A Positive and Negative Probability edge Techniques in Data Mining Association Rule

Monisha Roy[1],Rohit Miri[2], S R Tandan[4],
[1,4]Department of Computer Science & Engineering
Dr. C.V. Raman University, Kota, Bilaspur, India (C.G.)
[1]M.Tech scholar,Department of Computer Science & Engineering
Dr. C.V. Raman University, Kota, Bilaspur, India (C.G.)
[2,4]Department of Computer Science & Engineering

***Abstract*—**The fuzzy association techniques can be performs in different ways. In this paper, the fuzzy association rule is achieved . The positive and negative edge is calculated by the its membership values. The values is fall under the positive edge values if it membership values is greater than .5 and otherwise the value is fall under the negative edge values. Association rule between linguistic variable is calculated by using large item sets and candidate item sets by using the probability techniques..

***Keywords*—**Positive Edge, Negative Edge, Data mining

I.INTRODUCTION

The fuzzy association rule plays a vital role for decision making in the e commerce. In My work the crisp data set is converted in to the fuzzy values. The positive and negative values of each linguistic variable are calculated. The value is calculated by the highest membership function of fuzzy sets. The number of valid fuzzy values is calculated in each column and also the negative and positive edge is also calculated. The fuzzy association rule of linguistic variable is calculated by using the count values of positive and negative values by using the probability. in this work the large item set and candidate item set is do not require fuzzy values for the final association values of linguistic variables. It provides the easy way of finding the association rule and it's also reduces the complexity associated in the steps of finding the candidate item sets and large item sets using fuzzy values. the triangular membership function is uses for calculating the positive and negative values of fuzzy set.

II. LITERATURE REVIEW

Ruchi Bhargava et al[1] use the logic connectives to find the fuzzy association rule amongst the linguistic variables. They also use the weighted and confidence framework of data sets for finding the association rule. Its research work is also efficient in regarding execution time as well as for number of input sets. They also define the three set of fuzzy positive and three set of fuzzy negative rules for finding the association rule. Their work also does not requires to calculate the large item set as well as candidate item sets for their results. However they define the minimum supports.

The item sets that are frequently present in particular transaction id are chooses by Mohammed Al-Maoleg1[4] . Its algorithm works on low support and reduces the time complexity of the program very easily. Zhiyong ma [6] et al converts all the item sets into Boolean matrix by using CP tree method and reduces the time for the task.ArpnaShrivastava[7] et al has used the codes for all the items and removes the duplication by using data cleansing technique. This is also most efficient as compared to simple Apriori algorithm.

According to Thanh Minh Nguyen et al [2] the Fuzzy systems are of two types. In the first model is based on the linguistic on a collection of fuzzy rules. i.,e can be broadly categorized into two families. The first includes linguistic models based on a collection of fuzzy rules, The Mamdani model falls into this group. its contains the If-then rules.If A then do B . this type of rule that contains is fall on the category of positive fuzzy values because the something to do or action is performed.The second types of the rule has not been exploited much, includes *negative rules* (weight is negative), which prescribe actions to be avoided. Thus, in addition to the positive rules, it is possible to augment the rule-base with rules of the form, "IF *A*, Then do not do *B*

Sajid Mahmood et al [3]. they have done on the research the Negative and Positive Association Rules Mining from Text Using Frequent and Infrequent Itemsets. In this paper, they introduce an algorithm for finding the positive and negative association rules among frequent and infrequent itemsets. they identify associations among medications, symptoms, and laboratory results using state-of-the-art data mining technology.

This algorithms work on low support for finding the frequent item sets. The data sets with low supports are ignored by the proposed algorithms because it falls on the negative fuzzy rules which are not important for finding the association rule.

Neelukhare et al [8] has proposed the algorithm fuzzy association rule for mining multidimensional. Tzung-Pei Hong et al [11] An Effective Gradual Data-Reduction Strategy for Fuzzy Item set Mining,present an efficient mining approach to speed up the efficiency of finding fuzzy frequent item sets from dataset. Thomas sudkomp [9]is refinement in knowledge discovery are to produce rules that more accurately model the underlying data while maintaining rule interpretability. In this paper we introduce two refinement strategies for association rules with fuzzy temporal constraints.Usha Rani et al [12] it uses the multilevel approach model for finding the fuzzy association rule by using the top down approaches.

The other approaches which reduces the computational time by using artificial Bee colony optimization method (FABCO) is given by K. Sathesh Kumar and M. Hemalatha[5].Mehmet Kaya et al [18]has worked out an efficient algorithm by carrindout mining fuzzy clustering algorithm (CURE). They found out the centroid by CURE for triangular membership function ,so that they can range the fuzzy membership method correctly and also reduces the computational time.

Agrawal and his co-worker carried out some mining algorithm based on the large data sets, which also find association mining rule [10],[15][17]. This papers first defined the minimum threshold and minimum support for finding the frequent item sets. The fuzzy association rule is obtained by calculating the large item sets and candidates item sets.

machine learning and computing IPCSIT vol3 (2011) IACSIT PRESS SINGAPUR for comparison with proposed algorithm. They have done on fuzzy mining association rule to reduce the computational time. They all used the mining association rule for doing the task, the TRApriori mining association technique is used from the paper [14]E Ramaraj, K Rameshkumar, N Venkatesan"A better performed transaction Reduction algorithm for mining frequent item set from large voluminous database .

III. PROPOSED ALGORITHM

The algorithm finds the association between the linguistic variables without finding the fuzzy values of data sets. In place of fuzzy values, the positive and Negative edge values are calculated. The positive and negative values are obtained by triangular membership function values.

STEP-1: Take the input data set

STEP-2 Categorize each attributes as Low, Middle and High as fuzzy set categories it as linguistic variable.
STEP-3: Find out the range of low, middle and high linguistic variable .
STEP-4: Build Triangular Membership function according to the step 3.
STEP-5: find out the positive and negative fuzzy edge values. It is calculated by triangular membership functions. If it is giving more than .5 membership values than it is consider as positive edge. If it is less than that then it is fall under the negative edge values.
STEP-6: Find all the values of positive edge and negative edge values of all the linguistic variables (Low, Middle, High). The value of bothn the positive and negative edge values are 1. Negative edge value is indicated the left hand side and the positive edge value is indicated the right hand side.

STEP-7: Count the frequency of each item set(Frequency is the no of appearance of data set.) superlatively count the no of positive and negative edge count values.
STEP:8 Set the Minimum Support value.
STEP-09: find out the candidate item sets C1
STEP-10: Find L1 item sets
STEP-11: Generate the C2 candidate item sets
STEP-12: Repeat the step 6 to 7
STEP-13: Find the out the probability of each positive and negative edge.
STEP-14: Find the Association rule between linguistic variable.

IV. DATA ANALYSIS

For data analysis, the data is taken from KEEL Open data marketing repository. The data sets are mentioned are as follows:-

STEP-1: Working data sets

Income	Occupation	Household	Education	Marital Status
5	5	5	7	9
5	5	5	7	9
3	1	5	7	9
1	6	5	7	1
1	6	3	7	1
6	8	5	7	8
2	9	4	1	8
3	3	5	7	6
6	8	5	7	2
7	8	4	7	4
3	9	5	7	1
2	2	5	5	4

Table 1.1 KEEL Input Data set

STEP-2: Categorize each attributes as Low, Middle and High as fuzzy set categories it as linguistic variable.

STEP-3 STEP-3: Find out the range of low, middle and high linguistic variable .

STEP-4: Build Triangular Membership function according to the step 3.

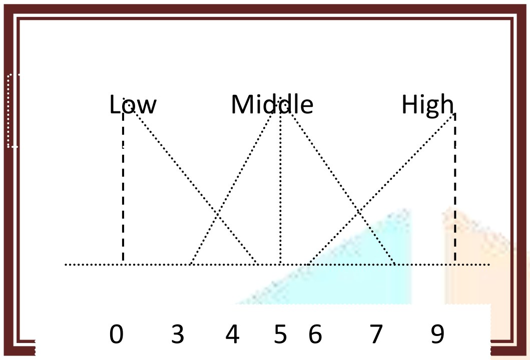

Fig 1.1 Membership Function

Low=0 Middle or mean= 5 and high=9

Range of all the Low Linguistic variables=0, 2, 4

Range of all the Middle Linguistic variables = 3, 5, 7

Range of all the High Linguistic variables =6, 7, 9

STEP-5: find out the positive and negative fuzzy edge values. It is calculated by triangular membership functions. If it is giving more than .5 membership values than it is consider as positive edge. If it is less than that then it is fall under the negative edge values.

STEP-6: : Find all the values of positive edge and negative edge values of all the linguistic variables (Low, Middle, High). The value of bothn the positive and negative edge values are 1. Negative edge value is indicated the left hand side and the positive edge value is indicated the right hand side.

SN	INC			OCT			HH			ED			MS		
	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
1	0|0	0|1	0|0	0|0	0|1	0|0	0|0	0|1	0|0	0|0	0|0	1|0	0|0	0|0	0|1
2	0|0	0|0	0|0	0|0	0|1	0|0	0|0	0|1	0|0	0|0	0|0	1|0	0|0	0|0	0|1
3	1|0	0|0	0|0	0|1	0|0	0|0	0|0	0|1	0|0	0|0	0|0	1|0	0|1	0|0	0|1
4	0|1	0|0	0|0	0|0	1|0	0|0	0|0	0|1	0|0	0|0	0|0	1|0	0|1	0|0	0|0
5	0|1	0|0	0|0	0|0	0|1	0|1	1|0	0|0	0|0	0|0	0|0	1|0	0|1	0|0	0|0
6	0|0	1|0	0|0	0|0	0|0	0|0	0|0	1|0	0|0	0|0	0|0	1|0	0|0	1|0	0|1
7	0|1	0|0	0|0	0|0	0|0	0|0	0|0	0|1	0|0	0|1	0|0	0|0	0|0	0|0	0|1
8	1|0	0|0	1|0	1|0	0|0	0|0	0|0	0|1	0|0	0|0	0|0	1|0	0|0	1|0	1|0
9	0|0	1|0	0|0	0|0	0|0	0|1	0|0	1|0	0|0	0|0	0|0	1|0	0|1	0|0	0|0
10	0|0	1|0	0|0	0|0	0|0	0|1	0|0	1|0	0|0	0|0	0|0	1|0	1|0	0|1	0|0
11	1|0	0|0	0|0	0|0	0|0	0|1	0|0	0|1	0|0	0|0	0|0	1|0	0|1	0|0	0|0
12	0|1	0|0	0|0	0|1	0|0	0|0	0|0	0|1	0|0	0|0	0|1	0|0	1|0	0|1	0|0

Table 1.2 Positive and Negative edge Values

STEP-7: Count the frequency of each item set (Frequency is the no of appearance of data set.) i.e negative or positive edge.

SN	INC			OCT			HH			ED			MS		
	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
1	0|0	0|1	0|0	0|0	0|1	0|0	0|0	0|1	0|0	0|0	0|0	1|0	0|0	0|0	0|1
2	0|0	0|0	0|0	0|0	0|1	0|0	0|0	0|1	0|0	0|0	0|0	1|0	0|0	0|0	0|1
3	1|0	0|0	0|0	0|1	0|0	0|0	0|0	0|1	0|0	0|0	0|0	1|0	0|1	0|0	0|1
4	0|1	0|0	0|0	0|0	1|0	0|0	0|0	0|1	0|0	0|0	0|0	1|0	0|1	0|0	0|0
5	0|1	0|0	0|0	0|0	0|1	0|1	1|0	0|0	0|0	0|0	0|0	1|0	0|1	0|0	0|0
6	0|0	1|0	0|0	0|0	0|0	0|0	0|0	1|0	0|0	0|0	0|0	1|0	0|0	1|0	0|1
7	0|1	0|0	0|0	0|0	0|0	0|0	0|0	0|1	0|0	0|1	0|0	0|0	0|0	0|0	0|1
8	1|0	0|0	1|0	1|0	0|0	0|0	0|0	0|1	0|0	0|0	0|0	1|0	0|0	1|0	1|0
9	0|0	1|0	0|0	0|0	0|0	0|1	0|0	1|0	0|0	0|0	0|0	1|0	0|1	0|0	0|0
10	0|0	1|0	0|0	0|0	0|0	0|1	0|0	1|0	0|0	0|0	0|0	1|0	1|0	0|0	0|0
11	1|0	0|0	0|0	0|0	0|0	0|1	0|0	0|1	0|0	0|0	0|0	1|0	0|1	0|0	0|0
12	0|1	0|0	0|0	0|1	0|0	0|0	0|0	0|1	0|0	0|0	0|1	0|0	1|0	0|0	0|0
Count	7	4	1	3	4	5	1	11	0	1	0	10	2	1	5
+ve	P =	P =	P =	P =	P =	P =	P =	P =	P =	P =	P =	P =	P =	P =	P =

-ve Count	4 N=3	1 N=3	0 N=1	2 N=1	3 N=1	5 N=0	0 N=1	9 N=2	0 N=0	1 N=0	0 N=0	0 N=10	2 N=0	0 N=1	5 N=0
Total	Positive Count= 32 and Negative Count=23														

Table 1.3 C1Positive and Negative values Count

STEP-8: Set the Minimum Support value as 5.

STEP-9: find out the candidate item sets C1

S.N	INC	HH	ED	MS
	L	M	H	H
1	0|0	0|1	1|0	0|1
2	0|0	0|1	1|0	0|1
3	1|0	0|1	1|0	0|1
4	0|1	0|1	1|0	0|0
5	0|1	0|0	1|0	0|0
6	0|0	1|0	1|0	0|1
7	0|1	0|1	0|0	0|1
8	1|0	0|1	1|0	0|0
9	0|0	1|0	1|0	0|0
10	0|0	1|0	1|0	0|0
11	1|0	0|1	1|0	0|0
12	0|1	0|1	0|0	0|0
Count	7	11	10	6
+ve -Ve Count	P=4 N=3	P=9 N=2	P=0 N=10	P=5 N=0

Table 1.4 C1 item sets

STEP: 10 select those item sets which has having the minimum support with having the minimum sum values. The data sets those having the minimum sum has the maximum fuzzy values.

STEP-11: Select the L1 item sets from C1

The large item sets L1 {INC-L,7} {HH-M,11}{ED-H,10},{MS-H,5 }

STEP-12: Generate the C2 candidate item sets

SN	INC	HH	Count
	L	M	
1	0|0	0|1	0|0
2	0|0	0|1	0|0
3	1|0	0|1	1|0
4	0|1	0|1	0|1
5	0|1	0|0	0|0
6	0|0	1|0	0|1
7	0|1	0|1	0|1
8	1|0	0|1	1|0
9	0|0	1|0	0|0
10	0|0	1|0	0|0
11	1|0	0|1	1|0
12	0|1	0|1	1|0
Count	7	11	7
Positive And Negative Count	P=4 N=3	P=9 N=2	P=3 N=4

Table 1.5 C2 item sets

S.N	C2 Candidate Item Sets	Counts
1	INC-L,HH-M	7 (P-3,N-4)
2	INC-L ,ED-H	5 (P=0,N=5)
3	INC-L,MS-H	2(P=1,N=1)
4	HH-M,ED-H	9(P=0,N=9)
5	ED-H,MS-H	5(P=0,N=5)

Table 1.6 Final C2 item sets

The item set which has having the minimum sum value has the maximum fuzzy values.

STEP-13 Find the l2 item Sets

(INC-L,HH-M,7),(INC-L ,ED-H,5),(HH-M,ED-H,9),(ED-H,MS-H,5)}

STEP:14 Find the C3 Item Sets

	INC	HH	ED	Count
	L	M	H	
1	0|0	0|1	1|0	0|0
2	0|0	0|1	1|0	0|0
3	1|0	0|1	1|0	1|0
4	0|1	0|1	1|0	1|0
5	0|1	0|0	1|0	0|0
6	0|0	1|0	1|0	0|0
7	0|1	0|1	0|0	0|0
8	1|0	0|1	1|0	1|0
9	0|0	1|0	1|0	0|0
10	0|0	1|0	1|0	0|0
11	1|0	0|1	1|0	1|0
12	0|1	0|1	0|0	0|0
Count	7	11	10	5
Positive And Negative Count	P=4 N=3	P=9 N=2	P=0 N=10	P=0 N=4

Table 1.7 C3 item sets

S.N	C2 Candidate Item Sets	Counts
1	INC-L,HH-M,ED-H	4 (P-0,N-4)
2	INC-L,HH-M,MS-H	4 (P-0,N-4)

Table 1.8 Final C3 item sets

STEP-14: Find the association rule.

S.N	C2 Candidate Item Sets	Counts	Probability
1	INC-L,HH-M	7 (P-4,N-4)	7 /12=58
2	INC-L ,ED-H	5 (P=0,N=7)	5 /12=44
4	HH-M,ED-H	9(P=1,N=9)	9 /12=75

5	ED-H,MS-H	5(P=5,N=1)	5 /12=44

Table 1.9 Final Association Rule

In Plain English, it means if House Hold is middle then Education is high with confidence or probabilty of 75%

V.CONCLUSIONS

The association between the linguistic variables is calculated in a easy way with the help of positive and negative fuzzy edges by using the probability techniques.

It reduces the complexity associated with the candidate item sets and large item sets of fuzzy values.

It also indicates the association amongst the variable is under positive edge or in a negative edge.

If number of positive edge is more than the number of negative edge than we may increase the support value .

REFERENCES

.[1] Ruchi Bhargava, Prof. Shrikant Lade, Prof. Daya Shnakar Panday "Mining Positive and Negative Fuzzy Association Rules with Item Cost", International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN: 2347-5552, Volume-2, Issue-4, July- 2014

[2 Thanh Minh Nguyen and Q. M. Jonathan Wu "**Fuzzy System with Positive and Negative Rules**", *Department of Electrical and Computer Engineering, University of Windsor Canada*

[3] Sajid Mahmood,[1,2] Muhammad Shahbaz,[1] and Aziz Guergachi "**Negative and Positive Association Rules Mining from Text Using Frequent and Infrequent Itemsets"** The Scientific World Journal Volume 2014 (2014), Article ID 973750, 11 pages http://dx.doi.org/10.1155/2014/973750

[4] Mohammed Al-Maolegi,Bassamarkok "An Improved Apriorialgorithm For Association Rules",International Journal On Natural Language Computing Volume 3 , No-1 , Febuary 2014.

[5] K Sathesh Kumar and M Hemlatha "An innovative potential on rule optimization using fuzzy artificial bee colony" Research journal of applied science , engineering and Technology.. volume 2627-2633 , ISSN:2040-7459, E-issn:2040-7467, April 05 2014.

[6]zhiyong MA , QingyingQiu and PeieenFeng "CPM algorithm for mining associaotion rules from databases of engineering design+n instances." Journal of information science and engineering vol30 , 463-481 (2014)

[7]Arpnashrivastava , R . C Jain and A K shrivatava "Generating 3rd level association rule using fast aprioriimplementation."British journal of mathematics and computer science. Volume 4(2):241-251, 2014

[8]Neelukhare ,neeruadlakha, K R pardasani "An algorithm for mining multidimentional fuzzy association rule" international journal of computer science and information security vol 5, no-1 ,2009

[9]Thomas sudkomp "Refinement of temporal constraints in fuzzy association " volume 48(2008) 14 June 2007

[10]Tzung-Pei Honga,Chan-Sheng Kuob,Shyue-Liang Wangc "A fuzzy AprioriTid mining algorithm with reduced computational time" Volume 5, Issue 1, December 2004, Pages 1–1

[11]Tzung-pei-Hang, Guo-Cheng lan, Yi-Hsinlin and Shing-tai pan. An effective gradual data reduction strategy for fuzzy item set mining international journal of fuzzy systems ,vol 15 no2 June 2013.

[12]Usha Rani, R VijayaPrakkash, Dr A. Gavardhan,Mining Multi-level association Rules using fuzzy, International journal of emerging technology and advanced engineering. Volume 3, issue 8, August 2013.

[13]JrShianchen hung liehchou, ChingHsue Cheng, jen-ya Wang CPDA Based Fuzzy association rules for learning achievement mining, 2009 international conference on machine learning and computing IPCSIT vol3 (2011) IACSIT PRESS SINGAPUR.

[14]E Ramaraj, K Rameshkumar, N Venkatesan"A better performed transaction Reduction algorithm for mining frequent item set from large voluminous database"Volume 5, Issue 1, December 2004, Pages 1–10

[15] C. H. Cai, W. C. Fu, C. H. Cheng and W. W. Kwong, "Mining association rules with weighted items," The International Database Engineering and Applications Symposium, 1998, pp. 68-77.

[16] T. P. Hong, C. S. Kuo, and S. C. Chi, "Mining association rules from quantitative data," Intelligent Data Analysis, Vol. 3, No. 5, 1999, pp. 363-376. A. Kandel, Fuzzy Expert Systems, CRC Press, Boca Raton, 1992, pp.8-19. Submitted manuscript 26

[17] S. Yue, E. Tsang, D. Yeung and D. Shi, "Mining fuzzy association rules with weighted items," The IEEE International Conference on Systems, Man and Cybernetics, 2000, pp. 1906-1911.

[18]Mehmet Kaya[1,*], Reda Alhajj[1,**], Faruk Polat[2], and Ahmet Arslan[3] "Efficient Automated Mining**ofFuzzyAssociation Rules** "DEXA 2002, LNCS 2453, pp. 133-142, 2002.