Volume 3, Issue 11, November 2013 **ISSN: 2277 128X**

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Soft Set Approach for Mining Quantitative Fuzzy Association Patterns in Databases

Saakshi Saraf *
Dept. Of Comp. Applications
MANIT,Bhopal (MP) India.

Neeru Adlakha
Dept. of Applied Mathematics and Humanities
SVNIT,Surat(Gujrat)India

Sanjay Sharma
Dept. Of Computer Applications
MANIT,Bhopal (MP) India.

Abstract: Association rule mining is an active data mining research area. In recent years, association rules from large databases have received considerable attention and have been applied to various areas such as marketing, retail and finance, etc. The traditional algorithms for mining frequent association patterns suffer from the problems of under prediction and over prediction of patterns. The main aim of the present paper is to develop a soft set approach for mining fuzzy quantitative association patterns in order to address the issues of under prediction and over prediction of these patterns. The proposed approach is illustrated with the help of a suitable example and experiment on a real world data set of air pollution. The transactional dataset is represented as soft set(fuzzy) using the concept of parameter co-occurrences in the transaction. The quantitative attributes are dealt with by fine –partitioning the value of each attributes and then creating new tables which represent each (fine) partition as a field. The results obtained by soft set approach, soft set fuzzy approach are compared with those obtained by Apriori algorithm without soft set approach. The significant differences have been observed in support and confidence levels of patterns obtained by three approaches.

Keywords: Association pattern mining, Quantitative data, Soft set, Soft set Fuzzy, Attribute partition.

I. INTRODUCTION

Association rule mining is an exploratory learning task to discover some hidden dependency relationships among items in transaction data. The process of finding frequent patterns, associations, correlations, or casual structures among sets of items or objects in transaction databases, relational databases, and other information repositories. Data mining has become the focus of research for computer scientists due to its wide applications in businesses engineering science and technology. Quantitative association rules denote association rules with both categorical and quantitative attributes. There have been several works on quantitative association rule mining such as the application of fuzzy techniques to quantitative association rule mining, the generalized association rule mining for quantitative association rules, and importance weight incorporation into association rule mining for taking into account the user's interest. In the past, several researchers have proposed mining algorithms for finding association rules in transactional data based on the concept of large itemsets[1-7]. Srikant et.al[8] have proposed a method for mining association rules from data sets using quantitative and categorical attributes. Their proposed method first determines the number of partitions for each quantitative attribute, and then maps all possible values of each attribute onto a set of consecutive integers. Other methods have also been proposed to handle numeric attributes and to derive association patterns. A number of algorithms [8-12] are reported in the literature for finding quantitative association rules. The fuzzy set theory introduced by Zadeh [29] is better than the interval method because fuzzy sets provide a smooth transition between member and non-member of a set. In these methods, each of quantitative attributes is replaced by a few other attributes that partition the range of the original one[13-15].

However, the transactional data in real world applications not only consists of quantitative values but also the uncertainty or impreciseness in data. This uncertainty in data poses new challenges for development of algorithms for mining association patterns in quantitative transactional data set. A number of theories like probability, fuzzy set, rough set, and vague set are reported in the literature to deal with uncertainty. These theories have been employed by earlier researchers for mining association patterns. However these theories are not completely capable of dealing with the inherent uncertainty in data. Molodtsov[16] defined soft set for dealing with uncertainty in data. Consequently Molodtsov[16] proposes a completely new approach for modeling vagueness and uncertainty called soft set theory which is free from the difficulty present in existing methods. With the establishment and development of soft set theory its applications gained momentum in recent years and are extended to data analysis[17],decision making[18-24],evaluation[25, 26],medical diagnosis[27] and so on. Herawan et. al.[28] have developed a soft set approach for mining association patterns in transactional datasets.we have extend the work development of soft set approach for quantitative association patterns[29] to fuzzy quantitative association patterns in transactional data set . In view of above in the present paper a soft set approach is developed for mining quantitative fuzzy association patterns in a transaction dataset. The approach is demonstrated with the help of suitable examples and experiments in real world dataset.

The paper is organized as follows. Section 2 describes fundamental concept of soft set theory, fuzzy soft set and association rules mining. .Section 3 shows that by using Soft set theory in the quantitative fuzzy association patterns are discovered .Section 4 shows the Experimental study on Air Pollution data set. Finally, the conclusion of this work.

II. PRELIMINARIES

A. Association rules [28]

Let I = {i_1, i_2, . . ., i|A|}, for |A| > 0 refers to the set of literals called set of items and the set D = {T_1, T_2, . . ., T|U|}, for |U| > 0 refers to the transactional dataset, where each transaction T$\in$D is a list of distinct items T = {i_1, i_2, . . ., i|M|}, $1 \leq |M| \leq |A|$ and each transaction can be identified by a distinct identifier TID. Let, a set X $\subseteq$ T $\subseteq$ I called an itemset. An itemset with k-items is called a k-Itemset. The support of an itemset X, denoted by sup (X) is defined as a number of transactions contained in X. An association rule between sets X and Y is an implication of the form X $\Rightarrow$Y, where X $\cap$ Y = $\varnothing$. The itemsets X and Y are called antecedent and consequent, respectively. The support of an association rule X$\Rightarrow$Y, denoted by sup (X$\Rightarrow$Y), is defined as a number of transactions in D containing X $\cup$ Y. The confidence of an association rule X$\Rightarrow$Y, denoted by cfi (X$\Rightarrow$Y) is defined as a ratio of the number of transactions in D containing X $\cup$Y to the number of transactions in D containing X. Thus,[1, 2]

Cfi (X $\Rightarrow$Y) = $\frac{\sup(X \cup Y)}{\sup(X)}$.

A huge number of association rules can be found from a transactional dataset. To find the interesting association rules in a transactional dataset, we must define a specified minimum support (called minsup) and specified minimum confidence (called minconf). The itemset X $\subseteq$ I is called frequent itemset if sup (X) $\geq$minsup.

It is known that a subset of any frequent itemset is a frequent itemset, a superset of any infrequent itemset is not a frequent itemset. Finally, the association rule X$\Rightarrow$Y holds if conf (X$\Rightarrow$Y) $\geq$minconf. The association rule is said to be strong if they meet the minimum confidence threshold.

B. Soft sets

In this section we present basic definition of soft sets which will be used to illustrate the proposed approach [16, 28, 30].

Definition 1 : Let U be an initial universe, P(U)be the power set of U,E be the set of all parameters and $A \subset E$ Then, a soft set (F,A) over U is a set defined by a function f(A) representing a mapping

$f_A : E \to P(U)$ Such that $f_A(x) = \varnothing$ if $x \notin A$

Here, f_A is called approximate function of the soft set f_A, and the value f_A (*x*) is a set called *x*-element of the soft set for all *x* $\subset$*E[16]* . It is worth noting that the sets f_A (*x*) may be arbitrary. Some of them may be empty, some may have nonempty intersection. Thus, a soft set f_A over *U* can be represented by the set of ordered pairs

$F_A = \{(x, f_A(x)) : x \in E, f_A(x) \in P(U)\}$

Note that the set of all soft sets over U will be denoted by S (U).

Example 1 : Let U={u1,u2,u3,u4,u5} be a universal set and E={x1,x2,x3,x4} be a set of parameters.

If A={x2, x3, x4} and then the soft set F_A is written by $F_A = \{(x2, \{u2, u4\}), (x4, U)\}$

C. Fuzzy sets

In this subsection, we present the basic definitions of fuzzy set theory [31] that is useful for subsequent discussions [32]. More detailed explanations related to this theory may be found in earlier studies[33].

Definition 2: Let *U* be a universe. A fuzzy set *X* over *U* is a set defined by a function μ_X representing a mapping

$\mu_X : U \to [0,1]$

Here, μ_X called membership function of *X* , and the value μ_X(u) is called the grade of membership of *u*$\in$ U .The value represents the degree of *u* belonging to the fuzzy set .*X* Thus, a fuzzy set *X* over *U* can be represented as follows, $X = \{(\mu_X(u)/u) : u \in U, \mu_X(u) \in [0,1]\}$

Note that the set of all the fuzzy sets over U will be denoted by F (U).

D. Fuzzy soft sets

The theory of fuzzy soft set, a more generalized concept, [34] is a combination of fuzzy set and soft set. Many researchers[18, 35] have contributed towards fuzzification of the notion of soft set.

Definition 3[36]: Let *U* be an initial universe set and *E* be a set of parameters. Let *F* (*U*) denotes the fuzzy power set of *U*. Let *A* ⊂ *E*. A pair (*F*, *A*) is called a fuzzy soft set over *U*, where *F* is a mapping given by *F*: *A*→*F* (*U*).

Example 2[36]**:** Suppose that *U* is the set of houses under consideration, *E* is the set of parameters where each parameter is a fuzzy word or a sentence involving fuzzy words, *E* = {expensive (e_1), beautiful (e_2)}. In the case, to define a fuzzy soft set means to point out expensive house, beautiful house. The fuzzy soft set (*F*, *E*) describes the “attractiveness of the houses” which Mr. X is going to buy.

Suppose that
$F(e_1) = \{(h_1, 0.5), (h_2, 1), (h_3, 0.4), (h_4, 1), (h_5, 0.3), (h_6, 0)\}$,
$F(e_2) = \{(h_1, 1), (h_2, 0.4), (h_3, 1), (h_4, 0.4), (h_5, 0.6), (h_6, 0.8)\}$,
The fuzzy set (F, E) is a parameterized family {F (e_i), i= 1, 2} and gives us a collection of approximate description of an object. The mapping F here is "house (.)" where dot (.) is to be filled up by a parameter $e \in E$. Therefore $F(e_1)$ means "house (expensive)" whose functional-value is the fuzzy set $\{(h_1, 0.5), (h_2, 1), (h_3, 0.4), (h_4, 1), (h_5, 0.3), (h_6, 0)\}$. Thus, we can view the fuzzy soft set (F,E) as a collection of fuzzy approximations (which are fuzzy sets) as below: (F,E) = {expensive houses = $\{(h_1, 0.5), (h_2, 1), (h_3, 0.4), (h_4, 1), (h_5, 0.3), (h_6, 0)\}$, beautiful houses =$\{(h_1, 1), (h_2, 0.4), (h_3, 1), (h_4, 0.4), (h_5, 0.6), (h_6, 0.8)\}$}.

E. Apriori Algorithm and Apriori Property

Apriori is an influential algorithm in market basket analysis for mining frequent item sets for Boolean association rules[1, 2].The name of apriori is based on the fact that the apriori uses prior knowledge of frequent item set properties. Aproiri employs an iterative approach known as a level-wise search, where k-item sets is found, denoted by L1.L1 is used to find L2,the set of frequent 2 item sets, which is used to find L3,and soon, until no more frequent K-item sets can be found.
Property: All non empty subsets of frequent item sets must be frequent[2] .

F. Transaction set[8]

A transaction T is a collection of one or more items. Let I = $\{i_1, i_2, \ldots, i_{|\bar{A}|}\}$, for $|\bar{A}| > 0$ refers to the set of items and the set D = {T1, T2, . . ., T|U|}, for |U| > 0 refers to the transactional dataset, where each transaction $T \in D$ is a list of distinct items T = $\{i_1, i_2, \ldots, i_{|M|}\}$, $1 \le |M| \le |\bar{A}|$ and each transaction can be identified by a distinct identifier TID. Let, a set $x \subseteq T \subseteq I$ called an itemset. An itemset with k-items is called a k-itemset[8].

$$T = \{x \mid \forall x \in I\}.$$

G. Quantitative transaction set[8]

Let P denote the set of positive integers. Let I_v denote the set $I \times P$. A pair $<x, v> \in I_v$ denotes the attribute x, with the associated value v.I_R denotes $\{\forall < x, l, u > \in I \times P \times P \mid l \le u\}$.

$T_q = \{< x, l, u > \mid \forall < x, l, u > \in I \times P \times P \mid l \le u\}$ If x is quantitative, l is lower limit and u is upper limit of P. A triple $<x, l, u> \in I_R$ denotes either a quantitative x with a value in the interval [l, u].

H. Soft transaction

A Soft transaction is denoted by $\bar{T}$. Let (F, E) be a soft set over the universe U and $X \subseteq E$. A set of attributes X is said to be supported by a transaction.

$$\bar{T} = \{(x, e) \mid \forall x \in I, e \in E\}.$$

I. Fuzzy Membership $\mu(e)$

The fuzzy membership $\mu(e)$ is defined as the possibility of items having parameter/property/characteristics e. Here $\mu(e)$ will always be associated with an item in the fuzzy soft set.

J. *Fuzzy Transaction*

A Fuzzy transaction denoted by $\bar{\bar{T}}$ is given by:

$\bar{\bar{T}} = \{(x, \mu(x) \mid \forall x \in I\}$ where $0 \le \mu(x) \le 1$, μ : I→[0,1], $\bar{\bar{T}} \subseteq$ D.

Where **D** be a universal set of transactions. $\mu(x)$ is degree of membership of x.

K. Fuzzy Soft Transaction

A Fuzzy soft transaction is denoted by T'' is given by:

$T'' = \{(x, (e, \mu(e))) \mid \forall x \in I, e \in E\}$, $0 \le \mu(e) \le 1$, μ : E→[0,1].

Where $\mu(e)$ is degree of membership of e(attribute) associated with an item x.

L. Soft quantitative transaction set

A Soft Quantitative transaction set is denoted by T_q'. Let (F, E) be a soft set over the universe U and $X \subseteq E$. A set of attributes X is said to be supported by a transaction.

$$T_q' = \{(< x, l, u >, e) \mid\mid \forall < x, l, u > \in I \times P \times P \mid l \le u, e \in E\}.$$

M. Soft quantitative fuzzy transaction set

A Soft Quantitative fuzzy transaction set is denoted by T_q''. Let (F, E) be a soft set over the universe U and $X \subseteq E$. A set of attributes X is said to be supported by a transaction.

$T_q'' = \{(<x,l,u>, e(\mu(e)) \| \forall <x,l,u> \in I \times P \times P \mid l \le u, e \in E\}$.

N. Association pattern

Association pattern represents the association of items. Pattern is denoted by P_i .A transaction T contains X, Y, a set of items in I, if *X, Y* ∈ T. An association rule is a pattern that states when *X* occurs, *Y* occurs with certain probability. It is expressed as:

$(P_i) = (I_{r_1} \cup I_{r_2} \cup I_{r_3} \cup .....I_{r_m})$, $r_j \ne r_u$, (j, u) =1(1)n ,

for m=1 $P_i = I_{r_1}, r_1 = 1(1)n$

for m=2 $P_i = I_{r_1} \cup I_{r_2}, r_1 \ne r_2$; r_1,r_2 =1(1)n.

and so on……

O. Soft association pattern

Soft Association Pattern represents the association of items with their parameter e. It is represented by $(P_i, e) = ((I_{r_1} \cup I_{r_2} \cup I_{r_3} \cup ......I_{r_m}), e)$.

P. Soft quantitative association pattern

Soft Quantitative Association Pattern represents the association of items with their parameter e. It is represented by $(P_i, e) = ((I_{r_1} \cup I_{r_2} \cup I_{r_3} \cup ......I_{r_m}), e)$ .Here each I_{r_1} represents $< x_{r_i}, l_{r_i}, u_{r_i} > \forall r_i = 1(1)n$

Q. Soft quantitative fuzzy Association Pattern

Fuzzy Soft Association Pattern represents the association of fuzzy items with their parameter e .It is calculated as $(P_i, (e_i, \mu(e_i))) = ((I_{r_1} \cup I_{r_2} \cup I_{r_3} \cup ......I_{r_m}), (e_i, \mu(e_i)))$

Here each I_{r_1} represents $< x_{r_i}, l_{r_i}, u_{r_i} > \forall r_i = 1(1)n$. $\mu(e_i) = Min\{\mu(e_{r_1}), \mu(e_{r_2}), ...\mu(e_{r_m})\}$ Based on the above concepts the soft set approaches for mining fuzzy associations are given in subsequent sections.

III. SOFT SET BASED MINING OF QUANTITATIVE FUZZY ASSOCIATION PATTERNS

In this section we present the applicability of soft set theory for mining quantitative fuzzy association patterns. The pre-requisite of using soft set approach for association pattern mining[28] in the transactional data set needs to be transformed into soft set(fuzzy),where each item is regarded as a parameter(attribute).Here improved Apriori algorithm, is proposed by considering that every item will have a relation (similarity) to others if they are purchased in the same transactions. An Illustrative example 3 is given to understand the concept of soft set based mining in quantitative data, for finding association patterns. The process is started from a given transactional database as shown in Table 1.

Steps of Algorithms:

Step-1.

Determine maximum number of items in a qualified transaction as defined:

M= {T|$\mathcal{M}$(T) ≤ δ, T ∈ **D**}, (1)

δ is a membership threshold; **D** be a universal set; $\mathcal{M}$ (T) is the number of items in transaction T

Step-2. Convert the transaction T, T⊆M into fuzzy transaction $\bar{\bar{T}}$.

Step-3. Transform Fuzzy transaction to Soft Quantitative Fuzzy Transaction set, Soft Quantitative Fuzzy transaction is denoted by T''.

Step-4.

Set *k*=1, where *k* is an index variable to determine the number of combination of items in item sets called k-itemsets.Here we denote P_k for k-itemset where P_k is an association Pattern.

Step-5. Determine minimum support for P_k, denoted by $\beta_k \in$ (0,| M|) as a minimum *threshold* of a combination *k* items appearing in the whole set of qualified transactions, where |**M**| is the number of qualified transactions. Here, β_k may have different value for every *k*.

Step-6.Select each patterns P_k from soft fuzzy transaction .

Step-7.Calculate the fuzzy membership of each pattern P_k for all T'' (soft quantitative fuzzy transaction) to obtain soft quantitative fuzzy association pattern .

Step-8. Compute sum of memberships of each Items in the T'' (soft quantitative fuzzy transaction) .

Step-9. Calculate Support for Every Pattern P_k .

Step-10. P_k will be stored in the set of frequent *k*-itemsets, L_k if and only if *support* (P_k) ≥ β_k.

Else if *support* (P_k) < β_k for all P_k then Goto 13.

Step-11. If k< 2 Goto 12.

else

Compute Confidence,Conf( ) = $\frac{Support(P_k)}{Support(P_{k-1})}$

Store confidence of P_k for frequent k itemsets L_k.

Step-12.

Set k=k+1, and if $k > \delta$, then go to Step-13.

Else goto step 5.

Step-13. Stop.

Example: 3 Suppose that δ arbitrarily equals to 3; that means qualified transaction is regarded as a transaction with no more than 3 items purchased in the transaction. There is a data set consisting of 10 transactions which contains two categories i.e., T= {Person profile, Items),

Where Profile = {Sex, Economic status}, Sex= {Male, Female}

Items= {Age, House, Furniture, Car, Shop, TV}.

Result of the step is a set of qualified transaction where M = { $T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9, T_{10}$ }

TABLE: 1 A QUALIFIED DATA TRANSACTION

Trans_ID	Person Profile with purchased items
T_1	{Male,11000,33,house,car}
T_2	{Female,10000,28,furniture,TV,car}
T_3	{32000,car}
T_4	{Female,15000,48,car,TV}
T_5	{Female,36000,29,house ,car ,TV}
T_6	{Male,40000,Shop,TV}
T_7	{Female,28000,41,Furniture,House}
T_8	{Female,40,Car,TV}
T_9	{Male,30,Furniture,Car}
T_{10}	{Male,42,Shop,House}

In table 1, there are 10 transactions; each transaction shows the profile of the person with items purchased. The T_1 shows that Person whose sex is Male, Economic status is salary Rs.11000 and Age is 33 years. He purchased items are House and Car. Therefore each transaction shows Person profile with their purchased items.

Here partitioning the attributes economic status and age is given in Table 2 and 3

Economic status and age with their consecutive values:

TABLE: 2 MAPPING OF ECONOMIC STATUS

Amount	***Values***
Less than 10000	E1
10000 …30000	E2
Greater than 30000	E3

TABLE: 3 MAPPING OF AGE

Age Intervals	**Values**
20….30	A1
31….45	A2
46….65	A3

After partitioning attributes in the transactions with their consecutive values in the database are:

TABLE: 4 A QUALIFIED DATA TRANSACTION

Trans_ID	Person Profile with purchased items
T_1	{Male,E2,A2,house,car}
T_2	{Female,E1,A1,furniture,TV,car}
T_3	{E3,car}
T_4	{Female,E2,A3,car,TV}
T_5	{Female,E3,A1,house ,car ,TV}
T_6	{Male,E3,Shop,TV}
T_7	{Female,E2,A2,Furniture,House}
T_8	{Female,A2,Car,TV}
T_9	{Male,A1,Furniture,Car}
T_{10}	{Male,A2,Shop,House}

Employing soft set theory and Apriori algorithm on transaction in Table 4, we calculate candidate itemset and frequent itemset.Finally we calculate Support and confidence of frequent itemsets.The itemsets and their supports are given below:

1 itemsets

$\{Male\} = \{T_1, T_6, T_9, T_{10}\} = 4, \{Female\} = \{T_2, T_4, T_5, T_7, T_8\} = 5$

$\{E1\} = \{T_2\} = 1, \{E2\} = \{T_1, T_4, T_7\} = 3, \{E3\} = \{T_3, T_5, T_6\} = 3$

$\{A1\} = \{T_2, T_5, T_9\} = 3, \{A2\} = \{T_1, T_7, T_8, T_{10}\} = 4$

$\{A3\} = \{T_4\} = 1, \{House\} = \{T_1(.5), T_5(.33), T_7(.5)\} = 1.33, \{Furniture\} = \{T(0.33)_2, T_7(0.5), T_9(0.5)\} = 1.33$

$\{Car\} = \{T_1(.5), T_2(.33), T_3(1), T_4(.5), T_5(.33), T_8(.5), T_9(.5)\} = 3.66, \{Shop\} = \{T_6(.5), T_{10}(.5)\} = 1$

$\{TV\} = \{T_2(.33), T_4(.5), T_5(.33), T_6(.5), T_8(.5)\} = 2.16$

{E1}, {A3}, {Shop} cannot be considered for further process because their support is less than threshold value is 1.33.

2 itemsets

$\{Male, House\} = \{T_1(.5), T_{10}(.5)\} = 1, \{Male, Furniture\} = \{T_9(.5)\} = .5, \{Male, Car\} = \{T_1(.5), T_9(.5)\} = 1$

$\{Male, TV\} = \{T_6(.5)\} = .5, \{Male, A1\} = \{T_9(1)\} = 1, \{Male, A2\} = \{T_1(1), T_{10}(1)\}, = 2\{Female, House\} = \{T_5(.33)\} = 0.33$

$\{Female, Furniture\} = \{T_2(.33), T_7(.5)\} = .83, \{Female, Car\} = \{T_2(.33), T_4(.5), T_5(.33), T_8(.5)\} = 1.66,$

$\{Female, TV\} = \{T_2(.33), T_4(.5), T_5(.33), T_8(.5)\} = 1.66, \{Female, A1\} = \{T_2(1), T(1)_5\} = 2$

$\{Female, A2\} = \{T_7(1), T_8(1)\} = 2, \{E2, House\} = \{T_1(.5), T_7(.5)\} = 1, \{E2, Furniture\} = \{T_7(.5)\} = .5$

$\{E2, Car\} = \{T(.5)_1, T_4(.5)\} = 1, \{E2, TV\} = \{T(.5)_4\} = .5, \{E2, A1\} = \{\} = 0, \{E2, A2\} = \{T(1)_1, T_7(1)\} = 2$

$\{E3, House\} = \{T_5(.33)\} = .33, \{E3, Furniture\} = \{\} = 0, \{E3, Car\} = \{T_3(1), T_5(.33)\} = 1.33$

$\{E3, TV\} = \{T_5(.33), T_6(.5)\} = .83, \{E3, A1\} = \{T_5\} = 1, \{E3, A2\} = \{\} = 0$

Those item sets cannot be considered for further process whose support is less than threshold value is 2.

3 itemsets:

$\{Male, House, Car\} = \{T_1(.5)\} = 1, \{Female, Furniture, Car\} = \{T_2(.33)\} = .33$

$\{Female, Furniture, TV\} = \{T_2(.33)\} = 0.33, \{Female, Furniture, A1\} = \{T_2(0.33)\} = 0.33$

$\{Female, Furniture, A2\} = \{T_7(.5\} = .5, \{Female, Car, TV\} = \{T_2(.33), T_4(.5), T_5(.33), T_8(.5)\} = 1.66$

$\{Female, Car, A1\} = \{T_2(.33), T_5(.33)\} = .66, \{Female, Car, A2\} = \{T_8(.5)\} = .5$

$\{Female, TV, A1\} = \{T_2(.33), T_5(.33)\} = .66, \{Female, TV, A2\} = \{T_8(.5)\} = .5$

$\{Female, A1, A2\} = \{\} = 0, \{E2, House, Car\} = \{T(.5)_1\} = .5, \{E3, Car, TV\} = \{T_5\} = .33$

Only those item sets are considered whose support is greater than or equal to 2.

4 itemsets:

$\{Female, Car, TV, A1\} = \{T_2(.33), T_5(.33)\} = .66$

Only those itemset are considered whose support is greater or equal to minimum support=2

TABLE 5: FREQUENT 1-ITEMSETS

Min_Sup =1.33	
1-itemsets	**Support**
{Male}	4
{Female}	5
{E2}	3
{E3}	3
{A1}	3
{A2}	4
{House}	1.33
{Furniture}	1.33
{Car}	3.66
{TV}	2.16

TABLE 6: FREQUENT 2-ITEMSETS

Min_Sup=1		
2-itemsets	**Support**	**Confidence**
(Male⇒House)	1	25%
(Male⇒Car)	1	25%
(Male⇒A1)	1	25%
(Male⇒A2)	2	50%
(Female⇒Car)	1.66	33.3%
(Female⇒TV)	1.66	33.3%
(Female⇒A1)	2	40%
(Female⇒A2)	2	40%
(E2⇒House)	1	33.3%
(E2⇒Car)	1	33.3%
(E2=A2)	2	66.6%
(E3⇒Car)	1.33	44.3%
(E3⇒A1)	1	33.3%

TABLE 7: FREQUENT 3-ITEMSETS

Min_Sup=.60		
3-itemsets	**Support**	**Confidence**
(Female, Car⇒TV)	1.66	33.3%
(Female, Car⇒A1)	.66	13%
(Female, TV⇒ A1)	.66	13%

TABLE 8: FREQUENT 4-ITEMSETS

Min_Sup=.60		
4-itemsets	**Support**	**Confidence**
{TV, Female,Car ⇒A1}	.66	30%

Table 8 shows that the female whose Age between 20 to 30, has purchased car and TV. Their confidence is 30%

IV. EXPERIMENT ON AIR POLLUTION DATASET[37]

We further explain the concept and proces of Quantitative soft set approach for association patterns by performing experiment on real data set of air pollution from Maharashtra pollution control board[37] of India.The data of air pollution of Pune city of India for the period of two years during 1January, 2010 to 31 December 2011 is used to perform the experiment[37].It is based on the observation of the air pollution data of two areas of Pune city,namely Chinchwad and Karve Road, that have 570 and 692 transactions respectively.It contains the data on concentration in Microgram per cubic meter($\mu g/m^3$) of three pollutants SO_2,NO_x and RSPM in the atmosphere of above two areas of Pune.The standards of acceptable limit of concentration of the above three pollutants is also given in the dataset available on websites of Maharashtra Pollution Control Board.The data is presented as the average amount of each data item per day.

Using the given acceptable limit, the taxonomy of data is prepared as given below:

T={Acceptable limit,Harmful limit,Very harmful}
Area={A1,A2} , A1= Chinchwad, A2=Karve

In this example there are three pollutants SO_2,NO_x and RSPM.In case of quantitative soft set approach the membership of pollutant present in a transaction is taken to be 1.The pollutant is either is in acceptable limit ,harmful and very harmful .the acceptable limit of three items SO_2,NO_x and RSPM are given in Table 9. The soft set and traditional approaches are employed to explore frequent association patterns. The association patterns obtained by two approaches are given in Table 10 along with support and confidence.

TABLE 9 : ACCEPTABLE LIMITS OF POLLUTANTS

	So_2	No_x	RSPM
Acceptable	<80	<80	<100
Harmful	80-100	80-100	100-200
Very harmful	>100	>100	>200

TABLE 10: ASSOCIATION PATTERNS FOR AIR POLLUTION DATA ALONG WITH THEIR SUPPORT AND CONFIDENCE.

Association Pattern	Soft Set approach		Traditional approach		Soft set(fuzzy quant)	
1 itemsets	Support	Confidence	Support	Confidence	Support	Confidence
{SO2,(Acceptable,A1)}	99.64%		99.84%		44.31%	
{SO2,(Harmful,A1)}	.35%		.15%		0.17%	
{SO2,(Very Harmful,A1)}	0		0		0	
{ NOx (Acceptable,A1)}	89.12%		85.81%		35.19%	
{ NOx,(Harmful,A1)}	8.59%		6.73%		5.61%	
{ NOx,(Very Harmful,A1)}	**2.28%**		7.44%		**2.10%**	
{ RSPM (Acceptable,A1)}	58.24%		47.78%		19.57%	
{ RSPM (Harmful,A1)}	36.6%		42.15%		33.85%	
{ RSPM (Very Harmful,A1)}	**5.08%**		10.06%		**4.56%**	
{SO2,(Acceptable,A2)}	100%		99.84%		51.58%	
{SO2,(Harmful,A2)}	0		.15%		0	
{SO2,(Very Harmful,A2)}	0		0		4.40%	
{ NOx (Acceptable,A2)}	83.09%		85.81%		34.96%	
{ NOx,(Harmful,A2)}	5.20%		6.73%		3.90%	
{ NOx,(Very Harmful,A2)}	11.70%		7.44%		8.95%	
{ RSPM (Acceptable,A2)}	39.16%		47.78%		13.07%	
{ RSPM (Harmful,A2)}	46.67%		42.15%		45.37%	
{ RSPM (Very Harmful,A2)}	14.16%		10.06%		11.63%	
2-Itemset	Support	Confidence	Support	Confidence		
{SO2,NOx(Acceptable,A1)}	89.12%	89.08%	82.72%	82.85%	34.84%	78.62%
{SO2,RSPM(Acceptable,A1)}	58.24%	58.4%	47.78%	47.85%	19.57%	44.16%
{Nox,RSPM(Accetable,A1)}	-	-	-	-	18.52%	52.62%
{SO2,NOx(Acceptable,A2)}	77.45%	77.4%	82.72%	82.85%	34.96%	67.77%
{SO2,RSPM(Acceptable,A2)}	39.16%	39.16%	47.78%	47.85%	13.07%	25.33%
{NOx,RSPM(Acceptable,A2)}	35.83%	43.13%	45%	52.44%	12.63%	36.12%
{NOx,RSPM (Very Harmful,A2)}	6.06%	51.85%	45%	52.44%	2.52%	28.22%
3 Itemset						
{SO2,NOx,RSPM(Acceptable,A1)}	56.14%	62.9%	45%	45%	18.52%	41.79%
{SO2NOx,RSPM(Acceptable,A2)}	35.83%	40.20%	45%	45%	12.63%	24.48%

TABLE 11: FREQUENT ASSOCIATION PATTERNS FOR AIR POLLUTION DATA ALONG WITH THEIR SUPPORT AND CONFIDENCE.

Association Pattern	Soft Set Approach(30%)		Traditional Approach(30%)		Fuzzy(Quantitative)30%	
1 itemsets	Support	Confidence	Support	Confidence	Support	Confidence
{SO2,(Acceptable,A1)}	99.64%	-	99.84%	-	44.31%	
{ NOx (Acceptable,A1)}	89.12%	-	85.81%	-	35.19%	
{ RSPM (Acceptable,A1)}	58.24%	-	47.78%	-	-	
{SO2,(Acceptable,A2)}	100%	-	99.84%	-	51.58%	
{ NOx (Acceptable,A2)}	83.09%	-	85.81%	-	34.96%	
{ RSPM (Acceptable,A2)}	39.16%	-	47.78%	-	-	
{ RSPM (Harmful,A2)}	46.67%	-	42.15%	-	45.37%	-
2-Itemset	Support	Confidence	Support	Confidence		
{SO2,NOx(Acceptable,A1)}	89.12%	89.08%	82.72%	82.85%	34.84%	78.62
{SO2,RSPM(Acceptable,A1)}	58.24%	58.4%	47.78%	47.85%	NF	-
{SO2,NOx(Acceptable,A2)}	77.45%	77.4%	82.72%	82.85%	34.96%	67.77
{SO2,RSPM(Acceptable,A2)}	39.16%	39.16	47.78%	47.85%	NF	-
{NOx,RSPM(Acceptable,A2)}	35.83%	43.13	45%	52.44%	NF	-
3 Itemset						
{SO2,NOx,RSPM(Acceptable,A1)}	56.14%	62.99%	45%	45%	-	-
{SO2NOx,RSPM(Acceptable,A2)}	35.83%	40.20%	45%	45%	-	-

A Table 11, we find significant difference in support and confidence level of association patterns discovered by the three approaches. In Table 11 shows the frequent association pattern for support threshold 30%. The patterns {RSPM,Acceptable,A1}and {RSPM,Acceptable,A2} is found to be frequent by traditional and soft set approach or in frequent by fuzzy soft approach. Further the patterns{SO2,RSPM,Acceptable,A1},{SO2,RSPM,Acceptable,A2} and {NO_x,RSPM,Acceptable,A2} are found to be frequent by traditional approach and soft set approach but infrequent by soft set(fuzzy) approach this is an example of over prediction of an association pattern by traditional approach and soft set approach. Thus we observe that the soft set (fuzzy) approaches quite useful in addressing the issues of under prediction and over prediction of association pattern a part of this a soft set(fuzzy) approach gives better picture of confidence level as compared to that in traditional approach and soft set approach which is visible in Table 11 as significant difference in confidence level of three approaches.

V. CONCLUSION

In this paper, a quantitative approach for mining fuzzy association patterns from transactional dataset using soft set theory was proposed. The soft set (fuzzy) approach applied on quantitative data has been successfully demonstrated using suitable example. Quantitative attributes are dealt with by fine partitioning the values of the attribute. Some key issue has been left for further research including (1)how to handle information lost due to partitioning (2) how to decide whether or not to partition a quantitative attributes (3) how to filter the rules if very large number of association rule are generated. Experimental results also show that it assists in obtaining small sets with good performance for most datasets. We are able to incorporate the uncertainty due to parameters in exploring association among the items or elements of soft set based fuzzy transaction. Thus we get a better picture of association relationship in soft set(fuzzy) approach for mining quantitative fuzzy association patterns.

REFERENCES

1. R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," *International conference on Management of data*, ACM SIGMOD pp. 207-216.
2. R. Agrawal, and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," *Proceeding VLDB '94 Proceedings of the 20th International Conference on Very Large Data Bases*, Morgan Kaufmann Publishers Inc. San Francisco, CA, pp. 487-499
3. J. Han, and Y. Fu, "Discovery of multiple-level association rules from large databases," *Book Discovery of multiple-level association rules from large databases*, Series Discovery of multiple-level association rules from large databases, ed., Editor ed.^eds., 1995, pp. 420-431.
4. A. Pandey, and K. Pardasani, "Rough Set Model for Discovery multidimensional Association rules.," *International Journal of computer science and network security*, vol. 9, no. 6, 2009, pp. 159-164.

5. R. Thakur, R. Jain, and K. Pardasani, "Mining level-crossing association rules from large databases.," *Journal of computer science*, vol. 2, no. 1, 2006, pp. 76-81.
6. R.S. Thakur, R.C. Jain, and K.R. Pardasani, "Fast Algorithm for mining multi-level association rules in large databases.," *Asian Journal of International Management*, vol. 1, 2007, pp. 19-26.
7. N. Khare, N. Adlakha, and K.R. Pardasani, "An Algorithm for Mining Multidimensional Fuzzy Association Rules " *International Journal of Computer Science and Information Security*, vol. 5, no. 1, 2009, pp. 72-76.
8. R. Srikant, and R. Agrawal, "Mining quantitative association rules in large relational tables," *In Proceedings of the ACM SIGMOD Conference on Management of Data*.
9. Y. Aumann, and Y. Lindell, "A statistical theory for quantitative association rules," *KDD '99 Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 261-270.
10. K.C.C. Chan, and W.-H. Au, " An effective algorithm for mining interesting quantitative association rules.," *In Proceedings of the 1997 ACM symposium on Applied computing*, pp. 88-90.
11. S. Haihong, T. Jing, J. Hong, and Y. Bingru, "Mining Quantitative Association Rules with Standard SQL Queries and it's Evaluation," *Journal of Computer Research and Development*, vol. 39, no. 1, 2002, pp. 307-312.
12. T.P. Hong, C.S. Kuo, and S.C. Chi, "A data mining algorithm for transaction data with quantitative values," *Intelligent Data Analysis*, vol. 3, no. 5, 1999, pp. 363-376.
13. A. Gyenesei, "Mining Weighted Association Rules for Fuzzy Quantitative Items," *Principles of Data Mining and Knowledge Discovery Lecture Notes in Computer Science*, vol. 1910, 2000, pp. 416-423.
14. T.-P. HONG, K.-Y. LIN, and B.-C. CHIEN, "Mining Fuzzy Multiple-Level Association Rules from Quantitative Data," *Applied Intelligence*, vol. 18, 2003, pp. 79–90.
15. P. Yan, and G. Chen, "Fuzzy Quantitative Association Rules and Its Applications," *Fuzzy Applications in Industrial Engineering Studies in Fuzziness and Soft Computing*, vol. 201, 2006, pp. 573-587.
16. D. Molodtsov, "Soft set theory-first results," *Computers and Mathematics with Applications,*, vol. 37., 1999, pp. 19–31.
17. Y. Zou, and Z. Xiao, "Data analysis approaches of soft sets under incomplete information," *Knowledge Based Systems*, vol. 21, 2008, pp. 941-945.
18. A.R. Roy, and P.K. Maji, "A fuzzy soft set theoretic approach to decision making problems," *Journal of Computational and Applied Mathematics*, vol. 203, 2007, pp. 412-418.
19. P.K. Maji, A.R. Roy, and R. Biswas, "An application of soft sets in a Decision making problem,," *Computers and Mathematics with Applications,*, vol. 44, 2002, pp. 1077– 1083.
20. Y. Jiang, Y. Tang, and Q. Chen, "An adjustable approach to intuitionistic fuzzy soft sets based decision making," *Applied Mathematical Modelling*, vol. 35, no. 2, 2011, pp. 824-836.
21. F. Feng, Y. Li, and V.L. Fotea, "Application of level soft sets in decision making based on interval-valued fuzzy soft sets," *Computers and Mathematics with Applications*, vol. 60, 2010, pp. 1756-1767.
22. F. Feng, Y.B. Jun, X. Liu, and L. Li, "An adjustable approach to fuzzy soft set based decision making," *Journal of Computational and Applied Mathematics*, vol. 234, no. 1, 2010, pp. 10-20.
23. N. Cagman, and S. Enginoglu, "Soft matrix theory and its decision making," *Computers and Mathematics with Applications*, vol. 59, 2010, pp. 3308-3314.
24. N. Cagman, and S. Enginoglu, "Soft set theory and uni-int decision making," *European Journal of Operational Research*, vol. 207, 2010, pp. 848-855.
25. Z. Xiao, W. Chen, and L. Li, "An integrated FCM and fuzzy soft set for supplier selection problem based on risk evaluation," *Applied Mathematical Modelling*, vol. 36, no. 4, 2012, pp. 444-1454.
26. Z. Xiao, X. Yang, Q. Niu, Y. Dong, K. Gong, S. Xia., and Y. Pang, " A new evaluation method based on D–S generalized fuzzy soft sets and its application in medical diagnosis problem," *Applied Mathematical Modelling*, vol. 36, no. 10, 2012, pp. 4592-4604.
27. R. Borgohain, and P.K. Das, "Application of Fuzzy Soft set in medical diagnosis using Fuzzy arithmetic operations on Fuzzy Number,," *SIBCOLTEJO*, vol. 05, 2010, pp. 107-116.
28. T. Herawan, and M.M. Deris, "A Soft Set approach for association rule mining," *Knowledge Based Systems*, vol. 24, 2011, pp. 186-195.
29. S. Saraf, N. Adlakha, and S. Sharma, "Soft Set Approach for Mining Quantitative Association Patterns in Databases," *IOSR Journal of Computer Engineering*, vol. 11, no. 6, 2013, pp. 55-65.
30. P.K. Maji, R. Biswas, and A.R. Roy, "Soft Set Theory," *Computer and Mathematics with Application*, vol. 45 2003, pp. 555-562.
31. L.A. Zadeh, "Fuzzy set," *Information and Control.*, vol. 8, 1965, pp. 338–353.
32. K. Atanassov, " Intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 20, 1986, pp. 87-96.
33. Y.L. Chen, and C.H. Weng, "Mining fuzzy association rules from questionnaire data," *Knowledge-Based Systems*, vol. 22 (1), 2009, pp. 46–56.
34. P.K. Maji, R. Biswas, and A.R. Roy, "Fuzzy soft sets," *Journal of Fuzzy Mathematics*, vol. vol. 9, no. 3, 2001, pp. 589–602.
35. N. Çagman, S. Enginoglu, and F. Citak, "Fuzzy soft set theory and its applications," *Iranian Journal of Fuzzy Systems*, vol. 8, no. 3, 2011, pp. 137-147.

36. B.X. Yao, J.L. Liu, and R.X. Yan, "Fuzzy Soft Set and Soft Fuzzy Set," *Proceedings of the IEEE Fouth International Conference on Natural Computation*.
37. MPCB,"http://mpcb.gov.in/envtdata/demoPage1.php,"*Book http://mpcb.gov.in/envtdata/demoPage1.php*, Series http://mpcb.gov.in/envtdata/demoPage1.php, ed., Editor ed.^eds., pp.

