INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

Smart Decision in intelligent Systems Using Weighted Frequent Itemset Mining Algorithm

GATRAM MALLIKHARJUNA RAO [1], Dr. Dr SIKHAKOLLI GOPI KRISHNA [2]

[1]Assistant professor, [2] professor

CSE Department, Sri Mittapalli College of Engineering, Guntur, Andhra Pradesh-522233

ABSTRACT

Intelligent decision is the key technology of smart systems. Data mining technology has been playing an increasingly important role in decision making activities. The introduction of weight makes the weighted frequent itemsets not satisfy the downward closure property any longer. As a result, the search space of frequent itemsets cannot be narrowed according to downward closure property which leads to a poor time efficiency. In this paper, the weight judgment downward closure property for weighted frequent itemsets and the existence property of weighted frequent subsets are introduced and proved first. The Fuzzy-based WARM satisfies the downward closure property and prunes the insignificant rules by assigning the weight to the itemset. This reduces the computation time and execution time. This paper presents an Enhanced Fuzzy-based Weighted AssociationRuleMining(E-FWARM) algorithm for efficient mining of the frequent itemsets. The pre-filtering method is applied to the input dataset to remove the item having low variance. Data discretization is performed and E-FWARM is applied for mining the frequent itemsets. The experimental results show that the proposed E-FWARM algorithm yields maximum frequent items, association rules, accuracy and minimum execution time than the existing algorithms.

Keywords : Frequent itemset mining, Weight judgment, Downward closure property,Intelligent decision, Smart system, Association Rule Mining (ARM),Data mining.

I. INTRODUCTION

Intelligent decision is the key technology of smart systems. Data mining technology has been playing an increasingly important role in decision making activities. FIM (Frequent Itemset Mining), as one of the most hot research topics in datamining, is an important approach to discover association rules in datasets, that is widely used in the field of precision marketing, personalized network optimization medical diagnosis and so on. So far, many perfect and mature FIM algorithms have been proposed for binary databases. However, with the rapid development of data acquisition and data processing technologies, various forms of complex data have emerged, like uncertain data.

Uncertain data means the existence of an item in a transaction is described by a likelihood measure or a probability. As is known, if we adopt a binary data model, then each item in a transaction can either be present or absent. However, in the uncertain data model, the existence of an item in a transaction can be indicated by a probability, thus it allows more information to be captured by the dataset which can lead to more accurate analytical results. However, each coin has two sides. Uncertain data model still have its drawbacks. The first disadvantage is that the size of the dataset would be much larger because of the storage of existence probability. Another disadvantage is that the mining algorithms for uncertain databases will be more complicated and time consuming. Thus, developing efficient mining algorithms for uncertain databases has become a hot research topic in recent years. Many algorithms have been developed to mine frequent itemsets in uncertain databases.

Most existing studies assume that all the items in uncertain databases have the same importance. However, in actual reality, the values and importances of various items are usually different to users. For example, the profits of a costly luxury goods and a cheap living goods cannot be mentioned in the same breath. Consequently, the mining based on only occurrence frequencies or existence probabilities without taking importances or values of items into account is insufficient to identify useful and meaningful patterns. To address this issue, a prominent solution is to let the users assign different weights to items to indicate their relative importances or values. The weight of items can be set by the users according to their professional domain knowledge or specific application requirements to indicate profits, risks, costs and so on. In this context, itemsets with high importances for the users will be discovered. Moreover, the introduction of weights of items can greatly reduce the number of frequent itemsets. However, the downward closure property used for mining frequent itemsets in uncertain databases would not hold any longer because different weights are assigned to items. This means an infrequent itemset may have a frequent superset. As a result, the searching space cannot be narrowed according to the downward closure property any longer which will lead to low time efficiency of FIM algorithms.

In this paper, on the basis of the weight judgment downward closure property, the E-FARM(Enhanced Fuzzy-based weight Association Rule Mining) algorithm is proposed to narrow the searching space of weighted frequent itemsets and improve the time efficiency. Consequently, more useful and meaningful weighted frequent itemsets in uncertain databases can be discovered. The main contributions of this paper are listed as following.

1. The weight judgment downward closure property and the existence property of weighted frequent subsets for uncertain databases are introduced and proved. The weight judgment downward closure property can be used to narrow the searching space of weighted frequent itemsets. The existence property of weighted frequent subsets can ensure all the weighted frequent itemsets be discovered.
2. The E-FWARM algorithm is proposed on the basis of weight judgment downward closure property to narrow the searching space of weighted frequent itemsets and improve the time efficiency.
3. A considerable amount of experiments are conducted on both real-life and synthetic datasets to evaluate the performance of the proposed E-FWARM algorithm in terms of runtime, number of patterns and memory consumption.

The remainder of this paper is organized as follows. Related works is reviewed in Section 2. The preliminaries and problem statement are given in Section 3. In section 4, the related properties are introduced and proved theoretically, and the E-FWARM algorithm is described in detail. Additionally, the completeness and time efficiency of E-FWARM algorithm are also analyzed in this section. Experimental results are discussed in Section 5. Finally, a conclusion is drawn and future work are discussed in Section 6.

II. RELATED WORKS

In this section, related works on frequent itemset mining in uncertain databases and weighted frequent itemset mining in binary databases are briefly reviewed.

A. Frequent Itemset Mining in Uncertain Databases

With the popular use of various data acquisition and communication technologies, a huge amount of data stored in a database may be inaccurate, imprecise, or incomplete in real-life applications, such as wireless sensor network applications or location-based services. To address this issue, developing efficient algorithms to mine patterns in uncertain databases has become a major research topic in recent years and many efficient FIM algorithms for uncertain databases have been proposed. These algorithms can be generally classified into two categories: candidate generate-and-test based uncertain frequent itemset mining and pattern-growth mining.

One way to mine frequent itemsets from uncertain data is to apply the candidate generate-and-test paradigm. For example, Chui et al. proposed U-Apriori algorithm which applies the candidate generate-and-test process to mine frequent itemsets from for uncertain data. Similar to Apriori algorithm for mining precise data, U-Apriori algorithm needs to scan the database frequently and generates a large number of candidate frequent itemsets. Chui and Kao applied the decremental pruning technique to further improve the efficiency of U-Apriori. MBP is an approximation method for uncertain frequent pattern mining based on statistical techniques. IMBP was proposed to more improve the mining speed and memory efficiency of MBP at the cost of losing accuracy.

An alternative to candidate generate-and-test based mining is pattern-growth mining, which avoids generating a large number of candidates. Commonly used pattern-growth mining paradigms are mostly based on hyperlinked structures or tree structures. For example, Aggarwal et al proposed a hyperlinked structure based algorithm called UH-mine to mine frequent patterns from uncertain data. Leung et al.proposed a tree-based mining algorithm called UF-growth which also constructs a tree structure to store the contents of the uncertain datasets, like its counterpart - the FP-growth algorithm for mining precise data. In order to reduce the tree size, Aggarwal et al.proposed the UFP-growth algorithm. To further reduce the tree size, Leung and Tanbeer proposed an uncertain frequent pattern mining algorithm called CUFgrowth, which builds a new tree structure called CUF-tree. Leung and Tanbeer introduced the concept of a prefixed item cap and proposed PUF-growth algorithm to mine uncertain frequent patterns which runs faster than CUF-growth. TPC-growth is an advanced version of PUF-growth. It employs an upgraded overestimation method that can tighten upper bounds to expected supports more than PUF-growth. CUFP-Mine is a method for mining exact uncertain frequent patterns without employing recursive call-based pattern growth manners. However, the larger the given database is, the worse the mining performance of CUFP-Mine becomes. AT-Mine is another tree-based efficient approach proposed to overcome the fatal problems of CUFP-Mine. It guarantees more efficient mining performance than that of CUFP-Mine, but it still has limitations in runtime and memory performance aspects. U-WFI is a tree-based approach that applies weight factors into uncertain pattern mining. Through weight constraints, the algorithm can find more meaningful uncertain frequent patterns but have limitations in the aforementioned aspects.

State-of-the-art algorithms based on tree structures can cause fatal problems in terms of runtime and memory usage according to the characteristics of uncertain databases and threshold settings because their own tree data structures can become excessively large and complicated in their mining processes. Various approaches have been suggested to overcome such problems. For example, Lee and Yun propose LUNA algorithm which is an exact, efficient algorithm for mining uncertain frequent patterns based on newly proposed list-based data structures and pruning techniques, which can also guarantee a complete set of uncertain frequent patterns to be mined more efficiently without pattern losses.

B. Weighted Frequent Itemset mining in Uncertain Databases

Traditional frequent itemset mining methods have a problem that it does not apply importance of each item obtained from the real world into the mining process. In order to discover more useful and interesting patterns, numerous algorithms have been developed for weighted frequent itemset mining. However, most of these algorithms are proposed for precise datasets or data streams, for example, WAR (Weighted Association Rules) algorithm , WARM (Weighted Association Rule Mining) algorithm , WFIM (Weighted Frequent Itemset Mining) algorithm , WSpan algorithm , WMFP-SW (Weighted Maximal Frequent Pattern mining over data streams based on Sliding Window model) algorithm, MWS (Maximal frequent pattern mining with Weight conditions over data Streams) algorithm , WEP (Weighted Erasable Patterns) mining algorithm and so on. Mining weighted frequent itemsets in uncertain databases have only a few researches. To our knowledge, only two algorithms are proposed to discover weighted frequent itemsets in uncertain datasets. Lee et al. suggested a new tree-based U-WFI (Uncertain Mining of Weighted Frequent Itemsets) algorithm which can mine uncertain frequent itemsets considering item weights from a given uncertain database. As a result, more meaningful itemsets with high importance and existential probabilities can be discovered effectively. Lin et al. proposed HEWI-Uapriori (High Expected Weighted Itemset) algorithm to mine high expected weighted itemsets based on high upperbound expected weighted downward closure property to early prune the search space and unpromising itemsets. Consequently, further research should be conducted to improve the efficiency of mining frequent itemset in uncertain databases.

III. PRELIMINARIES AND PROBLEM STATEMENT

In this section, preliminary definitions are given first and the problem of mining weight frequent itemsets are stated formally.

A. Preliminaries

Let be the uncertain dataset to be analyzed which is composed of a set of transactions, i.e.

DS={T1,T2,…,Tn}, n is the number of transactions in the uncertain dataset . There are a finite set of distinct items I={I1,I2,…,I_n} in the uncertain dataset Ds. Each transaction T_q∈DS , q∈{1,2,…,n } is a subset of I, and q is the unique identifier of transaction T_q, also called its TID (Transaction Identifier). According to the probabilistic model that commonly used for uncertain frequent pattern mining, the uncertainty can be expressed in terms of the existential probability $p(I_j,T_q)$, which indicates that the item I_j exists in T_q with a probability $p(I_j,T_q)$. The existential probability $p(I_j,T_q)$ ranges from a positive value close to 0 (indicating that I_j has an insignificantly low chance to be present in T_q) to a value of 1 (indicating that is definitely present in T_q). In order to assign different importances to distinct items in DS, a weight table is defined as wtable={$w(I_1)$,$w(I_2)$,..$w(I_m)$},where(I_j)∈[0,1], j∈{1,2,…,m} is the weight of the item I_j. If an itemset X contains kdistinct items, then X is called ak–itemset. If X⊆T_q, we can say that the itemset X is contained in the transaction T_q. The minimumexpected weighted support threshold is e∈(0,1]. Table 1 is an example of uncertain databases, which consists of 10 transactions and 6 items. Table 2 is the weight table of the items in Table 1.

Table 1: The uncertain dataset Ds.

TID	Transaction{(Item, Probability)}
1	(A, 0.8), (B, 0.4), (D, 1.0)
2	(B, 0.3), (F, 0.7)
3	(B, 0.7), (C, 0.9), (E, 1.0), (F,0.7)
4	(E, 1.0), (F, 0.5)
5	(A, 0.6), (C, 0.4), (D, 1.0)
6	(A, 0.8), (B, 0.8), (C, 1.0), (F, 0.3)
7	(A, 0.8), (C, 0.9), (D, 0.5), (E, 1.0)
8	(C, 0.6), (E, 0.4)
9	(A, 0.5), (D, 0.8), (F, 1.0)
10	(A, 0.7), (B, 1.0), (C, 0.9), (E, 0.8)

In this section, the example showed in Table 1 are use to explain the preliminary definitions. And the minimum expected weighted support threshold e is0.1

Definition 1 (Item weight) The item weight is a value used to describe the importance of an item, which is set by the users according to their preferences or application scenarios.

Table 2 : The weight table

Item	A	B	C	D	E	F
Weight	0.1	0.8	0.3	1.0	0.6	0.1

Definition 2 (Itemset weight) The weight of itemset X is denoted as w(X), which is the average of the weights of all items contained in itemset X. It can be defined as:

w(X)=$\sum_{Ij\in X} w(Ij)$/|k|

whereI_j is an item in itemset X, and |k| is the number of items in itemset X.

Definition 3 (Itemset probability in a transaction) The existential probability of an itemset X in a transaction T_q is denoted as p(X,T_q), which is the product of existential probabilities of the items contained in itemset X. It can be formally defined as:

$$p(X, T_q) = \prod_{Ij g X} p(Ij, Tq), \qquad (2)$$

where P(I_j,T_q) indicates the existential probability of item I_j in T_q.

Definition 4 (Expected support of an itemset in DS) The expected support of itemset X in dataset DS is denoted as expSup(X).It is the summation of existential probabilities of itemset X in all the transactions which contain itemset X. It can be formally defined as:

$$\text{expSup}(X) = w(X) * \sum_{X\subseteq Tq\wedge Tq\in DS} p(X, Tq) = \sum_{X\subseteq Tq\wedge Tq\in DS} \left(\prod_{Ij\in X} p(Ij, Tq)\right)$$

Definition 5 (Frequent itemset in DS) In uncertain dataset DS, if the expected support of itemset X is greater than or equal to the minimum expected support (the minimum expected support is the product of the minimum expected support threshold δ and the number of transactions in the uncertain dataset |DS|), i.e. expSup(X)≥δ×|DS| , then itemset X is a frequent itemset.

Definition 6 (Expected weighted support of an itemset in DS): The expected weight support of itemset X in dataset DS is denoted as expSup(X). It is the product of the expected support of itemset X and the weight of itemset X. It can be formally defined as:

$$\text{expwSup}(X) = w(X)\times \text{expSup}(X)$$

$$= \sum_{Ij\in X} w(Ij)\times \sum_{X\subseteq Tq\wedge Tq\in DS}\left(\prod_{Ij\in X} p(Ij, Tq)\right)$$

Table 3 : The frequent itemset and weighted frequent itemset table

Itemset	Weight	Frequency	expSup	isFI	expWSup	isWFI
(A)	0.10	6	4.20	Y	0.42	N
(B)	0.80	5	3.20	Y	2.56	Y
(C)	0.30	6	4.70	Y	1.41	Y
(D)	1.00	4	3.30	Y	3.30	Y

(E)	0.60	5	4.20	Y	2.52	Y
(F)	0.10	5	3.20	Y	0.32	N
(AB)	0.45	3	1.66	Y	0.747	N
(AC)	0.20	4	2.39	Y	0.478	N
(AD)	0.55	4	2.20	Y	1.21	Y
(AE)	0.35	2	1.36	Y	0.476	N
(BC)	0.55	3	2.33	Y	1.2815	Y
(BE)	0.70	2	1.50	Y	1.05	Y
(CE)	0.45	3	2.76	Y	1.242	Y
(EF)	0.35	1	1.20	Y	0.42	N
(ABC)	0.40	2	1.27	Y	0.508	N
(ACE)	0.33	2	1.2	Y	0.408	N
(BCE)	0.56	2	1.35	Y	0.765	N

B. Problem statement

Based on the above preliminary definitions, the problem of mining weighted frequent itemsets in uncertain databases can be formulated as following:

The uncertain database to be analyzed is, the user-specified weights of the items in are defined in, and the user-specified minimum expected weighted support threshold is Ɛ.The problem of mining weighted frequent itemsets in the uncertain database is to discover the weighted frequent itemsets considering both the weight and the existential probability constraints. An itemset X is a weighted frequent itemset if the expected weighted support of an itemset X is greater than or equal to the minimum expected weighted support, i.e. expwSup(X)≥Ɛ×|DS |. As mentioned above, the downward closure property cannot be applied directly to narrow the searching space of mining weighted frequent itemsets. Consequently, how to improve the time efficiency of mining weighted frequent itemsets is a major issue for urgent solution.

IV. FUZZY WEIGHTED ASSOCIATION RULE MINING

Let a dataset 'D' comprises a set of transactions $T=\{t_1,t_2,\ldots,t_n\}$with a set of items $I=\{i_1,i_2,\ \ldots\ ,i_{|I|}\}$. A Fuzzy dataset D' includes fuzzy transactions $T'=\{t`_1,t_{2`},\ldots,t_{n`}\}$with Fuzzy sets related with each item in I and identified by a set of linguistic labels $L=\{l_1,l_2,\ldots,l_{|L|}\}$. A weight 'w' is assigned to each linguistic label in the set. Each attribute $t_{i`}[l_j]$ is associated with several Fuzzy sets. A membership degree provides the degree of association in the range [0-1]. This indicates the correspondence between the value of each attribute and set of fuzzy linguistic labels.

Definition:

Fuzzy Item Weight (FIW) is a non-negative real number whose value ranges from 0 to 1. It is associated with each fuzzy set. The weight of a fuzzy set for an item is represented as Fuzzy Itemset Transaction Weight (FITW) is the combined weights of all the fuzzy sets associated to the items in the itemset present in a single transaction. The FITW for an itsemset is computed as

$$X = \prod_{k=1(\forall[i[l[w]]]\in X)}^{|L|} t'[ij[lk[w]]]$$

Fuzzy Weighted Support (FWS) is the aggregated sum of the FITW of all the itemsets in the transactions to the total number of transactions.

$$FWS(X)= \frac{\sum_{i=1}^{n} \prod_{k=1(\forall[i[l[w]]]\in X}^{|L|} \quad | \quad | \quad | \ |||}{}$$

Fuzzy weighted confidence is the ratio of the sum of votes satisfying X Y to the sum of votes satisfying X with Z=| X Y|. It is derived as

$$FWC\ (X\rightarrow Y) = \sum_{=1}^{n} \frac{\prod_{k=1(\forall[Z[w]]\in Z)}^{|Z|} \ `[\ [\ ]]}{\prod_{k=1(\forall[i[w]]\in X)}^{|\ |} \ `[\ [\ ]]}$$

FWARM

The FWARM algorithm belongs to the breadth first traversal group of the ARM algorithms. C_k denotes the set of candidate itemsets of the cardinality 'k', 'w' represents the weight of the items, 'F' indicates the set of frequent itemsets, 'R' indicates the set of potential rules and R` denotes the final set of Fuzzy weighted association rules.

FWARM Algorithm

Input: ‘T’ dataset Output:

Set of weighted association rules

Step 1: Initialize k=0; C_k=∅; $F_{k=}$∅

Step 2: C_k is the set of candidate itemsets

Step 3: k › ⇐1

Step 4: while

Step 5: if C_k= ∅ break

Step 6: ∀c∊ C_k

Step 7:

c.weightedSupport⇐

weightedsupportcount

Step 8: if c.weightedSupport>min_ws

Step 9: F⇐F U c

Step 10: k⇐ k + 1

Step 11: C_k = generateCandidate(F_{k-1})

Step 12: end while

Step 13: ∀ f ∈ F

Step14: Generate set of candidate rules{r_1,r_2,… ,r_n}

Step 15: R⇐RU {r_1,r_2,… ,r_n}

Step 16: ∀ r ∈ R

Step17:r.weightedConfidence⇐weighted confidence value

Step18: if r.weightedConfidence>min_wc R`⇐ R` U r to keep the city clean by informing about the garbage levels of the bins by providing graphical image of the bins via IOT Php web development platform.

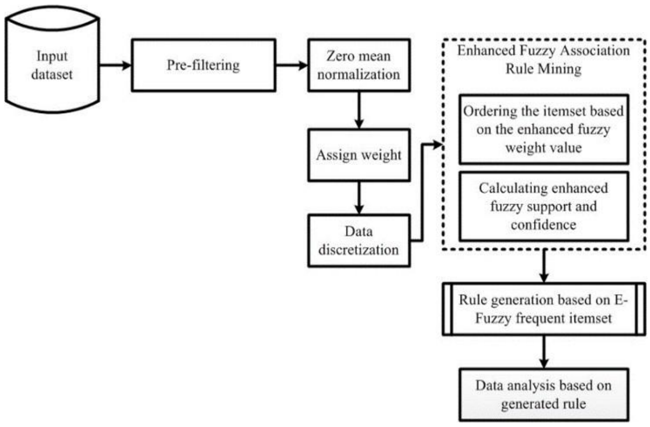


Fig.1 Overall Flow Diagram of The Proposed E-FWARM Algorithm

3.3 Date discretization

The input is the information matrix I[r, c]where ‘r’ is the shown data and ‘c’ is the shown samples. The transpose of the matrix is executed. The discretization for the matrix is required for applying the ARM algorithm.

3.4 E-FWARM Algorithm

The FCM is applied for clustering the data and determining the center of each fuzzy set and maximum and minimum values for each field of the input dataset. The triangular and trapezoid membership functions convert the dataset into a fuzzy dataset (Hong et al., 2004). The triangular membership function is described using the following equation

$$\mu(x,a,b,c,d) = \begin{cases} 0, & x < a, x > d \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{c-x}{c-b}, & c \le x \le d \end{cases}$$

Where ‘a’, ‘b’ and ‘c’ are the scalar parameters and ‘x’ is a vector. The parameters ‘a’ and ‘c’ represent the base of the triangle and parameter ‘b’ denotes the peak. The trapezoidal membership function is defined as

$$f(x;a,b,c) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases}$$

Where ‘a’ and ‘d’ represent the lower limit and upper limit and ‘b’ and ‘c’ denotes the lower limit and upper limit of the center. Fig.2 illustrates the triangular and trapezoid membership functions.

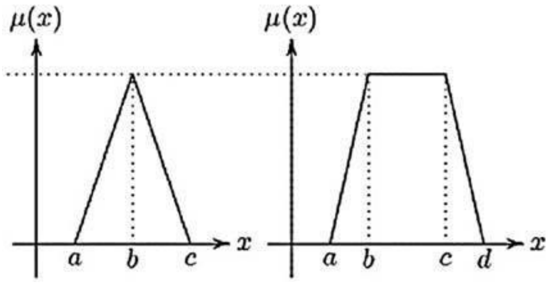


Fig.2 Triangular and trapezoid membership functions

A support value is computed for each item by aggregating the fuzzy membership functions for all data records. This aggregate value is stored in the primary candidate itemset C_1. The items that are greater than or equal to the minimum support min_sup are moved to large primary itemsets L_1.
The items are joinedand combined as

{{c[1],c[i]},{c[1],c[i+1]}, …. , {c[1],c[n]}}.

The items for each itemset do not belong to the same field. After every itemset is stored in the secondary candidate itemset C_2, the support value for each itemset is computed using a minimum operator for the fuzzy values of the items. The result of the minimum values in that itemset is added for all records. Finally, the added value is stored in the C_2. The itemsets whose value is greater than min_sup are moved to large secondary itemsets L_2. This combination is based on the every subitemset of the candidate itemset C_k. The candidate itemset should be a frequent itemset in the previous large itemsetL_k. The terms in the candidate itemset do not belong to the same field. The items are stored in the tertiary itemset C_3 and the support value is computed for each candidate itemset. The itemsets whose value is greater than or equal to the min_supare moved to the large itemset L_3. The itemsets are combined until the itemsetL_n is empty. The itemsets are pruned by selecting the itemsets including the target attribute. The itemsets are expressed as IF-THEN, the confidence value (CV) is computed as

$$CV = \frac{\sum[(IF)\cap(THEN)]}{\sum(\min(IF))} \tag{7}$$

The extracted rules are stored in the Knowledge Base (KB). The rules in the LB are inferred to the Fuzzy Inference System (FIS). The frequency of all the items in the database is assumed to be same, if the min_sup value is used for a whole database. The database contains high frequency items. Only few frequent itemsets are extracted, if the min _sup value is set too high. More number of frequent itemsets can be extracted, if the min_sup value is set too low. The FCM-Multiple Support (MS) Apriori model uses the FCM and MSApriori approach for extracting the highly frequent itemsets from the fuzzy datasets. The FCM-MSApriori inherits the benefits of both the FCM and MSApriori approach and provides more flexibility to the real-time applications.

FCM: {clustering dataset}
Find the fuzzy sets of the quantitative dataset
Calculate the sum of the membership values for each fuzzy term for all records
If sum ≥ min_sup then
Insert the fuzzy term into L_1
For k=2; $L_{k-1} \neq \emptyset$; k++ do
C_k=generate candidate from L_{k-1} ;
{
Insert into C_k
Select itemset; p.term$_1$, p.term$_2$, …,p·term$_{k-1}$, q.term$_{k-1}$

From p, q
Where p.term$_1$ = q.term$_2$, … , p.term$_{k-2}$ q.term$_{k-2}$,
p.term$_{k-1}$ ≠ q.term$_{k-1}$
}
For each itemset c$\in C_k$ do
Check all the sub-itemsets of all itemsets in C_k and it should be a frequent itemset in
For each k-1subset 's' of 'c' do
If s $\in L_{k-1}$ then
Delete c from C_k

End For
End For
For each itemset candidate in C_k do
Calculate the support value
If sum≥ min_sup then
Insert the fuzzy itemset into L_k
End For
End For
Select the frequent itemsets including the target attribute
Form the frequent itemsets that exist in L_2 to $L_n$under the form "If-Then"
For each rule
Calculate the confidence value for each rule
If CV≥ min_conf then
Accept the rule
End For
Check the rules for contradiction
Insert all the accepted rules in KB
Infer the generated rules in KB using FIS

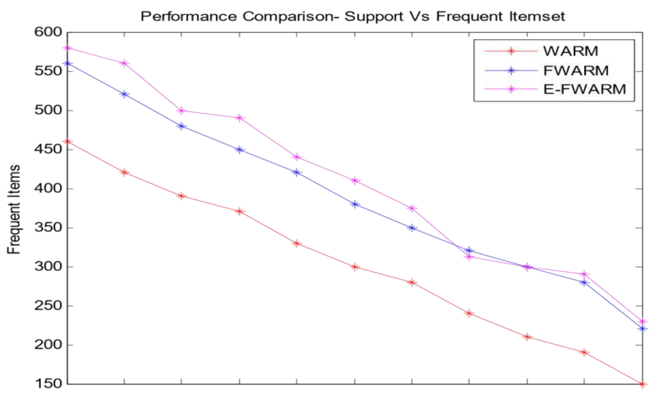
Input: 'D'- Dataset
'IW' - Itemset weight
'wsup' - Weighted support
'wconf' - Weighted confidence
'm' – itemset
C_m -Candiateitemset

F_m - Frequent itemset
c` - Number of candidate itemsets in C_m
Fz - Fuzzy Association rule
fs - Fuzzy itemset in fuzzy association rule
rs - Rules generated from C_m
WAR – Rules
min _wsup- Minimum weighted support
min _conf - Minimum weighted confidence
Output: WAR` - Set of weighted association rules
m = 0; C_m=Ø ; F_m =Ø
C_m =Set of 1 itemsets
m→1
Begin
if C_m =∅ break
∀c` ∈ C_m
c`.weightedsupport→
weightedsupportcount
if c`.weightedsupport>min_wsup
Fz → Fz U c`
m → m+1
C_{mk} = generate candidates(Fz_{m-1})
End
If fs ∈Fz_m
generate set of candidate rules {rs_1,... ,rs_n}
WAR → WAR U r
∀rs∈ WAR
rs.weighted confidence
→ weighted confidence value
If rs. weighted confidence >min_wconf
WAR` = WAR` U rs

V. PERFORMANCE ANALYSIS

The performance of the proposed work is evaluated by applying it in the groceries dataset on a system with Intel(R) Core i33220 x64-based processor and 8 GB capacity. The proposed work is simulated using Matlab 2017 software. The groceries dataset contains a collection of 9835 receipts of the purchased items. This dataset is involved in market basket analysis for finding the relationship between items purchased by the customers. This analysis helps the seller to develop their sales strategy based on the frequent items purchased together by the customers.

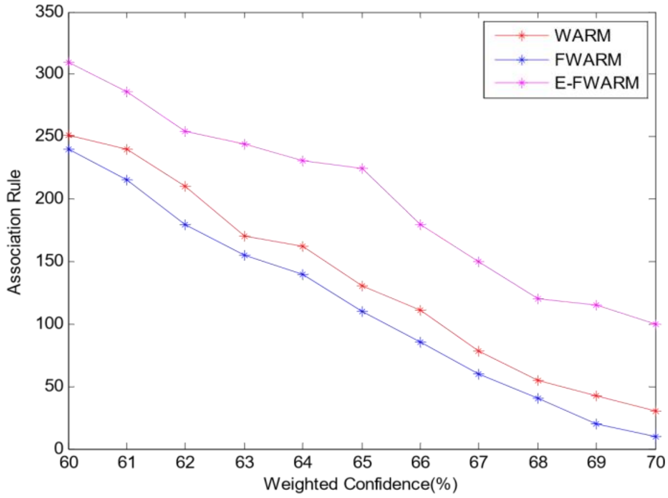
The proposed E-FWARM algorithm is compared with the WARM and FWARM (Vidya,2006). Fig.3 presents the comparative analysis of the number of frequent itemsets extracted by the proposed E-FWARM and existing WARM and FWARM. The proposed E-FWARM yielded maximum frequent items than the WARM and FWARM. There is a linear decrease in the number of frequent items with respect to the increase in the support value. Fig.4 shows the association rule rate analysis of the proposed E-FWARM and existing WARM and FWARM. The proposed E-FWARM algorithm extracts more association rules than the existing WARM and FWARM. There is a gradual decrease in the number of association rule with respect to the increase in the weighted confidence value. Fig.5 illustrates the accuracy analysis of the proposed E-FWARM and existing traditional Kmeans and Adaptive K-means algorithms (DeeptiAmbaselkar and Bagwan, 2016). The proposed EFWARM algorithm yields maximum accuracy of about 97%, while the traditional K-means and Adaptive K-means algorithms yield accuracy of about 70% and 75% respectively. The accuracy of the proposed E-FWARM is higher of about 22.68% and 27.83% than the Adaptive K-means and traditional K-means algorithms. Fig.6 depicts the execution time analysis of the proposed E-FWARM and existing traditional K-means and Adaptive Kmeans algorithms. The proposed E-FWARM algorithm requires execution time of about 2500 milliseconds (ms), while the traditional K-means and Adaptive K-means algorithms require about 3500 ms and 2800 ms respectively. The execution time of the proposed E-FWARM algorithm is about 28.57% and 10.71% than the traditional K-means and Adaptive K-means algorithms. Thus, the proposed E-FWARM algorithm is highly efficient for mining the frequent itemsets than the existing algorithms.



Support(%)

Figure 3: Frequent item rate analysis of the proposed E-FWARM and existing WARM and FWARM

Performance Comparison- Weighted Confidence Vs Association Rule



VI. CONCLUSION

In order to realize intelligent decision making in smart sys-tems, a weight judgment downward closure property based frequent itemset mining algorithm is proposed in this paper to narrow the searching space of weighted frequent itemsets and improve the time efciency. The weight judgment downward closure property for weighted frequent itemsets and the existence property of weighted frequent subsets are introduced and proved rst. Based on these two properties, the WD-FIM algorithm is described in detail. Moreover, the completeness and time efciency of WD-FIM algorithm are analyzed theoretically. Finally, the performance of the proposed WD-FIM algorithm is veried on both synthetic and real-life datasets.

VII. REFERENCES

[1]. Xuejian Zhao, Xihui Zhang, Pan Wang, Songle Chen and Zhixin Sun " A weighted frequent itemset mining algorithm,"2016

[2]. C. K.-S. Leung, M. A. F. Mateo, and D. A. Brajczuk, "A tree-based approach for frequent pattern mining from uncertain data," in Proc.PAKDD, 2008, pp. 653 6621

[3]. L. Wang, L. Feng, and M. Wu, "AT-Mine: An effcient algorithm of frequent itemset mining on uncertain dataset," J. Comput., vol. 8, no. 6, 1417 1427, Jun. 2013.

[4]. G. Lee, U. Yun, and K. H. Ryu, "Sliding window based weighted maximal frequent pattern mining over data streams," Expert Syst. Appl., vol. 41, no. 2, pp. 694 708, Feb. 2014.

[5]. L. Yue, "Review of algorithm for mining frequent patterns from uncer-tain data," Int. J. Comput.Sci. Netw. Secur., vol. 15, no. 6, pp. 17 21, Jun. 2015.

[6]. J. Pei, J. Han, and W. Wang, "Constraint-based sequential pattern min-ing: The pattern-growth methods," J. Intell. Inf. Syst., vol. 28, no. 2, 133 160, Apr. 2007.

[7]. C. C. Aggarwal and P. S. Yu, "A survey of uncertain data algorithms and applications," IEEE Trans. Knowl. Data Eng., vol. 21, no. 5, pp. 609 623, May 2009.

[8]. D. Suciu, "Probabilistic databases," ACM Sigact News, vol. 39, no. 2, 111 124, Jun. 2008.

[9]. C.-K. Chui, B. Kao, and E. Hung, "Mining frequent itemsets from uncer-tain data," in Proc. Paci c-Asia Conf. Knowl. Discovery Data Mining, 2007, pp. 47 58.

[10]. C. K. Chui and B. Kao, "A decremental approach for mining frequent itemsets from uncertain data," in Proc. PAKDD, 2008, pp. 64 75.

[11]. L. Wang, D. W.-L. Cheung, R. Cheng, S. D. Lee, and X. S. Yang, "Efcient mining of frequent item sets on large uncertain databases," IEEE Trans. Knowl. Data Eng., vol. 24, no. 12, pp. 2170 2183,Dec. 2012.

[12]. X. Sun, L. Lim, and S. Wang, "An approximation algorithm of mining frequent itemsets from uncertain dataset," Int. J. Advancements Comput.Technol., vol. 4, no. 3, pp. 42 49, Feb. 2012.

[13]. C. C. Aggarwal, Y. Li, J. Wang, and J. Wang, "Frequent pattern mining with uncertain data," in Proc. ACM KDD, 2009, pp. 29 38.

[14]. C. K.-S. Leung and S. K. Tanbeer, "Fast tree-based mining of frequent itemsets from uncertain data," in Proc. Int. Conf. Database Syst. Adv.Appl., 2012, pp. 272 287.

[15]. C. K.-S. Leung and S. K. Tanbeer, "PUF-Tree: A compact tree structure for frequent pattern mining of uncertain data," in Proc. Paci c-Asia Conf.Knowl. Discovery Data Mining, 2013, pp. 13 25.

[16]. C. K.-S. Leung, R. K. Mackinnon, and S. K. Tanbeer, "Tightening upper bounds to the expected support for uncertain frequent pattern min-ing," in Proc. 18th Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst., 2014, 328 337.

[17]. C. W. Lin and T. P. Hong, "A new mining approach for uncertain databases using CUFP trees," Expert Syst. Appl., vol. 39, no. 4, pp. 4084 4093, Mar. 2012.

[18]. G. Lee, U. Yun, and H. Ryang, "An uncertainty-based approach: Fre-quentitemset mining from uncertain data with different item importance," Knowl.-Based Syst., vol. 90, pp. 239 256, Dec. 2015.

[19]. G. Lee and U. Yun, "A new efcient approach for mining uncertain frequent patterns using minimum data structure without false positives," Future Gener. Comput. Syst., vol. 68, pp. 89 110, Mar. 2017.

[20]. W. Wang, J. Yang, and P. S. Yu, "Efcient mining of weighted association rules (war)," in Proc. 6th ACM SIGKDD Int. Conf. Knowl. Discovery DataMining, 2000, pp. 270 274.