IFERP
connecting engineers... developing research

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

Analysis of Cognitive Skill level of Students by Fuzzy FP Tree Growth Association to identify Frequent Pattern Mining

[1] D. Kalpanadevi, [2] Dr. M. Mayilvaganan
[1] Research Scholar, [2] Associate Professor
Department of Computer Science, PSG College of Arts and Science

Abstract: **- In this research, focuses on using fuzzy partition method and triangular membership function of quantitative value for each transaction item, for the generation of more realistic association using fuzzy intervals among quantitative attribute. Secondly, to implement Frequent Pattern Tree growth for deal with the process of data mining to analyze the frequent pattern item. The performance of cognitive skill of knowledge can be analyzed by the categories of logical Reasoning, Numerical Ability, and Perceptual Speed of ability and relates with parents' education, Standard of Medium study gender and specialty.**

Index Terms**: Apriori algorithm; FP-Tree growth algorithm; Fuzzy FP- Tree growth; fuzzy partition methods; triangular membership function; support count; confident.**

I. INTRODUCTION

By implication of association rules among the quantitative attributes and categorical attribute of a database employing fuzzy interval and frequent pattern tree growth algorithm. In addition, in order to understand the impact of Apriori algorithm [1] and FP-Tree growth [3], 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.

COLLECTION OF QUANTITATIVE DATA

As the data is collected and stored, rules of value can be found through association rules, which can be applied to the skill of Numerical ability, logical reasoning and Perceptual speed level of students and the brain dominant hemisphere, also related to the categories of gender, specialty, age, standard of study and parent education as shown in table. 1a and table 1b.

Table.1a Collection of Quantitative Training data

Attributes	Possible Values
Age	Age between 18 to 25
Gender	Male (m) and Female (f)
Logical Reasoning (L)	13 to 15 = L.Very High 10 to 12 = L.High 7 to 9 = L.medium 4 to 6 = L.Low 1 to 3 = L.Very Low
Numerical Ability (NA)	13 to 15 = N.Very High 10 to 12 = N.High 7 to 9 = N.medium 4 to 6 = N.Low 1 to 3 = N.Very Low
Perceptual Speed (P)	13 to 15 =P.Very High 10 to 12 = P.High 7 to 9 = P.Medium 4 to 6 = P.Low 1 to 3 = P.Very Low

Table.1b Collection of Quantitative Training data

Attributes	Possible Values
Left brain dominant (LBD)	-15 to -9
Moderate preference for the left (ML)	-8 to -5
Slight preference toward the left (SL)	-4 to -1
Whole-brain dominance (WB)	0
Slight preference toward the right dominance (SR)	+1 to +4
Moderate preference for the right (MR)	+5 to +8
Right brain dominant (RBD)	+9 to +15
Standard of Medium study	State Board/ Matriculation Board/ CBSE
Parent Education (PE)	Yes or No
Specialty	Computer Science- CS, Information Technology- IT, Computer Technology- CT, Computer Application –BCA, Commerce- CC

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

PREPROCESSING THE DATA BY CLASSICAL AND FUZZY INTERVALS

The training data of age, gender, specialty, parents' education and standard medium of studies to be considered as classical intervals, quantitative attributes belong to 0's and 1's, without overlap among intervals during pre-processing [4]. Here 41 training data are taken for experimental purpose. An interval of brain dominant values can be constructed by statistical approach. In this research, each quantitative data between -15 to-1 values is split into three intervals such as LBD, ML, SL. The value of '0' is considered as whole brain dominant which can be assumed as 1, and the data between 1 to 15 values is split into three intervals such as RBD, SR, MR respectively.

An intervals of Numerical, Logical Reasoning and Perceptual speed values of students that can be constructed by statistical approach by the category of very low, low, medium, high and very high with overlapping. The training set of 1000 instances of data taken in various colleges in Coimbatore district, the sample training data is shown in table 2.The application of this mechanism results in the compact tree structures that reduce computation time. The fuzzy frequent item sets, represented by linguistic terms, are derived from the fuzzy FP-tree [3].

Let us assign an initial value of intervals and fuzzy region of R1 (LBD) is created with parameter R1= {-15, -14, -13,-12, -11, -10, -9} and M1 =-12 i.e., (-15+-9)/2), R2 (ML) is created with the parameter {-9, -8 -7, -6 -5} and its modal value M2 = 7 and R3 (SL) is created with the parameter {-5, -4 , -3, -2, -1) and its modal value M3 =3).After initializing the value, find lower and upper border of fuzzy region of the attributes by using equations 1 to 6.In all cases the modal value M1, M2 and M3 cannot change.To find the first intervals, £= overlap (R1, R2) = 0.1, here M1=-12.

$$A_1^l = \min(quantitative\_attributes) = 0 \qquad (1)$$

$$R_1^U = \left(R_2^U - R_2^l\right) * £1 + R_1^U \qquad (2)$$

R_1^U = (-5+9)*0.1+ (-9) = 0.4-9= -8.6

Then we get, the fuzzy region of first interval A1 = {0, -12, -8.6}

To find the second interval, £1= overlap (R1, R2) = 0.1; £2= overlap (R2, R3) = 0.1, here M2=-7.

$$R_2^l = R_1^U - (R_1^U - R_1^l) * £1 \qquad (3)$$

R_2^l = -10-(-10+6)*0.1 =-10-0.4 =-10.4

$$R_2^U = \left(R_3^U - R_3^l\right) * £2 + R_3^l \qquad (4)$$

R_2^U = (-1+5) *0.1+ (-5) = -4.4

The fuzzy region of second interval A2= {-10.4, -7, -4.4}

To find the third interval, £2= overlap (R2, R3) = 0.1; here M3=-3.

$$R_3^l = R_2^U - (R_2^U - R_2^l) * £2 \qquad (5)$$

R_3^l= -5-(-5+9)*0.1= -5.4

$$A_3^U = \max(\text{Quantitative\_attribute}) = -1 \qquad (6)$$

The fuzzy region of third interval A3= {-5.4, -3, -1}

Finally to generating the fuzzy region R1, R2 and R3 for attribute of Left brain Dominant, Sight Left brain Dominant and Moderate, Left Brain Dominant, Right brain Dominant, Sight Right brain Dominant and Moderate Right Brain Dominant between the interval values of 1 to 15. To find the first intervals, £= overlap (B1, B2) = 0.1, here M1=3.

After finding the fuzzy region with overlaps, the membership degree of inputs can be computed for preprocessing of quantitative value into the intervals range between 0 and 1 by using triangular membership function [2] and the input values of training data pre-processed.

Here a linguistic variable of scoring values of LBD, ML,SL, SR, MR and RBD can be split into µ of as in equation 7, which can be used to determine the degree to which this input belongs to fuzzy set as below,

$$\mu_R(x) = \begin{cases} \frac{1}{v-u}(x-u), u \le x \le v \\ \frac{1}{w-v}(w-x), v \le x \le w \\ 0, otherwise \end{cases} \qquad (7)$$

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

Table 2. Mapping Table of Student Details from Fuzzy Region

T. Id	AGE	GENDER	DEPT	PE	STD	BD	N	LR	P	Membership degree
1	19	f	CS	Yes	State Board	SL	High	medium	medium	0.122
2	19	f	CS	Yes	State Board	SL	High	medium	high	0.38
3	19	f	CS	Yes	State Board	SR	Medium	High	Very high	0.122
4	19	f	CS	Yes	State Board	SR	Medium	very high	Very high	0.122
5	19	f	CS	Yes	State Board	SR	High	High	Very high	0.375
6	19	f	CS	Yes	State Board	SR	High	very high	Very high	0.166
7	19	f	CS	Yes	State Board	SL	High	medium	Very high	0.09
8	19	f	CS	Yes	State Board	SL	very high	medium	Very high	0.09
9	19	f	CS	No	State Board	ML	Medium	medium	Very high	0.122
10	19	f	CS	No	State Board	ML	High	medium	Very high	0.38
:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:
102	19	f	CS	Yes	State Board	SL	High	medium	Very high	0.163
103	19	f	CS	Yes	State Board	SL	very high	medium	high	0.163
104	19	f	CS	Yes	State Board	SL	very high	medium	Very high	0.163

As similarly, in brain dominant hemisphere data, the membership degree can be computed in R1, R2, R3, B1, B2, B3 attributes which belongs to fuzzy region with different degree in one transaction. The same process can be followed for finding the fuzzy region for Logical Reasoning, Numerical Ability and Perceptual Speed.

Therefore, the database of 41 tuples of quantitative data (training data) can be preprocessed into 104 fuzzy databases as shown in table 2.

The support of each item is determined by the membership degree of that item in every transaction or tuple as noted in table 2. Here, the minimum membership value [2] of each tuple is assigned as an overall membership degree of that tuple. Thus the database contains non-zero membership degree of every tuple.

The support count can be measured by the summing of all membership degrees of the required item in every transaction. The frequency of all items can be measured and is represented in table 3.

In this research, the assumed support is 23% for 104 transactions as given in table 5.21, then the required frequent items in table 4 is considered as a header table. The support count can be calculated by support count = total no of membership degree *23%. Here if the total membership degree of the fuzzy database is 21.086, then the fuzzy minimum support count 0.23 X 21.086 = 4.84978 is shown in table 4. The Remaining infrequent items are discarded from the transaction.

Table 3 Counting of the Frequency Item

SI. No	Items	Frequency Count
1	Age.19	21.086
2	gender.f	21.086
3	CS	16.446
4	P.Very high	15.27
5	PE.Yes	13.142
6	matriculation	7.944
7	std.state	9.49
8	L.Medium	9.481
9	BD.SL	8.368
10	N.Medium	8.752

11	PE.No	7.944
12	N.High	7.583
13	L.High	7.201
14	BD.ML	6.549
15	N.Very high	5.894
16	CT	4.641
17	L.Very High	4.404
18	P.High	3.514
19	BD.SR	3.528
20	p.medium	2.302
21	BD.RBD	0.707
22	BD.MR	0.468
23	BD.WBD	0.166
24	BD.LBD	0.244

Finally, each transaction can be sorted in descending order with respect to the frequent items, as shown in table 5.

Table 4 Header Table

SI. No.	Items	Possible Values	Count
1	Age.19	19	21.086
2	Gender. F	F	21.086
3	Speciality. CS	CS	16.446
4	P. Very high	PVH	15.27
5	PE. Yes	PEY	13.142
6	Std. Matriculation	MAT	7.944
7	Std. State	SB	9.49
8	L. Medium	LM	9.481
9	BD.SL	BDSL	8.368
10	N. Medium	NM	8.752
11	PE. No	PENO	7.944
12	N. High	NH	7.583
13	L. High	LH	7.201
14	BD.ML	BDML	6.549
15	N. Very high	NVH	5.894

BUILDING FUZZY FP TREE

According to the procedure, after crating a root termed "null", the first branch is constructed for transaction < 19 F CS PEY SB LM BDSL NH: 0.122>, where eight new nodes are created for items F CS PEY SB LM BDSL NH. And the node 19 is linked as the child of root. The root node termed as null. F is linked as the child of 19, CS is linked as the child of node F, PEY is linked as the child of node CS, SB is linked as the child of node of PEY, LM is linked as the child of node of SB, BDSL is linked as the child of node of LM, and finally NH is linked as the child of node of BDSL. As the next transaction < 19 F CS PEY SB LM BDSL NH: 0.38> does share common prefix with previous transaction, here membership degree is added with previous degree. Similarly, the remaining transactions are mapped and the tree structure is built. In this way all transactions are embedded into FP-tree and continue until all

Table 5 Pre-processed Fuzzy Databases (Ordered)

TID	Frequent Item (Ordered)	Membership degree
1	19 F CS PEY SB LM BDSL NH	0.122
2	19 F CS PEY SB LM BDSL NH	0.38
3	19 F CS PVH PEY SB NM LH	0.122
4	19 F CS PVH PEY SB NM	0.122
5	19 F CS PVH PEY SB NH LH	0.375
6	19 F CS PVH PEY SB NH	0.166
7	19 F CS PVH PEY SB LM BDSL NH	0.09
8	19 F CS PVH PEY SB LM BDSL NVH	0.09
9	19 F CS PVH SB LM NM PENO BDML	0.122
10	19 F CS PVH SB LM PENO NH BDML	0.38
:	:	:
:	:	:
104	19 F CS PVH PEY SB LM BDSL NVH	0.163

transactions are mapped to a path in the FP-tree and built a tree as shown in figure 1.

GENERATING FUZZY FREQUENT ITEM-SETS

After completing the FP-tree construction [4] [5], the conditional pattern base can be established by a condition FP-tree, and the frequent fuzzy item-sets contains 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 NVH contains fifteen prefix paths as shown in Figure 1. These are called conditional pattern base of "NVH". Similarly, the conditional base of all BDML, LH, NH, PENO, NM, BDSL, LM, SB, MAT, PEY, PVH, CS, F and 19 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 6.

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 6.

GENERATING FUZZY ASSOCIATION RULE

Here association rules are produced from frequent item-sets by given of support and confidence value. From the 41 instances of training data, an association rule can be generated from frequent items from fuzzy 104 transaction tuple. 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 find the confidence value.

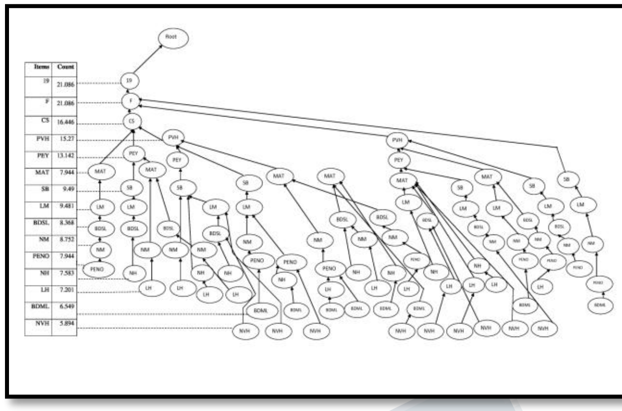

Table 6 Generating Frequent Item Sets

SI. No	Item	Conditional Pattern Base
1	NVH	{<BDSL LM SB PEY PVH CS F 19> <LM SB PEY PVH CS F 19> <PENO LM SB PVH CS F 19> <BDML MAT PEY PVH CS F 19><LH BDSL MAT PEY PVH F 19> <LH MAT PEY PVH F 19><BDSL MT PEY PVH F 19><MAT PEY PVH F 19> <BDSL LM SB PEY PVH F 19>}
2	BDML	{<PENO NM LM SB PVH CS F 19><NH PENO LM SB PVH CS F 19> <LH PENO NM MAT PEY PVH CS F 19><PENO NM MAT PEY PVH CS F 19> <NH MAT PEY PVH CS F 19><MAT PEY PVH CS F 19><PENO NM LM MAT PVH F 19><NH PENO NM MAT PVH F 19><PENO NM LM SB F 19>}
3	LH	{<NM MAT PEY CS F 19><NM SB PEY PVH CS F 19><NM BDSL MAT PEY CS F 19><NH SB PEY PVH CS F 19><PENO NM MAT PEY PVH CS F 19><NM BDSL MAT PEY PVH CS F 19><PENO NM BDSL MAT PEY PVH CS F 19><BDSL MAT PEY PVH F 19><MAT PEY PVH F 19><NH MAT PEY PVH F 19><PENO NM BDSL MAT PVH F 19>}
4	NH	{<BDSL LM SB PEY CS F 19><SB PEY PVH CS F 19><BDSL LM SB PEY PVH CS F 19><PENO LM SB PVH CS F 19><BDSL MAT PEY PVH CS F 19><MAT PEY PVH CS F 19><LM MAT PEY PVH F 19><MAT PEY PVH F 19> <BDSL LM SB PEY PVH F 19><PENO NM MAT PVH F 19>}
5	PENO	{<NM BDSL LM MAT CS F 19><NM LM SB PVH CS F 19><LM SB PVH CS F 19> <NM MAT PVH CS F 19><NM BDSL MAT PVH CS F 19><NM LM MAT PVH F 19> <NM BDSL LM MAT PVH F 19><NM MAT PVH F 19><NM BDSL MAT PVH F 19> <NM BDSL LM SB PVH F 19><NM LM SB F 19>}
6	NM	{<BDSL LM MAT CS F 19><MAT PEY CS F 19><SB PEY PVH CS F 19><LM SB PVH CS F 19><MAT PVH CS F 19><BDSL MAT PEY PVH CS F 19><BDSL MAT PVH CS F 19><BDSL LM SB PEY PVH F 19><LM MAT PVH F 19><BDSL LM MAT PVH F 19><MAT PVH F 19><BDSL MAT PVH F 19><BDSL LM SB PVH F 19> <LM SB F 19>}
7	BDSL	{<LM MAT CS F 19><LM SB PEY CS F 19><MAT PEY CS F 19><LM SB PEY PVH CS F 19><MAT PEY PVH CS F 19><MAT PEY PVH CS F 19><MAT PEY PVH F 19> <LM SB PEY PVH F 19><LM MAT PVH CS F 19><MAT PVH F 19><LM SB PVH F 19><SB F 19>}
8	LM	{<MAT CS F 19><SB PEY CS F 19><SB PEY PVH CS F 19><SB PVH CS F 19>

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

		<MAT PEY PVH F 19><SB PEY PVH F 19><MAT PVH F 19><SB PVH F 19> <SB F 19>}
9	SB	{<PEY CS F 19><PEY PVH CS F 19><PVH CS F 19><PEY PVH F 19><PVH F 19> <F 19>}
10	MAT	{<CS F 19><PEY CS F 19><PVH F 19>}
11	PEY	{<CS F 19><PVH CS F 19><PVH F 19>}
12	PVH	{<CS F 19><F 19>>}
13	CS	{<F 19>}
14	F	{<19>}
15	19	{Ø}

Table 7 Generating Fuzzy Association Rules

T.ID	**Rule Generation in Association**		**Confidence level**
	Premise	**Conclusion**	
1	NM	BDSL	0.5
2	BDSL	NM	0.5
3	NM	19, BDSL	0.5
4	19, NM	BDSL	0.5
5	BDSL	19, NM	0.5
6	19, BDSL	NM	0.5
7	MAT	NM, PENO	0.5087719298245614
8	MAT	19, NM, PENO	0.5087719298245614
9	19, MAT	NM, PE	0.5087719298245614
10	CS, MAT	BDSL	0.5217391304347826
11	BDSL	CS, MAT	0.5217391304347826
12	CS, MAT	19, BDSL	0.5217391304347826
13	19, CS, MAT	BDSL	0.5217391304347826
14	BDSL	19, CS, MAT	0.5217391304347826
15	19, BDSL	CS, MAT	0.5217391304347826
16	MAT	PE	0.5263157894736842

EXPERIMENT ANALYSIS

When compared with Apriori and Fuzzy FP tree algorithms, A priori produced 1098 rule bases by the occurrences of candidate generation. When combined with fuzzy set theory and FP Growth which deals with quantitative values. Fuzzy FP Tree can produce rule generations with respect to a minimum support count of 4.8 and confident level 50% efficiently as represented in table 7 which is sample output of the required 41 tuples.

When compared with FP Tree and Fuzzy FP Tree algorithms, FP tree maintained a classical data base of Boolean format 0's and 1's and Fuzzy FP Tree processed by Fuzzy set theory which handles the quantitative value between the intervals 0 to 1. Here FP Tree generates 134 rules.

By experimental of 1000 training data, the rule generation can be produced using different size of support count. Fuzzy FP-Tree algorithm proved the time efficiency and rule generation highly proved more efficient than Apriori and FP-Tree algorithm as shown in Figure 2 and table 8.

Table 8 Rule generation using different size of support count of 1000 transactions

Support Count	Apriori	FP Tree	Fuzzy FP Tree
0.1	2470	806	1926
0.2	991	563	794
0.3	458	367	356
0.4	224	167	198

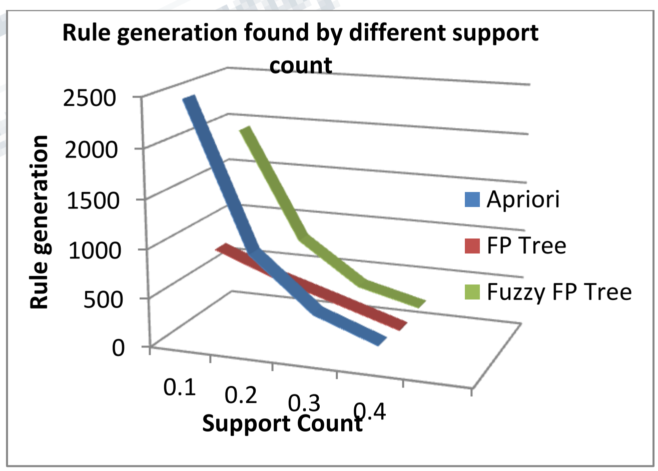

Figure.5.14 Rule generation with different support Count

III. CONCLUSION

In this paper, it can be concluded that the Fuzzy FP-Tree growth algorithm for analyzing the frequent pattern mining of the frequent item-set is placed as a consequent and the rest of the items are placed as antecedent in association rule. Totally, 41 tuples are taken as a training data set and the fuzzy intervals are constructed for the required data and forms 104 fuzzy transaction database. From this research, a fuzzy partition method is applied and triangular membership function of quantitative value for each transaction item is used. The generation of rule has produced more realistic association using fuzzy intervals among quantitative attributes, 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.

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] 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.

[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] 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