The Research of Fuzzy Classification Rules Applied to CRM

Chien-Hua Wang, Meng-Ying Chou and Chin-Tzong Pang

Abstract—In the era of great competition, understanding and satisfying customers' requirements are the critical tasks for a company to make a profits. Customer relationship management (CRM) thus becomes an important business issue at present. With the help of the data mining techniques, the manager can explore and analyze from a large quantity of data to discover meaningful patterns and rules. Among all methods, well-known association rule is most commonly seen. This paper is based on Apriori algorithm and uses genetic algorithms combining a data mining method to discover fuzzy classification rules. The mined results can be applied in CRM to help decision marker make correct business decisions for marketing strategies.

Keywords—Customer relationship management (CRM), Data mining, Apriori algorithm, Genetic algorithm, Fuzzy classification rules.

I. INTRODUCTION

CUSTOMER relationship management (CRM) is an important information system architecture as an enterprise enters e-Business [16]. Its purpose is to effectively collect, store, analyze data to come up with a core integration profile for customers through IT, such as the Internet, Data Warehouse, Data Mining etc. The results gained are to support enterprises dealing one on one interactions and customized marketing, sales and after-sales service. Through different media such as the Call Center, the Internet, Telephoning, Faxing and Salespeople, enterprises can interact, connect, trade and wait on their customers to attract new customers and secure old ones to enhance customer satisfaction, customer loyalty and profitability [23-24].

The relationship between customers and enterprises is a vital surviving factor for enterprises. When, enterprises accumulate some great amount of unhandled data such as customer transactions, one must know that this kind of data means little help to decision makers. In order to reach maximal efficient interactions between customers and enterprises, enterprises need to adopt data mining technology, such as association rules, classifications and clusters to discover useful information in databases, as well as acquire valuable information to provide enterprises with sound management skills and perfect core competitiveness in business [5].

Chien-Hua Wang is with the Department of Information Management, Yuan Ze University, Taoyuan, 32003, Taiwan, ROC (e-mail: thuck@saturn.yzu.edu.tw).

Meng-Ying Chou is with the Department of Information Management, Yuan Ze University, Taoyuan, 32003, Taiwan, ROC (e-mail: mengyingsmart@hotmail.com).

Chin-Tzong Pang is with the Department of Information Management, Yuan Ze University, Taoyuan, 32003, Taiwan, ROC (e-mail: imct-pang@saturn.yzu.edu.tw).

Additionally, in terms of decision making, one has to take users' perception and cognitive uncertainty of subjective decisions into consideration. Zadeh proposed the Fuzzy Theory [28] in 1965 to deal with cognitive uncertainty of vagueness and ambiguity [26]. Since linguistic variables and linguistic values [29-31] can be described with fuzzy concepts to subjectively correspond with the possible cognition of a decision maker, they are handy in carrying out analysis of decisions-making. Fuzzy data mining has then recently become an important research matter.

In this paper, we used a two-phase data mining technique to discover fuzzy rules for customer's classification problems [7]. This method is based on the Apriori algorithm. Firstly, the first-phase finds frequent fuzzy grids by dividing each quantitative attribute with a user-specified number of various linguistic values. Secondly, the second-phase generates effective fuzzy classification rules from those frequent fuzzy grids. In addition, the fuzzy support and the fuzzy confidence, which have been defined previously [6,11-12], are employed to determine which fuzzy grids are frequent and which rules are effective by comparison with the minimum fuzzy support (min FS) and the minimum fuzzy confidence (min FC), respectively.

Howver, both min FS and min FC are not easily user-specified for each classification problem. To solve this problem, the genetic algorithm (GA) [4] is incorporated into this algorithm to automatically determine those two parameters. A binary chromosome with sufficiently large length used in this paper is composed of two substrings: one for the min FS and the other for the min FC. Each generation of the GA can obtain the fitness value of each chromosome, which maximizes the classification accuracy rate and minimizes the number of fuzzy rules. When reaching the termination condition, a chromosome with the maximum fitness value is used to test the performance of this method.

The remaining parts of this paper are organized as follows. The customer relationship management and grid partition method are briefly introduced in Section II. The research method is introduced in Section III, which includes determining frequent fuzzy grids and effective fuzzy rules, and incorporating with the GA. In Section IV, the results are examined by performing on 3C hypermarket. he conclusion is given in Section V.

II. LITERATURE REVIEW

A. Customer relationship management (CRM)

CRM aims for enterprises to focus on customers, the most essential core of business operations, and tries to establish

a "Learning Relation" with customers. Through customers' responses to certain products and services, enterprises make customers their center to adjust management and operations effectively. By so doing, enterprises obtain the know-how to improve the quality of products and service [17]. In other words, via constant communication and thorough understanding, enterprises acquire new ways to secure existing customers, gain new customers to motivate customers' contribution and loyalty to enterprises [1, 15, 18].

1) Customer Satisfaction: Customer satisfaction is the satisfaction degree when products purchasing and service offering are taking place. Cardozo [2] painted out customer satisfaction increases customers' second purchasing behavior as well as the desire to purchase more products. Spreng [22] further thought customers perceive a state of sense and sensibility after the process of evaluating items and purchasing them. Customer satisfaction is to strengthen the relation with existing customers, which is a cost-saving approach to encourage "repurchasing tendency". Through their word-of-mouth, enterprises are likely to win new customers and pronounce prominent influence for profitability [26]. Some known factors to affect customer satisfaction are products, service, and corporate images. Products are divided into software and hardware, including quality, function, performance, efficiency, and price. Service is the serving attitude of servers, such as clothes, word usages, greetings, smiles and product-related knowledge. Corporate images include environmental awareness and the contribution to the society. The measuring method uses scales of service quality to proceed with customers' perception and expectancy of certain service and the degree of satisfaction is thus gained.

2) Customer Loyalty: Customer loyalty means the repeated purchasing behavior a customer demonstrates to particular companies, products or services. Jones and Sasser [13] assumed customer loyalty is customers' willingness to purchase the same products or services in the future. Customer loyalty can also be shown short-term and long-term. Short-termers represent customers waver when there are better alternatives whereas long-termers stay and seldom change. Because of this, enterprises learn only with good service and novelty can they secure customer loyalty.

B. Grid partition method

The concepts of linguistic variables were proposed by Zadeh [29-31]. Formally, a linguistic variable is characterized by a quintuple [19, 30] denoted by $(x, T(x), U, G, M)$. Here, x is the name of the variable; $T(x)$ denotes the set of names of linguistic values or terms, which are linguistic words or sentences in a natural language [3] of x; U denotes a universe of discourse; G is a syntactic rule for generating values of x, and M is a semantic rule for associating a linguistic value with a meaning. Using the grid partition methods, each attribute can be partitioned by various linguistic values. This grid partition method has been widely used in pattern recognition and fuzzy reasoning. For example, there are the applications to pattern classification by [9-10], to fuzzy neural networks by [14], and to the fuzzy rule generation by [25].

In the grid partition method, K-various linguistic values are defined in each quantitative attribute. K is also pre-specified before performing the proposed method. For example, $K = 3$ and $K = 4$ for x_1 that ranges from 0 to 60 are depicted in Fig. 1 and 2, respectively. If x_1 is "age", then linguistic values (i.e., A_{11}, A_{12}, A_{13}) with triangle-shaped membership functions depicted in Fig. 1 can be linguistically interpreted as *young*, *medium*, and *old*, and those (i.e., A_{11}, A_{12}, A_{13}) depicted in Fig. 2 can be interpreted as *young*, *mediumyoung*, *medium*, and *old*.
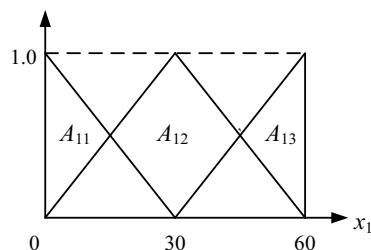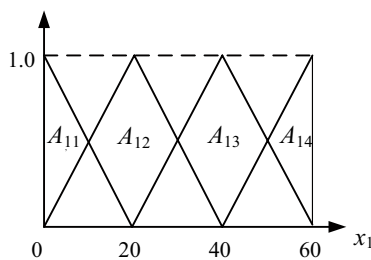
Fig. 1: $K = 3$ for "x_1"

Fig. 2: $K = 4$ for "x_1"

In this method, symmetric triangle-shaped linguistic values are used for simplicity. When a linguistic value is not yet determined if it is frequent, it is called a candidate 1-dim fuzzy grid. For a quantitative variable, say x_k, μ_{ki_k} is represented as follows [6-10]:

$$\mu_{ki_k}(x) = \max\{1 - \frac{|x - a_{i_k}|}{b^{K_i}}, 0\} \qquad (1)$$

where

$$a_{i_k} = \frac{m_i + (m_a - m_i)(i_k - 1)}{K_k - 1}, \qquad (2)$$

$$b^{K_i} = \frac{(m_a - m_i)}{K_k - 1} \qquad (3)$$

where m_a and m_i are the maximum and the minimum values of the domain interval of x_k, respectively. (A_{11}, A_{23}) is called a candidate 2-dim fuzzy grid that can be generated by using A_{11} and A_{23}. In other words, candidate 1-dim fuzzy grids can be further employed to generate the other candidate or frequent fuzzy grids with higher dimensions [7-8].

In regard to categorical attributes, each has a finite number of possible values, with no ordering among values (such as color or profession). If the distinct attribute values are n' (n'

is finite), this attribute can only be partitioned by n' linguistic values. For example, since the attribute "Class" is categorical, and the linguistic sentence of each linguistic value may be stated as follows:

$A_{2,1}^{Class\ label}$: class 1
$A_{2,2}^{Class\ label}$: class 2

where, it should be noted that the maximum number of dimensions for a single fuzzy grid d [7-8].

III. RESEARCH METHOD

In this section, we describe the individual phase of the method in section A. and B.. The GA part is described in section C.

A. Determining frequent fuzzy grids

Without loss of generality, given a candidate k-dim fuzzy grid $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k}$, where $1 \leq i_1, i_2, \ldots, i_k \leq K$, the degree which t_p belongs to this fuzzy grid can be computed as $\sum_{p=1}^{n} \mu_{A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k}}(t_p)$. The fuzzy support [5-7, 10-11] of $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k}$ is defined as follows:

$$FS(A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k})$$
$$= \sum_{p=1}^{n} \mu_{A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k}}(t_p)/n \quad (4)$$
$$= [\sum_{p=1}^{n} \mu_{K,i_1}^{x_1}(t_{p_1}) \cdot \mu_{K,i_2}^{x_2}(t_{p_2}) \cdot \ldots \cdot \mu_{K,i_k}^{x_k}(t_{p_k})]/n$$

The algebraic product uses a t-norm operator in the fuzzy intersection. When $FS(A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k})$ is larger than or equal to the user-specified min FS, that $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k}$ is a frequent k-dim fuzzy grid. for any two frequent grids, such as $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k}$ and $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k} \times A_{k,i_{k+1}}^{x_{k+1}}$, since $\mu_{A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k} \times A_{k,i_{k+1}}^{x_{k+1}}}(t_p) \leq \mu_{A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k}}(t_p)(1 \leq p \leq n)$, thus $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k} \times A_{k,i_{k+1}}^{x_{k+1}} \subseteq A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k}$ holds. But it should note that any subset of a frequent fuzzy grid must also be frequent. This is different from the Apriori property, but we may view it as a special property for mining frequent fuzzy grids. Next, utilizing the Table FGTTFS generates frequent fuzzy grids. FGTTFS consists of the following substructures [7-8].

1) Fuzzy grid table (FG): each row represents a fuzzy grid, and each column represents a linguistic value.
2) Transaction table (TT): each column represents (t_p), and each element records the membership function degree of the corresponding fuzzy grid.
3) Column FS: stores the fuzzy support corresponding to the fuzzy grid in FG.

An initial tabular FGTTFS is shown in TABLE I as an example, from which we can see that three are two sample

t_1 and t_2, with two attributes x_1 and x_2. Assume both x_1 and x_2 are divided into three linguistic values (i.e., $K = 3$), and x_2 is the attribute of class labels. Because each row of FG is a bit string consisting of 0 and 1, FG[u] and FG[v] (i.e., uth row and vth row of FG) can be paired to generate certain desired results by applying the Boolean operations. But any two linguistic values defined in the same attribute cannot be contained in the same candidate k-dim fuzzy grid ($k \geq 2$).

Further, a candidate k-itemset can be derived by joining two frequent (k-1) itemsets, and they share (k-2) itemsets in the Apriori algorithm. In this method, a candidate k-dim ($2 \leq k \leq d$) fuzzy grid can be viewed as to be derived by joining two frequent (k-1)-dim fuzzy grids, and share (k-2) linguistic values. However, this method will encounter one selection of many possible combinations to avoid redundant computations. To solve this problem, we apply exist integers $1 \leq e_1 \leq e_2 \leq \cdots e_k$, such that FG[$u, e_1$] = FG[$u, e_2$] = $\cdots$ = FG[u, e_{k-2}] = FG[u, e_{k-1}] = 1 and FG[v, e_1] = FG[v, e_2] = $\cdots$ FG[v, e_{k-2}] = FG[v, e_{k-1}] = 1, where FG[u] and FG[v] correspond to frequent (k-1)-dim fuzzy grids and FG[v, e_k] stands for the e_kth element of the vth rows of FG, thus FG[u] and FG[v] can be paired to generate a candidate k-dim fuzzy grid.

B. Determining effective fuzzy rules

The general type of one fuzzy classification rule denoted by R is stated as Eq.(5).

$$Rule\ R: A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k} \Rightarrow A_{C,i_\alpha}^{x_\alpha}$$
$$with\ CF(R) \quad (5)$$

where $x_\alpha(1 \leq \alpha \leq d)$ is the class label and CF(R) is the certainty grade of R. The above rule can be interpreted as: if x_1 is $A_{K,i_1}^{x_1}$ and x_2 is $A_{K,i_2}^{x_2}$ and $\cdots$ and x_k is $A_{K,i_k}^{x_k}$, then x_α is $A_{K,i_\alpha}^{x_\alpha}$ with certainty grade CF(R). Where the left-handed-side of "$\Rightarrow$" is the antecedence of R, and the right-handed-side is the consequence. Since $(A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k} \times A_{C,i_\alpha}^{x_\alpha}) \subseteq (A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k})$ holds, R can be generated by $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k} \times A_{C,i_\alpha}^{x_\alpha}$ and $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k}$. Moreover, the fuzzy confidence of R is defined as follows [6-8, 11-12]:

$$FC(R) = \frac{FS(A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k} \times A_{C,i_\alpha}^{x_\alpha})}{FS(A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k})} \quad (6)$$

When FC(R) is larger than or equal to the user-specified min FC, we can say that R is effective. FC(R) can further be used as the grade of certainty of R (i.e., CF(R) = FC(R)). Similarly, we also can use Boolean operations to obtain the antecedence and consequence of each rule.

However, some redundant rules must be eliminated in order to achieve compactness. Assume there exist two rules, such as R and S, having the same consequence, and the antecedence of R is contained in that of S, then R is redundant and can be discarded, whereas S is temporarily reserved. This is because the minimization of the number of antecedent conditions should be considered.

At the same time, Ishibuchi et al. [10] and Nozaki et al. [19] further demonstrated that the performance of fuzzy rule-based

TABLE I: Initial table FGTTFS for an example

Fuzzy grid	FG						TT		FS
	$A_{3,1}^{x_1}$	$A_{3,2}^{x_1}$	$A_{3,3}^{x_1}$	$A_{3,1}^{x_2}$	$A_{3,2}^{x_2}$	$A_{3,3}^{x_2}$	t_1	t_2	
$A_{3,1}^{x_1}$	1	0	0	0	0	0	$\mu_{3,1}^{x_1}(t_{1_1})$	$\mu_{3,1}^{x_1}(t_{2_1})$	$FS(A_{3,1}^{x_1})$
$A_{3,2}^{x_1}$	0	1	0	0	0	0	$\mu_{3,2}^{x_1}(t_{1_1})$	$\mu_{3,2}^{x_1}(t_{2_1})$	$FS(A_{3,2}^{x_1})$
$A_{3,3}^{x_1}$	0	0	1	0	0	0	$\mu_{3,3}^{x_1}(t_{1_1})$	$\mu_{3,3}^{x_1}(t_{2_1})$	$FS(A_{3,3}^{x_1})$
$A_{3,1}^{x_2}$	0	0	0	1	0	0	$\mu_{3,1}^{x_2}(t_{1_2})$	$\mu_{3,1}^{x_2}(t_{2_2})$	$FS(A_{3,1}^{x_2})$
$A_{3,2}^{x_2}$	0	0	0	0	1	0	$\mu_{3,2}^{x_2}(t_{1_2})$	$\mu_{3,2}^{x_2}(t_{2_2})$	$FS(A_{3,2}^{x_2})$
$A_{3,3}^{x_2}$	0	0	0	0	0	1	$\mu_{3,3}^{x_2}(t_{1_2})$	$\mu_{3,3}^{x_2}(t_{2_2})$	$FS(A_{3,3}^{x_2})$

systems can be improved by adjusting the grade of certainty of each rule. Therefore, it is possible to improve the classification ability of this method by incorporating the adaptive rules proposed by Nozzki et al [19]. into the proposed learning algorithm. Next, we determine the class label of t_p by applying fuzzy rules derived by the learning algorithm. Without loss of generality, if the antecedent part of a fuzzy associative classification rule R_τ is $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \cdots \times A_{K,i_k}^{x_k} \times A_{K_\tau,i_\tau}^{x_\tau}$, then we can calculate ω_τ of R_τ as Eq.(7).

$$\omega_\tau = \mu_{K_1,i_1}^{x_1}(t_{p1}) \cdot \mu_{K_2,i_2}^{x_2}(t_{p2}) \cdot \ldots \cdot \mu_{K_\tau,i_\tau}^{x_\tau}(t_{p\tau}) \cdot FC(R_\tau) \tag{7}$$

Then t_p can be determined to categorize the class label which is the consequent part of R_β, when

$$\omega_\beta = \max_j\{\omega_j | R_j \in TR\} \tag{8}$$

where TR is the set of fuzzy rules generated. The class label of t_p is determined and adaptive rules can be employed to adjust the fuzzy confidence of the "firing" rule R_τ. This is if t_p is correctly classified then R_τ is increased; otherwise, R_τ is decreased.

C. Genetic algorithm (GA)

Since min FS and min FC are difficult for users to appropriately give these two thresholds for each classification problem. Thus, GA is used to automatically generate the parameters (i.e., min FS and min FC) mentioned above.
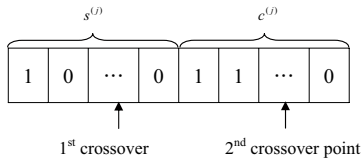
Fig. 3: The jth chromosomes.

In GA, the type of a binary chromosome is composed of min FS and min FC. It is as shown in Fig. 3, the jth chromosome is actually denoted by $s^{(j)}c^{(j)}$, with total length $(|s^{(j)}|+|c^{(j)}|)$, where $|s^{(j)}|$ and $|c^{(j)}|$ are lengths of $s^{(j)}$ and $c^{(j)}$, respectively. $s^{(j)}$ (or $c^{(j)}$) can be decoded by transforming the binary representation to an integer number, and this number is divided

by $2^{s^{(j)}}$ (or $2^{c^{(j)}}$). In addition, from Fig. 3, we can see that the two-point crossover operator [21] is used for exchanging partial information between two selected chromosomes, and two new chromosomes are generated at the same time to replace their parents. Two crossover points are randomly selected and lie in $s^{(j)}$ and $c^{(j)}$, respectively. It should be noted that both $|s^{(j)}|$ and $|c^{(j)}|$ should be sufficiently large; otherwise, it may be unnecessary to employ the GA to find min FS and min FC by the chromosomes [7].

In each generation of the GA, the fitness value of each chromosome can be obtained. Moreover, the fitness value $f(V^{(j)})$ of the jth chromosome is formulated as follows:

$$f(V^{(j)}) = W_{\text{CAR}} \cdot \text{CAR}(V^{(j)}) - W_V \cdot |V^{(j)}| \tag{9}$$

where $V^{(j)}$ denotes a set consisting of the effective fuzzy classification rules obtained by $s^{(j)}c^{(j)}$, and W_{CAR} and W_V are relative weights of the classification accuracy rate by $V^{(j)}$ and number of fuzzy rules in $V^{(j)}$, respectively. The chromosome that has the maximum fitness value in the final generation is further used to examine the classification performance of this method. That is, the acquisition of a compact fuzzy rule set with high classification accuracy rate is taken into account in the overall objective. In general, $0 < W_V \ll W_{\text{CAR}}$ holds the constraint since the classification power of a classification systems is more important than its compactness [9].

IV. RESEARCH RESULT

The transaction data set of this paper is obtained from a 3C hypermarket in Kinmen. The time is from April, 2011 to August, 2011. There are about 862 data. It is a transaction data set. The original data set contains as follows:

1) Name;
2) Sex;
3) ID number;
4) Birth data;
5) Telephone number;
6) Address;
7) Average year income;
8) Category.

Through data preprocessing, we clear data of noise and inconsistency and delete unnecessary items. Next, by data transformation, we discretize birth date and average year income. In addition, customer loyalty is obtained from times

TABLE II: Format of data transaction

TID	Age	Average year income	Customer satisfaction	Customer Loyalty	Category
1	$26 \sim 30$	$31K \sim 35K$	0.56	0.44	C
2	66 above	15K under	0.51	0.76	B
3	$26 \sim 30$	$31K \sim 35K$	0.89	0.83	A
4	$41 \sim 45$	$16K \sim 20K$	0.31	0.27	E
⋮	⋮	⋮	⋮	⋮	⋮

of purchase and customer satisfaction is obtained from the scale. And then we use the method of normalization. The data formats of generation are shown as TABLE II.

Customer satisfaction and customer loyalty are fuzzified, and the linguistic values are shown in Fig 3, and 4.
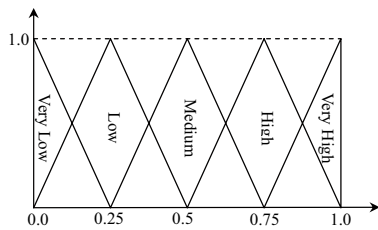

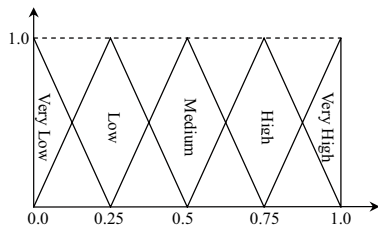Fig. 4: Linguistic values of customer satisfaction

Fig. 5: Linguistic values of customer loyalty

In addition, TABLE II consists of five classes (Class 1: A, Class 2: B, Class 3: C, Class 4: D and Class 5: E) and each class consists with four dimensions. Suppose that attribute x_1 is the age, attribute x_2 is the average year income, attribute x_3 is the customer satisfaction and attribute x_4 is the customer loyalty, and attribute x_5 is the class label (i.e., $d = 5$) to which $t_p = (t_{p1}, t_{p2}, \ldots, t_{p5})(1 \leq p \leq 862)$ belongs. The pairs (m_a, m_i) for x_1, x_2, x_3 and x_4 are (66, 15), (90, 15), (1, 0) and (1, 0), respectively. And x_5 should be noted that only five linguistic values; they are $A_{5,1}^{Classlabel}$: "Class 1", $A_{5,2}^{Classlabel}$: "Class 2", $A_{5,3}^{Classlabel}$: "Class 3", $A_{5,4}^{Classlabel}$: "Class 4" and $A_{5,5}^{Classlabel}$: "Class 5".

Parameter specifications used in this method are as follows:
1) Maximum iterations $J_{max} = 100$;
2) Population size, $N_{pop} = 30$;
3) $|s| = |c| = 10$;
4) $W_{CAR} = 10$;
5) $W_V = 1$;
6) Crossover probability, $\text{Prob}_C = 1.0$;

7) Mutation probability, $\text{Prob}_M = 0.05$;
8) Maximum number of generations, $t_{max} = 50$.

In the experiment, for the partition number k, we adopt Hu's suggestion which is known as set 5 [7]. The results are as follows:
1) IF Age = $41 \sim 45$ AND Average year income = 31K $\sim$ 35K AND Customer satisfaction = Very High AND Customer loyalty = High, THEN Class A with CF = 0.8122;
2) IF Age = $26 \sim 30$ AND Average year income = 31K $\sim$ 35K AND Customer satisfaction = High AND Customer loyalty = Medium, THEN Class B with CF = 0.6395;
3) IF Age = $21 \sim 25$ AND Average year income = 21K $\sim$ 25K AND Customer satisfaction = Medium AND Customer loyalty = Medium, THEN Class C with CF = 0.5297;
4) IF Age = $31 \sim 35$ AND Average year income = 31K $\sim$ 35K AND Customer satisfaction = Low AND Customer loyalty = Medium, THEN Class D with CF = 0.3583;
5) IF Age = $46 \sim 50$ AND Average year income = 26K $\sim$ 30K AND Customer satisfaction = Medium AND Customer loyalty = Low, THEN Class E with CF = 0.2974;

From the above five rules findings, the Customer satisfaction and Customer loyalty over "Medium", the CF is the least which is over 0.60. Next, the Customer satisfaction and Customer loyalty over "Low", the CF is the least which is over 0.29. Because CF increases the strength of the classified rule, when the value of CF is higher and higher, the category is classified with more certainty. Thus, CF is an index of threshold restriction. And, the five rules above are the output as meta- knowledge concerning the given transaction.

Also, the original transaction database is compared with that of the two-phase method, fuzzy c-mean and fuzzy k-nearest neighbor for accuracy, the results is shown as TABLE III. In TABLE III, the two-phase method is better than the other two fuzzy methods. The result corresponds with Hu's research [8].

TABLE III: The comparison of accuracy of three fuzzy methods

Methods	Two-phase method	Fuzzy c-mean	Fuzzy k-nearest neighbor
Accuracy rates	97.53%	91.51%	94.92%

V. CONCLUSIONS

In this competitive society, how to attract new customers and secure old customers play important roles in enterprises. The CRM is believed to utilize proper analytic tools with restricted resources to acquire the most valuable customers and increase their desires to purchase more.

In this paper, we use two-phased method to find fuzzy rules for classification problems which is based on the processing of the Apriori algorithm. First, the first-phase finds frequent fuzzy grids by dividing each quantitative attribute with a user-specified number of various linguistic values. And the second-phase generates effective fuzzy classification rules from those

frequent fuzzy grids. More, we use the GA to automatically find the appropriately min FS and min FC. The research results generate five fuzzy classification rules which are the output as meta-knowledge concerning the given transaction. In addition, this method compares to fuzzy classification methods which are fuzzy c-mean and fuzzy k-nearest neighbor. Finally, the comparison of results demonstrate that the two-phased method for fuzzy classification is better in accuracy rates.

Since fuzzy knowledge representation can facilitate interactions between the expert system and users, this method may be further viewed as a knowledge acquisition tool to discover fuzzy association rules to perform the Market Basket Analysis (MBA), which helps users make preferable decisions.

Acknowledgements

REFERENCES

[1] C.A., "Application of Data Warehouse for CRM," *Information and Computer*, Vol. 240, pp. 54-57, 2000.
[2] Cardozo, R. N., "An Experimental Study of Customer Effort, Expectation and Satisfaction," *Journal of Marketing Research*, Vol. 2, No. 3, pp. 244-249, 1965.
[3] S.M. Chen and W.T. Jong, "Fuzzy query translation for relational database system", *IEEE Transactions on System, Man, and Cybernetics*, Vol. 27, no. 4, pp. 714-721.
[4] D.E. Goldberg, *Genetic Algrithms in Search, Optimization, and Machine Learning*, Addison-Wesley, MA, 1989.
[5] J.W. Han, M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, San Francisco, 2001.
[6] Y.C. Hu, R.S. Chen and G.H. Tzeng, "Mining fuzzy association rules for classification problems," *Computers and Industrial Engineering*, Vol. 43, No. 4, pp.735-750, 2002.
[7] Y.C. Hu, R.S. Chen and G.H. Tzeng, "Finding fuzzy classification rules using data mining techniques," *Pattern Recognition Letters*, Vol. 24, pp.509-519, 2003.
[8] Y.C. Hu, "Finding useful fuzzy concepts for pattern classification using genetic algorithm", *Information Sciences*, Vol.175, pp.1-19, 2005.
[9] H. Ishibuchi, K. Nozaki, N. Yamamoto, H. Tanaka, "Selecting fuzzy if-then rules for classification problems using genetic algorithm", *IEEE Transactions on Fuzzy Systems*, Vol. 3, no. 3, pp.260-270, 1995.
[10] H. Ishibuchi, T. Nakashima, T. Murata, "Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 29, no. 5, pp.601-618, 1999.
[11] Ishibuchi, H., Yamamoto, T., Nakashima, T., "Fuzzy data mining: effect of fuzzy discretization," *In: Proceedings of the 1st IEEE International Conference on Data Mining*, San Jose, USA, pp.241-248, 2001[a].
[12] Ishibuchi, H., Nakashima, T., Yamamoto, T., "Fuzzy association rules for handling continuous attributes," *In: Proceedings of IEEE International Symposium on Industrial Electronics*, Pusan, Korea, pp.118-121, 2001[b].
[13] Jones, T. O. and W. E. Jr. Sasser, "Why Satisfied Customer Defect," *Harvard Business Review*, Vol. 73, No. 6, pp. 88-99, 1995.
[14] J.S.R. Jang, "ANFIS: adaptive-network-based fuzzy inference systems", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23, No. 3, pp.665-685, 1993.
[15] J.J. Lyu and Y.C. Lin, "Strategy application of CRM," *Quality Control Journal*, Vol. 37, No. 3, pp.20-23, 2001.
[16] T.C. Lin, "Customer Relationship Management (CRM) Research Frameworks and Some Essential Research Issues," *Journal of Information Management*, Vol 9, pp.31-56, 2002.
[17] Y.C. Liu, *From ERP, SCM, CRM to EC*, DrMarketing Co., Ltd, 2002.
[18] NCR, "Integrating Business Strategy and CRM," *e-Business Executive Report*, Nov. pp.20-25, 1999.
[19] Nozaki, K., Ishibuchi, H., Tanaka, H., "Adaptive fuzzy rule-based classification system," *IEEE Transaction on Fuzzy Systems*, Vol. 4, NO. 3, pp.238-250, 1996.
[20] W. Pedrycz, F. Gomide, *An Introduction of Fuzzy Sets: Analysis and Design*, MIT Press, Cambridge, 1998.

[21] Rooij, A.J.F., Jain, L., Johnson, R.P., *Neural Network Training Using Genetic Algirhtms*, World Scientific, Singapore, 1996.
[22] Spreng, R. A., "A Comprehensive Model of the Consumer Satisfaction Formation Process," *Dissertation Abstracts International*, Vol. 53, No.7, pp.2461-2462, 1993.
[23] Swift, R., *Accelerating Customer Relationships*, Prentice Hall, 2001.
[24] Tiwana, A., *The Essential Guide to Knowledge Management*, Prentice Hall PTR, Upper Saddle River, NJ, 2000.
[25] Wang, L.X., Mendel, J.M., "Generating fuzzy rules by learning from examples", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22, No.3, pp. 1414-1427, 1992.
[26] Wirtz, Jochen and John E. G. Bateson, "An Experimental Investigation of Halo Effects in Satisfaction Measures of Service Attributes," *International Journal of Service Industry Management*, No.6, pp.84-102, 1995.
[27] Yuan Y. and Shaw, M.J., "Induction of Fuzzy Decision Trees," *Fuzzy Sets and Systems*, Vol. 69, pp.125-139, 1995.
[28] L.A. Zadeh, "Fuzzy sets", *Information Control*, Vol.8 No.3, pp.338-353, 1965.
[29] L.A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning", *Information Science(part 1)*, Vol.8, No.3, pp.199-249, 1975[a].
[30] L.A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning", *Information Science(part 2)*, Vol.8, No.4, pp.301-357, 1975[b].
[31] L.A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning", *Information Science(part 3)*, Vol.9, No.1, pp.43-80, 1976.
[32] H. -J. Zimmermann, *Fuzzy sets, Decision Making, and Expert Systems*, Kluwer, Boston, 1991.