International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 11 (2015) pp. 29367-29383
© Research India Publications
http://www.ripublication.com

Efficient Association Rule Mining Based On Correlation Analysis

Dr.M.L.Valarmathi[1], Siji P D [2] and S.Mohana [3]

[1]Associate Professor Department of CSE,
Government College of Technology, Coimbatore, Tamil Nadu, India,
[2] Assistant Professor, Department of Computer science,
St. Josephs College Irinjalakuda, Thrissur, Kerala, India
[3] PG Research Scholar, Department of CSE,
Government College of Technology, Coimbatore, India
[1]drmlv@gct.ac.in,[2]srblessy@gmail.com
[3]mohana.apr7@gmail.com

Abstract

This paper makes a model of a fuzzy association rule mining based on correlation analysis. The existing algorithm (the Fuzzy C Means Multiple Support Apriori - FCM–MS- Apriori model) integrates adjustable membership and correlation analysis approach and will be given the modified Algorithm. This proposed algorithm uses to extract meaningful information for the road safety accident data. First fuzzy association rule are mined using FCM Apriori algorithm with multiple membership function. The obtained rules are further reduced by finding correlation using the predefined threshold measures. The result gave more efficient rules with better accuracy

Keywords— Fuzzy association rules, Clustering, FCM-Apriori algorithms, Multiple Membership, Multiple Support

I.INTRODUCTION

Association rule mining has been a accepted area in data mining (DM) research, more and more attracting the attention of researchers.[1] [2][3][4][5] are important works in this area. Association rules discovery presented in [6] intends to extract the characteristics, hidden association patterns and the correlation between the items (attributes) in a large database [7],[8]. The Apriori algorithm developed by [9] is a classic and popular algorithm for strong association rules (knowledge) extraction from a transaction database with high frequent itemsets using the pre-defined threshold measures. These thresholds are minimum support (minsupp) and minimum confidence (minconf). Association rules are formally written and presented in the form of ‘‘IF–Then’’ as follows: $X \rightarrow Y$, where X is called the antecedent and Y is called the consequence. Let $I = \{i_1, i_2, . . . , i_n\}$ be a set of distinct items (attributes). A collection of one or more items, i.e., any set of items is called an item-set. Let $D = \{t_1, t_2, . . . , t_m\}$ be a set of transaction IDs (TIDs). Each TID in D is formed from a set of items in I. The support count is the occurrence (frequency) of X and Y together, support (XUY), and the support value is the fraction of transactions that contains both X and Y.[34]

An item set whose support is greater than or equal to a minsupp threshold is called a frequent item set. The confidence value measures how often items in Y appear in transactions that contain X and is the ratio

Support(XUY)/Support(X) An association rule is an implication expression of the form (: $X \rightarrow Y$), where $X, Y \in I$ and $X \cap Y = \phi$ A strong association rule is that which has support and confidence greater than the user defined minsupp and minconf. The main task of the association rule discovery is to find all strong rules.

One of the advantages of association rule discovery is that it extracts explicit rules that are of practical importance for the user/ human expert to understand the application domain. Therefore this can be facilitated to adjust (extend) the rules manually with further domain knowledge, which is difficult to achieve with other mining approaches [10]. On paper [11] introduced the problem of extracting association rules from quantitative attributes by using the partitions method for these attributes. Some of the current association rule mining approaches for quantitative data neglected the values of the interval boundaries of the partitions. This causes sharpness of the boundary intervals which does not reflect the nature of human perception, justifiably argued by [12] 13]. Instead of using partition methods for the attributes, it is better to adopt the advantage of fuzzy set theory with a smooth transition between fuzzy sets. As a whole, the fuzzy approach is used for transforming quantitative data into fuzzy data. A variety of approaches has been developed in order to extract fuzzy association rules from quantitative data. sets [14]-[21].

In this paper investigates the problem of association rules extraction from quantitative data using fuzzy clustering techniques. Fuzzy clustering is a suitable method to transform quantitative data into fuzzy data, taking the advantage of fuzzy set theory over the partition method concerning the smooth transition among fuzzy sets. Fuzzy Association Rules (FARs) mining is adopted in this paper as a solution for extracting knowledge from the quantitative database.

The association rule mining aims to discover the relationships (rules) among the data attributes (features), which depend on minsupp and minconf. Consequently, large numbers of rules are anticipated, particularly if minsupp is set to be very low. Practically, a single minsupp is a vital parameter that controls the extracted number of association rules. The papers[22],[3] proposed an integrated data envelopment analysis based method to identify the most efficient association rules by ranking them using multiple criteria. Conventional association rule mining approaches like Apriori [9] and Frequent Pattern-Growth (FP-Growth) [23] are based on a single minsupp threshold. However, it was observed that using a single minsupp causes a dilemma called the ‘‘rare item problem’’ [24][23]

To resolve this rare item problem, author [8] developed a multiple support model called the Multiple Support Apriori (MS Apriori) algorithm. MS Apriori is based on the idea of setting a Mini-mum Item Support (MIS) for each item in a database, i.e., assigning multiple minsupp for different items in the database, instead of using a single minsupp for the whole database. Hence, MSaproiri is expressed as a generalization of the Apriori algorithm. Different MIS values can be assigned to assess different frequent items to facilitate the generation of frequent itemsets of rare items and prevent the production of uninteresting frequent itemsets [22] More recently, an approach has been developed to improve MS Apriori called Improved Multiple Support Apriori (IMSApriori) [8],[21].

This paper also proposes Fuzzy Association Rules (FARs) generated using Fuzzy clustering on quantitative data by adopting the multiple support approaches in order to deal with the limitations of using a single minsupp. FCM–Apriori model, is based on the integration of the Fuzzy C-Means (FCM) clustering algorithm and the Apriori approach for extracting FARs. FCM–MS Apriori model, is based on FCM and a multiple support thresholds approach.

Although the adoption of the MS idea from the classical partition case, the FCM–MSApriori model in the fuzzy case remains obstructed because it uses only one membership function without considering the price–quantity relation[25]. For example, In the Business field the implication of a pattern ‘‘Color Laser Printer with Low quantity‘‘ must be distinguished from that of another pattern ‘‘Printer Toner with Low quantity‘‘ although both patterns are assigned with a same fuzzy term. Managers may specify different definitions of Low quantity for Color Laser Printers and Printer Toner. Items with different prices result in different quantity demands; therefore, different membership functions must be dispatched to calculate their fuzzy term supports.

The rest of this paper is structured as follows. The next section presents the Existing algorithms and prediction models in the literature. Section 3 describes the proposed model with the case studies used to demonstrate the models. Experimental results of analysis are presented in Section 4. Finally, the conclusion are drawn in Section 5 with the key contribution of the research

II. RELATED WORK

The FCM–Apriori inherits benefits from FCM and Apriori and gives more flexibility to real-life applications, especially in business cases. The model acquires certain patterns in which the elements include rare items with higher profits, and excludes those that are trivial with lower profits. In recent years, classical extensions have been proposed in various applications. The paper [24] extended the idea to develop an algorithm for mining generalized association rules with multiple minimum supports. A method to address the issues of mining association rules with multiple minimum supports and maximum constraints is proposed in [26]. For fuzzy extensions, the paper [27] proposed an approach to find large-itemsets and association rules under the maximum constraint of multiple minimum supports. Subsequently, on integrated fuzzy set concepts, data mining, and multiple level taxonomy to find fuzzy association rules in a specified transaction dataset. In the line of mining sequential patterns, the concept [29] used the minimum spanning table method to find two-stage learning sequences in fuzzy sequential pattern mining. The author [30] proposed an idea on the absences of frequent fuzzy itemsets, and developed a method for mining negative fuzzy sequential patterns. Conversely, traditional fuzzy sequential pattern mining is referred to for mining positive patterns. There are three differing approaches in [31] used for the evaluation of the support, and extracted various levels of information for mining fuzzy sequential patterns. In [32] presented a multi-time-interval approach to discover fuzzy multi-time-interval sequential patterns. Several fuzzy inference systems developed by [33] for monitoring patient status; in particular, they included recursive fuzzy inference and non-recursive with sequential patterns as inputs

The author [34] proposed FCM–Apriori model extracts fuzzy rules for building a KB from a database, and our work is based on this paper. Their model utilizes the following two methods: FCM is used to transform the quantitative data set into fuzzy sets (terms). FCM is one of the fuzzy clustering algorithms based on an objective functioning method, developed by Bezdek [35] adopting the fuzzy set theory. In other words, it assigns a data object to more than one cluster. The Apriori approach is used for extracting fuzzy termsets (frequent itemsets) from a fuzzy data set based on interesting measures (minsupp and minconf). It is worth mentioning that the Apriori algorithm is adopted in order to deal with fuzzy data and therefore able to generate FARs. Throughout the rest of the paper, the term 'itemsets' corresponds to its termsets.

Fig. 1 shows the outline of the process: (i) getting from the database the data set, which is analyzed for consistency and any noisy data set will be removed, (ii) transforming the quantitative data set into fuzzy sets while using FCM and applying the Apriori approach to extract FARs (iii) then saving these rules in the KB, (iv) using a Fuzzy Inference System (FIS) to command the knowledge (rules) for a prediction and (v) testing the feasibility of the model in the case studies.

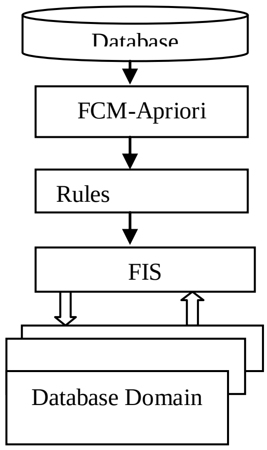


Fig. 1The FCM Apriori model

The following definitions (notations) are used in the model:

Field: Attribute (item or column) of the crisp input data.

Record (Case): Row with all fields.

Term: Fuzzy set class (fuzzy term).

x_{ij}: Value of the crisp input data.

$\mu(x)_{if}$: Fuzzy set membership value.

Sum_{jf}: Summation of each fuzzy term for all records.

Termset: A set of terms containing one term or more.

C_k: Contains candidate termsets, $1 \leq k \leq n$, where n = the maxi-mum number of fields.

L_k: Contains large termsets, $1 \leq k \leq n$, where n = maximum number of the fields.

minsupp: Minimum support threshold value (observing that minsupp = 2.45) minconf: Minimum confidence threshold value (observing that minconf = 0.4 for this value is selected based on many experiments run to find out the appropriate ones that enable us to extract useful rules.The FCM–Apriori works as follows:

Begin
FCM; {clustering data set}
Find the fuzzy sets of the quantitative data set, based on FCM
Calculate the sum of the membership values for each fuzzy term for all records, using Eq. (1)
IF sum_{jf}>minsupp Then Insert the fuzzy term intoL_1 , L_1={frequent termsets}
For (k= 2;$L_{k-1}\neq 0$; k++) **do**
C_k= generate candidate from L $_{k-1}$(join L_{k-1}called (p) with L_{k-1}called (q));
{
Insert into C_k
Select termset; From p.term1, p.term2, p.term$_{k-1,}$.. ,q. term $_{k-1}$ from p,q
Where p.term1=q.term1... p.termk-2,p.termk-2,p.termk-1≠q.termk-1
}
For each termset C €Ck **do**
Check all the sub-termsets of all termsets in L_{k-1} , and it should be a frequent termsets in
For each (k-1)subset of c **do**
IF s€ L_{k-1} Then
Delete from C_k
End
End
For each termset candidate in C_k **do**
Calculate the support value
IF $Sum_{jf} \geq$ minsuppThen
Insert the fuzzy termset into, L_k ={frequent termsets}
End
End

Select the frequent termsets including the target attribute (output attribute)
Form the frequent termsets (rules) that exist in to L_2 to L_n under the form “IF-Then”
For each rule
Calculate the confidence value for each rule
IF CV ≥ minconfThen
Accept the rule
End
Check the rules for contradiction
Insert all the accepted rules in KB
Infer the generated rules in KB using FIS
Stop

(1) FCM is used to cluster the data into terms and then to deter-mine the centre of each fuzzy set and the maximum and minimum value for each field of the input data set.
(2) The data set is converted into a fuzzy data set, using one of the standard membership functions (the triangular and trapezoid membership functions [14]
(3) A support value is calculated for each term by summing the fuzzy membership values in each term for all records using

$$Sum_{jf}=\sum_{i=1}^{n} \mu(x)\, ij \qquad \text{Eq. (1),}$$

then this summation value is stored in the candidate termset C_1.

(4) Terms which are greater than or equal to minsupp are moved to L_1.
(5) Terms are joined up and combined, as (L_1 join L_{1-}) = {{c[1], c[i]},{c[1], c[i + 1]} . . . {c[1], c[n]}}, where c[1] represents the first fuzzy term, c[2] indicating the second fuzzy term with c[n] indicating the last fuzzy term. Also c[1] ∩c[i] =Φ, c[1] ∩ c[i + 1] = ϕ, . . . c[1] ∩c[n] = ϕ (i.e., the terms for each termset do not belong to the same field).Once every termset is stored in the candidate termset C_2, the support value for each termset will be calculated using a minimum operator for the fuzzy values of the terms in the termset. Also the result of the minimum values in that termset is summed for all records. Finally, the results’ summations will be stored in the candidate termsets C_2.
(6) Termsets greater than minsupp are moved to L_2.
(7) Termsets are joined up and combined again as L_2 = p join L_2 = q, where p.term$_1$ = q.term$_1$ . . . p.term$_{k_2}$ = q.term$_{k_2}$, p.term$_{k_1}$ – q.term$_{k_1}$. This combination is based on every sub-termset of the candidate termset existing in C_k. The canididate termset should be a frequent termset in the previous large termset of L_{k_1}. Also the terms for each termset in C_k do not belong to the same field.
(8) Termsets are stored in the candidate termset C_3, then the support value is calculated for each candidate termset.
(9) Termsets and their support values in C_3 greater than or equal to minsupp are moved to L_3.

(10) Termsets are joined up and combined, until L_n is empty.

(11) Termsets are pruned by selection of the termsets including the target attribute. As a consequence, termsets are phrased in IF–Then form, then the Confidence Value (CV) is calculated based on Eq. (2). The rules that are greater than or equal to minconf are accepted. Then the contradiction rules are removed, based on the CV.

(12) $CV=\frac{\sum[(IF)\cap(Then)]}{\sum(\min(If))}$ eq(2)

2.2 The FCM–MS-Apriori Model

The use of a single minsupp for a whole database assumes that all items in the database have the same frequency. However, in real applications, the database contains some high frequency items, while others are of low frequency. The human expert, based on do-main knowledge, can set minsupp for a specific value in order to find the frequent itemsets. In that case, if minsupp is set too high it will extract a low number of frequent itemsets. Thus, the rare item problem will appear and cause a dilemma (called the rare item problem). On the other hand, if minsupp is set too low, it will extract a high number of frequent itemsets, which causes combinatorial explosions, i.e., all the possible associations will be found. Some of those frequent itemsets are uninteresting or insignificant [24]

To overcome the dilemma of the rare item problem, [24] proposed an algorithm called MSapriroi based on a multiple minimum support thresholds approach using MIS, where the number of generated rules depends on the control parameters used.

2.3 Fuzzy Correlation Analysis

support and confidence measures are insufficient at filtering out uninteresting association rules. To tackle this weakness, a correlation measure can be used to augment the support–confidence framework for association rules. Fuzzy correlation analysis illustrate the relationship between the attributes such as positive correlation, negative correlation or non-correlation. Fuzzy Correlation of each rule represents the correlated set of Sensitive Association Rule *(SAR)[33]* and on the basis of this correlation we will get the sensitive fuzzy association rule. The Pearson's correlation coefficient is often considered to evaluate how far the rules are linearly related in a sample.

The correlation coefficient ρ is defined using the formula:

$\rho = \underline{\sigma\ x,y}$

$\sigma x\ \sigma eq$ eq(3)

where σ x,y is the_support count both the items occurring together and σx is the count of X itemset and σx is the count of Y itemset. This leads to correlation rules of the form

A => B [support, confidence, correlation] , that is, a correlation rule is measured not only by its support and confidence but also by the correlation between itemsets A and B. Thus in this case, the more positive ρ is, the more positive is the association determined. This also indicates that when ρ is close to 1, an individual with a high value for one variable will likely have a high value for the other, and an individual with a lower value for one variable will likely to have a low value for the other.

On the other hand, the more negative ρ is , the more negative the association is, this also indicate that an individual with a high value for one variable will likely have a low value for the other when ρ is close to -1 and conversely. In this project road traffic accident dataset is used in which the Apriori algorithm is first applied to generate the fuzzy association rules. Then the Fuzzy correlation is done to generate more significant and meaningful rules.

III. THE PROPOSED FCM–MS-MM APRIORI MODEL (MULTIPLE MEMBERSHIP FUNCTION)

The proposed FCM–MS-MM Apriori model adopts this multiple minimum support concept [24] and multiple membership function [25] for different attributes depending upon the frequency. This model utilizes FCM and the MSapriori approach and is used for extracting FARs of rarely and highly frequent termsets from fuzzy data sets. The working is shown in Fig. 2.

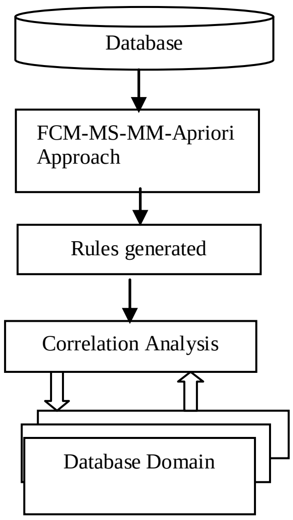


Fig 2.FCM-MS-MM apriori Model

III.1 Advantages of the proposed Algorithm

The list of advantages for new knowledge discovery model as follows:

1. The FCM-Apriori is more normal and appropriate in relation to human knowledge. The soft information or fuzzy linguistic terms which is discovered by the decision making process can be easily understand r

2. The FCM-MSApriori inherits benefits from FCM-Apriori and uses in real-life applications easily. The model acquires certain patterns in which the elements include rare items with higher priority, and excludes those that are trivial with less priority.
3. The idea of multiple membership functions is considered to offer fuzzy quantitative information, depending on the various membership functions. That is, although any two items may have the same fuzzy term, the meanings of the quantitative natures differ.
4. The fuzzy correlation analysis is applied on those obtained rules to effectively extract the strong association rules using Pearson Correlation Coefficient

This project is implemented using java Netbeans IDE 6.9.1. The Road Accident safety data is collected from the Department for Transportation (DFT) of UK- Great Britain government dataset which consists of about 2 lakhs record set with 30 attributes. These attributes are used based on the STATS19 road accident safety form of Great Britain where the form is to be filled by the respective victims and these are managed Transport administrators.

III.2 Multiple Membership Function

This study proposes a novel algorithm to discover association rules with multiple minimum supports -MMS). Although we adopt the MMS idea from the classical partition case, FCM-MS model in the fuzzy case remains obstructed because it uses only one membership function without considering the accident severity relation. For example, the implication of a pattern ‘‘Accident probability with time” must be distinguished from that of another pattern ‘‘Accident probability with peak time although both patterns are assigned with a same fuzzy term. Drivers/persons may specify different definitions of Accident probability quantity for time and peak time. Items with different time result in different aspects; therefore, to calculate their fuzzy term supports different membership functions must be dispatched.

IV. IMPLEMENTATION

Finding frequent patterns for this kind large volume of data is highly impossible or difficult to proceed since it is largely a time and memory consuming process. To handle this situation, sampling analysis is undertaken where the whole dataset is classified based on the class attribute such as accident severity. Once the sampling is done the dataset is effectively reduced to about 20000 records.

Fuzzy dataset is achieved through the conversion of numerical dataset such that all the values tend to be lie between 0 – 1. Since the fuzzy conversion is required to deal with uncertainty in the dataset. Fuzzy association rule mining approach is then applied on the sampled dataset to extract the rules using the pre-defined threshold measures. The support and confident values are calculated for each tuples in the record.

Thus the observed rules for Apriori algorithm involving the frequent itemset generation and rule generation are mined. Then the fuzzy correlation analysis is applied on those obtained rules to effectively extract the strong association rules using Pearson Correlation Coefficient.

The road safety accident data obtained after sampling process is made to get into the database as a pre-processing step of the algorithm. With this data, the further rule mining process willbe proceeded. Fig. 3 shows that Getting the Sampling data into Database.

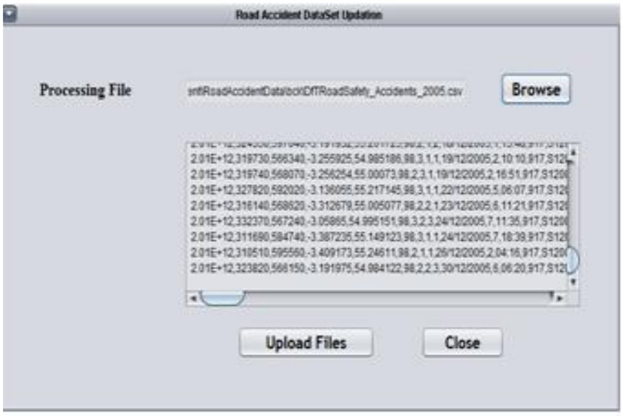


Fig. 3 Getting the Sampling data into Database

The road safety accident data are successfully uploaded to the database and view of the dataset is shown in Fig 4.

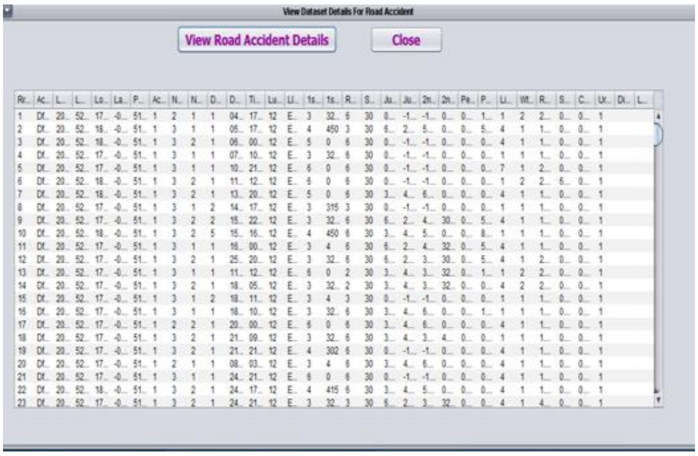


Fig. 4View of Uploaded Dataset

The different membership allotted to different attributes are given in Fig. 5

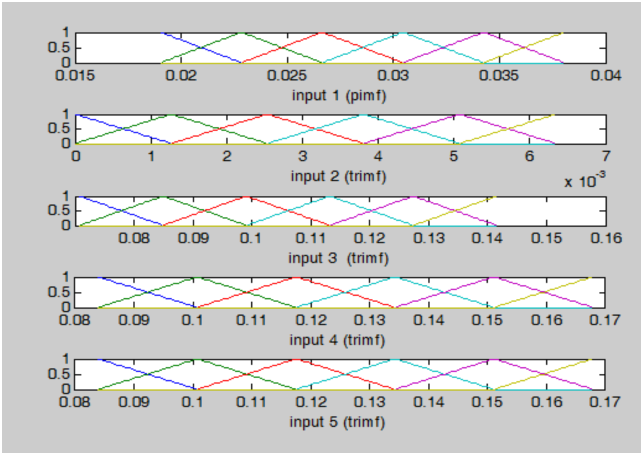


Fig. 5 multiple membership function

The dataset contains larger uncertainty and so these are first converted into numerical dataset by assigning numeric values to each tuples in the database.

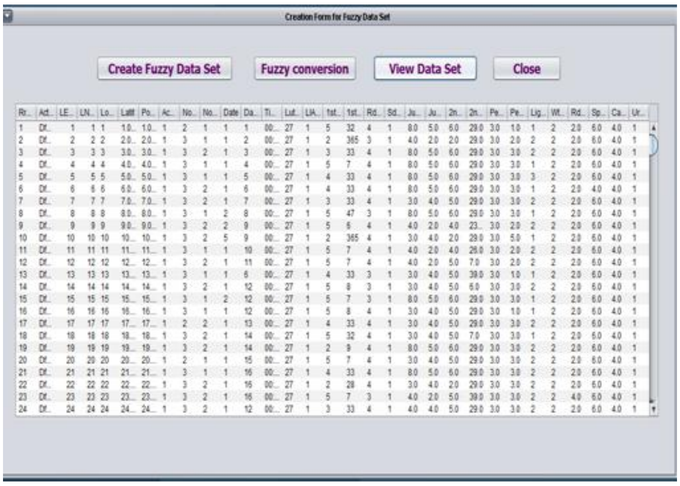


Fig. 6Fuzzy Conversion

The Fuzzy conversion is applied to the numeric dataset in order to handle the uncertainty of the data where the fuzzy values lies between 0 – 1 based on the number of occurences of each itemset in the dataset.

```
1 SELECT * FROM roadaccident.fuzzytable f;
```

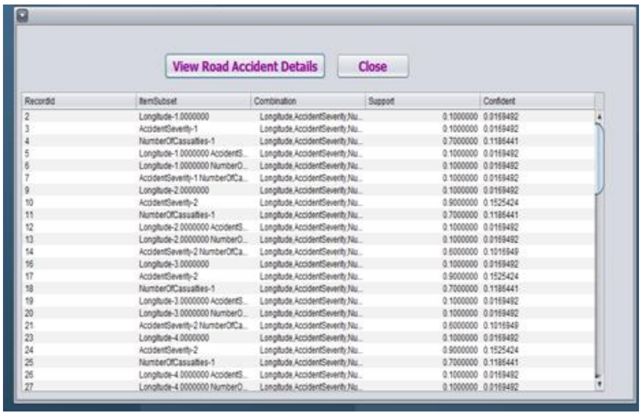
R...	File...	Ac...	Lo...	Lo...	Lo...	Lat...	P...	Ac...	Nu...	Nu...	Date	D...	Ti...	L...	L...	1...	1...	R...	Sp...	Ju...	Ju...	2...	2...	P...	P...	Li...	W...	R...	Sp...	C...	Ut...
1	res...	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
2	res...	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
3	res...	0.1	0.1	0.1	-0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
4	res...	0.1	0.1	0.1	-0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
5	res...	0.1	0.1	0.1	-0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
6	res...	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
7	res...	0.1	0.1	0.1	-0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
8	res...	0.1	0.1	0.1	-0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
9	res...	0.1	0.1	0.1	-0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
10	res...	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
11	res...	0.4	0.4	0.4	0.4	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
12	res...	0.4	0.4	0.4	0.4	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
13	res...	0.4	0.4	0.4	-0.2	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
14	res...	0.4	0.4	0.4	-0.2	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
15	res...	0.4	0.4	0.4	0.4	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
16	res...	0.4	0.4	0.4	0.4	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
17	res...	0.4	0.4	0.4	0.4	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
18	res...	0.4	0.4	0.4	-0.1	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
19	res...	0.4	0.4	0.4	-0.2	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
20	res...	0.4	0.4	0.4	-0.2	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
21	res...	0.4	0.4	0.4	-0.2	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
22	res...	0.4	0.4	0.4	-0.2	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
23	res...	0.4	0.4	0.4	-0.1	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
24	res...	0.4	0.4	0.4	-0.1	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0
25	res...	0.4	0.4	0.4	0.4	0.4	0.1	0.1	0.1	0.1	0.4	0.1	0.4	0.1	0.1	0.1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.0

Fig. 7 Created Fuzzy Table

The subset is created by calculating the support value and the candidate itemsets are generated until it satisfies the pre-defined minimum support threshold value.

Fig. 8 Subset Creation

The frequent itemsets are generated and the rules are generated based on these values and fuzzy association rules are obtained successfully.



RecordId	ItemSubset	Combination	Support	Confident
2	Longitude-1.0000000	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492
3	AccidentSeverity-1	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492
4	NumberOfCasualties-1	Longitude,AccidentSeverity,Nu...	0.7000000	0.1186441
5	Longitude-1.0000000 AccidentS...	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492
6	Longitude-1.0000000 NumberO...	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492
7	AccidentSeverity-1 NumberOfCa...	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492
9	Longitude-2.0000000	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492
10	AccidentSeverity-2	Longitude,AccidentSeverity,Nu...	0.9000000	0.1525424
11	NumberOfCasualties-1	Longitude,AccidentSeverity,Nu...	0.7000000	0.1186441
12	Longitude-2.0000000 AccidentS...	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492
13	Longitude-2.0000000 NumberO...	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492
14	AccidentSeverity-2 NumberOfCa...	Longitude,AccidentSeverity,Nu...	0.6000000	0.1016949
16	Longitude-3.0000000	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492
17	AccidentSeverity-2	Longitude,AccidentSeverity,Nu...	0.9000000	0.1525424
18	NumberOfCasualties-1	Longitude,AccidentSeverity,Nu...	0.7000000	0.1186441
19	Longitude-3.0000000 AccidentS...	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492
20	Longitude-3.0000000 NumberO...	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492
21	AccidentSeverity-2 NumberOfCa...	Longitude,AccidentSeverity,Nu...	0.6000000	0.1016949
23	Longitude-4.0000000	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492
24	AccidentSeverity-2	Longitude,AccidentSeverity,Nu...	0.9000000	0.1525424
25	NumberOfCasualties-1	Longitude,AccidentSeverity,Nu...	0.7000000	0.1186441
26	Longitude-4.0000000 AccidentS...	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492
27	Longitude-4.0000000 NumberO...	Longitude,AccidentSeverity,Nu...	0.1000000	0.0169492

Fig. 9 Support and Confident table

Fig. 10 shows that Support and Cofident value Table.These are the rules as a result of Apriori algorithm which frequently appears together in the dataset

RecordId	ItemSubset	support	confident	ItemCnt
1	PoliceForce,0.1 NumberOfVehicles,0.1	0.4795460	0.4796186	2
1	PoliceForce,0.1	0.4795460	1.0000000	1
1	PoliceForce,0.1 NumberOfVehicles,0.1 PoliceForce,0.1	0.4795460	0.4796186	3
1	NumberOfVehicles,0.1	0.9998487	1.0000000	1
1	PoliceForce,0.1 NumberOfVehicles,0.1 NumberOfVehicles,...	0.4795460	0.4796186	3
1	Date,0.1	0.0233039	1.0000000	1
1	PoliceForce,0.1 NumberOfVehicles,0.1 Date,0.4	0.0239092	0.0239128	3
1	Date,0.4	0.0462547	1.0000000	1
1	PoliceForce,0.1 NumberOfVehicles,0.1 Date,0.7	0.0349054	0.0349107	3
1	Date,0.7	0.0710719	1.0000000	1
1	PoliceForce,0.1 NumberOfVehicles,0.1 Date,0.8	0.0710719	0.0710826	3
1	Date,0.8	0.1496595	1.0000000	1
1	PoliceForce,0.1 NumberOfVehicles,0.1 Date,0.9	0.3393695	0.3393695	3
1	Date,0.9	0.7097100	1.0000000	1
1	PoliceForce,0.4	0.1385120	1.0000000	1
1	PoliceForce,0.1 PoliceForce,0.1 NumberOfVehicles,0.1	0.4795460	0.4796186	3
1	PoliceForce,0.1 PoliceForce,0.1 Date,0.4	0.0239092	0.0476382	3
1	PoliceForce,0.1 PoliceForce,0.1 PoliceForce,0.1	0.4795460	1.0000000	3
1	PoliceForce,0.1 PoliceForce,0.1 Date,0.7	0.0349054	0.0676839	3
1	PoliceForce,0.1 PoliceForce,0.1 Date,0.8	0.0710719	0.1273385	3
1	PoliceForce,0.1 PoliceForce,0.1 Date,0.9	0.3393695	0.3993115	3
1	PoliceForce,0.1 Date,0.4 PoliceForce,0.1	0.0239092	0.0476382	3
1	PoliceForce,0.1 Date,0.4 NumberOfVehicles,0.1	0.0239092	0.0239128	3

Fig. 10 Support and Cofident value Table

The generated FAR's are then analyzed for Fuzzy Correlation to reduced the rules further to generate mor potential rules.

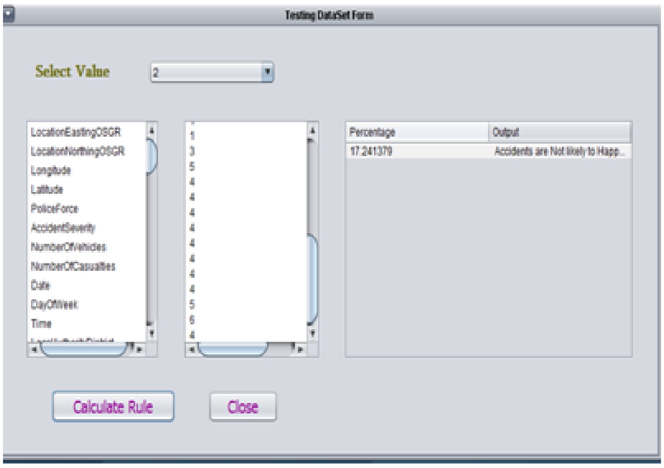
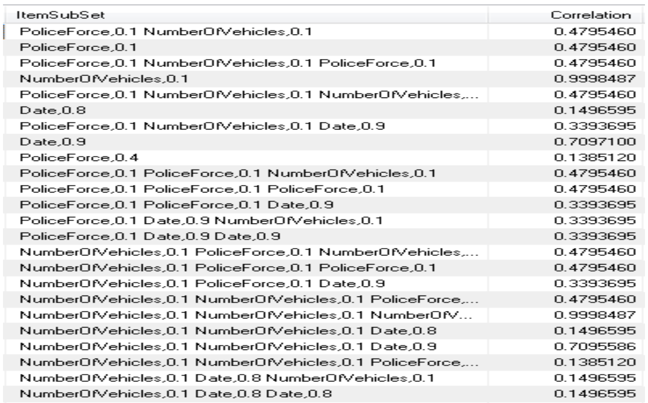


Fig. 11 Fuzzy Correlation on FARs

Fig. 12 shows that Fuzzy Correlation Rules with Values.The potentially Correlated rules are displayed with Correlation values which is the resultant ruleset of the Road Safety Accident dataset.



ItemSubSet	Correlation
PoliceForce,0.1 NumberOfVehicles,0.1	0.4795460
PoliceForce,0.1	0.4795460
PoliceForce,0.1 NumberOfVehicles,0.1 PoliceForce,0.1	0.4795460
NumberOfVehicles,0.1	0.9998487
PoliceForce,0.1 NumberOfVehicles,0.1 NumberOfVehicles,...	0.4795460
Date,0.8	0.1496595
PoliceForce,0.1 NumberOfVehicles,0.1 Date,0.9	0.3393695
Date,0.9	0.7097100
PoliceForce,0.4	0.1385120
PoliceForce,0.1 PoliceForce,0.1 NumberOfVehicles,0.1	0.4795460
PoliceForce,0.1 PoliceForce,0.1 PoliceForce,0.1	0.4795460
PoliceForce,0.1 PoliceForce,0.1 Date,0.9	0.3393695
PoliceForce,0.1 Date,0.9 NumberOfVehicles,0.1	0.3393695
PoliceForce,0.1 Date,0.9 Date,0.9	0.3393695
NumberOfVehicles,0.1 PoliceForce,0.1 NumberOfVehicles,...	0.4795460
NumberOfVehicles,0.1 PoliceForce,0.1 PoliceForce,0.1	0.4795460
NumberOfVehicles,0.1 PoliceForce,0.1 Date,0.9	0.3393695
NumberOfVehicles,0.1 NumberOfVehicles,0.1 PoliceForce,...	0.4795460
NumberOfVehicles,0.1 NumberOfVehicles,0.1 NumberOfV...	0.9998487
NumberOfVehicles,0.1 NumberOfVehicles,0.1 Date,0.8	0.1496595
NumberOfVehicles,0.1 NumberOfVehicles,0.1 Date,0.9	0.7095586
NumberOfVehicles,0.1 NumberOfVehicles,0.1 PoliceForce,...	0.1385120
NumberOfVehicles,0.1 Date,0.8 NumberOfVehicles,0.1	0.1496595
NumberOfVehicles,0.1 Date,0.8 Date,0.8	0.1496595

Fig. 12 Fuzzy Correlation Rules with Values

V.CONCLUSION AND FUTURE ENHANCEMENTS

In this paper, the implementation of fuzzy correlation technique to extracts strong rules by eliminating redundancies and uninteresting rules from the fuzzy association rules. Thus efficient rules are mined from the large dataset through uncovering the hidden characteristics, frequent patterns and associations. Hence knowledge discovery process is achieved with higher accuracy and completeness.

From the implementation it is observed that the Fuzzy Association Rules using Apriori approach generates 220 rules from the 20000 records and Fuzzy Correlation generates 137 rules which are more meaningful. As compared with fuzzy association rule mining approach, fuzzy correlation achieves higher accuracy by improving performance.

As a result more interesting and strong rules are extracted which is very useful for the traffic administrators to obtain more knowledge about the traffic safety and can be able to prevent and manage the accidents in a better way.

As a future work, the classification approaches can be investigated to enhance prediction accuracy and performance further. Also attribute reduction techniques can be applied to obtain more dominating rules with minimal set of candidate generation and with reduced time for database scanning.

REFERENCES

[1] Jain, V., Benyoucef, L., & Deshmukh, S. G. (2008). "A new approach for evaluating agility in supply chains using fuzzy association rules mining", Engineering Applications of Artificial Intelligence, 21(3), pp-367–385.

[2] Toloo, M., & Nalchigar, S. (2011). "On ranking discovered rules of data mining by data envelopment analysis: some models with wider applications", In K. Funatsu & K. Hasegawa (Eds.), New Fundamental Technologies in Data Mining (pp. 425–446). InTech publisher.

[3] Toloo, M. Sohrabi, B., & Nalchigar, S. (2009),". A new method for ranking discovered rules from data mining by DEA", Expert Systems with Applications, 36(4), pp- 8503–8508.

[4] Ho, G. T. S., Ip, W. H., Wu, C. H., & Tse, Y. K. (2012). "Using a fuzzy association rule mining approach to identify the financial data association.", Expert Systems with Applications, 39(10),pp-9054–9063.

[5] Chiu, H. P., Tang, Y. T., & Hsieh, K. L. (2012), " Applying cluster-based fuzzy association rules mining framework into EC environment", Journal of Applied Soft Computing, 12(8), pp-2114–2122

[6] Agrawal, R., Imielinski, T. & Swami, A. (1993), " Mining association rules between sets of items in large databases", In Proceedings of the 1993 ACM SIGMOD international conference on management of data, Washington, DC, United States.

[7] Kannan, S., & Bhaskaran, R. (2009), " Association rule pruning based on interestingness measures with clustering", International Journal of Computer Science Issues (IJCSI), 6(1), pp-35–43.

[8] Kiran, R., & Reddy, P. (2010), " Mining rare association rules in the datasets with widely varying items' frequencies", In Proceedings of 15th international conference on database systems for advanced applications (DASFAA 2010).

[9] Agrawal, R. & Srikant, R. (1994), " Fast algorithms for mining association rules in large databases", In Proceeding of 20th international conference on very large databases (VLDB), Santiago, Chile.

[10] Gedikli, F., & Jannach, D. (2010)," Neighborhood-restricted mining and weighted application of association rules for recommenders", Lecture Notes in Computer Science ,Vol. 6488, pp. 157–165).

[11] Srikant, R. & Agrawal, R. (1996), “Mining quantitative association rules in large relational tables”, In Proceedings of the 1996 ACM SIGMOD international conference on management of data, Montreal, Quebec, Canada.

[12] Kuok, C. M., Fu, A., & Wong, M. H. (1998), “ Mining fuzzy association rules in databases”, ACM SIGMOD Record, 27(1),pp- 41–46.

[13] Kaya, M. & Alhajj, R. (2003), ”Facilitating fuzzy association rules mining by using multi-objective genetic algorithms for automated clustering”, In Proceedings of the third IEEE international conference on data mining (ICDM’03).

[14] Hong, T. P., Kuo, C. S., & Wang, S. L. (2004), “ A fuzzy Apriori Tid mining algorithm with reduced computational time”, Applied Soft Computing Journal, 5(1), pp-1–10.

[15] Zhang, L., Shi, Y. & Yang, X. (2005), “A fuzzy mining algorithm for association-rule knowledge discovery”, In Proceedings of the eleventh Americas conference on information systems, Omaha, NE, USA.

[16] Huang, M. J., Tsou, Y. L., & Lee, S. C. (2006), “ Integrating fuzzy data mining and fuzzy artificial neural networks for discovering implicit knowledge”, Knowledge-Based Systems, 19(6), pp-396–403.

[17] Lei, Z. & Ren-hou, L. (2007), “An algorithm for mining fuzzy association rules based on immune principles”, In Proceedings of the 7th IEEE international conference on bioinformatics and bioengineering, Boston, MA.

[18] Pach F.P Gyenesei, A., & Abonyi, J. (2008), “ Compact fuzzy association rule-based classifier”, Expert Systems with Applications, 34(4), pp-2406–2416.

[19] Chen, C. H., Tseng, V. S., & Hong, T. P. (2008), ” Cluster-based evaluation in fuzzy- genetic data mining”, IEEE Transactions on Fuzzy Systems, 16(1), pp-249–262.

[20] Ashish, M. & Vikramkumar, P. (2010). FPrep: Fuzzy clustering driven efficient automated pre-processing for fuzzy association rule mining. In Proceedings of IEEE international conference on Fuzzy systems, India (pp. 1–8).

[21] Palacios, A. M., Gacto, M. J., & Alcala-Fdez, J. (2010), “ Mining fuzzy association rules from low-quality data”, Soft Computing. doi:10.1007/s00500-011-0775-3.

[22] Hu, Y-H., & Chen, Y-L. (2006), “Mining association rules with multiple minimum supports: a new mining algorithm and a support tuning mechanism.”, Decision Support Systems (vol. 42, pp. 1–24).

[23] Han, J., Pei, J., & Yin, Y. (2000), ” Mining frequent patterns without candidate generation”, In Proceedings of the 2000 ACM SIGMOD international conference on management of data. Dallas, Texas, United States: ACM.

[24] Liu, B., Hsu, W., & Ma, Y. (1999), ” Mining association rules with multiple minimum supports”, In Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99). San Diego, California, United States: AC

[25] Tony Cheng-Kui Huang (2013), " Discovery of fuzzy quantitative sequential patterns with multiple minimum supports and adjustable membership functions", Information sciences 222,pp-126-146-Elsevier

[26] Y.C. Lee, T.P. Hong, W.Y. Lin(2004), "Mining fuzzy association rules with multiple minimum supports using maximum constraints", The Eighth International Conference on Knowledge-based Intelligent Information and Engineering Systems 3214, pp-1283–1290.

[27] Y.C. Lee, T.P. Hong, W.Y. Lin,(2005), "Mining association rules with multiple minimum supports using maximum constraints", International Journal of Approximate Reasoning 40, pp- 44–54.

[28] Y.C. Lee, T.P. Hong, T.C. Wang (2008), "Multi-level fuzzy mining with multiple minimum supports", Expert Systems with Applications 34 , pp- 459–468.

[29] Y.C. Hu, G.H. Tzeng, C.M. Chen (2004), "Deriving two-stage learning sequences from knowledge in fuzzy sequential pattern mining", Information Sciences 159, pp- 69–86.

[30] N.P. Lin, H.J. Chen, W.H. Hao, H.E. Chueh, C.I. Chang (2007), "Mining negative fuzzy sequential patterns", in: Proc. of the 7th WSEAS International Conference on Simulation, Modelling and Optimization, Beijing, China, pp. 52–57.

[31] C. Fiot, A. Laurent, M. Teisseire (2007), "From crispness to fuzziness: three algorithms for soft sequential pattern mining", IEEE Transaction on Fuzzy Systems (6) pp-1263–1277.

[32] T.C.K. Huang (2010), "Knowledge gathering of fuzzy multi-time-interval sequential patterns", Information Sciences 180 (170), pp-3316–3334.

[33] Wang Huajin ; Sch. of Inf. Eng., Jiangxi Univ. of Sci. & Technol., Ganzhou, China ; Yi Chengfu (2012), "Privacy-Preservation Association Rules Mining Based on Fuzzy Correlation", 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery FSKD 2012, pp-757-760

[34] Lilly P.L,Siji P.D,(2013) "A New Featured Fuzzy Clustering Algorithm Based On Adaptive Clustering", VISTAS Vol.2,No.1,ISSN- ISSN: 2319-577, pp7-11