COMPARISON OF APRIORI, FP-TREE GROWTH AND FUZZY FP-TREE GROWTH ALGORITHM FOR GENERATING ASSOCATION RULE MINING OF COGNITIVE SKILL

Dr. M. Mayilvaganan, D. Kalpanadevi
Associate Professor, Research Scholar
Department of Computer Science, PSG College of Arts and Science
kalpanapani@gmail.com

Abstract— *:* In this research, focuses on implication of association rules among the quantitative attributes and categorical attribute of a database employing fuzzy logic and Frequent Pattern Tree growth algorithm. In the first stage, apply fuzzy partition methods and use triangular membership function of quantitative value for each transaction item, for the generation of more realistic association using fuzzy intervals among quantitative attribute. In second stage, implement Frequent Pattern Tree growth for deal with the process of data mining to analyze the frequent pattern item. In addition, in order to understand the impact of Apriori algorithm and FP-Tree growth algorithm on the execution time, accuracy of best rule found from the frequent pattern mining and the number of generated association rules, the experiment can be performed by using different sizes of support count. In third stage, an experiment results shows Fuzzy FP- Tree growth algorithm is more efficient than existing methods of Apriori and FP Tree growth algorithm.

Keywords— Apriori algorithm; FP-Tree growth algorithm; Fuzzy FP- Tree growth; fuzzy partition methods; triangular membership function; support count; confidence

INTRODUCTION

In data mining, the association rules are used to finds for the associations between the different items of the transactions database. The research focuses the cognitive skill analysis from the students in Numerical Ability, Logical Reasoning and Perceptual Speed. In this paper, the scope of the research is the question type is split as simple, moderate and complicated and the scoring result under the different categories of questions in numerical ability and Logical reasoning. Each should be split into three intervals such as Low, Medium and High can be implemented in frequent pattern mining in Fuzzy FP tree growth for generating fuzzy association rule and analyse the cognitive complexity level of student.

MATERIALS AND METHODS

APRIORI ALGORITHM

In Apriori algorithm, an association rule generation process can be represented in two steps [1] [7][10]:

1. In the First, minimum support is applied for find out the satisfied frequent item-sets.
2. In the Second, minimum confidence can be found out from the required frequent item-sets and constraints are used to mine all association rules.

Apriori algorithm works continuously repeated to scan the database. It uses breadth-first search and a tree structure to count candidate item sets efficiently. Let us assume that it generates candidate item sets of length i from item sets of length i − 1. Then it prunes the candidates which have an infrequent sub pattern. Finally find out all possible combination of frequent item-sets until it does not generate candidate item. Downward closure property of support give the guarantee of frequent item-set contains all frequent i-length item sets [11].

Apriori does not filter prior candidate item-set, it helps to reduce the amount of candidate item-sets to scan. Therefore it needs more time to complete the scanning a database. By implementing process it is not completely efficient.

The support supp(X) of an item-set X is defined as the proportion of transactions in the data set which contain the item-set which can be calculated by the equation (1) [9][1].

$$\text{Supp (X)} = \frac{\text{(no. of transactions which contain the itemset X)}}{\text{(total no.of transactions)}} \qquad (1)$$

The confidence of a rule is defined to interpreted as an estimate of the probability P(Y | X), the probability of finding the right hand side of the rule in transactions under the condition that these transactions also contain the left hand side from the equation 2.

$$\text{Confidence(X=>Y)} = \frac{\text{support}(X \cup Y)}{\text{support}(X)} * 100\% \qquad (2)$$

In this research, to construct the fuzzy intervals [8] and explore the Frequent Pattern Mining algorithm, apply to the sample database of students scored data such as marks attained in numerical and logical reasoning which is illustrated in table 5.1. The model parameter of output variable of cognitive skill of required attributes are broken down into 3 fuzzy sets low, medium and high is The fuzzy frequent item sets, represented by linguistic terms are derived from the fuzzy FP-tree. The count of a fuzzy item set obtained by a fuzzy intersection (minimum) operator can be easily achieved without scanning the database.

FP-TREE GROWTH ALGORITHM

FP- Tree Growth is an algortihm [3] that generates frequent itemsets from an FP-Tree [7] [9] by Divide and Conquer Strategy. Insert sorted items by frequency into a pattern tree. It construct conditional frequent pattern tree and insert sorted item by conditional pattern base which satisfy minimum support. To discover the frequent itemset without candidate item generation which is constructed using 2 passes over the dataset, In pass 1 to scan the transaction data form database and find support for each item,then gradually increase the support and discarded the infrequent items, then sort in decreasing order [12]. In pass 2, FP growth read transaction at a time and maps it to a path. To allocate each item node in tree, each item can link with each node based on single linked list. Fixed order is used, so, path may overlap when transactions share the item. Finally, the header table mines the conditional pattern tree which finds out all frequent item-sets in recursive manner. It does not need association length to proceed phases which generate candidate item-sets in Apriori [1].

PROPOSED METHODOLOGY -FUZZY FP-TREE GROWTH ALGORITHM

In this research, implication of association rules among the quantitative attributes and categorical attributes of a database employing fuzzy logic and frequent pattern growth algorithm. Before applying FP tree growth algorithm, to partitioning the attributes is split into three intervals using triangular membership function. In fuzzy set, the membership degree of each element is any value in between 0 and 1, where the membership degree of each fuzzy class is characterized by using triangular membership functions. These fuzzy intervals lead to the generation of more meaningful and right association rules than classical intervals.

DATA BASE COLLECTION OF NUMERICAL ABILITY AND LOGICAL REASONING SCORING BY THE CATEGORY OF QUESTION TYPE

The representation of numerical and logical reasoning can be split into simple, moderate and complicated, and each should be split into three intervals such as Low, Medium and High for generating fuzzy association rule.

In this research, to construct the fuzzy intervals and explore the Frequent Pattern Mining algorithm [4], apply to the sample database of students scored data such as marks attained in numerical and logical reasoning which is illustrated in table 1. The model parameter of output variable of cognitive skill of required attributes are broken down into 3 fuzzy sets low, medium and high is shown in table 2.

Table 1: Training Database for students scored in Numerical Ability and Logical Reasoning

Transaction ID	Transaction Items					
	Numerical Simple	**Numerical Moderate**	**Numerical Complicate**	**Logical Simple**	**Logical Moderate**	**Logical Complicate**
1	5	4	2	4	2	2
2	5	3	2	5	4	3
3	5	5	2	3	0	4
4	5	3	2	3	1	3
5	3	3	2	5	4	3
6	5	4	2	5	5	4
7	4	2	2	5	4	3
8	5	4	3	5	0	3
9	5	4	5	5	3	1
10	5	4	4	5	1	2
11	5	4	3	2	4	4
12	2	2	3	5	4	2
13	5	3	4	5	5	3
14	2	2	4	5	4	1
15	5	2	3	5	5	1
16	5	5	2	4	1	2
17	4	3	2	5	5	3
18	5	4	2	2	5	4
19	5	5	4	5	1	2
20	5	5	5	5	2	2
21	3	3	2	4	5	2
22	5	3	4	5	1	2
23	3	2	2	5	3	2
24	4	4	2	5	5	2

Table 2: Numerical Ability question level Split by Simple, Moderate and Complicated for Associate rule mining of identifying the scoring of students

Ranges	Total Marks	Constructing an Interval of fuzzy Region	Ranges for Numerical Ability	Ranges for Logical Reasoning
Simple	5	0 - 1.6 0.7 - 3.4 2.6 – 6	Low -NS.Low Medium -NS.Medium High -NS.High	Low -LS.Low Medium -LS.Medium High - LS.High
Moderate	5		Low -NM.Low Medium-NM.Medium High -NM.High	Low -LM.Low Medium-LM.Medium High -LM.High
Complicate	5		Low -NC.Low Medium -NC.Medium High -NC.High	Low -LC.Low Medium - LC.Medium High -LC.High

CONSTRUCTING FUZZY INTERVALS

The construction of fuzzy Intervals of quantitative attribute is split into three fuzzy intervals by applying statistical approach. The lower border of first intervals is the minimum value over domain of the quantitative attributes. The higher border is computed by using the mean and standard deviation of the value of quantitative attributes [6].

In this research, triangular membership function is used to construct the membership function [2] specified by three parameters (boundaries): Lower limit value (left vertex), upper limit value (right vertex) and modal value (center vertex). In the fuzzy region, the representation of triangular membership function with highest is 1 at the center of the fuzzy region. Where x can be represents the ranges of input value in numerical type of the attributes. A_i^l denotes lower limit value, A_i^u denotes upper limit value and M_i denotes modal value in fuzzy region. The modal value M_i kept in any region which can be computed as $M_i = \frac{A_i^l + A_i^u}{2}$.

Before constructing fuzzy interval [8], measure the proper vector of interval values in fuzzy region in the universe of discourse [5]. Let X= {1, 2, 3, 4, 5} be the universe of discourse of the attributes simple, Moderate and complicated. The attributes are split into three intervals such as {low, medium and high}.

From this given universe of discourse of attributes X, the minimum value= 1 and maximum value = 5. Based on measuring the increment unit in the axis of simple, Moderate and Complicated can be defined as equation 5.1.

$$IncreC = \left[\frac{A_i^{\max} + A_i^{\min}}{|A_i|} * w \right] \quad (5.1)$$

Where $A_i^{\max}$ the maximum is value of A_i, $A_i^{\min}$ is the minimum value of A_i and $|A_i|$ is the number of distinct value of A_i and w is the positive integer user-defined weight for control the number of regions needs to be created. It helps to define the vector of interval values of fuzzy regions in the universe of discourse. Similarly, increment unit of Moderate and Complicated attributes can be measured.

In this research, suppose an attribute of A contains minimum value = 0 and maximum value =5 and assumes the weight of the increment unit w= 1 and substitute in the equation 5.1 and computed as 5.2, then we get the increment unit of the interval is

$$Incre(simple_i) = \frac{0+6}{6} * 1 = 1$$

$$Incre(Moderate_i) = \frac{0+6}{6} * 1 = 1$$

$$Incre(Complicate_i) = \frac{0+6}{6} * 1 = 1 \quad (5.2)$$

So, the vector unit of each Attribute - Simple, Moderate and Complicated is {0, 1, 2, 3, 4, 5, 6}. Here the last unit of 6 with k= 1 satisfies the condition 4< $A_i^{\max}$ =5 $\leq$ 6.

Constructing the fuzzy region, to shape the membership functions of an input attribute is affected by the result of overlap measure by the equation of 5.3.

$$Overlap(A_l, A_S) = \frac{\left|F^i_{com}(A_l, A_S)\right|}{\left|F^i_{all}(A_l, A_S)\right|} \quad (5.3)$$

Where $\left|F^i_{com}(A_l, A_S)\right|$ represents the value of $\{a_{ij}, a_{ik, ..} a_{in}\}$ and in the same region belong to the required set of classes in A_l and the value of $\{\{a_{ip}, a_{iq, ..}a_{iu}\}\}$ and in the same region belong to the required set of classes in A_S. $\left|F^i_{all}(A_l, A_S)\right|$ represent all values of classes in A_l and A_s. The proposed measure for computing the degree of overlap between the adjacent fuzzy regions A_l, A_S is computed.

Initial parameters of attribute Simple, Moderate and Complicates fuzzy region created as A_1, A_2 and A_3.

Let us assign an initial value of intervals and fuzzy region of A_1 (Low) is created with parameter A_1= {0,1} and M_1 =0.5 i.e., ((0+1)/2), A_2 (Medium) is created with the parameter={1,2,3} and its modal value M_2 = 2 i.e., ((1+3)/2)) and A_3 (High) is created with the parameter={3,4,5} and its modal value M_3 =4 i.e., ((3+5)/2) .

The degree of overlap between adjacent functions of new output attribute is assumed as 50%. Let to find out the lower limit and upper limit of each interval, by measuring the degree of overlap of adjacent fuzzy region referred as £, where 0<£ <1, i.e., $£_1$= overlap (A_1, A_2) and $£_2$= overlap (A_2, A_3) then the upper border value of fuzzy region A_1 will be change by the equation 5.5, due to the percentage amount of overlap between the regions A_1^U is shift inside the region of A_2; Further the lower border value of fuzzy region A_2 is calculated by equation 5.6, here the percentage amount of is shift inside of A_1 region. The value of A_2^l will change from the assigning value and updated by overlapping. Similarly, to find the upper and lower border of A_2^U, A_3^l, A_3^U, where A_3^U = Max (quantitative_ attribute) by the equation of 5.7 to 5.9. In all cases the modal value M_1, M_2 and M_3 cannot change.

To find the first intervals, £= overlap (A_1, A_2) = 0.3, here M_1=1.

$$A_1^l = \min(quantitative\_attributes) \quad (5.4)$$

$$= 0$$

$$A_1^U = \left(A_2^U - A_2^l\right) * £_1 + A_1^U \quad (5.5)$$

$$A_1^U = (3-1)*0.3+1 = 0.6+1= 1.6$$

Then we get, the fuzzy region of first interval A_1 = {0, 1, 1.6}

To find the second interval, $£_1$= overlap (A_1, A_2) = 0.3; $£_2$= overlap (A_2, A_3) = 0.2, here M_2=2.

$$A_2^l = A_1^U - (A_1^U - A_1^l) * £_1 \quad (5.6)$$

$$A_2^l = 1-(1-0)*0.3 =1-0.3 = 0.7$$

$$A_2^U = \left(A_3^U - A_3^l\right) * £_2 + A_3^l \quad (5.7)$$

$$A_2^U = (5-3)*0.2+3=3.4$$

The fuzzy region of second interval A_2= {0.7, 2, 3.4}

To find the third interval, $£_2$= overlap (A_2, A_3) = 0.2; here M_3=4.

$$A_3^l = A_2^U - (A_2^U - A_2^l) * £_2 \quad (5.8)$$

$$A_3^l = 3-(3-1)*0.2= 2.6$$

$$A_3^U = \max(\text{Quantitative \_ attribute}) \quad (5.9)$$

$$= 6$$

The fuzzy region of third interval A_3= {2.6, 4, 6}

Finally to generating the fuzzy region A_1, A_2 and A_3 for attribute of Numerical Simple, Numerical Moderate and Numerical Complicated and Logical Simple, Logical Moderate and Logical Complicated with overlapping the regions successfully as shown in Figure 1.

After finding the intervals, the membership degree of input which represents 'x', by following triangular membership function to computing the degree of each numeric input of attribute which belong to the region A_i function.

This interval has been characterized using the triangular membership function shown in equation 5.10,

$$\mu(x, A_i^l, M_i, A_i^u) = \max\left(0, \min\left(\frac{x - A_i^l}{M_i - A_i^l}, \frac{A_i^u - x}{A_i^u - M_i}\right)\right) \quad (5.10)$$

The model parameter of output variable of cognitive skill of required attributes are broken down into 3 fuzzy sets low, medium and high.

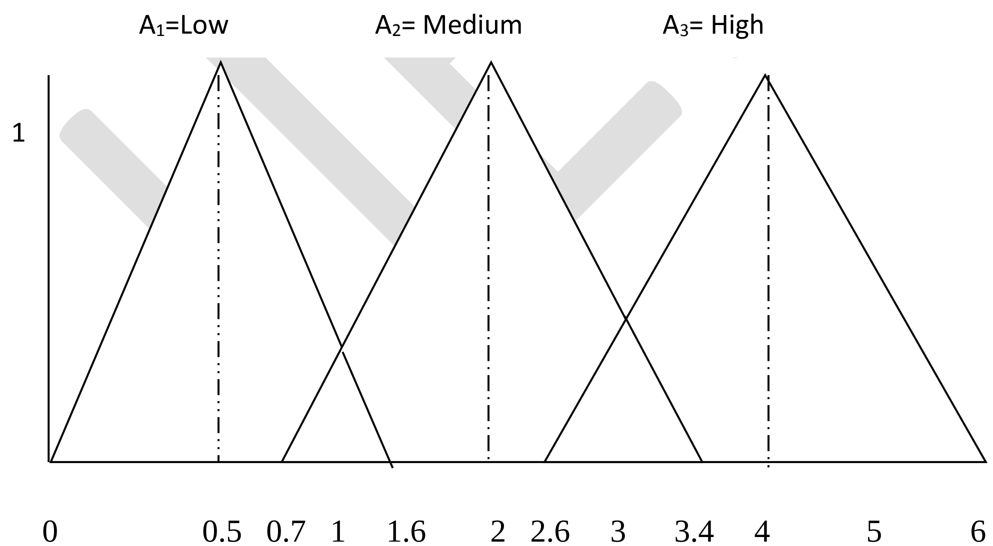

Figure 1 Generating fuzzy regions A_1, A_2 and A_3 attributes of Simple Moderate and Complicated of Numerical and Logical Reasoning with overlapping

Here an linguistic variable of scoring values of simple, moderate and Complicated of Numerical and Logical Reasoning can be split into μ, which can be used to determine the degree to which this input belongs to fuzzy set as below,

$$A_1 = \mu_{Low}(x) = \begin{cases} 0, x < 0 \\ \dfrac{(x-0)}{0.5}, 0 < x \le 0.5 \\ \dfrac{(1.6-x)}{1.1}, 0.5 < x < 1.6 \end{cases} \tag{5.11}$$

$$A_2 = \mu_{Medium}(x) = \begin{cases} \dfrac{(x-0.7)}{1.3}, 0.7 \le x \le 2 \\ \dfrac{(3.4-x)}{1.4}, 2 < x < 3.4 \end{cases} \tag{5.12}$$

$$A_3 = \mu_{High}(x) = \begin{cases} \dfrac{(x-2.6)}{1.4}, 2.6 < x \le 4 \\ \dfrac{(6-x)}{2}, 4 < x < 6 \end{cases} \tag{5.13}$$

Preprocessing the Domain Values for Associating the Related Data for Research

As the data collected and stored, rules of value can be found through association rules, which can be applied to skill of Numerical Ability and Logical Reasoning question. The question level can Split by Simple, Moderate and Complicated for Associate rule mining as shown in table 1 to identify the scoring of students.

For instance, when x= 1, the membership degree can be computed, the value of mark in any simple, moderate and Complicated attribute which belong to fuzzy regions A_1 and A_2 with different degree. Such as μ(1, A_1)= μ(1, 0, 0.5, 1.6) = 1.2 and μ(1, A_2) = μ(1, 0.7, 2, 3.4) = 0.23. Similarly, when X=3, the membership degree can be computed, the value of mark in any simple, moderate and Complicated attribute which belong to fuzzy regions A_2 and A_3 with different degree. Such as μ(3, A_2)= μ(3, 0.7, 2, 3.4) = 0.285 and μ(3, A_3) = μ(3, 2.6, 4, 6) = 0.285.

An overall mapping table from quantitative attribute of table 3, the membership degree of each datum of each domain in fuzzy data base of degree in between 0 and 1. From the below table 3, illustrates the sample training data of 100 tuples scoring simple, moderate and complicated type of numerical ability and logical reasoning.

For instance, to take first 30 tuple from the dataset, it produces 120 fuzzy transaction tuples by fuzzy partitioning as shown in table 4 with membership degree.

The frequency of all items can be measured and represent in table 5. The support of each item is determined by membership degrees of that item in every transaction or tuple from table 4. Here, the minimum membership value of each tuples assigned as an overall membership degree of that tuple. Thus the database contains non-zero membership degree of that every tuple.

The support count can be measured by the summing of all membership degree of that required item in every transaction. For example, the support of NS.High is 28.145; i.e., the membership degree of NS.High is 0.5+0.285+0.285+0.285+0.285+0.285+0.23+0.285+0.285+...+0.285+0.285=28.145 which can be summed in the fuzzy set of each tuple. Similarly, the remaining support of all items can be found in the table 4 respectively. In this research, assumed support is 23% for 120 transactions as given in the table 3, then the required frequent items in table 6 considered as a header table. Remaining infrequent items are discarded from the transaction. According to the selection of support count, NS.Medium, LS.Medium, LS.Low, and LC.Low are infrequent items and frequent items can be representing in table 6. Finally, an each transaction of can sort in descending order with respect to frequent item as shown in table 7 respectively. Here transaction data of '0' consider as not attended the corresponding type of question during the test.

Table 3 Fuzzy Set transformed after partition the quantitative data from the data transaction

T.Id	Numerical Simple	Numerical Moderate	Numerical Complicate	Logical Simple	Logical Moderate	Logical Complicate
1	0.5/NS.High	1/NM.High	1/NC.Medium	1/LS.High	1/LM.Medium	1/ LC.Medium
2	0.5/NS.High	0.285/NM.Medium and 0.285/NM.High	1/NC.Medium	0.5/LS.High	1/LM.High	0.285/ LC.Medium and 0.285/LC.High
3	0.5/NS.High	0.5/NM.High	1/NC.Medium	0.285/LS.Medium and 0.285/LS.High	1.2/LM.Low and 0.23/LM.Medium	1/LC.High
4	0.5/NS.High	0.285/NM.Medium and 0.285/NM.High	1/NC.Medium	0.285/LS.Medium and 0.285/LS.High	1.2/LM.Low and 0.23/LM.Medium	0.285/ LC.Medium and 0.285/LC.High
5	0.285/NS. Medium and 0.285/High	0.285/NM.Medium and 0.285/NM.High	1/NC.Medium	1.2/LS.Low and 0.23/LS. Medium	1/LM.High	0.285/ LC.Medium and 0.285/LC.High
6	0.5/NS.High	1/NM.High	1/NC.Medium	1/LS.Medium	0.5/LM.High	1/LC.High
7	1/NS.High	1/Medium	1/NC.Medium	0.5/LS.High	1/LM.High	0.285/ LC.Medium and 0.285/LC.High
8	0.5/NS.High	1/NM.High	0.285/NC.Medium and 0.285/NC.High	0.5/LS.High	1.2/LM.Low and 0.23/LM.Medium	1.2/ LC.Low and 0.23/ LC.Medium
9	0.5/NS.High	1/NM.High	0.5/NC.High	1.2/LS.Low and 0.23/ LS.Medium	0.285/LM.Medium and 0.285/LM.High	1.2/ LC.Low and 0.23/ LC.Medium
10	1/NS.High	1/NM.High	1/NC.High	0.5/LS.High	1.2/LM.Low and 0.23/LM.Medium	1/ LC.Medium
11	0.5/NS.High	1/NM.High	0.285/NC.Medium and 0.285/NC.High	1/LS.Medium	1/LM.High	1/LC.High
12	1/NS.Medium	1/NM.Medium	0.285/NC.Medium and 0.285/NC.High	0.5/LS.High	1/LM.High	1/ LC.Medium
13	0.5/NS.High	0.285/NM.Medium and 0.285/NM.High	1/NC.High	0.5/LS.High	1.2/LM.Low and 0.23/LM.Medium	0.285/ LC.Medium and 0.285/LC.High
14	1/NS.Medium	1/NM.Medium	1/NC.High	0.5/LS.High	1.2/LM.Low and 0.23/LM.Medium	1.2/ LC.Low and 0.23/ LC.Medium
15	0.5/NS.High	1/NM.Medium	0.285/NC.Medium and 0.285/NC.High	0.5/LS.High	0.5/LM.High	1.2/ LC.Low and 0.23/ LC.Medium
16	0.5/NS.High	0.5/NM.High	1/NC.Medium	1/LS.High	1.2/LM.Low and 0.23/LM.Medium	1/ LC.Medium

T.Id	**Numerical Simple**	**Numerical Moderate**	**Numerical Complicate**	**Logical Simple**	**Logical Moderate**	**Logical Complicate**
17	1/NS.High	0.285/NM.Medium and 0.285/NM.High	1/NC.Medium	0.5/LS.High	0.5/LM.High	0.285/ LC.Medium and 0.285/LC.High
18	0.5/NS.High	1/NM.High	1/NC.Medium	1/LS.Mediu m	0.5/LM.High	1/LC.High
19	0.5/NS.High	0.5/NM.High	1/NC.High	0.5/LS.High	1.2/LM.Low and 0.23/LM.Medium	1/ LC.Medium
20	0.5/NS.High	0.5/NM.High	0.5/NC.High	0.5/LS.High	1/LM.Medium	1/ LC.Medium
21	0.285/NS.Medi um and 0.285/High	0.285/NM.Medium and 0.285/NM.High	1/NC.Medium	1.2/LS.Low and 0.23/LS.Med ium	1.2/LM.Low and 0.23/LM.Medium	1/ LC.Medium
22	0.5/NS.High	0.285/NM.Medium and 0.285/NM.High	1/NC.High	0.5/LS.High	1.2/LM.Low and 0.23/LM.Medium	1/ LC.Medium
23	0.285/NS. Medium and 0.285/High	1/Medium	1/NC.Medium	0.5/LS.High	0.285/LM.Medium and 0.285/LM.High	1/ LC.Medium
24	1/NS.High	1/NM.High	1/NC.Medium	0.5/LS.High	0.5/LM.High	1/ LC.Medium
25	0.5/NS.High	0.5/NM.High	1/NC.Medium	1/LS. Medium	0.5/LM.High	1/ LC.Medium
26	0.5/NS.High	1/NM.High	1/NC.Medium	1/LS.High	1/LM.Medium	1/ LC.Medium
27	0.5/NS.High	1/NM.High	1.2/NC.Low and 0.23/NC.Medium	0.5/LS.High	0.285/LM.Medium and 0.285/LM.High	0.285/ LC.Medium and 0.285/LC.High
28	0.285/NS.Medi um and 0.285/High	1/NM.High	1/NC.Medium	0.5/LS.High	1/LM.High	0.285/ LC.Medium and 0.285/LC.High
29	0.5/NS.High	1/NM.High	0.285/NC.Medium and 0.285/NC.High	0.5/LS.High	0	0.285/ LC.Medium and 0.285/LC.High
30	0.5/NS.High	0.5/NM.High	1/NC.High	1/LS. Medium	1/LM.Medium	0.285/ LC.Medium and 0.285/LC.High

Table 4 Fuzzy Database

T. ID	FUZZY SET						MEM. DG
	NS	NM	NC	LS	LM	LC	
1	0.5/NS.High	1/NM.High	1/NC.Medium	1/LS.High	1/LM.Medium	1/ LC.Medium	0.5
2	0.5/NS.High	0.285/NM.Medium	1/NC.Medium	0.5/LS.High	1/LM.High	0.285/ LC.Medium	0.285
3	0.5/NS.High	0.285/NM.Medium	1/NC.Medium	0.5/LS.High	1/LM.High	0.285/LC.High	0.285
4	0.5/NS.High	0.285/NM.High	1/NC.Medium	0.5/LS.High	1/LM.High	0.285/ LC.Medium	0.285
5	0.5/NS.High	0.285/NM.High	1/NC.Medium	0.5/LS.High	1/LM.High	0.285/LC.High	0.285
6	0.5/NS.High	0.5/NM.High	1/NC.Medium	0.285/LS.Medium	1.2/LM.Low	1/LC.High	0.285
7	0.5/NS.High	0.5/NM.High	1/NC.Medium	0.285/LS.Medium	0.23/LM.Medium	1/LC.High	0.23
8	0.5/NS.High	0.5/NM.High	1/NC.Medium	0.285/LS.High	1.2/LM.Low	1/LC.High	0.285
9	0.5/NS.High	0.5/NM.High	1/NC.Medium	0.285/LS.High	0.23/LM.Medium	1/LC.High	0.285
10	0.5/NS.High	0.285/NM.Medium	1/NC.Medium	0.285/LS.Medium	1.2/LM.Low	0.285/ LC.Medium	0.285
11	0.5/NS.High	0.285/NM.Medium	1/NC.Medium	0.285/LS.High	1.2/LM.Medium	0.285/LC.High	0.285
12	0.5/NS.High	0.285/NM.High	1/NC.Medium	0.285/LS.Medium	1.2/LM.Low	0.285/ LC.Medium	0.285
13	0.5/NS.High	0.285/NM.High	1/NC.Medium	0.285/LS.High	1.2/LM.Medium	0.285/LC.High	0.285
14	0.5/NS.High	0.285/NM.Medium	1/NC.Medium	0.285/LS.High	1.2/LM.Low	0.285/ LC.Medium	0.285
15	0.5/NS.High	0.285/NM.Medium	1/NC.Medium	0.285/LS.High	1.2/LM.Low	0.285/ LC.Medium	0.285
16	0.5/NS.High	0.285/NM.Medium	1/NC.Medium	0.285/LS.Medium	1.2/LM.Low	0.285/LC.High	0.285
17	0.5/NS.High	0.285/NM.Medium	1/NC.Medium	0.285/LS.High	1.2/LM.Medium	0.285/ LC.Medium	0.285
18	0.5/NS.High	0.285/NM.Medium	1/NC.Medium	0.285/LS.Medium	1.2/LM.Medium	0.285/ LC.Medium	0.285
19	0.285/NS.Medium	0.285/NM.Medium	1/NC.Medium	1.2/LS.Low	1/LM.High	0.285/ LC.Medium	0.285
20	0.285/NS.Medium	0.285/NM.High	1/NC.Medium	1.2/LS.Low	1/LM.High	0.285/ LC.Medium	0.285
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
114	0.285/NS.Medium	1/NM.High	1/NC.Medium	0.5/LS.High	1/LM.High	0.285/LC.High	0.285
115	0.285/High	1/NM.High	1/NC.Medium	0.5/LS.High	1/LM.High	0.285/ LC.Medium	0.285
116	0.285/High	1/NM.High	1/NC.Medium	0.5/LS.High	1/LM.High	0.285/LC.High	0.285
117	0.5/NS.High	1/NM.High	0.285/NC.Medium	0.5/LS.High	0	0.285/ LC.Medium	0.285
118	0.5/NS.High	1/NM.High	0.285/NC.Medium	0.5/LS.High	0	0.285/LC.High	0.285
119	0.5/NS.High	1/NM.High	0.285/NC.High	0.5/LS.High	0	0.285/ LC.Medium	0.285
120	0.5/NS.High	1/NM.High	0.285/NC.High	0.5/LS.High	0	0.285/LC.High	0.285

Table 5 Counting of the Frequency Item

SI.No	Items	Frequency Count
1	NS.High	28.145
2	NS.Medium	6.44
3	NM.High	21.59
4	NM.Medium	12.655
5	NC.Medium	23.385
6	NC.High	10.2
7	LS.High	21.525
8	LS.Medium	7.48
9	LS.Low	5.02
10	LM.Low	8.89
11	LM.Medium	10.56
12	LM.High	13.92
13	LC.Low	3.115
14	LC.Medium	21.32
15	LC.High	9.65

Table 6 Header Table

SI. No	Items	Possible Value	Count
1	NS.High	NSH	28.145
2	NC.Medium	NCM	23.385
3	NM.High	NMH	21.59
4	LS.High	LSH	21.525
5	LC.Medium	LCM	21.32
6	LM.High	LMH	13.92
7	NM.Medium	NMM	12.655
8	LM.Medium	LMM	10.56
9	NC.High	NCH	10.2
10	LC.High	LCH	9.65
11	LM.Low	LML	8.89

Table 7 Pre-processed Fuzzy Database

TID	Frequent Item (Ordered)	Mem DG
1	NSH, NCM, NMH, LSH, LCM, LMM	0.5
2	NSH, NCM, LSH, LCM, LMH, NMM	0.285
3	NSH, NCM, LSH, LMH, NMM, LCH	0.285
4	NSH, NCM, NMH, LSH, LMH, LCH	0.285
5	NSH, NCM, NMH, LSH, LMH, LCH	0.285
6	NSH, NCM, NMH, LCH, LML	0.285

7	NSH, NCM, NMH, LMM, LCH	0.23
8	NSH, NCM, NMH, LSH, LCH, LML	0.285
9	NSH, NCM, NMH, LSH, LMM, LCH	0.285
10	NSH, NCM, LCM, NMM, LML	0.285
11	NSH, NCM, LSH, NMM, LMM, LCH	0.285
12	NSH, NCM, NMH, LCH, LML	0.285
13	NSH, NCM, NMH, LSH, LMM, LCH	0.285
14	NSH, NCM, LSH, LCM, NMM, LML	0.285
15	NSH, NCM, LSH, LCM, NMM, LML	0.285
16	NSH, NCM, NMM, LCH, LML	0.285
17	NSH, NCM, LSH, LCM, NMM, LMM	0.285
18	NSH, NCM, LCM, NMM, LMM	0.285
19	NCM, LCM, LMH, NMM	0.285
20	NCM, NMH, LCM, LMH	0.285
21	NCM, LCM, LMH, NMM	0.285
22	NCM, LMH, NMM, LCH	0.285
23	NCM, NMH, LMH, LCH	0.285
24	NCM, NMH, LMH, LCH	0.285
25	NSH, NCM, LCM, LMH, NMM	0.285
26	NSH, NCM, NMH, LCM, LMH	0.285
27	NSH, NCM, LCM, LMH, NMM	0.285
28	NSH, NCM, LMH, NMM, LML	0.285
29	NSH, NCM, NMH, LMH, LCH	0.285
30	NSH, NCM, NMH, LCM, LCH	0.285
:	:	:
:	:	:
117	NSH, NCM, NMH, LSH, LCM	0.285
118	NSH, NCM, NMH, LSH, LCH	0.285
119	NSH, NMH, LSH, LCM, NCH	0.285
120	NSH, NMH, LSH, NCH, LCH	0.285

Let us assume the item in transaction NS.High as NSH, NS.Medium as NSM and other items are termed as in table 6.

An FP-tree is basically a prefix tree in which each path represents a set of transactions that share the same prefix. Then the procedure root= node (), i.e., null which is applied on ordered frequent items from the header table 5.6 to build up FP-tree, which is shown in Figure 2. The header file is build during the construction of FP tree because each item point's first occurrence is in tree via node link. Nodes with same name are linked in sequence via such node links as shown in Figure 1.

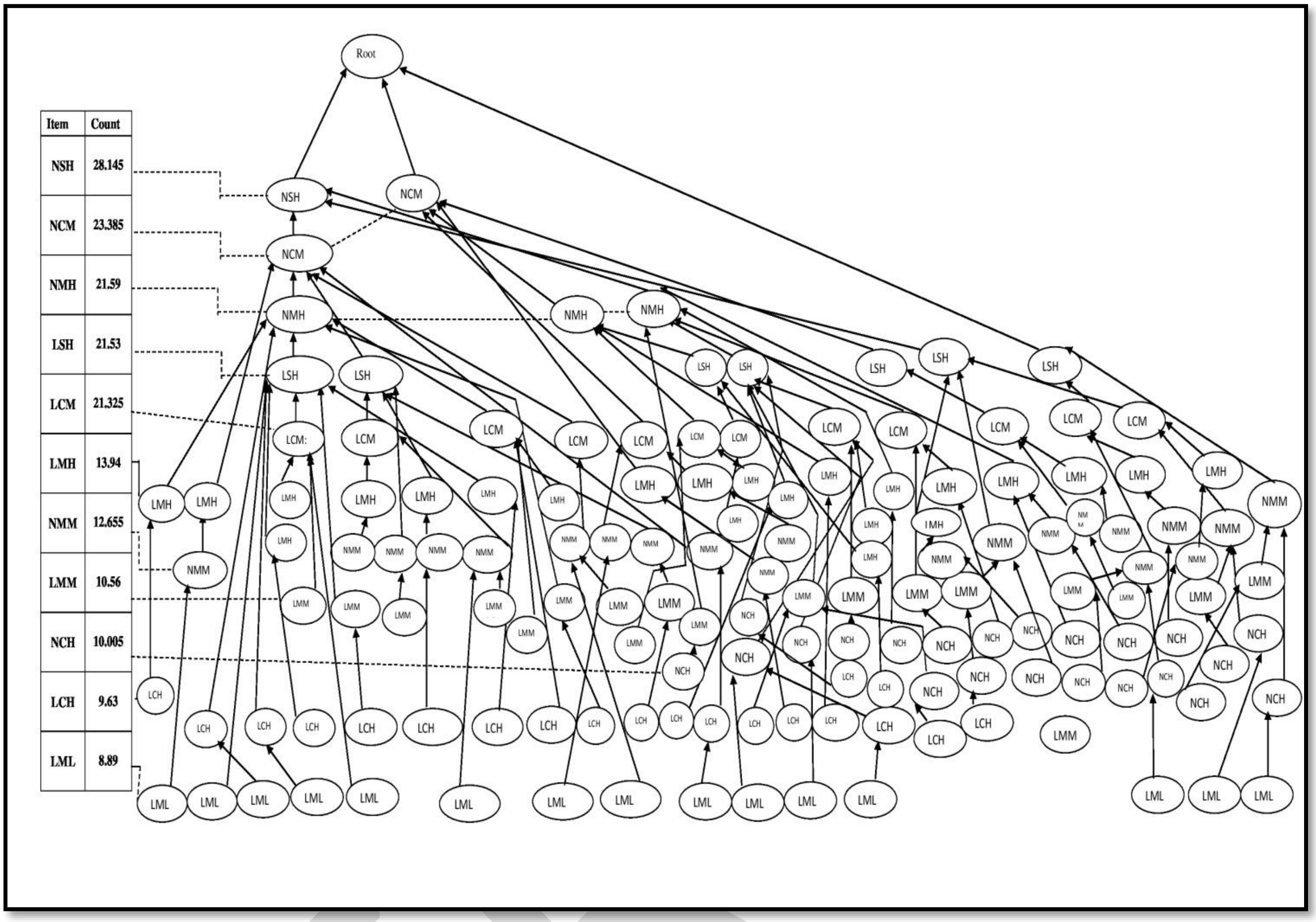

Figure 1 Building Fuzzy FP Tree

According to the procedure, after crating a root termed "null", the first branch is constructed for transaction < NSH, NCM, NMH LSH, LCM, LMM: 0.5>, where six new nodes are created for items NCM, NMH, LCM, LSH, LMM. And the node NSH is linked as the child of root. The root node termed as null. NCM is linked as the child of NSH, NMH is linked as the child of node NCM, LSH is linked as the child of node NMH, LCM is linked as the child of node of LSH and finally LMM is linked as the child of node of LCM. As the next transaction < NSH, NCM, LSH, LCM, LMH, NMM: 0.285> does share common prefix with previous transaction up to NCM. At NCM, a new child node is created and link to LSH and LCM is linked as the child of node LSH, LMH is linked as the child of node LCM and NMM is linked as the child of node LMH. In case, the membership degree of NSH, NCM is added with previous degree in each transaction. Similarly, the remaining transactions are mapped and the tree structure is built. The nineteenth transaction has common prefix < NCM, LCM, LMH, NMM: 0.285>, there is no common nodes in previous transaction, so the second branch is constructed from the root. In this way all transactions are embedded into FP-tree and continue until all transactions are mapped to a path in the FP-tree.

GENERATING FUZZY FREQUENT ITEM-SETS

After completing the FP-tree construction, the conditional pattern base can establish a condition FP-tree, the frequent fuzzy item-sets containing more than one fuzzy region can be found in a way to similar to the FP growth Mining algorithm.

The process of this algorithm, the conditional pattern base from the child node LML and finishes at the root node "Null". Hence the node of LML contains fifteen prefix paths as shown in Figure 2. These are called conditional pattern base of "LML". Similarly, the conditional base of all LCH, NCH, LMM, NMM, LMH, LCM, LSH, NMH, NCM and NSH patterns base can be analyzed from the prefix based on bottom up process in one by one. Finally all conditional pattern bases can be found as shown in table 8.

Table 8 Generating Frequent Item Sets

Item	Conditional Pattern Base	Conditional Frequent item set
LML	{<NMM, LMH, NCM, NSH: 0.57> <LSH, NMH, NCM, NSH:0.695> <LCM, LSH, NMH, NCM, NSH: 1.53> <LCH, NMH, NCM, NSH: 0.8> <NMM, LCM, LSH, NCM, NSH:0.57> <NMM, LCM, NCM, NSH: 0.57> <LCH, NMM, NCM, NSH: 0.285> <NCH, NMM, LCM, LSH, NSH:0.57> <LCH, NCH, LMM, LSH, NMH,,NSH: 1.025> <NCH, LSH, NMH, NSH: 0.515> <NCH, NMH, LSH, NSH:0.515> <NCH, LCM, LSH, NMH, NSH:1.245>}	<NSH LML: 8.89>
LCH	{<LMH, NMH, NCM, NSH: 0.285> <LMH, NCM, NSH: 0.57> <LMH, LSH, NMH, NCM, NSH: 0.57> <NMH, NCM, NSH : 3.4> <NMM, LMH, LSH, NCM, NSH : 0.57> <LCM, NMH, NCM, NSH: 0.285> <LMM, LSH, NMM, NCM, NSH: 0.715> <LMM, NMM, LSH, NCM, NSH: 0.57> <NMM, NCM, NSH: 0.285> <NMM, LMH, NCM: 0.285> <LMH, NMH, NCM: 0.57> <LMH, NMH, NCM, NSH: 0.8> <NCH, NMM, LSH, NSH: 0.23> <NCH, LMM, NMM, LSH, NSH: 0.515>}	Ø
NCH	{<LMM, NMM, LCM, LSH, NSH :0.8> <NMM, LMH, LCM, LSH :0.855> <LMM, NMM, LCM, LSH, NSH:0.8> <NMM, LMH, NMH, NSH: 0.57> <LMH, NMH, NSH: 0.8> <LMM, NMM, LSH, NSH: 0.515> <NMM, LSH, NSH: 0.515> <LMM, LSH, NMH, NSH :0.46> <LCM, LSH, NMH, NSH: 1.475> <LMM, LCM, LSH, NMH, NSH:0.69> <LCH, LCM, NMH, NSH: 1.03> <LMM, NMH, NSH : 1.03> <LMM, LCM, LSH, NMH, NSH: 0.69 >}	<NSH NCH : 3.115>
Item	**Conditional Pattern Base**	**Conditional Frequent item set**
LMM	{<LCM, LSH, NMH, NCM, NSH: 1.76> <NMM, LCM, LSH, NCM, NSH:0.855> <LSH, NMH, NCM, NSH: 2.11> <NMM, LSH, NCM, NSH: 0.57> <NMM, LCM, LSH, NSH: 1.315> <NMM, LSH, NSH: 0.515>	< NSH LMM: 10.56 >

	<LSH, NMH, NSH: 1.19> <LCM, LSH, NMH, NSH:1.19> <NMH, NSH: 1.055>}	
NMM	{<LMH, NCM, NSH: 0.57> <LMH, LCM, LSH, NCM, NSH: 1.14> <LCM, LSH, NCM, NSH:1.86> <LMH, LSH, NCM, NSH: 0.8> <LSH, NCM, NSH: 0.7> <LCM, NCM, NSH: 1.060> <LMH, LCM, LSH, NCM: 7.345> <LMH, LCM, NCM, NSH:0.57> <LMH, LCM, NCM: 0.57> <NCM, NSH: 0.285> <LMH, NCM: 0.285> <LMH, LCM, LSH: 0.57> <LCM, LSH, NSH: 1.03> <LMH, NMH, NSH:0.855> <LSH, NSH: 1.015>}	Ø
LMH	{<NMH, NCM, NSH: 0.285 > <NCM, NSH: 0.57> <LSH, NMH, NCM, NSH: 0.57> <LCM, LSH, NCM, NSH: 0.855> <LSH, NCM, NSH: 0.57> <LCM, NMH, NCM, NSH: 0.285> <LCM, LSH, NCM: 1.085> <LCM, NCM, NSH: 1.14> <LCM, NCM: 0.855> <LCM, NMH, NCM: 0.285> <NCM: 0.285> <NMH, NCM: 1.425> <LCM, LSH: 0.8> <NMH, NCM, NSH: 2.335> <NMH, NSH: 1.655> <LCM, NMH, NSH: 0.92>}	<NSH NCM LMH: 0.855>
LCM	{<LSH, NMM, NCM, NSH: 6.272> <LSH, NCM, NSH: 5.418> <NMH, NCM, NSH: 0.57> <NCM, NSH: 1.995> <LSH, NCM: 0.57> <LSH, NMH, NSH: 2.845> <NMH, NSH: 0.855> <NCM:0.57> <NMH, NCM: 0.285> <LSH: 0.57> <LSH, NSH: 1.37>}	Ø
Item	**Conditional Pattern Base**	**Conditional Frequent item set**
LSH	{<NMH, NCM, NSH: 14.05> <NCM, NSH: 2.120> <NCM: 0.57> <LSH:0.57> <NSH: 1.37> <NMH, NSH: 2.845>}	<NMH NCM NSH: 14.05>

NMH	{<NCM, NSH:15.21> <NCM: 0.855> <NSH: 5.525>}	<NCM NSH NMH: 15.21>
NCM	{<NSH: 21.105>	<NSH NCM: 21.105>
NSH	{Ø}	Ø

Finally, the frequent fuzzy item set can be generated by the recursive approach of FP Tree growth. Here intersection operation can be performed for finding the minimum item-set from the conditional base pattern tree as shown in table 8.

GENERATING FUZZY ASSOCIATION RULE

Generate the Association rule from frequent item-set with the support and confidence. From the 30 instances of transaction data of student skill analysis, the frequent item can be produced as in table 5.8. For rule generation, one of the rules of the frequent item-set is placed as a consequent and the rest of the items are placed as antecedent in association rule. Then the confidence value is determined by the equation 5.3.

$$Sup(A \to B) = \sum_{x,y \in N} \min(A(x)B(y))$$

$$conf\,(A \to B) = \frac{\sum_{x,y \in N} \min(A(x)B(y))}{\sum_{x \in N} A(x)} \qquad (5.3)$$

Let us consider, the frequent item-set <NSH NCM LMH: 0.855> where the support of this association is 3.115. The candidate rules are NSH NCM ➔LMH, NSH LMH→NCM and LMH NCM→NSH. The confidence of NSH NCM → LMH is

$$\frac{\sup port(NSH \cup NCM \cup LMH)}{\sup port(NSHNCM)} *100\% = \frac{0.855}{6.61} = 0.13{=}13\%$$

and similarly, to find confident of all frequent item-set as shown in table 9.

Table 9 Experiment Result

SI.No.	Generate Fuzzy Association Rules		Confidence %
	A(x)	**B(y)**	
1	NSH, LSH	NMH	51.3
2	NSH	LSH	56.5
3	LCM	LSH	57.1
4	LCH	NSH	57.7
5	NSH	LSH	58.5
6	LCM	NSH	60.7
7	LCH	NCM	61.5
SI.No.	**Generate Fuzzy Association Rules**		**Confidence %**
	A(x)	**B(y)**	
8	LCM	NCM	64.3
9	LCH	LSH	65.4
10	LSH	NSH	62.9
11	LSH	NSH	74.1

12	NMH	LSH	65.9
13	LSL	NCM	77.8
14	LCM	NCM	78.8
15	NMH	NSH	82.9
16	NMH	NCM	83.3
17	LMH	NCM	84.6
18	NMM	NCM	100
19	NSM	NCM	100
20	NCH	LSH	100

From this experiment, the result of rule generation of the given transaction of 1120 tuples produced fuzzy association rule from frequent item-set in table 9, illustrates the rule generation of frequent item set and its confident level is measured by the support of association of items evolved in the pattern. From this analysis, by assuming the confident level is 0.5, 20 association rule can be generated with respect of the support count 0.23. The table 5.9, the rule generation of NMH→ NSH shows high confident value of 82.9%. The count of a fuzzy item set obtained by a fuzzy intersection (minimum) operator can be easily achieved without scanning of the database and analyzing the rule generation from the example transaction by given the support count, the frequent pattern item-set can be found efficiently with the confident value of 50% to100%.

The rule of NSH→NMH indicates that if the student attained high in simple question of Numerical ability, male or female poses high in moderate question of Numerical and the possibility is 82.9%. The rule of NMH→NSH indicates that if the student attained high in numerical moderate question, the student of male or female poses high in numerical simple question and again the confident value is 82.9%.

Table 10 Rule generation using different size of support count of 1000 transactions

Support Count	**Apriori**	**FP Tree**	**Fuzzy FP Tree**
0.1	502	148	356
0.2	145	38	176
0.3	67	20	55
0.4	34	7	39
0.5	22	4	26

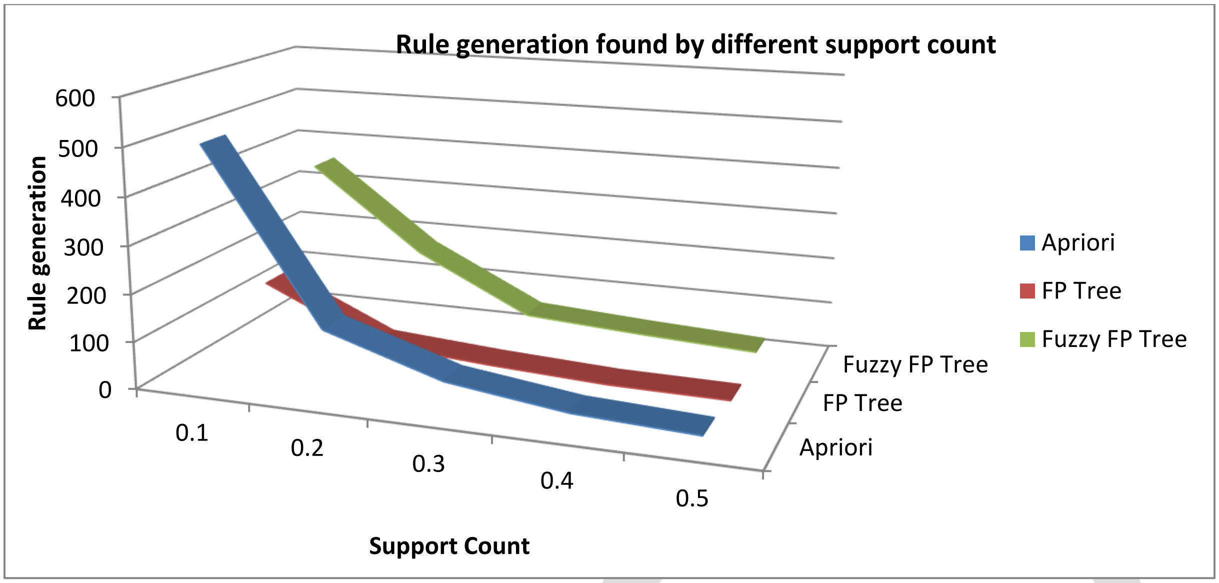

Figure 5.2 Rule generation found by different support count

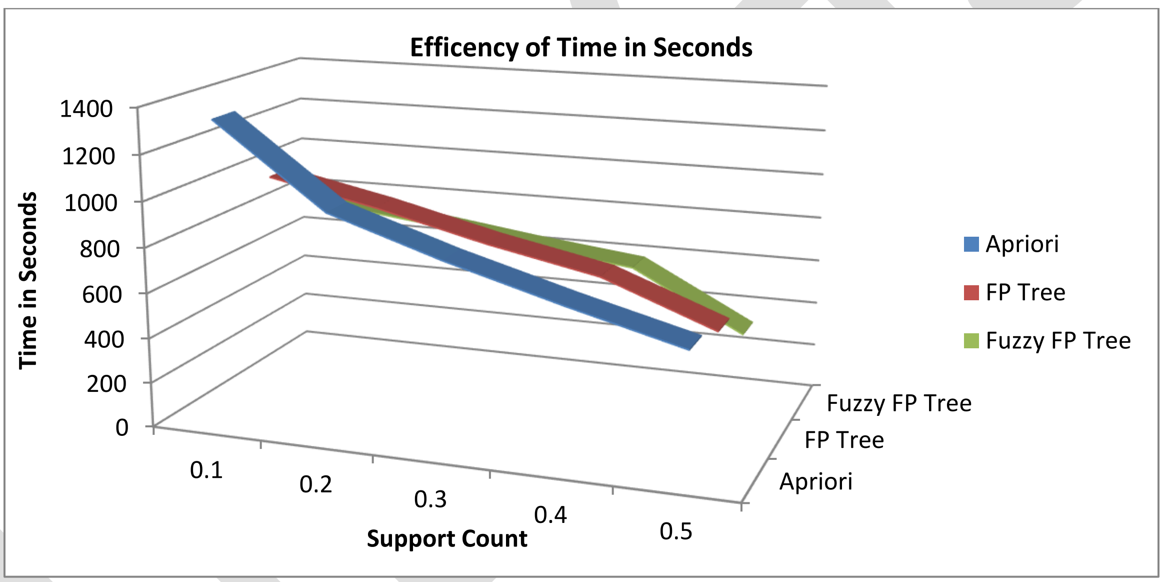

Figure 3 Execution time of rule generation by comparing three algorithms by different support count

Similarly, table 10 represents the rule generation of behavior of student's skill level of simple, moderate and complicated question in numerical and logical reasoning of 1000 training data for analyzing by this association. The number of rules generated in association with frequent pattern mining; the experiment can be performed by using different sizes of support count and analyzed an efficiency of three algorithms. From this analysis, time efficiency and accuracy of rule generation can be experimented. By comparison of (Apriori, FP-Tree and Fuzzy FP Tree growth algorithm) Fuzzy FP-Tree algorithm proved the time efficiency and rule generation is highly more efficient than Apriori, FP-Tree algorithm as shown in Figure 2 and 3.

CONCLUSION

In this paper, it can be concluded that by the comparison of three algorithms (Apriori, FP Tree and Fuzzy FP tree algorithm) obtained the rule generation by given the support of 23% and confident is 50%. A priori produced rule bases by the occurrences of candidate generation and produced more rules. FP tree maintained a classical data base of Boolean format and Fuzzy set theory which handles the quantitative value between the intervals 0 to 1, therefore fuzzy FP Tree growth can combined with fuzzy set theory and FP Growth which deals with quantitative values for each attributes measured in training tuples which helps to speed up the learning phase.

The count of a fuzzy item set obtained by a fuzzy intersection (minimum) operator can be easily achieved without scanning of the database and analyze the rule generation by given different size of support count from 0.1 to 0.5 by the confident level of 0.5.

Finally, Fuzzy FP-Tree algorithm proved the time efficiency and rule generation is highly more efficient than Apriori, FP-Tree algorithm. The frequent pattern mining can be obtained with respect to given numeric quantitative data of 1000 training data as shown in table 10.

REFERENCES:

[1] Jiwei Han, Jian Pei, Yiwen Yen and Runying Mao: "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach, Data Mining and Knowledge Discovery, 8, 53–87, 2004, 2004 Kluwer Academic Publishers

[2] Takagi T. and M. Sugeno, "Fuzzy identification of system and its applications to modeling And control", Proc. IEEE Trans. on System Man and Cybernetics, Vol. SMC-15, No. 1, pp. 116-132, 1985.

[3] Rakesh Agrawal; Tomasz Imielinski and Arun N. Swami, "Mining Association Rules Between Sets of Items in Large Databases", Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-216, Washington, D.C., May 1993.

[4] A.H.M. Sajedul Hoque, Rashed Mustafa, Sujit Kumar Mondal and Md. Al-Amin Bhuiyan," A Fuzzy Frequent Pattern-Growth Algorithm For Association Rule Mining" International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.5, September 20

[5] H. M. Sajedul Hoque, Sujit Kumar Mondol, Tassnim Manami Zaman, Dr. Paresh Chandra Barman, Dr. Md. Al-Amin Bhuiyan, "Implication of Association Rules Employing FP-Growth Algorithm for Knowledge Discovery", 14th International Conference on Computer and Information Technology 2011 (ICCIT 2011), 22-24 December, Dhaka, Bangladesh.

[6] George J. Klir and Bo Yuan: Fuzzy Sets and Fuzzy Application: Theory and Applications, Prentice Hall of India Private Limited, 2005.

[7] Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, 2nd Ed., Morgan Kaufmann Publishers, 2006.

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

[9] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction to Data Mining", Pearson Education Inc., Fourth Edition, 2009.

[10] Heikki, Mannila, "Data mining: machine learning, statistics, and databases", IEEE, 1996.

[11] Arun K Pujari, "Data mining techniques", University Press (India) Private Limited.

[12] Usama Fayyad, "Data Mining and Knowledge Discovery in Databases: Implications for Scientific Databases" Proceedings of the Ninth International Conference on Scientific and Statistical Database Management IEEE Computer Society Washington, DC, USA, 1997.