iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 Issue: VI Month of publication: June 2016

DOI:

Volume 4 Issue VI, June 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Positive and Negative Approach for Association Rule in Data Mining Techniques

Devashish Tamrakar[1],Rohit Miri[2],Asha Miri[3] , 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*
[3]*Department of Computer Science & Engineering*
Govt. Polytechnic College, Bilaspur (C.G)

Abstract—when we talk about the fuzzy association rule then its takes lots of calculation or it required software tools to calculate the complex result. The replacement of fuzzy association rule is the Novel approach by Positive and negative factor of data sets. Here the range of linguistic variable is calculated by finding a standard deviation and mean. The positive and negative factor of data set is calculated. In this Approach we do not required to calculate fuzzy values of data sets. It means that no need to calculate the membership function.

Keywords—Positive Edge, Negative Edge, Data mining

I. INTRODUCTION

It is a complex task to calculate the fuzzy association rule. It requires building the membership function values and then each membership values of item sets are calculated by using the software tools. Later steps the fuzzy association concept is applied. In this novel approach no need to build the membership function and also no need to calculate the membership values. We will find the range of linguistic variables by using standard deviation and mean. The positive and negative difference edge is calculated. In this approach we also able to calculate the majority of data sets or fuzzy values. In a triangular membership function the triangle is divided into two parts. The left part of the triangle and the right part of the triangle. In this approach the left part of the triangle indicate the negative edge and the right side of the triangle indicates the positive edge factor.In place of fuzzy values, the positive and Negative edge values are calculated. The positive and negative values are obtained by subtracting highest values of triangular membership function.

II. LITERATURE REVIEW

Different approaches have been used by the researchers to reduce the time complexity of fuzzy association rule.

The item sets that are frequently present in particular transaction id are chooses by Mohammed Al-Maoleg1[1] . Its algorithm works on low support and reduces the time complexity of the program very easily. Zhiyong ma [3] et al converts all the item sets into Boolean matrix by using CP tree method and reduces the time for the task.ArpnaShrivastava[4]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.

Neelukhare et al [5] has used algorithm for mining multidimensional fuzzy association rule mining. Tzung-Pei Hong et al [8]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[6]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 [9] implement a model of mining multilevel association rules .It use top down approach and fuzzy boundary value in place of sharp boundary intervals.

Another approach reduces the operational time carried out by Apriori algorithm by using artificial Bee colony optimization method (FABCO) is given by K. Sathesh Kumar and M. Hemalatha[2].Mehmet Kaya et al [15]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 [7]. These algorithms break the mining steps into two phases. In the first phase candidate of item sets are obtained and counted by scanning the transactions. The number of item set must support the minimum pre-defined threshold value called minimum support.

Volume 4 Issue VI, June 2016
ISSN: 2321-9653

Then later we make the pair of item sets and apply the association rule for getting the required output. Srikant and Agrawal also proposed partitioned based mining association algorithm. Cai at al proposed weighted mining rule of data sets [12]. Yue et al, extended the fuzzy concept based on vectors [14].Most of them are find out the range of triangular fuzzy membership function directly , means they assumed the range of linguistic variable. But on my paper We have find out the range of linguistic variable by using mean and standard deviation.

Mining association rule was also performed by the [13]T. P. Hong, C. S. Kuo, and S. C. Chi, "Mining association rules from quantitative data," Intelligent Data Analysis. The aim of his research is digging out the essential or useful item from very large data set. Authors want to improve the data mining algorithm. We used [7]Tzung-Pei Honga,Chan-Sheng Kuob,Shyue-Liang Wangc "A fuzzy AprioriTid mining algorithm with reduced computational time and [10]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 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 [11]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 subtracting highest values of triangular membership function.

STEP-1: Take a Data Sets
STEP-2: Find out the mean and Standard deviation
STEP-3 Categorize each attributes as Low, Middle and High as fuzzy set categories it as linguistic variable.
STEP-4: Find out the range of low, middle and high linguistic variable by using standard deviation and mean.
STEP-5: Again categories the linguistic variable as its positive edge and negative edge. According to their range. Output is positive then it goes to positive edge and if output is negative then it goes to negative edge.
STEP-6: Find all the values of positive edge and negative edge of linguistic variables (Low, Middle, High)
STEP-7: Count the frequency of each item set(Frequency is the no of appearance of data set.) i.e negative + positive edge.
STEP-8: Find the sum of all the linguistic variables.
STEP:9 Set the Minimum Support for data sets
STEP-10: find out the candidate item sets C1
STEP-11: Select the L1 item sets from C1
STEP-12: Generate the C2 candidate item sets
STEP-13: Repeat the step 6 to 7
STEP-14: Find the association rule.

IV. DATA ANALYSIS

For data analysis, the data is taken from Final Year computer science and engineering students. The marks of five subjects have been taken for the data analysis.

STEP-1: Take a Data Sets

Student No	AI	ERP	NN	CC	DM
1	88	65	65	73	71
2	61	79	81	77	61
3	61	89	86	79	89
4	73	86	79	84	62
5	70	81	87	72	79
6	81	77	86	87	87
7	67	87	75	71	80

8	66	63	84	84	86
9	75	85	79	61	61
10	74	83	85	85	90

STEP-2: Find out the mean and Standard deviation

Find the mean and standard deviation for the above mentioned whole data set that will be useful for finding the range of linguistic variables. i. e. Low, middle and High

Mean=77 and standard Deviation is 9

STEP-3 Categorize each attributes as Low, Middle and High as fuzzy set categories it as linguistic variable. i.e AI-L, AI-M and AI-H

STEP-4: Find out the range of low, middle and high linguistic variable by using standard deviation and mean.

Low=61 Middle or mean=77 and high=90

Range of all the Low Linguistic variables=52, 61, 72

Range of all the Middle Linguistic variables = 68, 77, 87

Range of all the High Linguistic variables =80, 90, 99

STEP-5: Again categories the linguistic variable as its positive edge and negative edge. Output is positive then it goes to positive edge and if output is negative then it goes to negative edge.

STEP-6: Find all the values of positive edge and negative edge of linguistic variables (Low, Middle, High)

SN	AI			ERP			NN			CC			DM		
	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
1	0|0	0|0	2|0	0|5	0|4	0|0	0|4	0|0	0|0	0|0	4|0	0|0	0|10	6|0	0|0
2	0|0*	0|0	0|0	0|0	0|2	0|0	0|0	0|4	9|0	0|0	0|0*	0|0	0|0*	0|0	0|0
3	0|0*	0|0	0|0	0|0	0|0	1|0	0|0	0|9	4|0	0|0	0|2	0|0	0|0	0|0	1|0
4	0|0	4|0	0|0	0|0	0|9	3|0	0|0	0|2	0|0	0|0	0|7	6|0	0|1	0|0	0|0
5	0|9	7|0	0|0	0|0	0|4	8|0	0|0	0|10	2|0	0|11	5|0	0|0	0|0	0|2	0|0
6	0|0	0|4	8|0	0|0	0|0*	0|0	0|0	0|9	4|0	0|10	0|10	3|0	0|0	0|10	2|0
7	0|6	0|0	0|0	0|0	0|10	2|0	0|0	2|0	0|0	0|0	6|0	0|0	0|0	0|3	10|0
8	0|5	0|0	0|0	0|2	0|0	0|0	0|0	0|7	6|0	0|0*	0|7	6|0	0|0	0|9	4|0
9	0|0	2|0	0|0	0|0	0|8	4|0	0|0	0|2	0|0	0|0	0|0	0|0	0|0*	0|0	0|1
10	0|0	3|0	0|0	0|0	0|4	6|0	0|0	0|8	4|0	0|20	0|8	5|0	0|0	0|0	0|0*

The negative edge and positive edge are denoted by the A|B respectively. The less the number of positive and negative edge has more the value of fuzzy. It means that the value positive and negative is inversely proportion to fuzzy values. The positive and negative values are obtained from subtracting the range of linguistic variable that holds the fuzzy value one from the triangular membership function. The count value is the no of frequency or support value for the data sets.

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	AI			ERP			NN			CC			DM		
	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
1	0|0	0|0	2|0	0|5	0|4	0|0	0|4	0|0	0|0	0|0	4|0	0|0	0|10	6|0	0|0
2	0|0*	0|0	0|0	0|0	0|2	0|0	0|0	0|4	9|0	0|0	0|0*	0|0	0|0*	0|0	0|0
3	0|0*	0|0	0|0	0|0	0|0	1|0	0|0	0|9	4|0	0|0	0|2	0|0	0|0	0|0	1|0
4	0|0	4|0	0|0	0|0	0|9	3|0	0|0	0|2	0|0	0|0	0|7	6|0	0|1	0|0	0|0
5	0|9	7|0	0|0	0|0	0|4	8|0	0|0	0|10	2|0	0|11	5|0	0|0	0|0	0|2	0|0
6	0|0	0|4	8|0	0|0	0|0*	0|0	0|0	0|9	4|0	0|10	0|10	3|0	0|0	0|10	2|0

7	0|6	0|0	0|0	0|0	0|10	2|0	0|0	2|0	0|0	0|0	6|0	0|0	0|0	0|3	10|0
8	0|5	0|0	0|0	0|2	0|0	0|0	0|0	0|7	6|0	0|0*	0|7	6|0	0|0	0|9	4|0
9	0|0	2|0	0|0	0|0	0|8	4|0	0|0	0|2	0|0	0|0	0|0	0|0	0|0*	0|0	0|1
10	0|0	3|0	0|0	0|0	0|4	6|0	0|0	0|8	4|0	0|20	0|8	5|0	0|0	0|0	0|0*
Count	5	5	2	2	8	6	1	8	6	4	9	4	4	5	6

STEP-8: Find the sum of all the linguistic variables.

SN	AI			ERP			NN			CC			DM		
	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
1	0|0	0|0	2|0	0|5	0|4	0|0	0|4	0|0	0|0	0|0	4|0	0|0	0|10	6|0	0|0
2	0|0*	0|0	0|0	0|0	0|2	0|0	0|0	0|4	9|0	0|0	0|0*	0|0	0|0*	0|0	0|0
3	0|0*	0|0	0|0	0|0	0|0	1|0	0|0	0|9	4|0	0|0	0|2	0|0	0|0	0|0	1|0
4	0|0	4|0	0|0	0|0	0|9	3|0	0|0	0|2	0|0	0|0	0|7	6|0	0|1	0|0	0|0
5	0|9	7|0	0|0	0|0	0|4	8|0	0|0	0|10	2|0	0|11	5|0	0|0	0|0	0|2	0|0
6	0|0	0|4	8|0	0|0	0|0*	0|0	0|0	0|9	4|0	0|10	0|10	3|0	0|0	0|10	2|0
7	0|6	0|0	0|0	0|0	0|10	2|0	0|0	2|0	0|0	0|0	6|0	0|0	0|0	0|3	10|0
8	0|5	0|0	0|0	0|2	0|0	0|0	0|0	0|7	6|0	0|0*	0|7	6|0	0|0	0|9	4|0
9	0|0	2|0	0|0	0|0	0|8	4|0	0|0	0|2	0|0	0|0	0|0	0|0	0|0*	0|0	0|1
10	0|0	3|0	0|0	0|0	0|4	6|0	0|0	0|8	4|0	0|20	0|8	5|0	0|0	0|0	0|0*
Count	5	5	4	2	8	6	1	8	6	4	9	4	4	5	6
Sum	20	20	10	6	41	26	5	53	29	41	49	20	11	40	18

STEP-9: Define the minimum support.

Minimum support=6

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.

The candidate C1 item sets { ERP-M(8/41),NN-M(8/53),CC-M(9/49),ERP-H(6/28),DM-H(6/28)}

STEP-11: Select the L1 item sets from C1

The item sets ERP-M(7/42) and ERP-H(7/36) . Both the item sets are having the minimum support of 7 , so the DB-H is selected in C1 item sets and ERP-M is ignored because if both the item sets are having the same support then we will select which has minimum sum values. The fuzzy value of minimum sum values is highest.

The candidate L1 item sets { ERP-M(8/41),NN-M(8/53),CC-M(9/49),DM-H(6/18)}

STEP-12: Generate the C2 candidate item sets

S.N	ERP-M(1)	NN-M(2)	CC-M(3)	DM(4)	12	13	14	23	24	34
1	4	0	4	0	0	4	0	0	0	0
2	2	4	0*	0	4	2	0	0	8	0
3	0	9	2	1	0	0	0	9	2	2
4	9	2	7	0	9	9	0	0	0	0
5	4	10	5	0	10	5	0	0	9	0
6	0*	9	10	2	9	10	2	9	2	10
7	10	2	6	10	10	10	10	10	0	10
8	0	7	7	4	0	0	0	7	9	7
9	8	2	0	1	8	8	8	2	0	0
10	4	8	8	0*	8	8	4	8	0	8
SUM					7/58	8/56	4/24	6/45	5/30	5/37

The item set which has having the minimum sum value has the maximum fuzzy values. From the above C2 candidate item sets the combination 12 has the lowest summation, it means having with the highest fuzzy values.

STEP-13: Repeat the step 6 to 7

STEP-14: Find the association rule.

There is association between 12, ERP-M and NN-M. If we find the probability of both the combination then the combination 12 has highest probability so the combination ERP-M and NN-M has more confidence.

V. CONCLUSIONS

The proposed algorithm finds the fuzzy association rule of data sets without building the membership function and also without calculating the fuzzy values.

It is easy way to find the fuzzy association rule.

The output is similar to output of fuzzy association rule.

REFERENCES

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

[2] 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.

[3] 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)

[4] 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

[5] 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

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

[7] 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

[8] 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.

[9] 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.

[10] 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.

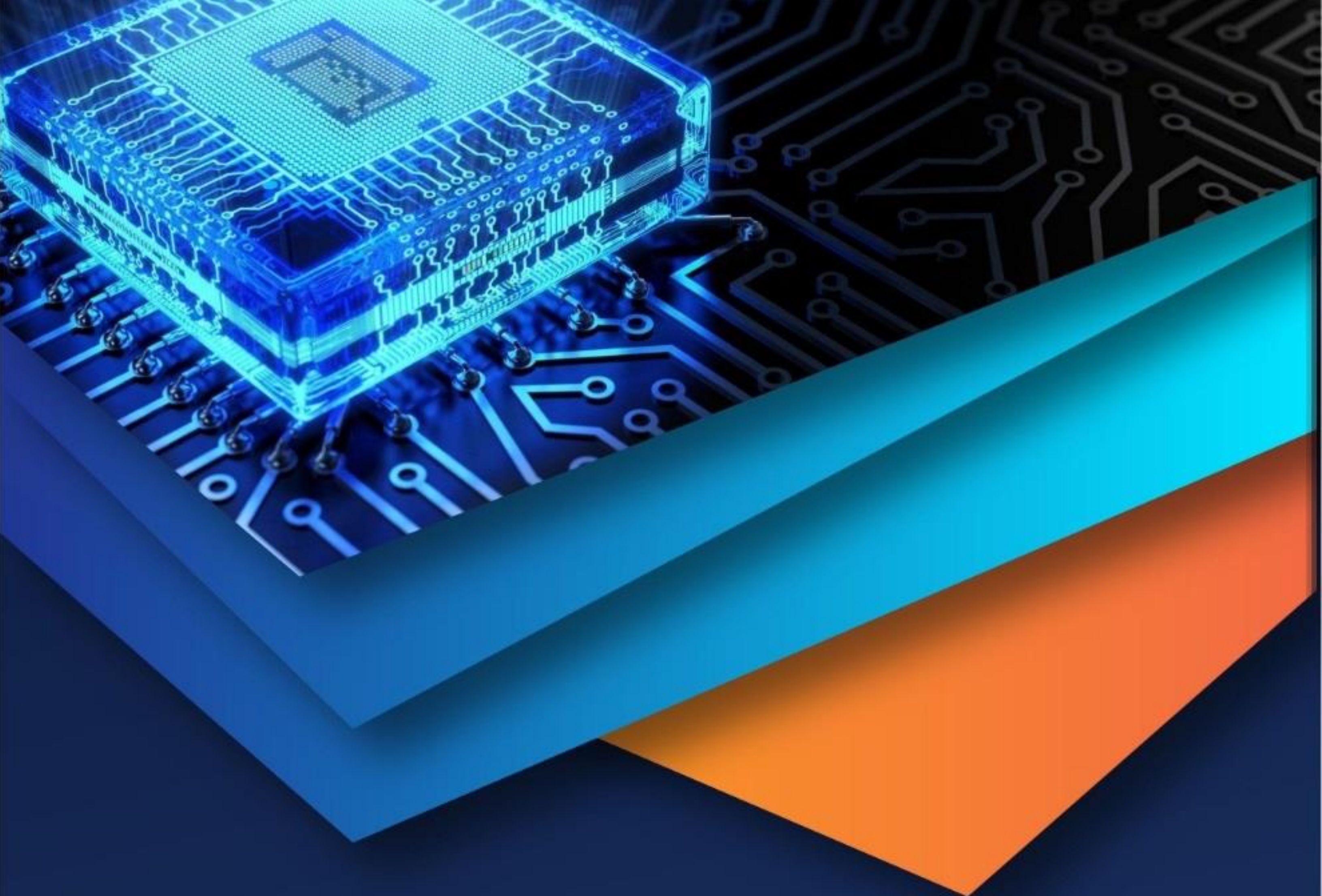
[11] E Ramaraj, K Rameshkumar, N Venkatesan"A better performed transaction Reduction algorithm for mining frequent item set from large voluminous database"Volume 5, Issue 1, December 2004, Pages 1–10

[12] 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.

[13] 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

[14] 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.

[15] Mehmet Kaya1,*, Reda Alhajj1,**, Faruk Polat2, and Ahmet Arslan3 "Efficient Automated MiningofFuzzyAssociation Rules "DEXA 2002, LNCS 2453, pp. 133-142, 2002.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY