International Journal of Engineering Research and Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 11, Issue 07 (July 2015), PP.84-95

A Fuzzy Soft Set Approach for Mining Association Patterns

Saakshi Saraf[1], Neeru Adlakha[2], Sanjay Sharma[3]
[1]*Department Of Computer Applications,Maulana Azad National Institute Of Technology,Bhopal (MP) India.*
[2]*Department of Applied Mathematics and Humanities,S.V. National Institute Of Technology, Surat, Gujarat India.*
[3]*Department of Computer Applications,Maulana Azad National Institute Of Technology, Bhopal (MP) India.*

Abstract: In this paper, a fuzzy soft set approach is proposed for mining association patterns from a transactional data set.This approach is proposed to capture the uncertain item relationships in data sets and enhance the precision in association rule mining. The transactional data set is represented as soft set using the concept of parameter co-occurrences in the transaction. Thus, each transaction is transformed into soft transaction to generate soft transactional dataset. For the fuzzy parameters present in the transactions, the fuzzy membership is determined and used to transform the soft transactional data set into fuzzy soft transactional dataset. The mining of association patterns is performed on the fuzzy soft transactional data set. The fuzzy soft set approach has been illustrated with the help of an example and experiment on a real world data set. The results of fuzzy soft set approach have been compared with those obtained by traditional, fuzzy set and soft set approaches for mining association patterns. The fuzzy soft set approach gives better picture of association relationship, confidence levels and is helpful in addressing the issues of under-prediction and over-prediction of association patterns.

Keywords*: Fuzzy Soft set, Soft Set, Fuzzy soft transaction, Fuzzy soft Association pattern*

I. INTRODUCTION

The databases of organizations today consist of large volumes of data piled up due to every day transactions. The managers, scientists, engineers and other decision makers of these organizations are deeply concerned with extracting information and knowledge from these databases for decision making and other applications. A number of data mining techniques are reported in the literature (Atanassov 1994) for extracting information and knowledge from these databases. Association rule mining techniques are widely used by various decision makers of the organizations to explore association patterns in these databases. A good number of algorithms are reported in the literature (Agrawal et al. 1993; Agrawal and Srikant 1994; Mustafa et al. 2006) for mining association rules in databases under deterministic conditions. However the real world data in the fields of engineering, science, technology, biology, medicine and business etc. contains lot of uncertainty, vagueness and impreciseness. The presence of different types of uncertainty in these data sets poses challenges for decision making and mining patterns in these data sets.Various types of theories Eg. theory of Probability, fuzzy sets(Chen and Weng 2009; Zadeh 1965), intuitionitic fuzzy sets(Atanassov 1986), vague sets(Gau and Buehrer 1993) and Rough set (Pawlak 1982) are being employed for dealing with the uncertainties. These theories have been used to deal with different types of uncertainty in data for decision making as well as development of data mining approaches. A number of attempts are reported for development of fuzzy set (Intan 2006), rough set (Guan et al. 2003) and vague set (Lu et al. 2007) approaches for association rule mining .However these rough set, fuzzy set, vague set and probability theories have their inherent difficulties in handling various kinds of uncertainty.Consequently, (Molodtsov 1999) proposed a completely new approach for modeling vagueness and uncertainty called soft set theory which is free from the difficulties present in existing methods was reported in the literature. Some researchers have initiated (Molodtsov 1999; Xiao et al. 2005; Maji et al. 2003)the work with some proposition on soft set operation (Ali et al. 2009) pointed out some new assertions on soft set. With the establishment and development of soft set theory, its wide applications are reported in recent years and are extended to data analysis (Zou and Xiao 2008), decision-making (Maji et al. 2002; Roy and Maji 2007; Cagman and Enginoglu 2010), evaluation(Xiao et al. 2012), medical diagnosis(Borgohain and Das 2010). In soft set theory, membership is decided by adequate parameters, rough set theory employs equivalence classes, whereas fuzzy set theory depends upon grade of membership. Although theories are quite distinct yet deal with vagueness. Joint application of these theories may result in a fruitful way (Lin 1996).Especially on the issue of combination between fuzzy sets, soft sets, rough set, Intuitionistic fuzzy set in these field many researchers have proposed some new concepts on it. Research worker have established mathematical approaches of vagueness such as fuzzy soft set(Maji et al. 2001; Cagman et al. 2010; Çagman et al. 2011),soft groups (Aktas and Cagman 2007), Soft set relations and functions(Babitha and Sunil 2010), soft matrix theory(Cagman and Enginoglu 2010), soft lattices(Karaaslan et al. 2012), Bijective soft set (Gong et al. 2010) and application of soft set to lattices(Nagarajan and Meenambigai 2011). Due to various kinds of uncertainty in data and their relationship. The association rule mining algorithms suffers from the problems of under-prediction and over-prediction of association patterns.Recently (Herawan and Deris 2011) association rule mining approach for transactional data set which can be represented as a soft set is reported in the literature. The research workers have used the concept of parameter co-occurrences in a transaction to define the notion of regular and maximal association rule between two sets of parameters. There algorithms is able to take care of uncertainty arising due to non consideration of parameters for association rule mining .However ,there is a great scope for further improvement by addressing the other issues of uncertainties like imprecision due to non-consideration of grade membership of items in a transaction and other subjective, objective uncertainty involve in a data and their relationship.In order to enhance the precision in soft set approach for association rule mining(Herawan and Deris 2011) an extension is proposed to capture the uncertainty,item relationship in data sets in this paper.Two sources of uncertainity are considered the degree of individual items importances and degree of association among the items in addition to uncertainity arising due to ignorance of parameters. The first is the degree of individual item importance, or item multiplicity or quantity. The item multiplicity or quantity can be classified as randomness, since the number of items occurring in one transaction can be different from other transactions. We allow the multiplicity/quantity of items to affect the outcome of the association rule mining process. The second is the degree of inter-relationships among the items. Degree of interrelationships can be considered as vagueness and inconsistency, since their values can be set differently due to different data sources. In this paper, such differentiation in the data sets is provided, thus making the association rule mining process more generalized.

In view of above the soft set approach developed by earlier research workers (Herawan and Deris 2011) is modified and extended to develop fuzzy soft set approach for mining association patterns to address the issues of under and over prediction of association rules. The paper is organized as follows. Section 2 describes fundamental concept of soft set theory and association rules mining. Section 3 describes soft set approach for mining association patterns .Section 4 shows that by using Fuzzy soft set theory, association patterns are discovered and also shows comparison of soft set and fuzzy soft set association patterns. Section 5 shows the results of experiment on real data set of air pollution. Finally, the conclusion of this work.

2. Preliminaries

2.1. Association rules (Herawan and Deris 2011)

Let $I = \{i_1, i_2, \ldots, i|A|\}$, for $|A| > 0$ refers to the set of literals called set of items and the set $D = \{T_1, T_2, \ldots, T|U|\}$, for $|U| > 0$ refers to the transactional dataset, where each transaction $T \in D$ is a list of distinct items $T = \{i_1, i_2, \ldots, i|M|\}$, $1 \leq |M| \leq |A|$ 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,

Cfi $(X \Rightarrow Y) = \dfrac{\sup(X \cup Y)}{\sup(X)}$ **.(Agrawal et al. 1993; Agrawal and Srikant 1994)**

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 rules are said to be strong if it meets the minimum confidence threshold.

2.2. Soft sets

In this section U refers to an initial universe, E is a set of parameters; P (U) is the power set of U (Atanassov 1986; Gau and Buehrer 1993; Molodtsov 1999).

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 FA 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 $x \subset E$. 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 $F_A = \{(x2, \{u2, u4\}), (x4, U)\}$

2.3 Fuzzy sets

In this subsection, we present the basic definitions of fuzzy set theory (Zadeh 1965) that is useful for subsequent discussions (Atanassov 1986). More detailed explanations related to this theory may be found in earlier studies(Chen and Weng 2009).

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).

2.4 Fuzzy soft sets

The theory of fuzzy soft set, a more generalized concept, (Maji et al. 2001) is a combination of fuzzy set and soft set. Many researchers(Roy and Maji 2007; Çagman et al. 2011) have contributed towards fuzzification of the notion of soft set.

Definition 3(Yao et al. 2008) 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 $A \subset E$. A pair (*F*, *A*) is called a fuzzy soft set over *U*, where *F* is a mapping given by *F*: $A \to F$ (*U*).

Example 2(Yao et al. 2008)**:** 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 =$\{(h_1, 1), (h_2, 0.4), (h_3, 1), (h_4, 0.4), (h_5, 0.6), (h_6, 0.8)\}$.

2.5. Transaction

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.

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

2.5.1. 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.

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

2.5.2. 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.

2.5.3. 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.

2.5.4. 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.

2.5.5. 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……

2.5.6 Fuzzy Association Pattern

Fuzzy Association pattern represents the association of fuzzy items and the fuzzy membership of each Association pattern is calculated for each item. It is represented by

$(P_i, \mu(P_i)) = ((I_{r_1} \cup I_{r_2} \cup I_{r_3} \cup .....I_{r_m}), \mu(P_i))$,

$\mu(P_i) = Min\{\mu(I_{r_1}), \mu(I_{r_2}),.......\mu(I_{r_m})$

Where P_i =association of items

$\mu(P_i)$ =membership of each Items in Pattern P_i.

2.5.7 Soft Association Pattern

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

2.5.8 Fuzzy Soft Association Pattern

Fuzzy Soft Association Pattern represents the association of fuzzy items with their parameter e .It is calculated as

$(P_i, (e_i, \mu(e_i))) = ((I_{r_1} \cup I_{r_2} \cup I_{r_3} \cup ......I_{r_m}), (e_i, \mu(e_i)))$

$\mu(e_i) = Min\{\mu(e_{r_1}), \mu(e_{r_2}),.......\mu(e_{r_m})\}$

Based on the above concepts the soft set approaches for mining associations are given in subsequent sections.

III. SOFT SET APPROACH FOR MINING ASSOCIATION PATTERNS

An Illustrative Example 3 is given below to understand well the concept of the soft set approach for mining association patterns.

Example 3: There is a data set consisting of 10 transactions which contains quality of houses.

Where Quality = {Beautiful, Modern} and there are 2 parameters Beautiful and Modern denoted by $e_1$and e_2.

Table: 1 A Qualified Data Transaction (M)

Transaction_Id	Houses, Green surrounding, Community hall with their parameters
T_1	$\{(h_1, e_1, e_2)\}$
T_2	$\{(h_2, e_1, e_2)\}$
T_3	$\{(h_3, e_1, e_2), (g_3, e_1', e_2')\}$
T_4	$\{(h_4, e_1, e_2), (g_4, e_1', e_2'), (c_4, e_1'', e_2'')\}$
T_5	$\{(h_5, e_1, e_2), (g_5, e_1', e_2')\}$
T_6	$\{(h_6, e_1, e_2), (c_6, e_1'', e_2'')\}$
T_7	$\{(h_7, e_1, e_2), (g_7, e_1', e_2')\}$
T_8	$\{(h_8, e_1, e_2), (g_8, e_1', e_2'), (c_8, e_1'', e_2'')\}$
T_9	$\{(h_9, e_1, e_2), (g_9, e_1', e_2'), (c_9, e_1'', e_2'')\}$
T_{10}	$\{(h_{10}, e_1, e_2), (g_{10}, e_1', e_2'), (c_{10}, e_1'', e_2'')\}$

In Table 2, the fuzzy membership values of parameters $e_1, e_2, e_1', e_2', e_1'', e_2''$ is given.

Table: 2 Fuzzy values of parameters

	House		Green_Surrounding		Community Hall	
Trans_Id	e_1	e_2	e_1'	e_2'	e_1''	e_2''
T_1	0.4	1.0	0	0	0	0
T_2	0.6	1.0	0	0	0	0
T_3	1.0	1.0	0.8	0.2	0	0
T_4	1.0	0.5	0.5	0.5	0.6	0.4
T_5	0.3	0.7	0.4	0.6	0	0
T_6	0.6	0.6	0	0	1.0	0.5
T_7	1.0	1.0	0.6	0.3	0	0
T_8	0.4	0.6	0.5	0.4	0.5	0.4
T_9	0.2	1.0	0.5	0.6	0.6	0.4
T_{10}	1.0	0.7	1.0	1.0	1.0	1.0

The process is started from a given transactional database as shown in Table 1.The set of qualified transactions is given by M = $\{T_1, T_2, T_3, T_4 ......T_{10}\}$ Employing soft set theory and Apriori algorithm on transactions in Table 1 we calculate candidate itemsets and their support . The support threshold is used to determine frequent itemsets and then the confidence of frequent itemsets is calculated. The support threshold is denoted by β_k for k-itemsets, k=1(1)3. β_k can be taken same or different for each I_k , k=1(1)3. The support and confidence of frequent itemset are given in Tables 3, 4, 5 respectively.

Table 3: Frequent 1-Itemsets

Min_Sup =1	
1-itemsets	**Support**
$(house, e_1)$	4
$(house, e_2)$	5
$(green\_surr, e_1)$	1
$(green\_surr, e_2)$	1
$(comm\_hall, e_1)$	2
$(comm\_hall, e_2)$	1

Table 4: Frequent 2-Itemsets

Min_Sup=1		
2-itemsets	**Support**	**Confidence**
$(house \to green\_surr, (e_1))$	1	25%
$(house \to comm\_hall, (e_1))$	1	25%

Table 5: Frequent 3-Itemsets

Min_Sup=1		
3-itemsets	**Support**	**Confidence**
$(house, green\_surr \to comm\_hall, (e_1))$	1	100%

This approach has been modified to develop Fuzzy soft set approach for mining association patterns which is presented in the next section.

IV. FUZZY SOFT SET APPROACH

In this section we present the applicability of Fuzzy soft set theory for mining association patterns. We recall the basic notions of soft sets and fuzzy soft sets. The pre-requisite of using Fuzzy soft set approach for mining transaction is that the transactional data set needs to be transformed into soft transaction, to generate soft transactional dataset. The soft transactional data set is transformed into Fuzzy soft transaction using the concept of fuzzy membership of parameter for each item in the transaction to generate fuzzy soft transactional data set. In Real life applications of fuzzy soft set approach for decision making, decision maker has to choose threshold value in advance. Let U be an initial universe of objects and E(U) (simply denoted by E) the set of parameters in relation to objects in U. The (Intan 2006) algorithm and (Herawan and Deris 2011) are employed to propose the algorithm for mining fuzzy soft association pattern. The steps of the proposed algorithm are described as given below:

Step-1:

Determine $\delta \in$ {1, 2, 3,..,n} (maximum number of items in a qualified transaction).δ is a membership threshold to determine maximum number of items in a transaction by which the transaction may or may not be considered in the process of generating rules. In this case, the process just considers all transactions with the number of items along with their membership in the transactions less than or equal to δ .Formally, let **D** be a universal set of transactions. **M** $\subseteq$ **D** is considered as a subset of qualified transactions for generating rules that the number of items along with their membership in its transactions is no greater than δ as defined:

$$\mathbf{M} = \{T \mid \mathcal{M}(T) \le \delta,\ T \in \mathbf{D}\}, \qquad (1)$$

Where $\mathcal{M}$(T) is the number of items in transaction T.

Step 2: Identify the parameter e ∈ E, to determine soft transactions, is denoted by $\bar{T}$. Let (F, E) be a soft set over the universe U and $X \subseteq E$.

Step 3: Convert the transaction T, T⊆M into Soft transaction. $\mathrm{M}\ (x)$ =Total quantitative measure of all x∈T.

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

The soft transaction is given by

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

Step 4: Transform Soft transaction to Fuzzy Soft Transaction set T'' as given below.

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

Where $\mu(e)$ is degree of membership of e (attribute) for each item in $T''(e)$

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

Step-5:

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-6:

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 7: Given the fuzzy membership of each pattern P_k for all T'' (Fuzzy Soft transaction) select fuzzy Soft association pattern from T'' as given below:

$(P_i,(e_i,\mu_j(e_i))) = ((I_{r_1} \cup I_{r_2} \cup I_{r_3} \cup ......I_{r_m}),(e_i,\mu_j(e_i)))$, $r_j \neq r_u$, (j,u)=1(1)m

$\mu_j(e_i) = Min\{\mu(e_{r_1}),\mu(e_{r_2}),.......\mu(e_{r_m})\}$, j=1(1)N.

Here, for k=1 $P_1 = I_{r_1}, r_1 = 1(1)m$

For k=2 $P_2 = I_{r_1} \cup I_{r_2}, r_1 \neq r_2$, r_1, r_2=1(1)m and so on.

where N is the total number of transactions.

Step 8: Calculate Support for Every Pattern P_k by computing sum of memberships of each patterns as given below:

Fuzzy $\sup port(P_k) = \sum_{j=1}^{N} \mu_j(P_k)$, ∀ T ∈M

Step-9:

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 go to 12.

Step 10:

If k<2 Goto 11

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

Store confidence of P_k for frequent k-itemsets L_k.

Step-11:

Set *k*=*k*+1, and if *k* > δ, then go to Step-12.

Else goto step 6.

Step-12: Stop.

The transaction set in Table 1 of example: 3 is used to explain the concept of proposed algorithm for mining fuzzy soft associations as given below:

Example 4: Table 1 transactional Data set is taken: U= {house, green surrounding, community hall} and the set of parameters is given by A= {Beautiful, Modern} be consisting of the parameters that Mr. X is interested in buying a house. That means out of available houses with quality in U, Mr. X want to buy the house which qualifies with the attributes in A to the utmost extent. Now all the available information on houses under consideration can be formulated as a Fuzzy soft set F = (F, A) describing "attractiveness of houses" that Mr. X is going to buy. In Real life applications of fuzzy soft set based decision making, decision maker has to decide threshold value in advance and represent their requirement using membership given in Table 2.

Step 1:

There is a data set consisting of 10 transactions which contains two categories i.e. T= {Quality, House options),
Where Quality = {Beautiful, Modern} And
House options= {house, green surrounding, community hall}.

Step 2:

Identifying the parameter e ∈ E from table 1. In this example only 2 parameters e_1, e_2 are associated with each item house, green surrounding, community hall.

Step 3:

Let soft transaction denoted by $\overline{T_i} = \{\psi_i, \phi_i\}$ ψ_i =set of Items in i[th] transaction, ϕ_i =set of parameter in the i[th] transaction.

$$\overline{\overline{T}} = \{(x,\mu(x) \mid \forall x \in I\}$$

$\overline{\overline{T}}_1=\{\psi_1,\phi_1\}, \overline{\overline{T}}_2=\{\psi_2,\phi_2\}, \overline{\overline{T}}_3=\{\psi_2,\phi_3\}, \overline{\overline{T}}_4=\{\psi_4,\phi_4\}, \overline{\overline{T}}_5\{(\psi_5,\phi_5\}, \overline{\overline{T}}_6=\{\psi_6,\phi_6\}$

$\overline{\overline{T}}_7=\{\psi_7,\phi_7\}, \overline{\overline{T}}_8=\{\psi_8,\phi_8\}, \overline{\overline{T}}_9=\{\psi_9,\phi_9\}, \overline{\overline{T}}_{10}=\{\psi_{10},\phi_{10}\}$

Step 4: Transform Soft transaction to Fuzzy Soft Transaction set, Fuzzy Soft transaction denoted by T''.

$$T''\{house(Beautiful)\}=\{(\psi_1,\mu(e_1)),(\psi_2,\mu(e_1))(\psi_3,\mu(e_1))(\psi_4,\mu(e_1))(\psi_5,\mu(e_1))(\psi_6,\mu(e_1))(\psi_7,\mu(e_1))(\psi_8,\mu(e_1))(\psi_9,\mu(e_1))(\psi_{10},\mu(e_1))\}$$

$$T''\{house(Modern)\}=\{(\psi_1,\mu(e_2)),(\psi_2,\mu(e_2)),(\psi_3,\mu(e_2)),(\psi_4,\mu(e_2)),(\psi_5,\mu(e_2)),(\psi_6,\mu(e_2)),(\psi_7,\mu(e_2)),(\psi_8,\mu(e_2)),(\psi_9,\mu(e_2)),(\psi_{10},\mu(e_2))\}$$

$$T''\{green\_surr(Beautiful)\}=\{(\psi_3,\mu(e_1')),(\psi_4,\mu(e_1')),(\psi_5,\mu(e_1')),(\psi_7,\mu(e_1')),(\psi_8,\mu(e_1')),(\psi_9,\mu(e_1')),(\psi_{10},\mu(e_1'))\}$$

$$T''\{green\_surr(Modern)\}=\{(\psi_3,\mu(e_2')),(\psi_4,\mu(e_2')),(\psi_5,\mu(e_2')),(\psi_7,\mu(e_2')),(\psi_8,\mu(e_2')),(\psi_9,\mu(e_2')),(\psi_{10},\mu(e_2'))\}$$

$$T''\{comm\_hall(Beautiful)\}=\{(\psi_4,\mu(e_1'')),(\psi_6,\mu(e_1'')),(\psi_8,\mu(e_1'')),(\psi_9,\mu(e_1'')),(\psi_{10},\mu(e_1''))\}$$

$$T''\{comm\_hall(Modern)\}=\{(\psi_4,\mu(e_2'')),(\psi_6,\mu(e_2'')),(\psi_8,\mu(e_2'')),(\psi_9,\mu(e_2'')),(\psi_{10},\mu(e_2''))\}$$

So on…..

Step 5: The process is started by looking for support of P_k itemsets for which k is set equal to 1.

Step-6: Since $\delta=3$, then $k \in \{1,2,3\}$.It is arbitrarily given that the system just considers support of k-itemsets threshold value of each parameter for k=1,2,3.The collections of Fuzzy approximation of values according to each transaction, the support for quality of each house are shown in Table 2.

Step-7: Given the fuzzy membership of each pattern P_k for all T''(Fuzzy Soft transaction) select fuzzy Soft association pattern from T''.we have decided the threshold value of each parameter according to their priority:

When k=1;

For 1-Itemsets:

$$\{house,(beautiful)\}=\{(0.4/T1),(0.6/T2),(1.0/T3),(1.0/T4),(0.3/T5),(0.6/T6),(1.0/T7),(0.4/T8),(0.2/T9),(1.0/T10)\}=6.5$$

$$\{house(Modern)\}=\{(1.0/T1),(1.0/T2),(1.0/T3),(0.5/T4),(0.7/T5),(0.6/T6),(1.0/T7),(0.6/T8),(1.0/T9),(0.7/T10)\}=8.1$$

$$\{green\_surr(Beautiful)\}=\{(0.8/T3),(0.5/T4),(0.4/T5),(0.6/T7),(0.5/T8),(0.5/T9),(1.0/T10)\}=4.3$$

$$\{green\_surr(Modern)\}=\{(0.2/T3),(0.5/T4),(0.6/T5),(0.3/T7),(0.4/T8),(0.6/T9),(1.0/T10)\}=3.6$$

$$\{comm\_hall(Beautiful)\}=\{(0.6/T4),(1.0/T6)(0.5/T8)(0.6/T9)(1.0/T10)=3.7$$

$$\{comm\_hall(Modern)\}=\{(0.4/T4),(0.5/T6)(0.4/T8)(0.4/T9)(1.0/T10)=2.7$$

In 1 Itemsets, only those patterns are considered for further processing which satisfy the threshold value greater than equal to 1.

When k=2;

For 2-Itemsets:

$$\{house,green\_surr,(e_1)\}=\{(0.8/T3),(0.5/T4),(0.3/T5),(0.6/T7),(0.4/T8),(0.2/T9),(1.0/T10)\}=3.8$$

$$\{house,comm\_hall,(e_1)\}=\{(0.6/T4),(0.6/T6),(0.4/T8),(0.2/T9),(1.0/T10)\}=2.8$$

$$\{green\_surr,comm\_hall,(e_1)\}=Nil$$

$$\{house,green\_surr(e_2)\}=\{(0.2/T3),(0.5/T4),(0.6/T5),(0.3/T7),(0.4/T8),(0.6/T9),(0.7/T10)\}=3.3$$

$$\{house,comm\_hall(e_2)\}=\{(0.6/T4),(0.6/T6)(0.4/T8)(0.2/T9)(1.0/T10)=2.8$$

$$\{green\_surr,comm\_hall(e_2)\}=\{(0.4/T4),(0.5/T6)(0.4/T8)(0.4/T9)(1.0/T10)=Nil$$

In 2-Itemsets, only those patterns are considered for further processing which satisfy the threshold value greater than equal to 1.

When k=3,

For 3-Itemsets:

$$\{house,green\_surr,comm\_hall(e_1)\}=\{(0.5/T4),(0.4/T8),(0.2/T9),(1.0/T10)\}=2.1$$

$$\{house,green\_surr,comm\_hall(e_2)\}=\{(0.4/T4),(0.4/T8),(0.4/T9),(0.7/T10)\}=1.9$$

In 3-Itemsets, only those patterns are considered for further processing which satisfy the threshold value greater than equal to 1.

Step 8: Support for every Pattern P_k are calculated as given:

The fuzzy support of all P_k are calculated in table 6, 7, 8:

Step-9: If the fuzzy support of pattern P_k is greater than equal to B_k then it is frequent .The frequent fuzzy soft association patterns are given in Table 6:

Table 6: Frequent Item sets.

1-itemsets	Support	Confidence
$(house, e_1)$	6.5	-
$(house, e_2)$	8.1	-
$(green\_surr, e_1)$	4.3	-
$(green\_surr, e_2)$	3.6	-
$(comm\_hall, e_1)$	3.7	-
$(comm\_hall, e_2)$	2.7	-
2-Itemsets		
$(house \rightarrow green\_surr, (e_1))$	3.8	58%
$(house \rightarrow comm\_hall, (e_1))$	2.8	43%
$(house \rightarrow green\_surr, (e_2))$	3.3	40%
$(house \rightarrow comm\_hall, (e_2))$	2.8	34%
3-Itemsets		
$\{house, green\_surr \rightarrow comm\_hall(e_1)\}$	2.1	55.2%
$\{house, green\_surr \rightarrow comm\_hall(e_2)\}$	1.9	57.5%

Step 10: The confidence of these frequent association patterns is given in table 6.
Step-11: This step is just to increment the value of K in which if k > δ , then the process goes to step-12. Else goto step 6.
Step-12: Stop.
Comparison of results of Soft Set and Fuzzy Soft Set association Patterns is given in Table 7 below:

Table 7: Comparison of soft set and fuzzy soft set based frequent patterns.

Frequent itemsets	Fuzzy Soft set		Soft set	
2-Itemsets	**Supp**	**Conf**	**Supp**	**Conf**
$(house \rightarrow green\_surr, (e1))$	3.8	58%	1	25%
$(house \rightarrow comm\_hall, (e1))$	2.8	43%	1	25%
$(house \rightarrow green\_surr, (e2))$	3.3	40%	-	-
$(house \rightarrow comm\_hall, (e2))$	2.8	34%	-	-
3- Itemsets	**Supp**	**Conf**	**Supp**	**Conf**
$\{house, green\_surr \rightarrow comm\_hall(e1)\}$	2.1	55.2%	1	100%
$\{house, green\_surr \rightarrow comm\_hall(e2)\}$	1.9	57.5%	-	-

The comparisons of above results show us very significant difference in support and confidence of the two approaches. This implies that soft set based association rule mining does not give us the true picture of confidence when the vagueness/impreciseness of parameters are not taken into account. In comparison to fuzzy soft set approach the results by simply soft set approach show under prediction of association patterns in the present example. This implies that incorporating degree of relationship into account is helpful in addressing the issues of under prediction of association patterns by soft set only.

V. EXPERIMENT ON AIR POLLUTION DATASET

We further explain the concept and proces of soft set and fuzzy soft set approach for mining association rules by performing experiment on real data set of air pollution from Maharashtra pollution control board (M P C B) 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(M P C B).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 Micrograms per cubic meter ($\mu g/m^2$) 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.

According to enviornment , acceptable limit of So2, Nox, RSPM is 80,80,100 respectively
For calculating Fuzzy soft set of Good and Dangerous condition on the dataset following formulas are used:

Membership of 1-Itemsets	Good Condition	Dang Condition
So2	$\mu_s(x)=(80-x)/80$	$\mu_s(x)=(x-80)/80$
Nox	$\mu_n(y)=(80-y)/80$	$\mu_n(y)=(y-80)/80$
RSPM	$\mu_r(z)=(100-z)/100$	$\mu_r(z)=(z-100)/100$

For 2 Itemsets:

$x+y <=80$	Good Condition: $G(x,y)= 80-(x+y)/80$	
	Dangerous Condition: $D(x,y)=$	0
$x+y >=80$	Good Condition: $G(x,y)=0;$	
	Dangerous Condition: $D(x,y)=(x+y)-80/80$	
$(x+z)$ or $(y+z) <=180/2$	Good Condition: $G(x,z)= 90-(x+y)/90$	
	Dangerous Condition: $D(x,z)=$	0
$(x+z)$ or $(y+z) >=180/2$	Good Condition: $G(x,z)=0;$	
	Dangerous Condition: $D(x,z)=(x+z)-90/90$	

Using the given acceptable limit, the taxonomy of data is prepared as given below:
T={Good Condition, Dangerous Condition, Area}
Area={A1,A2,A3}
A1= Chinchwad, A2=Karve, A3=All(Chinchwad + Karve)
Taking threshold value of fuzzy soft set as 33% and soft set as 50% we get the frequent association patterns as shown in Table 8.

The Table: 8 the frequent association patterns obtained by soft set and fuzzy soft set approaches.

	Soft set(σ=50%)		Fuzzy soft set(σ=33%)	
1 itemsets	**Support**	**Confidence**	**Support**	**Confidence**
{SO2,(Good,A1)}	99.65%	-	60.78%	-
{NOx,(Good,A1)}	89.47%	-	38.02%	-
{RSPM,(Good,A1)}	**58.59%**	**-**	-	-
{SO2,(Good,A2)}	100%	-	81.67%	-
{NOx,(Good,A2)}	83.23%	-	38%	-
{RSPM,(Dang,A2)}	60.26%	-	42.72%	
{SO2,(Good,A3)}	99.85%	-	72.23%	-
{NOx,(Good,A3)}	86.05%	-	38.01%	-
{RSPM,(Dang,A3)}	51.74%	-	33.72%	-
2-itemsets	**Support**	**Confidence**	**Support**	**Confidence**
{SO2,NOx,(Good,A1)}	88.94%	89.26%	63.12%	100%
{SO2,RSPM,(Good,A1)}	58.42%	58.62%	-	-
{NOx,RSPM,(Good,A1)}	56.31%	62.94%	-	-
{SO2,NOx,(Good,A2)}	83.23%	83.23%	81.67%	100%
{SO2,NOx,(Good,A3)}	85.81%	85.95%	84.49%	100%
3-itemsets	Support	Confidence	Support	Confidence
{SO2,NOx,RSPM,(Good,A1)}	**56.31%**	**56.51%**	-	-

Comparing the results by both the approaches we observe a significant change in confidence levels and support levels. In Table 8, entries in the third and fifteenth rows containing patterns {RSPM,(Good,A1)},{SO2,NOx,RSPM,(Good,A1)}are found to be frequent by soft set approach. This implies over prediction of association patterns by soft set approach. In this example, in comparison to fuzzy soft set approach, the results obtained by simply soft set approach show over prediction of association patterns. Thus it is observed that fuzzy soft set approach is helpful in addressing the issues of over prediction of association rules in this experiment.

VI. CONCLUSION

In this paper, we have proposed a fuzzy soft set approach for mining association patterns from a transactional dataset, which has been a successfully demonstrated using suitable example. We are able to incorporate the impreciseness of parameters in terms of fuzzy sets in a soft set transaction. Thus we get a better picture of support and confidence levels by fuzzy soft set approach for mining association patterns. Also the fuzzy soft set approach is quite effective in addressing the issues of under and over prediction of association patterns. Similar approaches can be developed to incorporate degree of relationships among the items of the set and impreciseness of parameters to get better results.

REFERENCES

[1]. Agrawal R, Imieliński T, Swami A Mining association rules between sets of items in large databases. In: International conference on Management of data, 1993. ACM SIGMOD pp 207-216
[2]. Agrawal R, Srikant R Fast Algorithms for Mining Association Rules in Large Databases. In: Proceedings of 20th International Conference on Very Large Data Bases, 1994. pp 487-499
[3]. Aktas H, Cagman N (2007) Soft Sets and soft Groups. Information Sciences 177:2726-2735
[4]. Ali MI, Feng F, Liu X, Min WK, Shabir M (2009) On some new operations in soft set theory. Computers and Mathematics with Apllications 57:1547-1553
[5]. Atanassov K (1986) Intuitionistic fuzzy sets. Fuzzy Sets and Systems 20:87-96
[6]. Atanassov K (1994) Operators over interval valued intuitionistic fuzzy sets. Fuzzy Sets and Systems 64:159–174
[7]. Babitha KV, Sunil JJ (2010) Soft set relations and functions. Computer and mathematics with Applications 60 (7):1840-1849
[8]. Borgohain R, Das PK (2010) Application of Fuzzy Soft set in medical diagnosis using Fuzzy arithmetic operations on Fuzzy Number, . SIBCOLTEJO 05:107-116
[9]. Cagman N, Citak F, Enginoglu S (2010) Fuzzy parameterized fuzzy soft set theory and its applications. Turkish Journal of Fuzzy Systems 1 (1)
[10]. Cagman N, Enginoglu S (2010) Soft matrix theory and its decision making. Computers and Mathematics with Applications 59:3308-3314
[11]. Çagman N, Enginoglu S, Citak F (2011) Fuzzy soft set theory and its applications. Iranian Journal of Fuzzy Systems 8 (3):137-147
[12]. Chen YL, Weng CH (2009) Mining fuzzy association rules from questionnaire data. Knowledge-Based Systems 22 (1):46–56.
[13]. Chen Z (2001) Data Mining and Uncertain Reasoning: An Integrated Approach. Wiley-Inter-Science,
[14]. Gau WL, Buehrer DJ (1993) Vague sets. IEEE Trans System Man Cybernet 23 (2):610–614
[15]. Gong K, Xiao Z, Zhang X (2010) The bijective soft set with its operations. Computers and Mathematics with Applications 60:2270–2278
[16]. Guan JW, Bell DA, Liu DY The rough set approach to association rule mining. In: The proceedings of the Third IEEE International Conference on Data mining(ICDM'03), 2003. pp 529-532
[17]. Herawan T, Deris MM (2011) A Soft Set approach for association rule mining. Knowledge Based systems 24:186-195
[18]. Intan R (2006) An Algorithm For Generating Single Dimensional Fuzzy Association Rule Mining JURNAL INFORMATIKA 7 (1):61 - 66
[19]. Karaaslan F, Cagman N, Enginoglu S (2012) Soft Lattices. Journal of New Results in Science 1:5-12
[20]. Lin TY (1996) A set theory for soft computing a unified view of fuzzy sets via neighborhoods. in Proceedings of the IEEE International Conference on Fuzzy Systems 2:1140-1146
[21]. Lu A, Ke Y, Cheng J, Ng W Mining vague association rules In: Lecture Notes in Computer Science, 2007. pp 891-897
[22]. M P C B http://mpcb.gov.in/envtdata/demoPage1.php. Accessed 1 January 2010
[23]. Maji PK, Biswas R, Roy AR (2001) Fuzzy soft sets. Journal of Fuzzy Mathematics vol. 9, no. 3:589–602
[24]. Maji PK, Biswas R, Roy AR (2003) Soft Set Theory. Computer and Mathematics with Application 45 555-562.
[25]. Maji PK, Roy AR, Biswas R (2002) An application of soft sets in a Decision making problem, . Computers and Mathematics with Applications, 44:1077– 1083.
[26]. Molodtsov D (1999) Soft set theory-first results. Computers and Mathematics with Applications, 37.:19–31.
[27]. Mustafa MD, Nabila NF, Evans DJ, Saman MY, Mamat A (2006) Association rules on significant rare data using second support. International Journal of Computer Mathematics 83 (1) 69–80

[28]. Nagarajan EKR, Meenambigai G (2011) An application of soft sets to Lattices. Kragujevac Journal of Mathematics 35 (1):75-87
[29]. Pawlak Z (1982) Rough sets. International Journal of computer and Information Science 11:341–356
[30]. Roy AR, Maji PK (2007) A fuzzy soft set theoretic approach to decision making problems. Journal of Computational and Applied Mathematics 203:412-418
[31]. Xiao Z, Chen L, Zhong B, Ye S Recognition for soft information based on the theory of soft sets. In: Proceedings of ICSSSM '05, 2005. IEEE, pp 1104-1106
[32]. Xiao Z, Yang X, Niu Q, Dong Y, Gong K, Xia. S, Pang Y (2012) A new evaluation method based on D–S generalized fuzzy soft sets and its application in medical diagnosis problem. Applied Mathematical Modelling 36 (10):4592-4604
[33]. Yao BX, Liu JL, Yan RX Fuzzy Soft Set and Soft Fuzzy Set. In: Proceedings of the IEEE Fouth International Conference on Natural Computation, 2008.
[34]. Zadeh LA (1965) Fuzzy set. Information and Control 8:338–353
[35]. Zou Y, Xiao Z (2008) Data analysis approaches of soft sets under incomplete information. Knowledge Based Systems 21:941-945