International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:3, No:5, 2009

Finding Fuzzy Association Rules Using FWFP-Growth with Linguistic Supports and Confidences

Chien-Hua Wang and Chin-Tzong Pang

***Abstract*—**In data mining, the association rules are used to search for the relations of items of the transactions database. Following the data is collected and stored, it can find rules of value through association rules, and assist manager to proceed marketing strategy and plan market framework. In this paper, we attempt fuzzy partition methods and decide membership function of quantitative values of each transaction item. Also, by managers we can reflect the importance of items as linguistic terms, which are transformed as fuzzy sets of weights. Next, fuzzy weighted frequent pattern growth (FWFP-Growth) is used to complete the process of data mining. The method above is expected to improve Apriori algorithm for its better efficiency of the whole association rules. An example is given to clearly illustrate the proposed approach.

***Keywords*—**Association Rule, Fuzzy Partition Methods, FWFP-Growth, Apiroir algorithm

I. Introduction

DATA mining is a methodology for the extraction of new knowledge from data. This knowledge may relate to problem that we want to solve [14]. Thus, data mining can ease the knowledge acquisition bottleneck in building prototype systems [6, 7]. If data mining extraction can effectively be applied on all varieties of analysis, is will assist the process of decision-making in business.

In common transactions, association rules (X→Y) is the most popular mean to be applied. The purpose is to search for the relation exists among items of database. The relation reflects that when items (X) appear, other items (Y) are likely to appear as well [4]. For instance, when a customer purchases bread, one might also get milk along with. Thus, association rules can assist decision makers to scope out the possible items that are likely to be purchased by consumers. Meanwhile, it facilitates planning marketing strategies [2].

The well-known Apriori algorithm [1], applies two-phases approach to look for the association rules from database. It collects all the frequent itemsets during phase one. During the second phase, it applies frequent itemsets to generate effective association rules. In the approach of Apriori algorithm, each item is treated as Boolean variable, and support and confidence [4] are to respectively evaluate the accessibility and reliability of association rules. By these, we can determine the quantity of association rules. Nevertheless, when processing huge database, Apriori algorithm cannot effectively handle the matter. Thus, there are related techniques were proposed, such as Frequent-Pattern Growth algorithm, Dynamic Itemset Counting algorithm, Direct Hashing pruning algorithm, and etc.

Next, regarding to the matter of decision making, it takes the consideration of user's perception and cognitive uncertainty of subjective decisions. Zadeh proposed fuzzy theory in 1965 that deals with cognitive uncertainty of vagueness and ambiguity [18]. Due to linguistic variables and linguistic value [19, 20, 21] were described to fuzzy concepts to correspond with the possible cognition of a decision maker subjectively. It helps proceed analysis of decisions. Thus, fuzzy data mining has recently become an important topic to research.

Besides, most conventional data-mining algorithms set the minimum support and minimum confidence at numerical values. Linguistic minimum support and minimum confidence value are, more natural and understandable. In this paper, we used Fuzzy Weighted FP-Growth to deal with problems of quantitative transactions such as the ones with linguistic minimum support and minimum confidence value. Also, items may have different importance, which is evaluated by managers or experts, e.g. linguistic terms.

And we transform importance of items, quantitative value of transactions, minimum supports and minimum confidences into fuzzy set, and filters weighted large items out by fuzzy operations. Finally, we used FP-Tree algorithm to mine fuzzy weighted association rules, which with linguistic supports and confidences.

This paper is organized as follows. Literature review in Section 2 discusses FP-Growth and Fuzzy Partition Methods. The notation used in this paper is defined in Section 3. FWFP-tree algorithm for managing quantitative transactions, linguistic minimum supports and linguistic minimum confidences is proposed in Section 4. An example to illustrate the proposed mining algorithm is given in Section 5. Conclusion is presented in Sections 6.

C.-H. Wang is with the Department of Information Management, Yuan Ze University, Taoyuan, Taiwan, 32003 (e-mail:thuck@saturn.yzu.edu.tw).
C.-T. Pang is with the Department of Information Management, Yuan Ze University, Taoyuan, Taiwan, 32003 (e-mail:imctpang@saturn.yzu.edu.tw).

II. Literature Review

A. FP-Growth (Frequent-Pattern Growth)

FP-Growth was proposed by Han, Pei and Yin [5], the method is mainly performed by data structure of FP-Tree.

FP-Growth focuses on the disadvantages of Apriori algorithm that doesn't deal with huge data efficiently, and to improve algorithm. For instance, a mining work of association rule needs to find 100 patterns of frequency, $\{a_1, a_2, \ldots, a_{100}\}$, among the range, when applying Apriori algorithm, 10^{30} candidate items shall at least be generated. In addition, hundreds of scan in database is needed to complete the process. However, the time complexity along the whole process increases when a larger amount of data is encountered.

In comparison, FP-Growth just solved this problem. It mines frequent itemsets but doesn't generate the method of candidate itemsets, it adopts the concept of divide-and-conquer which compresses frequent itemset of length 1 L_1 to in a FT-Tree, but still retains association information of itemsets. Thus, after compressing the data, the conditional FP-Tree is formed. Also, keep mining and recursion backward to the FP-Tree. Frequent patterns are generated as frequent itemsets.

In FP-Tree construct process, it needs to scan database twice. The first scan is to figure out the supports of each item. The second scan is to construct FP-tree according to the support items of scan in descending order. In addition, FP-Tree structure includes a frequent-item-header table. Here, it records all frequent patterns and pointer points out item nodes of first occurrence.

An advantage of FP-Growth is that the constructing FP-Tree has great performance of compression, and its process of mining can reduce the cost of rescanning data. In addition, it applies conditional FP-Tree on avoiding generating candidate item and testing examining process. About its disadvantage, mining process needs extra processing time and space to store that it continuously generates large conditional bases and conditional FP-Trees.

B. Fuzzy Partition Methods

The concepts of linguistic variables were proposed by Zadeh [19, 20, 21] and it is reasonable that we view each attribute as a linguistic variable. Formally, a linguistic variable is characterized by a quintuple [16, 22] denoted by (x, $T(x)$, U, G, M), in which 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 simple fuzzy partition methods, each attribute can be partitioned by various linguistic values. The simple fuzzy partition methods have been widely used in pattern recognition and fuzzy reasoning. For example there are the applications to pattern classification by Ishibuchi et al. [11, 12], to fuzzy neural networks by [13], and to the fuzzy rule generation by [17].

In the simple fuzzy partition methods, K various linguistic values are defined in each quantitative attribute. K is also pre-specified before executing the proposed methods. Triangular membership functions are usually used for the linguistic values. For example, $K = 3$ and $K = 4$ for the attribute "Width" (denoted by x_1) that ranges from 0 to 60 are shown as Figs.1 and 2, respectively. That is, three (i.e., small, middle and large) and four (i.e., small, middle small, medium large and large) various linguistic values are defined in Figs.1 and 2, respectively [8, 9].

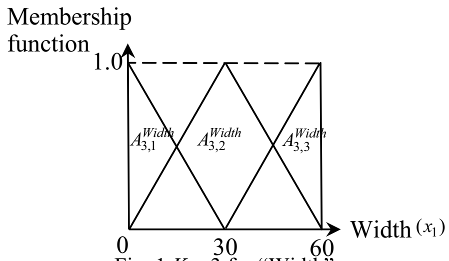

Fig. 1 $K = 3$ for "Width"

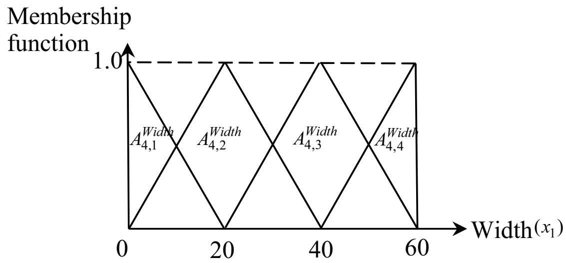

Fig. 2 $K = 4$ for "Width"

In the method, each linguistic value is actually viewed as a candidate 1-dim fuzzy grid. Then A_{K,h_1}^{Width} can be represented as follows [8, 9]:

$$\mu_{K,h_1}^{Width}(x) = \max\{1 - |x - a_{h_1}^K| / b^K, 0\} \tag{1}$$

where

$$a_{h_1}^K = mi + (ma - mi)(z_1 - 1)/(K - 1) \tag{2}$$

$$b^K = (ma - mi)/(K - 1) \tag{3}$$

III. Notation

The notation used in this paper is defined as follows.

n	the total number of transaction data;
m	the total number of items;
d	the total number of managers, $1 \le d$;
u	the total number of membership functions for item importance;
z	number of grids used to describe each sample item, where $1 \le z$;
D_i	the i-th transaction datum, $1 \le i \le n$;
k	dimension of one fuzzy grid, where $1 \le k \le z$;
K	maximal number of fuzzy partitions in each

	item;
A_j	the j-th item, $1 \le j \le m$;
A^j_{K,h_k}	h_k-th linguistic value of K linguistic value in the item, where $1 \le h_k \le K$ and $1 \le j \le m$;
μ^j_{K,h_k}	membership function of A^j_{K,h_k}
ma	the maximum value of membership function of item quantities
mi	the minimum value of membership function of item quantities
$count_{jh}$	the summation μ^j_{K,h_k} of the same grid;
$max\text{-}count_j$	the maximum count value among $count_{jh}$ values;
$max\text{-}R_j$	the fuzz grid of A^j_{K,h_k} with $max\text{-}count_j$;
W_{jg}	the transformed fuzzy weight for importance of item A_j, evaluated by the g-th manager, $1 \le g \le d$;
W^{ave}_j	the fuzzy average weight for importance of item A_j;
α	the predefined linguistic minimum support value;
β	the predefined linguistic minimum confidence value;
I_t	the t-th membership function of item importance, $1 \le t \le u$;
I^{ave}	the average weight of all possible linguistic terms of item importance;
$wsup_j$	the fuzzy weighted support of item A_j;
$wconf_R$	the fuzzy weighted confidence of rule R;
$minsup$	transforming fuzzy set from the linguistic minimum support valueα;
$wminsup$	the fuzzy weighted set of minimum supports;
$minconf$	transforming fuzzy set from the linguistic minimum confidence valueβ;
$wminconf$	the fuzzy weighted set of minimum confidences;
C_r	the set of candidate weighted itemsets with r items;
L_r	the set of large weighted itemsets with r items.

IV. The Proposed Algorithm

The description of the algorithm and structure is as follows:

The algorithm:

INPUT: a. Including set of n quantitative transaction data;

b. A set of m items with their importance evaluated by d managers;

c. Four sets of membership functions respectively for item quantities, item importance, minimum support and minimum confidence;

d. A pre-defined linguistic minimum support value α;

e. A pre-defined linguistic minimum confidence value β.

OUTPUT: A set of weighted fuzzy association rules.

Step 1: Transform each linguistic term of importance for item A_j, $1 \le j \le m$, which is evaluated by the g-th manager into a fuzzy set W_{jg} of weights, $1 \le g \le d$, using the given membership functions of item importance.

Step 2: Calculate the fuzzy average weight W^{ave}_j of each item A_j by fuzzy addition as:

$$W^{ave}_j = \frac{1}{d} \times \sum_{g=1}^{d} W_{jg} \tag{4}$$

Step 3: Transform the quantitative value μ^j_{K,h_k} of each item A_j in each transaction datum D_i (i=1 to n, j=1 to m), into a fuzzy set, using (1) (2) (3) to show as

$$\left(\frac{\mu^j_{K,h_1}}{A^j_{K,h_1}} + \frac{\mu^j_{K,h_2}}{A^j_{K,h_2}} + \cdots + \frac{\mu^j_{K,h_k}}{A^j_{K,h_k}}\right)$$

where K is maximal number of fuzzy partitions for A_j, h_k-th linguistic value of K linguistic value in the item, where $1 \le h_k \le K$ and $1 \le j \le m$.

Step 4: Calculate the count of each fuzzy grid A^j_{K,h_k} in the transaction data as:

$$count_{jh} \quad \sum_{i=1}^{n} \mu^j_{K,h_k} \tag{5}$$

Step 5: Find $max\text{-}count_j$. Let $max\text{-}R_j$ be the grid with $max\text{-}count_j$ for item A_j.

$$max\text{-}count_j \quad \max_{j=1}^{K} (\, count_{jh}), \text{ for } j = 1 \text{ to } m \tag{6}$$

where m is the number of items.

Step 6: Calculate the fuzzy weighted support $wsup_j$ of each item A_j as:

$$wsup_j \quad \frac{1}{n} \times (\max - R_j \times W^{ave}_j) \tag{7}$$

where n is the numbers of transactions.

Step 7: Transform the given linguistic minimum support value α into a fuzzy set (denoted $minsup$) of minimum supports, using the given membership functions for minimum supports.

Step 8 : Calculate the fuzzy weighted set ($wminsup$) of the given minimum support value as:

$wminsup = minsup$ × (the gravity of I^{ave})

where

$$I^{ave} \quad \frac{1}{u} \times \sum_{t=1}^{u} I_t \tag{8}$$

with u being the total number of membership functions for item importance and I_t being the t-th membership. I^{ave} re- presents the fuzzy average weight of all possible linguistic terms of importance.

Step 9: Set $r = 1$, where r is used to store the number of items kept in the current itemset.

Step 10: During scan one, find out the correspondences of minimum support and the length of one itemsets. Next, establish a descending data table by the length of each transaction.

Step 11: According to the header table, rebuild another fuzzy set table. Next establish a MFFP-Tree (Membership function FP-tree) during the second scan.

Step 12: Mine the itemsets of header table ascendingly. And set up the conditional pattern base of each node in a MFFP-Tree. Next, establish conditional MFFP-Tree.

Step 13: Repeatedly mine conditional MFFP-Tree, and gradually increase frequency pattern base. If one single path is included in conditional MFFP-Tree, all patterns can be listed.

Step 14: After mining, each pattern must be larger or equal to the fuzzy average weighted. Fuzzy association rules are formed.

Step 15: Transform the given linguistic minimum confidence value β into a fuzzy set (*minconf*) of minimum confidences, using the given membership function for minimum confidences.

Step 16: Calculate the fuzzy weighted set (*wminconf*) of the given minimum confidence value as:

$$wminsup = minsup \times (\text{the gravity of } I^{ave})$$

where I^{ave} is the same as that calculated in Step 8.

Step 17: Check the association rules from FP-tree each large weighted q-itemset s with items $(s_1, s_2, \ldots, s_q)$, $q \geq 2$, using the following substeps:

(a) Generate all possible fuzzy association rules as follows:

$$s_1 \wedge \ldots \wedge s_{j-1} \wedge s_{j+1} \wedge \ldots \wedge s_q \rightarrow s_j, j=1 \text{ to } q$$

(b) Calculate the fuzzy weighted confidence value $wconf_R$ of each possible fuzzy association rule R as:

$$wconf_R = \frac{count_s}{count_{s-s_j}} \times \mathrm{W_s} \tag{9}$$

where

$$\mathrm{count_s} = \sum_{i=1}^{n} (\mathop{Min}_{k=1}^{q} f_{is_k}) \quad \text{and } \mathrm{W_s} = \mathop{Min}_{i=1}^{q} W_{s_i}^{ave}$$

(c) Check whether the fuzzy weighted confidence $wonf_R$ of fuzzy association rule R is greater than or equal to the fuzzy weighted minimum confidence *wminconf* by fuzzy ranking. If $wconf_R$ is greater than or equal to *wminconf*, keep rule R in the interesting rule set.

Step 18: For each rule R with fuzzy weighted support $wsup_R$ and fuzzy weighted confidence $wconf_R$ in the interesting rules set find the linguistic minimum support grid S_i and the linguistic minimum confidence grid C_j with $wminsup_{i-1} \leq wsup_R < wminsup_i$ and $wminconf_{i-1} \leq wconf_R < wminconf_j$ by ranking, where:

$$wminsup_i = minsup_i \times (\text{the gravity of } I^{ave}),$$

$$wminconf_i = minconf_j \times (\text{the gravity of } I^{ave}),$$

$minsup_i$ is the given membership function for S_i and $minconf_j$ is the given membership function for C_j. Output Rule R with linguistic support value S_i and linguistic confidence value C_j.

V. An Example

In this section, an example is given to illustrate the proposed fuzzy weighted frequent pattern-tree algorithm. This is a simple example to show how the proposed algorithm can be used to generate weighted fuzzy association rules from a set of quantitative transactions. The data set includes six quantitative transactions, as show in Table 1.

TABLE 1
The data set used in this example

TID	ITEMS
1	(A, 4), (B, 4), (E, 9)
2	(B, 3), (C, 5), (F, 3)
3	(B, 2), (C, 3), (D, 2), (E, 8)
4	(A, 7), (C, 7), (E, 9)
5	(C, 2), (D, 2), (F, 1)
6	(A, 4), (B, 3), (C, 5), (F, 2)

Each transaction is composed of a transaction identifier and items purchased. There are six items, respectively being A, B, C, D, E and F, to be purchased. Each item is represented by a tuple (item name, item amount). For example, the first transaction consists of four units of A, four units of B and nine units of E.

Also assume that the membership functions for item quantities are the same for all the items and are shown in Fig 3. In this example, amounts are represented by three fuzzy grids (Low, Middle and High) are produced for each item amount according to the predefined membership functions. The importance of the items is evaluated by three managers as shown in Table 2.

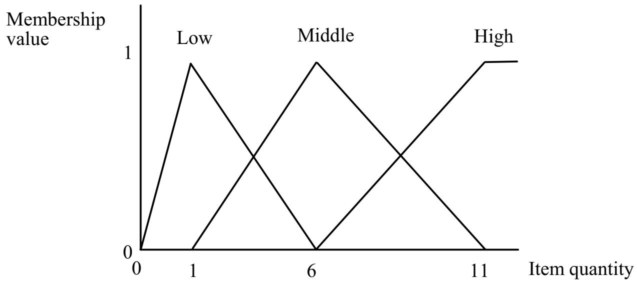

Fig. 3 The membership function for item quantities in this example

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:3, No:5, 2009

TABLE II
THE ITEM IMPORTANCE EVALUATED BY THREE MANAGERS

ITEM \ MANAGER	MANAGER 1	MANAGER 2	MANAGER 3
A	Important	Ordinary	Ordinary
B	Very Important	Important	Important
C	Ordinary	Important	Important
D	Unimportant	Unimportant	Very Uniportant
E	Important	Important	Important
F	Unimportant	Unimportant	Ordinary

Similar, assume the membership functions for item importance are given in Fig 4.

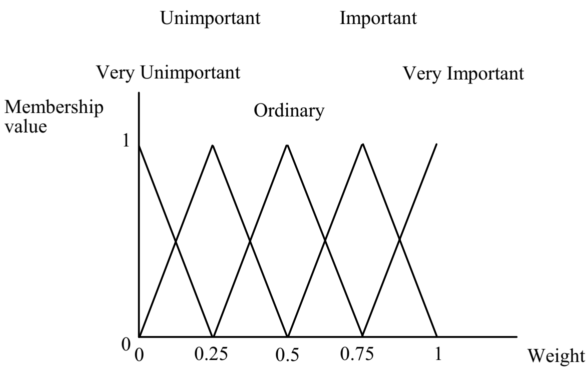

Fig. 4 The membership function for item importance used in this example

In Fig 4, item importance is partitioned into five fuzzy grids: Very Unimportant, Unimportant, Ordinary, Important and Very Important. Each fuzzy grid is represented by a membership function. The membership functions in Fig 4 can be represented as follows:

Very Unimportant (VU): (0, 0, 0.25),
Unimportant (U): (0, 0.25, 0.5)
Ordinary (O): (0.25, 0.5, 0.75)
Important (I): (0.5, 0.75, 1)
Very Important (VI): (0.75, 1, 1)

For the transaction data given in Table 1, the proposed fuzzy mining algorithm proceeds as follows.

Step 1: The linguistic terms for item importance given in Table 2 are transformed into fuzzy sets by the membership functions in Fig 4. For example, item A is evaluated to be important by Manager 1, and can then be transformed as a triangular fuzzy set (0.5, 0.75, 1) of weights. The transformed results for Table 2 are shown in Table 3.

TABLE III
THE FUZZY WEIGHTS TRANSFORMED FROM THE ITEM IMPORTANCE IN TABLE 2

ITEM \ MANAGER	MANAGER 1	MANAGER 2	MANAGER 3
A	(0.5, 0.75, 1)	(0.25, 0.5, 0.75)	(0.25, 0.5, 0.75)
B	(0.75, 1, 1)	(0.5, 0.75, 1)	(0.5, 0.75, 1)
C	(0.25, 0.5, 0.75)	(0.5, 0.75, 1)	(0.5, 0.75, 1)
D	(0, 0.25, 0.5)	(0, 0.25, 0.5)	(0, 0, 0.25)
E	(0.5, 0.75, 1)	(0.5, 0.75, 1)	(0.5, 0.75, 1)
F	(0, 0.25, 0.5)	(0, 0.25, 0.5)	(0.25, 0.5, 0.75)

Step 2: The average weighted of each item is calculated by fuzzy addition. Take Item A as an example. The three fuzzy weights for Item A are respectively (0.5, 0.75, 1), (0.25, 0.5, 0.75) and (0.25, 0.5, 0.75). The average weight is then ((0.5, 0.75, 1) + (0.25, 0.5, 0.75) + (0.25, 0.5, 0.75))/3, which is derived as (0.33, 0.58, 0.83). The average fuzzy weights of all items are calculated, with results shown in Table 4.

TABLE IV
THE AVERAGE FUZZY WEIGHTS OF ALL THE ITEMS

Item	Average Fuzzy Weight
A	(0.333, 0.583, 0.833)
B	(0.583, 0.833, 1)
C	(0.417, 0.667, 0.917)
D	(0, 0.167, 0.417)
E	(0.5, 0.75, 1)
F	(0.083, 0.333, 0.583)

Step 3: The quantitative values of the items in each transaction are represented by fuzzy sets. Take the first item in Transaction 1 as an example. The amount '4' of A is applied formula (1), (2), (3) to form the fuzzy set (0.4/A.Low + 0.6/A.Middle) using the given membership functions (Fig 3). The step is repeated for the other items, and the results are shown in Table 5, where the notation *item.term* is called a fuzzy grid.

TABLE V
THE FUZZY SETS TRANSFORMED FROM THE DATA IN TABLE 1

TID	Fuzzy Sets
1	$\left(\frac{0.4}{A.Low}+\frac{0.6}{A.Middle}\right),\left(\frac{0.4}{B.Low}+\frac{0.6}{B.Middle}\right),\left(\frac{0.4}{E.Middle}+\frac{0.6}{E.High}\right)$
2	$\left(\frac{0.6}{B.Low}+\frac{0.4}{B.Middle}\right),\left(\frac{0.2}{C.Low}+\frac{0.8}{C.Middle}\right),\left(\frac{0.6}{F.Low}+\frac{0.4}{F.Middle}\right)$
3	$\left(\frac{0.8}{B.Low}+\frac{0.2}{B.Middle}\right),\left(\frac{0.6}{C.Low}+\frac{0.4}{C.Middle}\right),\left(\frac{0.8}{D.Low}+\frac{0.2}{D.Middle}\right),\left(\frac{0.6}{E.Middle}+\frac{0.4}{E.High}\right)$
4	$\left(\frac{0.8}{A.Middle}+\frac{0.2}{A.High}\right),\left(\frac{0.8}{C.Middle}+\frac{0.2}{C.High}\right),\left(\frac{0.4}{E.Middle}+\frac{0.6}{E.High}\right)$
5	$\left(\frac{0.8}{C.Low}+\frac{0.2}{C.Middle}\right),\left(\frac{0.8}{D.Low}+\frac{0.2}{D.Middle}\right),\left(\frac{1.0}{F.Low}\right)$
6	$\left(\frac{0.4}{A.Low}+\frac{0.6}{A.Middle}\right),\left(\frac{0.6}{B.Low}+\frac{0.4}{B.Middle}\right),\left(\frac{0.2}{C.Low}+\frac{0.8}{C.Middle}\right),\left(\frac{0.8}{F.Low}+\frac{0.2}{F.Middle}\right)$

Step 4: The scalar cardinality of each fuzzy grid in the transactions is calculated as the count value. Take the fuzzy grid A.Low as an example. Its scalar cardinality = (0.4 + 0 + 0 + 0 + 0 +0.4) = 0.8. The step is repeated for the other grids, and the results are shown in Table 6.

TABLE VI
THE COUNTS OF THE FUZZY GRIDS

ITEM	COUNT	ITEM	COUNT	ITEM	COUNT
A.Low	0.8	C.Low	1.8	E.Low	0
A.Middle	2.0	C.Middle	3.0	E.Middle	1.4

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:3, No:5, 2009

A.High	0.2	C.High	0.2	E.High	1.6
B.Low	2.4	D.Low	1.6	F.Low	2.4
B.Middle	1.6	D.Middle	0.4	F.Middle	0.6
B.High	0	D.High	0	F.High	0

Step 5: The fuzzy grid with the highest count among the three possible grids for each item is found. Take item A as example. Its count is 0.8 for Low, 2.0 for Middle, and 0.2 for High. Since the count for Middle is the highest among the three counts, the grid Middle is used to represent the item A in later building MFFP-Tree process. The number of item.grids is thus the same as that of the original items, making the processing time reduced. This step is repeated for the other items. Thus, "Low" is chosen for B, "Middle" is chosen for C, "Low" is chosen for D, "High" is chosen for E and "Low" is chosen for F.

Step 6: The fuzzy weighted support of each item is calculated. Take Item A as an example. The average fuzzy weight of A is (0.333, 0.583, 0.833) from step 2. Since the grid Middle is used to represent the item A and its count is 2.0, its weighted support is than (0.333, 0.583, 0.833) * 2.0 / 6, which is (0.111, 0.194, 0.278). Results for all the items are shown in Table 7.

TABLE VII
THE FUZZY WEIGHTED SUPPORTS OF ALL THE ITEMS

Item	Average Fuzzy Weight
A	(0.111, 0.194, 0.278)
B	(0.233, 0.333, 0.4)
C	(0.208, 0.333, 0.458)
D	(0, 0.044, 0.111)
E	(0.133, 0.2, 0.267)
F	(0.033, 0.133, 0.233)

Step 7: The given linguistic minimum support value is transformed into a fuzzy set of minimum supports assume the membership functions for minimum supports are given in Fig 5, which are the same as those in Fig 4.

Also assume the given linguistic minimum support value is "*Low*". It is then transformed into a fuzzy set of minimum supports, (0, 0.25, 0.5), according to the given membership function in Fig 5.

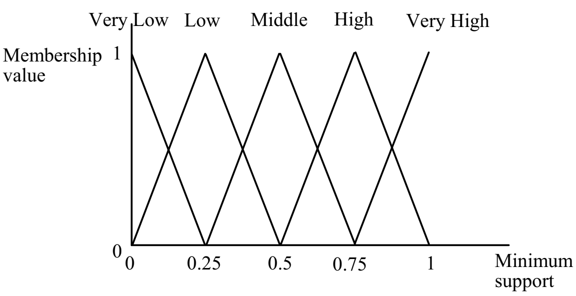

Fig. 5 The membership functions of minimum supports

Step 8: The fuzzy average weight of all possible linguistic terms of importance in Fig 5 is calculated as:

$$I^{ave} = [(0, 0, 0.25) + (0, 0.25, 0.5) + (0.25, 0.5, 0.75) + (0.5, 0.75, 1) + (0.75, 1, 1] / 5 = (0.3, 0.5, 0.7).$$

The gravity of I^{ave} is then (0.3 + 0.5 + 0.7) / 3, which is 0.5. The fuzzy weighted set of minimum supports for "Low" is then (0, 0.25, 0.5) × 0.5, which is (0, 0.125, 0.25).

Step 9: r is set at 1, where r is used to store the number of items kept in the current itemsets.

Step 10: During scan one, find out the correspondences of minimum support larger than zero and generate the length of one itemsets. Next, establish a descending data table by the length of each transaction (such as header table), shown as Table 8.

TABLE VIII
HEADER TABLE

Item	Support
C.Middle	5
B.Low	4
A.Middle	3
E.High	3
F.Low	3
D.Low	2

Step 11: According to header table, rebuild another fuzzy set table, shown as Table 9. And establish a MFFP-Tree during the second scan, shown as Fig 6.

TABLE IX
THE NEW FUZZY SETS FROM TABLE 5

TID	Fuzzy Sets
1	$\frac{0.4}{B.Low}, \frac{0.6}{A.Middle}, \frac{0.6}{E.High}$
2	$\frac{0.8}{C.Middle}, \frac{0.6}{B.Low}, \frac{0.6}{F.Low}$
3	$\frac{0.4}{C.Middle}, \frac{0.8}{B.Low}, \frac{0.4}{E.High}, \frac{0.8}{D.Low}$
4	$\frac{0.8}{C.Middle}, \frac{0.8}{A.Middle}, \frac{0.6}{E.High}$
5	$\frac{0.2}{C.Middle}, \frac{1.0}{F.Low}, \frac{0.8}{D.Low}$
6	$\frac{0.8}{C.Middle}, \frac{0.6}{B.Low}, \frac{0.6}{A.Middle}, \frac{0.8}{F.Low}$

Step 12: Mine the itemsets of header table ascendingly. And set up the conditional pattern base of each node in a MFFP-Tree. Next establish conditional MFFP-Tree, shown as Table 10.

Step 13: Repeatedly mine conditional MFFP-Tree, and gradually increase frequency pattern base. If one single path is included in conditional MFFP-Tree, all patterns can be listed.

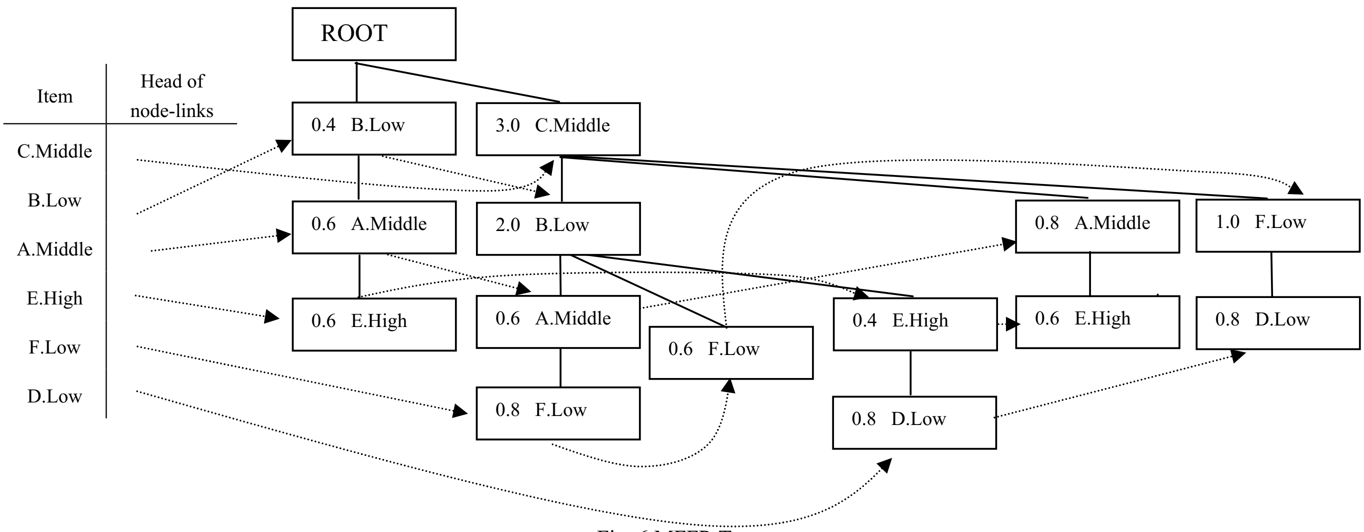

Fig. 6 MFFP-Tree

TABLE X
GENERATE FREQUENT PATTERNS

Item	Conditional Pattern-Base	Conditional MFFP-Tree	Generate frequent patterns
F.Low	<(C.Middle),(B.Low), (A.Middle)>, <(A.Middle)>, <(C.Middle), (B.Low)>, <(C.Middle)>	<(C.Middle):2, (B.Low):2, (A.Middle):1>	{(A.Middle),(F.Low)}, {(C.Middle),(F.Low)}, {(B.Low),(F.Low)}, {(C.Middle),(B.Low),(F.Low)}
E.High	<(B.Low),(A.Middle)>, <(C.Middle),(B.Low)>, <(C.Middle),(A.Middle)>	<(C.Middle):2, (B.Low):2, (A.Middle):2>	{(C.Middle),(E.High)}, {(B.Low),(E.High)}, {(A.Middle),(E.High)}, {(C.Middle),(B.Low),(A.Middle),(E.High)}
A.Middle	<(B.Low)>, <(C.Middle)>, <(C.Middle, B.Low)>	<(C.Middle):2, (B.Low):2>	{(C.Middle),(A.Middle)}, {(B.Low),(A.Middle)}, {(C.Middle),(B.Low),(A.Middle)}
B.Low	<(C.Middle)>	<(C.Middle):1>	*
C.Middle	*	*	*

Step 14: Finally, each pattern must be larger or equal to the fuzzy average weighted. Here, generate rules of corresponding (A.Middle, C.Middle), (A.Middle, E.High) and (B.Low, C.Middle), which are (0.078, 0.137, 0.194), (0.078, 0.133, 0.183) and (0.139, 0.211, 0.261).

Step 15: The given linguistic minimum confidence value is transformed into a fuzzy set of minimum confidences. Assume the membership functions for minimum confidence values are shown in Fig 7, which are similar to those in Fig 5.

Also assume the given linguistic minimum confidence value is “Middle”. It is then transformed into a fuzzy set of minimum confidences, (0.25, 0.5, 0.75), according to the given membership function in Fig 7.

Step 16: The fuzzy average weight of all possible linguistic terms of importance is the same as that found in Step 8. Its gravity is thus 0.5. The fuzzy weighted set of minimum confidences for “Middle” is then (0.25, 0.5, 0.75) × 0.5, which is (0.125, 0.25, 0.375).

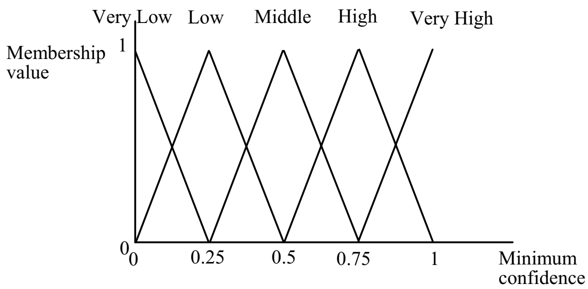

Fig. 7 The membership functions of minimum confidences

Step 17: The fuzzy association rules from conform fuzzy weighted support are checked by using the following substeps.

(a) All possible association rules are formed as follows:

IF *A.Middle*, then *C.Middle*;
IF *C.Middle*, then *A.Middle*;
If *A.Middle*, then *E.High*;
If *E.High*, then *A.Middle*;
If *B.Low*, then *C.Middle*;
If *C.Middle*, then *B.Low*.

(b) The weighted confidence value for the above possible fuzzy association rules are calculated. Take the first possible fuzzy association rule as an example. The fuzzy count of A.Middle is 2.0. The fuzzy count of A.Middle $\cap$ C.Middle is 1.4. The minimum average weight of (A.Middle, C.Middle) is (0.333, 0.583, 0.833). The weighted confidence value for the association rule "*IF A.Middle, then C.Middle*" is:

$$\frac{1.4}{2.0}\times(0.333,0.588,0.833)=(0.233,0.408,0.583).$$

The weighted confidence value for the other fuzzy association rules can be similarly calculated.

(c) The weighted confidence of each association rule is compared with the fuzzy weighted minimum confidence by fuzzy ranking. Assume the gravity ranking approach is adopted in this example. Take the association rule "If A.Middle, then C.Middle" as an example. The average height of the weighted confidence for this association rule is (0.233 + 0.408 + 0.583) / 3, which is 0.408. The average height of the fuzzy weighted minimum confidence for "Middle" is (0.125 + 0.25 + 0.375) /3, which is 0.25. Since 0.408 > 0.25, the association rule "If A.Middle, then C.Middle" is thus put in the interesting rule set. In this example, the following six rules are put in the interest rule set:

1. If a middle number of A is bought then a middle number of C is bought;
2. If a middle number of C is bought then a middle number of A is bought;
3. If a middle number of A is bought then a high number of E is bought;
4. If a high number of E is bought then a middle number of A is bought;
5. If a low number of B is bought then a middle number of C is bought;
6. If a middle number of C is bought then a low number of B is bought.

Step 18: The linguistic support and confidence values are found for each rule R. Take the interesting association rule "If A.Middle, then C.Middle" as an example. Its fuzzy weighted support is (0.078, 0.137, 0.194) and fuzzy weighted confidence is (0.233, 0.408, 0.583). Since the membership function for linguistic minimum support region "Low" is (0, 0.25, 0.5) and for "Middle" is (0.25, 0.5, 0.75), the weighted fuzzy set for these two regions are (0, 0.125, 0.25) and (0.125, 0.25, 0.375). Since (0, 0, 0.125) < (0.078 0.137, 0.194) < (0.125, 0.25, 0.375) by fuzzy ranking, the linguistic support value for Rule R is then Low. Similarly, the linguistic confidence value for Rule R is High.

The interesting linguistic association rules are then output as:

1. If a middle number of A is bought then a middle number of C is bought, with a low support and a high confidence;
2. If a middle number of C is bought then a middle number of A is bought, with a low support and a middle confidence;
3. If a middle number of A is bought then a high number of E is bought, with a low support and a middle confidence;
4. If a high number of E is bought then a middle number of A is bought, with a low support and a high confidence;
5. If a low number of B is bought then a middle number of C is bought, with a low support and a high confidence;
6. If a middle number of C is bought then a low number of B is bought, with a low support and a middle confidence.

The six rules above are thus output as meta-knowledge concerning the given weighted transactions.

I. Conclusion

In this research, we combine the concept of fuzzy weight, fuzzy partition methods in data mining, and use FP-Growth to propose FWFP-Growth algorithm mining all association rules. Hong et al.[10] proposed combining the concept of fuzzy set with association rule, due to they used Aprioir algorithm. If it executed mining in huge database, all efficiency would be wrong and a huge mining cost might be needed. This paper applied FWFP-Growth algorithm to solve the disadvantages of Apriori algorithm in repeating scanning database and the waste of reducing time.

In addition, definition in linguistic value, managers can select his own preference, and refer to past experiences and relate cognitive to design number of linguistic values and shapes, such as Gaussian distribution and trapezoid membership function. So, it confirms cognitive of manager in subject. In fact, Pedrycz [15] had pointed out triangle membership function constructs usefulness and validation in fuzzy system. This is also an important fact of using triangle membership function in this research. In the future, we will attempt to design other fuzzy data mining models for various problem domains.

References

[1] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. & Verkamo, A. I. "Fast discovery of association rules", in U. M. Fayyad, G. Piatesky-Shapiro, P., Smyth, and R. Uthursamy, *Advances in Knowledge Discovery and Data Mining* , AAAI Press, Menlo Park, pp.307-328, 1996.

[2] Berry, M. and Linoff, G., *Data Mining Techniques: for marketing, sales, and customer support*. John Wiley & Sons, NY, 1997.

[3] Chen, S.M., Jong, W.T., "Fuzzy query translation for relational database systems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol.27, no.4, pp.714-721, 1997.

[4] Han, J. W. and Kamber, M. *Data: Concepts and Techniques*. Morgan Kaufmann, San Francisco, 2001.

[5] Han, J., Pei, J., and Y., "Mining frequent patterns without candidate generation," *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp.1-12, 2000.

[6] Hong, T. P., & Chen, J. B., "Find relevant attributes and membership functions", *Fuzzy Sets and Systems*, vol.103, no.3, pp. 389-404, 1999.

[7] Hong, Tp., Wang, T. T., Wang, S. L., & Chien, B. C., "Learning a coverage set of maximally general fuzzy rules by rough sets," *Expert Systems with Applications*, vol.19, no.2, pp.97-103, 2000.

[8] Hu, Y. C., Chen, R. S. & Tzeng, G. H., "Finding fuzzy classification rules using data mining techniques ," *Pattern Recognition Letters*, vol.24, pp.509-519, 2003.

[9] Hu. Y. C., "Mining association rules at a concept hierarchy using fuzzy partition," *Journal of Information Management*, vol.13, no.3, pp.63-80, July, 2006.

[10] T.P. Hong, M. J. Chiang & S. L. Wang, "Fuzzy weighted data mining from quantitative transactions with linguistic minimum supports and confidences," *International Journal of Fuzzy Systems*, vol8, no4, pp. 173-182, Dec. 2006.

[11] Ishibuchi, H., Nakashima, T., & Murata, T., "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.

[12] Ishibuchi, H., Nozaki, K., Yamamoto, N. and Tanaka, H., "Selecting fuzzy if-then rules for classification problems using genetic algorithms," *IEEE transactions on fuzzy systems*, vol.3, no.3, pp.260-270, 1995.

[13] Jang, J. S. R., Sun, C. T. and Mizutani, E., *Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence*. Prentic-Hall, NJ, 1997.

[14] Myra, S., "Web useage mining for web site evaluation", *Communications of the ACM*, vol.43, no.8, pp.127-134, 2000.

[15] Pedrycz, W. "Why triangular membership functions?," *Fuzzy Sets and Systems*, vol.64, pp.21-30, 1994.

[16] Pedrycz, W., Gomide, F., *An introduction of fuzzy sets: analysis and design*. MIT Press, Cambridge, 1998.

[17] Wang, L. X. and Mendel, J. M. "Generating fuzzy rules by learning from examples," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 6,pp. 1414-1427, 1992.

[18] Yuan Y. and Shaw., M. J. "Induction of fuzzy decision trees," *Fuzzy Sets and Systems* , vol.69 , pp.125-139, 1995.

[19] Zadeh, L. A. "The concept of linguistic variable and its application to approximate reasoning," *Information Science* (Part 1), vol. 8, no. 3, pp.199-249, 1975a.

[20] Zadeh, L. A. "The concept of linguistic variable and its application to approximate reasoning," *Information Science* (Part 2), vol. 8, no. 4, pp.301-357, 1975b.

[21] Zadeh, L. A. "The concept of linguistic variable and its application to approximate reasoning," *Information Science* (Part 3), vol. 9, no.1, pp:43-80, 1976.

[22] Zimmermann, H.-J., *Fuzzy sets, Decision making and expert systems*. Kluwer, Boston, 1991.

Chien-Hua Wang received his M.S. degree in information management from Hsuan Chuang University, Hsinchu, Taiwan, R.O.C., in 2006. He is currently a Ph. D student in Yuan Ze University. His current research interests include Data Mining, Electronic Commerce and Decision Making.

Chin-Tzong Pang received the Ph.D. degree in mathematics from National Central University, Taiwan, R.O.C., in 1997. He is a Professor in the Department of Information Management, Yuan Ze University, Taiwan. His current research interests in clued matrix analysis, fuzzy control, fuzzy matrix analysis, and fuzzy database. He has published papers in *Linear Algebra and Its Applications, SIAM Journal on Matrix Analysis and Applications,* IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Fuzzy Sets and Systems,* and *Computers and Mathematics with Applications.*