner level. Therefore, our OFARM algorithm will eventually doscover optimized association rules after repetitively improving fuzzy-set partitions and finding frequent item-sets.

A. A Multi-objective Optimization Scheme

Our OFARM tries to optimize the partition points of the fuzzy sets with respect to multiple objectives at the same time, where the objectives include:

1) the supports of the rules,
2) the confidences, and
3) the certainty factors.

However, it is inconvenient to optimize the partition points for all these objectives over all association rules at the same time, so we construct some reasonable approximate

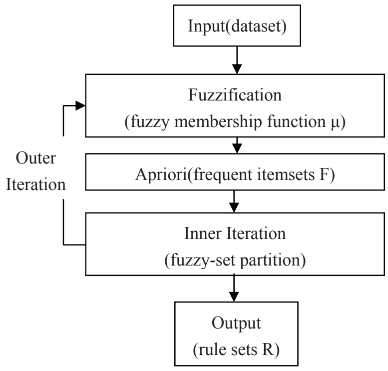


Fig. 1. Flowchart of Optimized Fuzzy Association Rule Mining (OFARM)

objectives. First, for a given association rule r, the following function

$$\varphi(r) = \min \left\{ \begin{array}{l} \mathrm{Supp}(r) - \mathrm{minSupp} \\ \mathrm{Conf}(r) - \mathrm{minConf} \\ \mathrm{CF}(r) - \mathrm{minCF} \end{array} \right\} \tag{5}$$

is used for rule quality control, where $\mathrm{minSupp}$, $\mathrm{minConf}$ minCF are predefined thresholds. It is clear that a larger value of $\varphi(r)$ indicates a better quality of r, so based on this rule quality control function, we define the following objectives for optimization:

1) $\Phi_1 = \max_r \varphi(r)$, i.e., the quality of the best rule,
2) $\Phi_3 = \max_{r_1, r_2, r_3} \sum_{i=1}^{3} \varphi(r_i)$, i.e., the sum of the quality of the best three rules (if there are not enough rules, the worst rule is duplicated for multiple times to complete this sum),
3) Φ_5, the sum of the quality of the best five rules,
4) Φ_{10}, the sum of the quality of the best ten rules, and
5) $\Phi_{n/2}$, the sum of the good half of the rules we have mined, where n is the number of these rules.

Now our OFARM tries to **optimize these five objectives simultaneously**, by searching in the domain of the partition points along a direction that is computed by the algorithm in Table I.

The idea of this direction-computation algorithm is as follows. Given a fixed input of partition points, for each objective function, its gradient at the input determines a unique plane that takes this gradient as its normal vector and contains this input. This plane separates the domain of the partition points into two halves. When searching along any direction located within the half pointed by the gradient vector, the value of objective function will increase at the neighbourhood of the input vector. For convenience, we call such directions **positive directions** of this objective function, and call the half of the domain containing positive directions the **positive half space** of the objective function. Note that it is unnecessary to confine the search along the gradient direction, searching along any positive directions will work as well. Now we need a direction that is a positive direction of all objective functions. Obviously this can be fulfilled by finding the intersection of the positive half spaces of these objective functions.

TABLE I
COMPUTE DIRECTION.

Algorithm 1

Input: minimum support(minSupp); minimum confidence (minConf); minimum certainty factor (minCF); a given vector of partition points $\mathcal{X}_0$, the number of rules (n), and the gradients of objective functions at $\mathcal{X}_0$:

$$\mathbf{g}_1 = \nabla\Phi_1,\ \mathbf{g}_2 = \nabla\Phi_3,\ \mathbf{g}_3 = \nabla\Phi_5,$$

$$\mathbf{g}_4 = \nabla\Phi_{10},\ \mathbf{g}_5 = \nabla\Phi_{n/2},$$

Output: the direction $\boldsymbol{\gamma}$ along which hopefully all objective functions increase at the neighbourhood of $\mathbf{x}$.

Initialization: $\boldsymbol{\gamma} \leftarrow \{0, \ldots, 0\}$.
for $i = 1, \ldots, 5$ **do**
  $\boldsymbol{\beta} \leftarrow \mathbf{g}_i$;
  for $j \neq i$ **do**
    if $\langle \boldsymbol{\beta}, \mathbf{g}_j \rangle < 0$ **then**
      $\boldsymbol{\beta} \leftarrow \boldsymbol{\beta} - \frac{\langle \boldsymbol{\beta}, \mathbf{g}_j \rangle}{\langle \boldsymbol{\beta}, \boldsymbol{\beta} \rangle} \mathbf{g}_j$;
    end if
  end for
  $\boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma} + \boldsymbol{\beta}$;
end for
Return $\boldsymbol{\gamma}$;

Since the optimization objective function is with respect to the parameters of the fuzzy-set membership functions, let's consider the fuzzy sets and their membership functions.

B. Fuzzy Sets and Membership Functions

Suppose that all transactions in a dataset take a common set of quantitative attributes, also suppose that the data of each quantitative attribute will be converted to categorical data by dividing the range of that attribute into three fuzzy sets. In such a case, let FS_L, FS_M, FS_R denote these three fuzzy sets, with the subscripts L, M, R indicating their position in the range of the attribute data (i.e., Left, Middle, Right).

To generate these three fuzzy sets for each quantitative attribute, we first partition the range of the attribute data into five intervals, with each of the intervals containing almost equal number of transaction data items. Actually, to make this partition, we sort the transaction data of this attribute by their value, and then compute four dividing points that separate these five intervals as follows:

$$DP_i = 0.5\left(s_{\lfloor i \cdot m/5 \rfloor} + s_{\lfloor i \cdot m/5 \rfloor + 1}\right), \quad i = 1, 2, 3, 4,$$

where DP_i is the i-th dividing point, s_k is the k-th sorted value of the transaction data items, and m is the number of transactions. Now we assign the whole 1st (i.e. the leftmost) interval to FS_L, so that within this interval, the membership function of FS_L is 1 and those of other fuzzy-sets are 0. Similarly, we assign the whole 3rd interval to FS_M, the whole 5th interval to FS_R. For another two intervals, we let interval 2 partially belongs to FS_L and partially belongs to FS_M, and let interval 4 partially belongs to FS_M and partially belongs to FS_R. This partitioning process is illustrated by Figure 2.

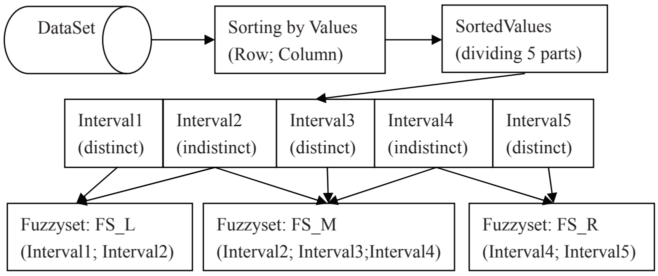


Fig. 2. Flow Chart of Generating the Fuzzy Sets

Obviously, the central task of the construction of the membership functions is at interval 2 and interval 4 as mentioned above, since each of these two intervals are shared between two fuzzy sets. So let's formulate the membership functions.

Take interval 2 as example, it is shared by FS_L and FS_M. Let μ_L, μ_M denote the membership function of FS_L, FS_M respectively, then for all $x \in [DP_1, DP_2]$ (i.e., interval 2), by the definition of membership function,

$$\mu_L(x) + \mu_M(x) = 1,$$
$$\mu_L(x) \geq 0, \quad \mu_M(x) \geq 0.$$

Now let's make the membership functions and their derivatives continuous at the dividing points of the intervals, then

$$\mu_L(DP_1) = 1,\ \mu_L(DP_2) = 0,\ \mu'_L(DP_1) = \mu'_L(DP_2) = 0,$$

$$\mu_M(DP_1) = 0,\ \mu_M(DP_2) = 1,\ \mu'_M(DP_1) = \mu'_M(DP_2) = 0.$$

For simplicity and without loss of generality, let $DP_1 = 0$ and $DP_2 = 1$, then the only degree-3-polynomial configuration of the membership functions satisfying the relations above is to let

$$\mu_M(x) = x^2(-2x + 3),\ \mu_L(x) = 1 - \mu_M(x). \quad (6)$$

However, this configuration does not give us the flexibility of adjusting the membership. Hence, a small modification is made:

$$\mu_M(x) = r(x)^2(-2r(x) + 3),\ \mu_L(x) = 1 - \mu_M(x), \quad (7)$$

where

$$r(x) = x^{\alpha}, \quad (8)$$

such that the parameter α is given by

$$r(x_0) = 0.5, \quad (9)$$

where x_0 is an adjustable point in (DP_1, DP_2), and it is the parameter which will be used in the optimization objective function.

Similarly, we define another parameter x_1 for interval 4, so that the two fuzzy sets sharing interval 4 (i.e., FS_M and FS_R) are adjustable.

Note that if there are K quantitative attributes in data sets, we will need one pair of x_0, x_1 for each of these attributes, that means we need $2K$ of such parameters. For the convenience of reference, let $\mathcal{X}$ denotes the set comprised of all these parameters at the rest of this paper.

According to this measure proposed above, the number of fuzzy sets of the categorical attributes derived by dividing quantitative attributes can easily extend to q. Suppose q to be an odd number, then we will partition the range of the attribute data into $(q-1)\times 2+1$ intervals. The situations of the first fuzzy sets and the last fuzzy sets are corresponding the situations of fuzzy sets FS_L and FS_R, which come from the first two intervals and the last two intervals respectively. Then all the fuzzy membership function of other fuzzy sets including the median fuzzy sets can be defined as the same as it is in the fuzzy set FS_M, that is, they are all generated from three corresponding intervals. Likewise, the definition of fuzzy membership function could be accomplished by dividing the range of the attribute data into $2q$ intervals without the median fuzzy sets, if the q were an even number.

C. Strong Rules

An association rule r is called strong rule if and only if the certainty factor and confidence of $X \to Y$ are greater than the two user-defined thresholds minCF and minConf, and the support of frequent itemset $X \cup Y$ is greater than minSupp. Note that this means

$$\varphi(r) > 0,$$

where φ is the rule-quality control function as defined in (5).

D. The Inner Iteration

The inner iteration aims at finding optimized fuzzy-set partitions with the objective function mentioned above. For this purpose, it will use the association rules from last iteration, except that at the first inner iteration, which uses the association rules generated by the unoptimized fuzzy-sets. The pseudo-code is shown in Table II. From the $Line$ 2 to $Line$ 14 is the whole inner iteration process, $Line$ 4-9 updates the three user-defined thresholds. Then the $Line$ 10-13 show the step to optimize the objective function, $Line$ 12 computes the search direction like the algorithm of Table I.

E. The Outer Iteration

The outer iteration searches for appropriate association rule sets and frequent item-sets with the fuzzy-set partition parameters improved by the inner iteration. The pseudo-code is shown in Table III. All of the steps of the outer iteration are shown from $Line$ 1 to $Line$ 8. After initialization step including $Line$ 1-5, the $Line$ 6-8 illustrate the optimization step for partition points, which is just the inner iteration step of Table II.

TABLE II

OPTIMIZE FUZZY-SET PARTITION PARAMETERS.

Algorithm 2

Input: initial (or previous) partition parameters $\mathcal{X}_0$, the maximum number of iterations I;

Output: optimized partition parameters $\mathcal{X}$; the optimized set of frequent item-sets F and the optimized set of Association Rules R;

1: Initialization: $\mathcal{X} \leftarrow \mathcal{X}_0$;
2: Generate frequent item-sets F and association rules R, make sure R contains only strong rules;
3: **for** $i = 0$ to I **do**
4:     **for** $fs \in F$ **do**
5:         Update $\mathrm{Supp}(fs, \mathcal{X})$;
6:     **end for**
7:     **for** $r \in R$ **do**
8:         Update $\mathrm{Supp}(r, \mathcal{X})$, $\mathrm{Conf}(r, \mathcal{X})$, and $\mathrm{CF}(r, \mathcal{X})$;
9:     **end for**
10:     Compute the objective functions;
11:     Richardson extrapolation: calculate derivatives;
12:     Use Algorithm 1 to compute a search direction $\boldsymbol{\gamma}$;
13:     Do a line-search along $\boldsymbol{\gamma}$ for step size λ, update

$$\mathcal{X} \leftarrow \mathcal{X} + \lambda\boldsymbol{\gamma}$$

to make the values of the objective functions larger;
14: **end for**
15: Return $\mathcal{X}, F, R$;

TABLE III

SEARCH FOR ASSOCIATION RULE SETS.

Algorithm 3

Input: a database D; minimum support(minSupp); minimum confidence (minConf); the maximum number of outer iterations J and the maximum number of inner iterations I;

Output: the optimized set of frequent item-sets F and the optimized set of Association Rules R;

1: Initialization: load D and compute $DP_1, \ldots, DP_4$ for each quantitative attributes, where DP_i's are dividing points of the intervals as mentioned in section B;
2: **for** $\mathcal{X}$ components x_0, x_1 of each numeric attribute (fuzzy-set partition parameters) **do**
3:     $x_0 \leftarrow 0.5 * (DP_0 + DP_1)$;
4:     $x_1 \leftarrow 0.5 * (DP_2 + DP_3)$;
5: **end for**
6: **for** $j = 0$ to J **do**
7:     Algorithm:Optimize Fuzzy-set Partition Parameters, with input parameters $\mathcal{X}$ and I;
8: **end for**
9: Return F, R obtained from $Line$ 7;

IV. Experiments

A. Corresponding Method and Experimental Datasets

In this section, we evaluate optimized fuzzy association rule mining (OFARM) we proposed by comparing it with a general fuzzy association rule mining (GFARM) algorithm in [9]. The membership function of GFARM and our OFARM is defined in section III-B and strong rules they using are defined in section III-C. Three datasets, including "Wisconsin Diagnostic Breast Cancer (WDBC) ", "Wisconsin Prognostic Breast Cancer (WPBC)", "Pima Indians Diabetes" from UCI (University of California at Irvine) repository, have been used to demonstrate the effectiveness and efficiency of our OFARM algorithm. The objective functions $\Phi_1, \Phi_3, \Phi_5, \Phi_{10}$ and $\Phi_{n/2}$ are defined in section III-A. The maximum number of outer iterations and the number of fuzzy sets are set to be $J = 5$, $q = 3$ respectively. The pruning method [14] is added to monitor association rules, to avoid huge number of fuzzy-set rules.

The details of our OFARM are shown in Figure 1. To ensure the integrity and quality of results, different thresholds of minSupp, minConf and minCF are selected for three datasets. The higher the thresholds are, the better the quality of the discovered association rules are. So if lower thresholds are chosen, there will be numerous of low quality rules, however, when too high thresholds are chosen, the number of rules will be hard to improve. Thus, we should set a suitable threshold to adapte to datasets.

B. Output of Strong Rules

In this section, we discover strong rules (related with diabetes) from "Pima Indians Diabetes", which comprises 768 instances from UCI. All of the attributes are described in Table IV, and which shows the examples of strong rules defined in section III-C.

TABLE IV

ATTRIBUTES IN PIMA INDIANS DIABETES DATASET.

Attribute ID	Attribute Description
0	Number of times pregnant
1	Plasma glucose concentration a 2 hours in an oral glucose tolerance test
2	Diastolic blood pressure (mm Hg)
3	Triceps skin fold thickness (mm)
4	2-Hour serum insulin (mu U/ml)
5	Body mass index (weight in kg/(height in m)^2)
6	Diabetes pedigree function
7	Age (years)
8	Class variable (0 or 1)

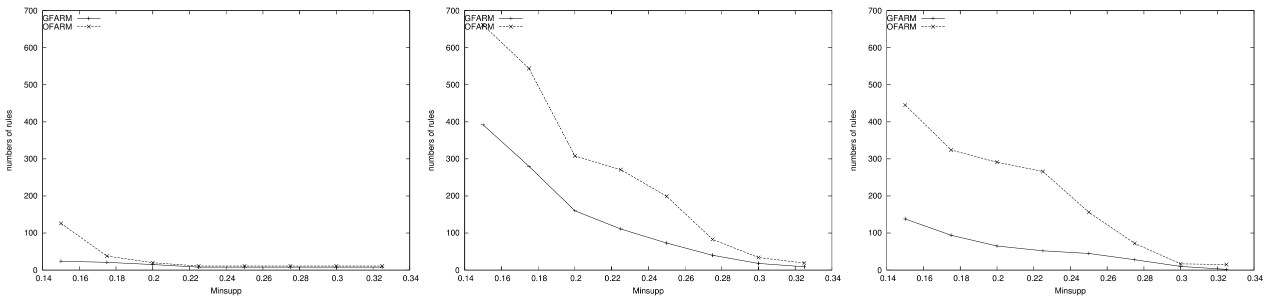
We group attributes 0-7 (non-categorical data objects) into three fuzzy sets. Attribute 8 (binary data object: 0 denotes non-diabetes and 1 denotes diabetes) is the most important one. Next, we fix three user-defined thresholds as $minSupp = 0.1, minConf = 0.5$ and $minCF = 0.1$. Then, every rule in this section is given in two parts, antecedent (on the left side of $\rightarrow$) and consequent (on the right side). The antecedent is represented by a sequence of pairs (attribute ID, fuzzy set ID), while the consequent is represented by a single pair (attribute ID, fuzzy set ID). In this section, the strong rules will be shown only if their consequent is (8, 1), which denotes diabetes relating with the most important attribute 8 . As mentioned above, the attribute IDs are ranged from 0 to 8, while the fuzzy set IDs are ranged from 0 to 2. Fuzzy set ID being 0 means the value on this is very small, while fuzzy set ID being 2 implies the value is great. Note that the concrete partition points are shown in the third column of Table V. For instance, $M_{L,0} = 1.5, M_{R,0} = 5.5$ shows the 0-th attribute is divided into three fuzzy sets with these two partitioning points $(-\infty, 1.5], (1.5, 5.5], (5.5, \infty)$, which just corresponds to the order number $0, 1, 2$.

TABLE V

STRONG RULES COMPARISON ON PIMA INDIANS DIABETES DATASET.

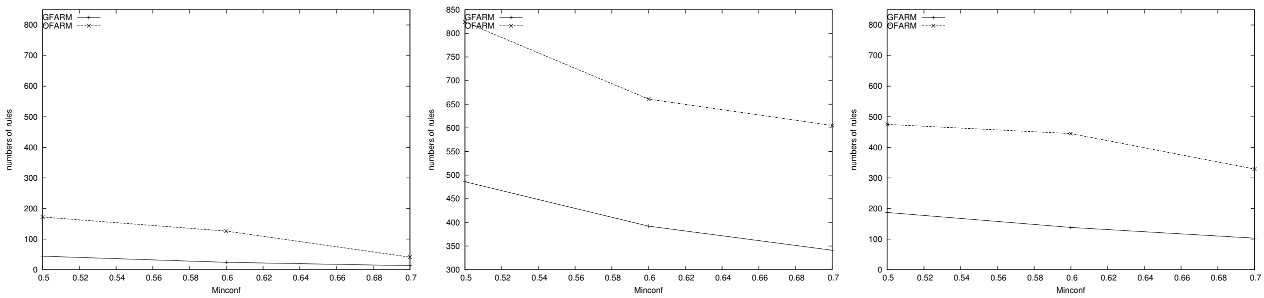
Model	Strong Rules	Partitioning Points(Fuzzy Sets)
GFARM	$(0,2) \rightarrow (8,1)$	$M_{L,0} = 1.5, M_{R,0} = 5.5;$
	$(7,2) \rightarrow (8,1)$	$M_{L,1} = 102, M_{R,1} = 136;$
	$(0,2)(7,2) \rightarrow (8,1)$	$M_{L,2} = 66, M_{R,2} = 78;$
	$(1,2) \rightarrow (8,1)$	$M_{L,3} = 25.5, M_{R,3} = 32.5283;$
	$(4,2) \rightarrow (8,1)$	$M_{L,4} = 121.372, M_{R,4} = 168.519;$
	$(1,2)(4,2) \rightarrow (8,1)$	$M_{L,5} = 28.3, M_{R,5} = 35.75;$
	$(4,2)(7,2) \rightarrow (8,1)$	$M_{L,6} = 0.2615, M_{R,6} = 0.572;$
		$M_{L,7} = 25, M_{R,7} = 38;$
OFARM	$(6,1)(7,2) \rightarrow (8,1)$	
	$(0,2)(7,2) \rightarrow (8,1)$	$M_{L,0} = 1.00417, M_{R,0} = 4.05981;$
	$(2,2)(7,2) \rightarrow (8,1)$	$M_{L,1} = 108.782, M_{R,1} = 125.105;$
	$(1,2) \rightarrow (8,1)$	$M_{L,2} = 69.9052, M_{R,2} = 74.1035;$
	$(1,2)(3,1) \rightarrow (8,1)$	$M_{L,3} = 23.1856, M_{R,3} = 34.7484;$
	$(1,2)(6,1) \rightarrow (8,1)$	$M_{L,4} = 137.612, M_{R,4} = 178.725;$
	$(4,2) \rightarrow (8,1)$	$M_{L,5} = 30.3625, M_{R,5} = 35.2095;$
	$(1,2)(7,2) \rightarrow (8,1)$	$M_{L,6} = 0.220179, M_{R,6} = 0.688003;$
	$(0,2)(1,2) \rightarrow (8,1)$	
	$(1,2)(4,2) \rightarrow (8,1)$	
	$(1,2)(5,2) \rightarrow (8,1)$	

From table V, we can observe that the GFARM discovers 7 rules in all, while the OFARM we proposed gets 11 rules with 4 conjunct strong rules. Different from these common rules, the GFARM has three rules with antecedent (0,2), (7,2), (4,2)(7,2), while our OFARM discovers other 7 antecedents (6,1)(7,2),(2,2)(7,2), (1,2)(3,1), (1,2)(6,1), (1,2)(7,2), (0,2)(1,2), (1,2)(5,2). Our OFARM finds more specific and more non-high (attribute belonging to the middle fuzzy sets) antecedents. Actually, the more specific the rules are, the easier we find useful rules in real-world applications. Generally, the higher the values of attributes, the more possible individuals catch diseases. However, it is difficult to predict/diagnose diseases from individuals, if the values of their attributes are not higher than normal values. So, one of the advantages of our OFARM algorithm is that it can find two rules with low values of antecedents, while the GFARM find nothing. Also, according to the pruning step in [14], the less antecedents rules with less confidence are filtered, so the GFARM performs not good since it find less specific antecedents. In summary, our OFARM produces not only more strong rules but also more useful and effective rules.



(a) Number of Rules of Diabetes Databset (b) Number of Rules of WDBC Databset (c) Number of Rules of WPBC Databset

Fig. 3. Comparison in Number of Rules Between GFARM and OFARM with Fixed minConf.



(a) Number of Rules of Diabetes Dataset (b) Number of Rules of WDBC Dataset (c) Number of Rules of WPBC Dataset

Fig. 4. Comparison in Number of Rules Between GFARM and OFARM with Fixed minSupp.

TABLE VI

PERFORMANCE COMPARISON OF PIMA INDIANS DIABETES DATASET.

Minsupp,Minconf = 0.6	Model	Φ_{10}	Φ_5	Φ_3	Φ_1
Minsupp=0.1	GFARM	0.385484	0.411058	0.424363	0.427242
	OFARM	0.452308	0.495134	0.512719	0.525517
Minsupp=0.125	GFARM	0.359073	0.389162	0.40556	0.422856
	OFARM	0.438575	0.473307	0.490337	0.501061
Minsupp=0.15	GFARM	0.325793	0.363516	0.38465	0.422856
	OFARM	0.417732	0.458473	0.466242	0.476415
Minsupp=0.175	GFARM	0.287954	0.3329	0.361152	0.422856
	OFARM	0.38062	0.419432	0.42623	0.433716
Minsupp=0.2	GFARM	0.243655	0.298565	0.329458	0.411097
	OFARM	0.30872	0.364397	00.399143	0.423923
Minsupp=0.225	GFARM	0.195612	0.256611	0.288037	0.377087
	OFARM	0.220501	0.28893	0.327691	0.39604

C. Comparisons and Analysis

The number of rules is a critical metric to evaluate the association rules. In this subsection, we collect the number among different minsup with fixed minConf $= 0.6$, and fixed minCF $= 0.1$, to compare the number of rules between the GFARM and OFARM we proposed. Figure 3(a)-3(c) show that the number of strong rules discovered by GFARM gradually increases with minSupp while that discovered by our OFARM algorithm ascends more quickly; meanwhile, our OFARM is persistently greater than the GFARM. That is, our OFARM can filter more rules. So we can see from Figure 3(a)-3(c), the curves of our OFARM are always over the curves of GFARM, though the gaps between every two curves varies. For instance, the numbers of rules discovered by our OFARM algorithm show great upward trends than GFARM when minSupp is low, however, only slight upward trends are described when minSupp increases. That is, if the minSupp is lower, our OFARM algorithm shows an outstanding increase in the number of strong rules. However, if the minSupp is higher, the improvement of our OFARM algorithm will be lower but still outperforms the result of GFARM. The reason is that the original strong rules are not exact and frequently changed with even a slight adjustment when minSupp is small. However, with a higer minSupp, the original strong rules which already have high quality, the adjustment become difficult and unnecessary, so the number of rules rise slightly. Supposing minSupp $= 0.15$ in Figure 3(b), the result of GFARM is around 400, while the result of our OFARM algorithm is better by rising to over 600. However, if the minSupp rises up to 0.3, the improvement is only around 20.

Our OFARM algorithm also excels GFARM in terms of gradual increasing confidence with fixed minSupp and minCF. As it is demonstrated in Figure 4(a)-4(c) with fixed minSupp $= 0.15$, minCF $= 0.1$, the number of strong rules are also always greater in our OFARM algorithm than in GFARM for the three datasets. Since only three points (minConf $= 0.5, 0.6, 0.7$) are chosen, the gaps between curves of our OFARM and GFARM are more stable in Figure 4(a)-4(c). Meanwhile, the gaps still illustrate slowly downward trends as the value of minConf goes up. For instance, Figure 4(a) shows that the gap between two curves is more than 100 with the minConf $= 0.5$; however, when the minConf is equal to 0.7, the rule numbers of our OFARM only slightly increase (approximating to 30).

We use four metrics to further verify our OFARM algorithm. The four metrics Φ_1, Φ_3, Φ_5 and Φ_{10} as the assessments of strong rules for dataset "Pima Indians Diabetes" are collected in Table VI. Then Table VI shows that the results of our OFARM algorithm invariably exceed the counterpart algorithm GFARM. Though the improvement varies with minSupp (from 0.1 to 0.225), that is, these four values decrease when the minConf increase in both two algorithms. Aslo, our OFARM witnesses great increase in all above four metrics when minSupp is low, then the increase becomes slow when minSupp increase. To be more specific, at the point of minSupp $= 0.1$, our OFARM algorithm achieves the slowest increase in metric Φ_{10}. The Φ_{10} of our OFARM performs 0.066824 better than that of GFARM; while the highest increment point is 0.098275 on Φ_1; When minSupp rise to 0.225, the Φ_1 shows the lowest increment 0.018953, while the greatest one is 0.039654 shown at the metric: Φ_3. That is, the increment of our OFARM at minSupp $= 0.1$ is twice of that with minSupp $= 0.225$.

In summary, comparing with the GFARM algorithm, our OFARM algorithm demonstrates greater number of strong rules and higher value of assessing metrics that can represent the quality of strong rules. That means, our OFARM algorithm outperforms the corresponding algorithm in terms of both quantity and quality of the number of strong rules. That is, optimizing partition points increases the quantity of the rules, while the multiple objective function enhancing the quality of rules as a whole even with varible minSupp and minConf thresholds.

V. Conclusions

In this paper, an optimized fuzzy-association-rule mining algorithm based on a generic measure has been proposed. We have shown that the features of the multiple objective function optimization make the proposed model easy to formulate and use for continuous data. Taking the two-level iteration processes into account, the fuzzy association rules and the frequent item-sets are optimized by improving the fuzzy-set partition parameters repeatedly. The experiment also demonstrates that the algorithm is capable of balancing the weight of optimization between the quantity and the quality of strong rules; that is, our OFARM algorithm outperforms the counterpart algorithm GFARM in both rule quantity and rule quality. Nevertheless, the experimental results about gradual changing in minSupp and minConf shows stability and robustness of the algorithm.

References

[1] Y. Aumann and Y. Lindell, "A statistical theory for quantitative association rules," *Journal of Intelligent Information Systems*, vol. 20, no. 3, pp. 255-283, May. 2003.

[2] H. Kalia, S. Dehuri, A. Ghosh, "A Survey on Fuzzy Association Rule Mining," *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 9, no. 1, pp. 1-27, 2013.

[3] W.J. Frawley, G. Piatetsky-Shapiro, C.J. Matheus,"Knowledge discovery in databases: an overview," *AI magazine.*, vol. 13, no. 3, pp.57, 1992.

[4] D.W.-L. Cheung, L. Wang, S.-M. Yiu and B. Zhou, "Density-based mining of quantitative association rules," *PAKDD, Lecture Notes in Computer Science, Springer*, vol. 1805, pp.257-268, 2000.

[5] K. Wang, S.H.W. Tay and B. Liu, "Interestingness-based interval merger for numeric association rules," *In KDD.*, pp. 121-128, 1998.

[6] Y. Ke, J. Cheng and W. Ng, "An information-theoretic approach to quantitative association rule mining," *Knowl Inf Syst*, vol. 16, no. 2, pp. 213-244, 2008.

[7] A. Salleb-Aouissi, C. Vrain and C. Nortet, Quantminer, "A genetic algorithm for mining quantitative association rules," *IJCAI*, vol. 7, 2007.

[8] J.H. Lee and H.L. Kwang, "An extension of association rules using fuzzy sets," *In Proceedings of the Seventh IFSA World Congress*, pp. 399-402, 1997.

[9] M. Delgado, N. Marłn and D. Snchez, "Fuzzy association rules: general model and applications," *Fuzzy Systems, IEEE Transactions*, vol. 11, no. 2, pp. 214-225, 2003.

[10] D. Dubois, E. Hllermeier and H. Prade, "A systematic approach to the assessment of fuzzy association rules," *Data Mining and Knowledge Discovery* , vol. 13, no. 2, pp. 167-192, 2006.

[11] M. De Cock, C. Cornelis, E. E. Kerre, " Fuzzy Association Rules: a Two-Sided Approach," *Proc. of Internal. Conf. on Fuzzy Information Processing-Theories and Applications*, pp.385-390, 2003.

[12] C.L. Chen, FSC. Tseng, T. Liang, "An integration of WordNet and fuzzy association rule mining for multi-label document clustering," *Data and Knowledge Engineering*, vol. 69, pp. 1208-1226, 2010.

[13] J. Alcal-Fdez, R. Alcala, "A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning," *Fuzzy Systems, IEEE Transactions*,vol. 19, no. 5, pp. 857-872, 2011.

[14] J. Han, M. Kamber, J. Pei," *Data mining: concepts and techniques*, Morgan kaufmann, pp. 229-242, 2006.

[15] E.H. Shortliffe and B.G. Buchanan, "A model of inexact reasoning in medicine," *Mathematical biosciences*, vol. 23, No. 3, pp. 351-379, 1975.

[16] F. Berzal, I. Blanco, D.S.ánchez and M.A. Vila, "Measuring the accuracy and interest of association rules: A new framework," *Intelligent Data Analysis*, vol. 6, No. 3, pp. 221-235, 2002.

[17] S. Brin, R. Motwani, J.D. Ullman and S. Tsur, "Dynamic itemset counting and implication rules for market basket data," *In ACM SIGMOD Record*, vol. 26, No. 2, pp. 255-264, 1997.

[18] C. Silverstein, S. Brin and R. Motwani, "Beyond market baskets: Generalizing association rules to dependence rules," *Data mining and knowledge discovery*, vol. 2, no. 1, pp. 39-68, 1998.

[19] J. Fliege, B.F. Svaiter, " Steepest descent methods for multicriteria optimization," *Mathematical Methods of Operations Research*, vol. 51, no. 3, pp. 479-494, 2000.