**Journal of Theoretical and Applied Information Technology**
31st January 2018. Vol.96. No 2


ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# E-FWARM: ENHANCED FUZZY-BASED WEIGHTED ASSOCIATION RULE MINING ALGORITHM

**[1]K. MANGAYARKKARASI, [2] M. CHIDAMBARAM**

[1]Research Scholar, Research and Development Centre, Bharathiar University, Coimbatore, Tamilnadu, India

[2]Assistant Professor, Rajah Serfoji College, Bharathidasan University, Thanjavur, Tamilnadu, India

E-mail: [1]kanthimangai@gmail.com, [2]chidsuba@gmail.com

**ABSTRACT**

In the Association Rule Mining (ARM) approach, equal weight is assigned to all itemsets in the dataset. Hence, it is not appropriate for all datasets. The weight should be assigned based on the significance of each itemset. The WARM reduces extra steps during the generation of rules. As, the Weighted ARM (WARM) uses the significance of each itemset, it is applied in the data mining. 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 Association Rule Mining (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:** *Association Rule Mining (ARM), Data Mining, Frequent Itemset Mining, Fuzzy-based Weighted Association Rule Mining (FWARM)*

## 1. INTRODUCTION

Data mining is an emerging technique that addresses the problem of restructuring the data into the useful information. The rule generation methods are used in the area of data mining to determine the relationship between different items. The Association Rule Mining (ARM) is widely used for finding data patterns that reveals the combination of events occurring simultaneously based on the associations among a large set of data items. The Apriori algorithm (Agrawal and Srikant, 1994) is a popular algorithm for extracting high frequent itemsets from a database using the predefined threshold measures such as minimum support and minimum confidence (Sowan et al., 2013).

In the ARM model, it is assumed that all data items have same significance without considering their weight and difference between the transactions and importance of each item set. The Weighted ARM (WARM) algorithm works only based on the binary attributes and not on the databases using the significance of each item set. The ARM aims to discover the relationships among the data attributes depending on minimum support and minimum confidence value. The rare item problem occurs if a single minimum support is used (Liu et al., 1999, Hu and Chen, 2006, Kiran and Reddy, 2010). Liu et al. (Liu et al., 1999) developed Multiple Support (MS) Apriori algorithm to solve the rare item problem based on the concept of assigning Minimum Item Support (MIS) values for each item in the database. Despite of using single minimum support, various minimum supports are used for different items in the database. Different MIS values are assigned to compute different frequent items. This enables the generation of rare frequent itemsets and prevents the generation of tedious frequent itemsets (Hu and Chen, 2006). The Fuzzy-based weighted ARM (FWARM) algorithms (Jain et al., 2008, Toloo and Nalchigar, 2011, Toloo et al., 2009, Ho et al., 2012, Chiu et al., 2012) play a significant role in the field of data mining.

### 1.1 Motivation and Research Objectives

In the previous work, a weighted rule mining approach is applied to extract the frequent itemsets. The datasets are ranked based on the queries and demands of the cloud users. It is found that the selected datasets are frequently used by most of the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

users. A secure cloud data mining model is designed to identify the best frequently used datasets for their queries (Mangayarkkarasi and Chidambaram). This generates very less number of itemsets and requires minimum time consumption than the traditional ARM algorithm.

In this paper, E-FWARM algorithm is applied for efficient extraction of the frequent itemsets from the database. The main objective of the research work is to extract maximum frequent items without requiring more complexity. The pre-filtering method is applied to the input dataset to remove the items having low variance. Then, zero-mean normalization is applied and the weight is allocated to each data item. Data discretization is performed and E-FWARM algorithm is applied for mining the frequent itemsets. The rule is generated based on the proposed algorithm. Finally, the data analysis is performed based on the generated rule. The proposed E-FWARM algorithm yields maximum frequent items, association rules, accuracy and minimum execution time than the existing algorithms.

The sections of the paper are systematized as follows: Section II describes an overview of the existing ARM approaches. Section III provides the basic definitions of the FWARM algorithm. Section IV explains the proposed work including normalization, weight assignment, data discretization and E-FWARM algorithm. Section IV presents the performance analysis including comparative analysis of the proposed E-FWARM algorithm, existing WARM and FWARM algorithms. The conclusion of the proposed work is discussed in Section V.

## 2. RELATED WORKS

Sowan et al. (Sowan et al., 2013) combined Fuzzy C-Means (FCM) and Apriori approach to identify the Fuzzy Association Rule (FAR) for improved prediction of the road traffic. The knowledge can be extracted from a database to predict the future value efficiently. Mallik et al. (Mallik et al., 2015) proposed a Rank-Based WARM algorithm for ranking the rules based on the weighted condensed support and weighted condensed confidence. The proposed RANWAR generates minimum number of frequent itemsets and requires minimum execution time than the existing ARM algorithms.

Pears et al. (Pears et al., 2013) automated the weight assignment process by expressing a linear model that captures relationships between the items. The Valency model is extended by increasing the field of communication beyond immediate neighborhoods. Azadnia et al. (Azadnia et al., 2013) presented a combined approach of WARM and Genetic Algorithm (GA) to reduce the total tardiness. The best suitable path can be identified by combining the GA with the traveling salesman problem.

Nithya and Duraiswamy (Nithya and Duraiswamy, 2015) applied average ranking feature selection approach and FWARM classifier for efficient diagnosis of the healthcare dataset. The classification accuracy is improved and computational time is minimized by reducing the number of rules. Galárraga et al. (Galárraga et al., 2013) developed a rule mining model inspired by the ARM and introduced a novel confidence measure. The mining efficiency and coverage of the proposed model is better than the existing approaches.

Vo et al. (Vo et al., 2013) proposed Weighted Itemset-Tidset tree data structure for effective mining of the weighted frequent itemsets. A Diffset strategy is described for computing the weighted support for the itemsets. The memory consumption and computation time can be reduced. Altuntas et al. (Altuntas et al., 2013) presented FWARM approach for efficiently solving the facility layout problem in the cellular manufacturing system. Cai et al. (Cai et al., 2013) proposed WARM model for mining the weighted frequent items in the mobile computing environment. Gosain and Bhugra (Gosain and Bhugra, 2013) presented a comparative survey of different association rules on the quantitative data based on various parameters such as weighted measure, support and confidence factors.

Rubia and Sasikala (Rubia and Sasikala, 2014) described the comparative study of fuzzy logic and WARM for efficient mining of frequent datasets. The fuzzy logic and WARM required minimum time and memory consumption. Bansal et al. (Bansal et al., 2017) applied Weighted Fuzzy Privacy Preserving Mining (WFPPM) to extract the sensitive association rules based on the weight of the rule. The actual sensitive rules can be found out while maintaining the quality of the released data.

Datta and Bose (Datta and Bose, 2015) proposed identification of the frequent sets through the weighted support and confidence for the generation of useful positive rule. The chances of rule discovery are increased while maintaining the same level of minsupp and minconf in both cases. Jafarzadeh et al. (Jafarzadeh et al., 2015) developed

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

an enhanced Apriori algorithm for mining the fuzzy association rules. Vidya (Vidya, 2014) proposed genetic based fuzzy WARM for providing appropriate results according to the user requirements. The proposed approach is efficient than the existing algorithm.

Chen et al. (Chen et al., 2016) proposed a fuzzy temporal ARM algorithm based on the scalar cardinality of each linguistic term of each item. Agarwal and Nanavati (Agarwal and Nanavati, 2016) introduced an ARM scheme using the hybridization of the GA and PSO algorithm. High quality rules are obtained by maintaining optimal tradeoff between the interpretability and accuracy with quicker convergence rate. Sheng et al. (Sheng et al., 2016) developed a data-driven technique by integrating the Apriori algorithm and probabilistic graphical model. The prediction accuracy is improved by using the association rules. Sarkar et al. (Sarkar et al., 2017) introduced a novel model ARM and GA for the automatic selection of the optimal Support and Confidence value to generate useful rules. Mansha et al. (Mansha et al., 2016) proposed a novel algorithm for mining the association rules from uncertain data based on Self-Organizing Map (SOM) clustering. Effective generation of the frequent itemsets and association rules is ensured by maintaining the feasibility of the neural network.

(Faridi et al., 2017a) proposed an agricultural intelligent decision support system for mining weighted fuzzy spatial rules and constructing the fuzzy inference system. The proposed system offered well informed decisions for the efficient utilization of the wastelands. (Paul et al., 2017) developed an automatic fuzzy diagnostic system based on the Genetic Algorithm (GA) and a Modified Dynamic Multi-Swarm Particle Swarm Optimization (MDMS-PSO) to optimize the membership functions of the Fuzzy diagnostic system for predicting the risk level of heart disease. The proposed system yielded better accuracy on different datasets. (Faridi et al., 2017b) introduced a new algorithm for weighted fuzzy spatial association rule mining for reclaiming the wasteland and ranking the rules based on the faith and weight-area score. The proposed algorithm is efficient and scalable when compared to the Apriori algorithm. (Bansal et al., 2017) applied Weighted Fuzzy Privacy Preserving Mining (WFPPM) for extracting complex association rules. Actual sensitive rules are obtained by considering the weight of individual parameter.

The conventional ARM algorithms using the crisp set are used for handling the Boolean data. But, there is a need to discover knowledge and handle huge amount of quantitative data in real life. The dataset should be partitioned and converted into Boolean type, for extracting the association rules from the quantitative data. As a result, it may suffer from the sharp boundary problem (Kalia et al., 2013). The Fuzzy association rules are developed to solve this problem. In this paper, an E-FWARM algorithm is proposed for the efficient mining of the frequent itemsets

## 3. FUZZY WEIGHTED ASSOCIATION RULE MINING

Let a dataset 'D' comprises a set of transactions $T = \{t_1, t_2, \dots, t_n\}$ with a set of items $I = \{i_1, i_2, \dots, i_{|I|}\}$ (Muyeba et al., 2008). A Fuzzy dataset $D'$ includes fuzzy transactions $T' = \{t'_1, t'_2, \dots, t'_n\}$ with Fuzzy sets related with each item in I and identified by a set of linguistic labels $L = \{l_1, l_2, \dots, l_{|L|}\}$. A weight 'w' is assigned to each linguistic label in the set. Each attribute $t'_i[i_j]$ 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 $i_j$ is represented as $i_j[l_k[w]]$.

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

$$X = \prod_{k=1\,(\forall[i[l[w]]]\in X)}^{|L|} t'\left[i_j[l_k[w]]\right] \tag{1}$$

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

$$FWS(X) = \frac{\sum_{i=1}^{n} \prod_{k=1\,(\forall[i[l[w]]]\in X)}^{|L|} t'_i\left[i_j[l_k[w]]\right]}{n} \tag{2}$$

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

$$FWC(X \to Y) = \sum_{i=1}^{n} \frac{\prod_{k=1\,(\forall[Z[w]]\in Z)}^{|Z|} t'_i[z_k[w]]}{\prod_{k=1\,(\forall[i[w]]\in X)}^{|X|} t'_i[x_k[w]]} \tag{3}$$

**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:** $\boldsymbol{R'} = \textit{Set of weighted association rules}$

Step 1: Initialize $k = 0; C_k = \emptyset; F_k = \emptyset$

Step 2: $C_k$ is the set of candidate itemsets

Step 3: $k \Leftarrow 1$

Step 4: while

Step 5: if $C_k = \emptyset$ break

Step 6: $\forall c \in C_k$

Step 7: $\mathrm{c.weightedSupport} \Leftarrow \mathrm{weighted\ support\ count}$

Step 8: if $c.weightedSupport > \min\_ws$

Step 9: $\mathrm{F} \Leftarrow \mathrm{F} \cup \mathrm{c}$

Step 10: $k \Leftarrow \mathrm{k} + 1$

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

Step 12: end while

Step 13: $\forall \mathrm{f} \in F$

Step 14: Generate set of candidate rules $\{r_1, \dots, r_n\}$

Step 15: $\mathrm{R} \Leftarrow \mathrm{R} \cup \{r_1, \dots, r_n\}$

Step 16: $\forall \mathrm{r} \in R$

Step 17: $\mathrm{r.weightedConfidence} \Leftarrow \mathrm{weighted\ confidence\ value}$

Step 18: if $\mathrm{r.weightedConfidence} > \min\_\mathrm{wc}$ $R' \Leftarrow R' \cup r$

## 4. E-FWARM ALGORITHM

The input dataset contains a large amount of data. The pre-filtering method is applied on the dataset to remove the data having low variance. The data with low variance is insignificant. The general contrast in the information for each previous data should be inspected and the data is filtered. The separated data should be normalized. This normalization changes the information to a standard scale. The zero-mean normalization is utilized to change the information such that the mean value of each zero becomes zero and standard deviation of each data becomes one. Fig.1 shows the overall flow diagram of the proposed E-FWARM algorithm.

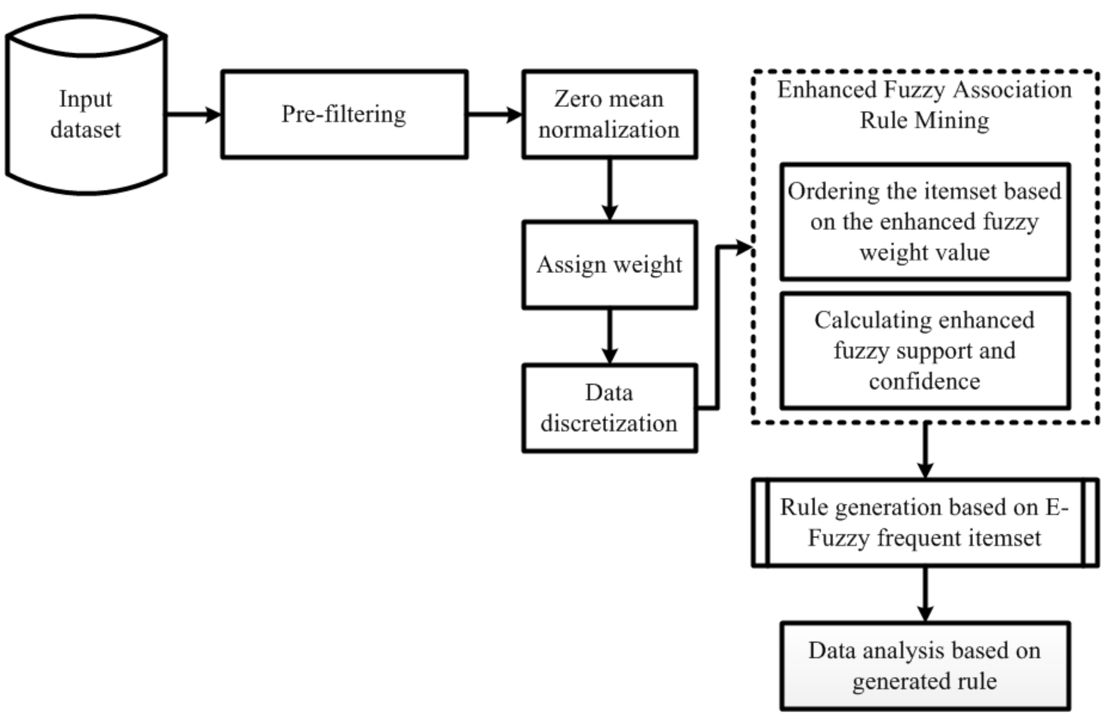


*Fig.1 Overall Flow Diagram Of The Proposed E-FWARM Algorithm*

### 3.1 Zero Mean Normalization

Normalization is the process of scaling the individual data samples so to lie within a small specified range. The zero-mean normalization (Bandyopadhyay et al., 2014) converts data into a specific format such that mean value of the data becomes zero and standard deviation becomes one. It is formulated as

$$x_{ij}^{norm} = \frac{x_{ij}-\mu}{\sigma} \tag{4}$$

Where $x_{ij}$ and $x_{ij}^{norm}$ are the values of the ith data at the jth sample before and after normalization. $\mu$ and $\sigma$ denote the mean and standard deviation of the data. When zero-mean normalization is applied, all the data are made to slide vertically so that the mean value of the data is zero.

### 3.2 Weight assignment

The weight is allocated to each data according to their ranking. The weight of all data is calculated such that variation in the weight of any consecutively ranked data is similar. The weight of the data that is ranked first is one. The range of assigned weight lies in the range between zero and one. If the total number of data in the system is 'n', the weight of each data $1 \le i \le n$ is computed from the rank function $r_i$ and number of data.

### 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

$$f(x;a,b,c) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases} \tag{5}$$

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

$$\mu(x,a,b,c,d) = \begin{cases} 0, & x < a, x > d \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{c-x}{c-b}, & c \le x \le d \end{cases} \tag{6}$$

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.

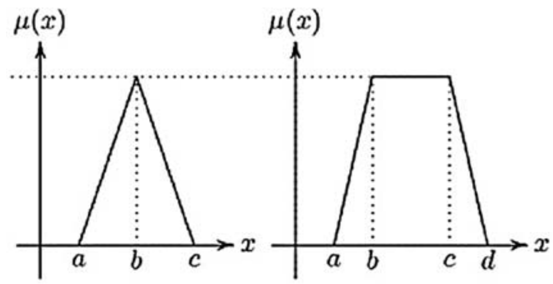


*Fig.2 Triangular and trapezoid membership functions*

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 joined and combined as $\{\{c[1], c[i]\}, \{c[1], c[i+1]\}, \ldots, \{c[1], c[n]\}\}$. 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 evergy sub-itemset of the candidate itemset $C_k$. The candidate itemset should be a frequent itemset in the previous large itemset$L_{k_1}$. 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 $\min\_sup$ are moved to the large itemset $L_3$. The itemsets are combined until the itemset $L_n$ is empty. The itemsets are pruned by selecting the itemsets including the target attribute. The itemsets are

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

expressed as IF-THEN, the confidence value (CV) is computed as

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

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 (Siji and Valarmathi).

---

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 $sum \geq \min\_sup$ 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; $p \cdot term_1, p \cdot term_2, \ldots, p \cdot term_{k-1}, q \cdot term_{k-1}$
From p, q
Where $p \cdot term_1 = q \cdot term_2, \ldots, p \cdot term_{k-2} = q \cdot term_{k-2}, p \cdot term_{k-1} \neq q \cdot term_{k-1}$
}
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 $L_{k-1}$
For each $(k-1)$ subset '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 $sum \geq \min\_sup$ 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 $CV \geq \min\_conf$ 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
$C_m$- Candidate itemset
$F_m$- 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
$\min\_wconf$ - Minimum weighted confidence
Output: $WAR'$ - Set of weighted association rules
$m = 0;\ C_m = \emptyset; F_m = \emptyset$
$C_m = Set\ of\ 1\ itemsets$
$m \rightarrow 1$
Begin
if $C_m = \emptyset$ break
$\forall c' \in C_m$
$c' \cdot weighted\ support \rightarrow weighted\ support\ count$
if $c' \cdot weighted\ support > \min\_wsup$
$Fz \rightarrow Fz \cup c'$
$m \rightarrow m + 1$
$C_{mk} = generate\ candidates\ (Fz_{m-1})$
End
if $fs \in Fz_m$
generate set of candidate rules $\{rs_1, \ldots, rs_n\}$
$WAR \rightarrow WAR \cup rs$
$\forall\ rs \in WAR$
$rs.weighted\ confidence \rightarrow weighted\ confidence\ value$
if $rs.weighted\ confidence > \min\_wconf$
$WAR' = WAR' \cup rs$

---

## 5. PERFORMANCE ANALYSIS

The performance of the proposed work is evaluated by applying it in the groceries dataset

(Marafi, 2014) on a system with Intel(R) Core i3-3220 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 K-means and Adaptive K-means algorithms (Deepti Ambaselkar and Bagwan, 2016). The proposed E-FWARM 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 K-means 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.

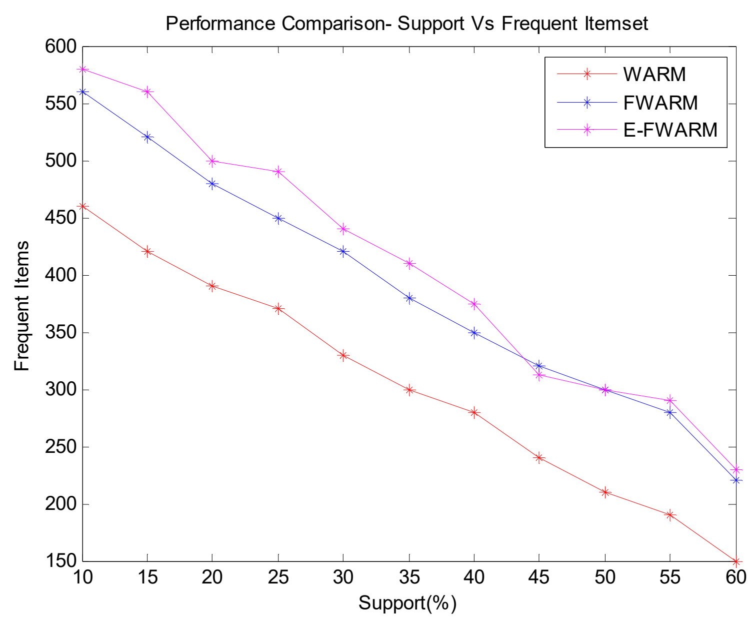


*Figure 3: Frequent item rate analysis of the proposed E-FWARM and existing WARM and FWARM*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

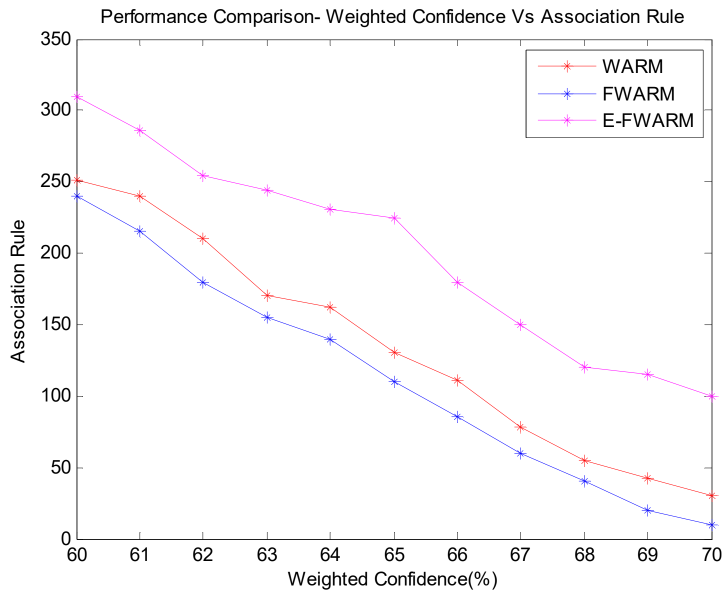


*Figure 4: Association rule rate analysis of the proposed E-FWARM and existing WARM and FWARM*

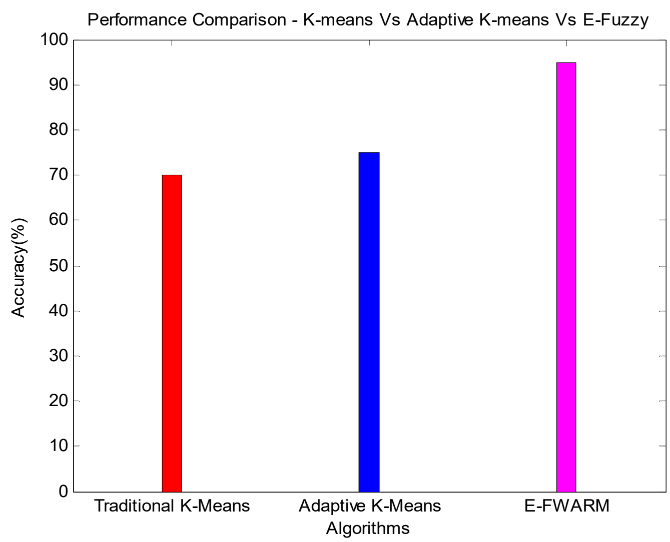


*Figure 5: Accuracy analysis of the proposed E-FWARM and existing K-means and Adaptive K-means algorithms*

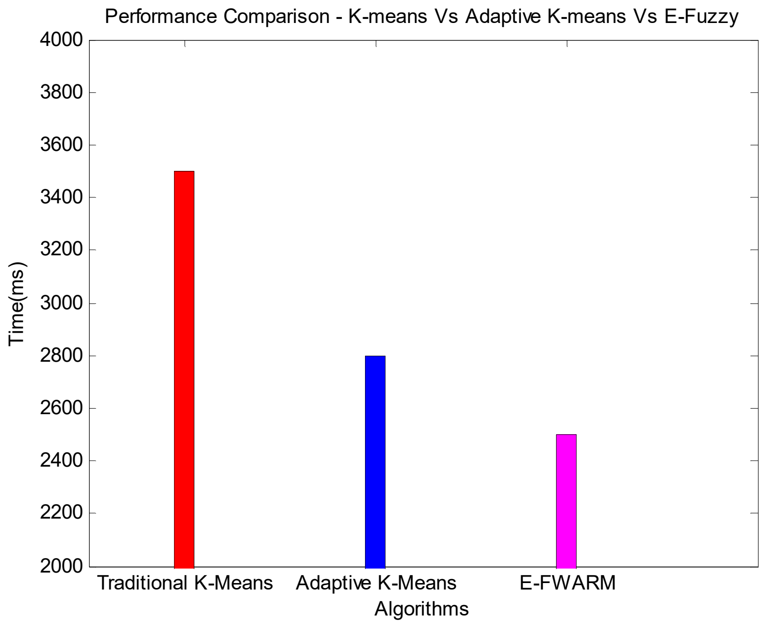


*Fig.6 Execution time analysis of the proposed E-FWARM and existing K-means and Adaptive K-means algorithms*

## 6. CONCLUSION

The ARM algorithms usually extract a huge quantity of rules and do not guarantee about the relevancy of all the extracted rules. Also, the drawbacks of the existing ARM algorithms are mining of non-interesting rules and huge number of discovered rules. It requires multiple scanning of the database and generates many candidate itemsets. An E-FWARM algorithm is proposed for the efficient mining of the frequent itemsets. The significance of each item is considered by assigning weight to each item. This results in the efficient mining of frequent items and association rules. The performance of the proposed E-FWARM algorithm is evaluated by comparing it with the WARM, FWARM, traditional K-means and adaptive K-means algorithms. The proposed E-FWARM algorithm extracted more number of frequent items and association rules when compared to the WARM and FWARM algorithms. The accuracy of the proposed E-FWARM algorithm is higher than the traditional K-means and Adaptive K-means algorithms. The proposed E-FWARM algorithm consumes minimum execution time than the traditional K-means and Adaptive K-means algorithms. In future, the Positive Association Rules (PARs) and Negative Association Rules (NARs) are to be mined by using the Fuzzy-based algorithm.

## REFRENCES:

[1] AGARWAL, A. & NANAVATI, N. Association rule mining using hybrid GA-PSO for multi-objective optimisation. IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2016. IEEE, 1-7.

[2] AGRAWAL, R. & SRIKANT, R. Fast algorithms for mining association rules. Proc. 20th int. conf. very large data bases, VLDB, 1994. 487-499.

[3] ALTUNTAS, S., DERELI, T. & SELIM, H. 2013. Fuzzy weighted association rule based solution approaches to facility layout problem in cellular manufacturing system. *International Journal of Industrial and Systems Engineering,* 15, 253-271.

[4] AZADNIA, A. H., TAHERI, S., GHADIMI, P., MAT SAMAN, M. Z. & WONG, K. Y. 2013. Order batching in warehouses by minimizing total tardiness: a hybrid approach of weighted association rule mining and genetic algorithms. *The Scientific World Journal*, 2013.

[5] BANDYOPADHYAY, S., MALLIK, S. & MUKHOPADHYAY, A. 2014. A survey and comparative study of statistical tests for identifying differential expression from microarray data. *IEEE/ACM transactions on computational biology and bioinformatics,* 11**,** 95-115.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195


[6] BANSAL, M., GROVER, D. & SHARMA, D. 2017. Sensitivity Association Rule Mining using Weight based Fuzzy Logic. *Global Journal of Enterprise Information System,* 9**,** 1-9.

[7] CAI, J.-H., ZHAO, X.-J. & XUN, Y.-L. 2013. Association rule mining method based on weighted frequent pattern tree in mobile computing environment. *International Journal of Wireless and Mobile Computing,* 6**,** 193-199.

[8] CHEN, C.-H., LAN, G.-C., HONG, T.-P. & LIN, S.-B. 2016. Mining fuzzy temporal association rules by item lifespans. *Applied Soft Computing,* 41**,** 265-274.

[9] CHIU, H.-P., TANG, Y.-T. & HSIEH, K.-L. 2012. Applying cluster-based fuzzy association rules mining framework into EC environment. *Applied Soft Computing,* 12**,** 2114-2122.

[10] DATTA, S. & BOSE, S. Discovering association rules partially devoid of dissociation by weighted confidence. IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS), 2015. IEEE, 138-143.

[11] DEEPTI AMBASELKAR & BAGWAN, D. A. B. 2016. Adaptive K-means clustering for association rule mining from gene expression data. *International Journal of Engineering and Technical Research (IJETR),* 5**,** 25-28.

[12] FARIDI, M., VERMA, S. & MUKHERJEE, S. An Agricultural Intelligence Decision Support System: Reclamation of Wastelands Using Weighted Fuzzy Spatial Association Rule Mining. International Conference on Information and Communication Technology for Intelligent Systems, 2017a. Springer, 551-559.

[13] FARIDI, M., VERMA, S. & MUKHERJEE, S. 2017b. A novel algorithm of weighted fuzzy spatial association rule mining (WFSARM) for wasteland reclamation. *Journal of Information and Optimization Sciences***,** 1-17.

[14] GALÁRRAGA, L. A., TEFLIOUDI, C., HOSE, K. & SUCHANEK, F. AMIE: association rule mining under incomplete evidence in ontological knowledge bases. Proceedings of the 22nd international conference on World Wide Web, 2013. ACM, 413-422.

[15] GOSAIN, A. & BHUGRA, M. A comprehensive survey of association rules on quantitative data in data mining. IEEE Conference on Information & Communication Technologies (ICT), 2013. IEEE, 1003-1008.

[16] HO, G. T., IP, W., WU, C.-H. & TSE, Y. K. 2012. Using a fuzzy association rule mining approach to identify the financial data association. *Expert Systems with Applications,* 39**,** 9054-9063.

[17] HONG, T.-P., KUO, C.-S. & WANG, S.-L. 2004. A fuzzy AprioriTid mining algorithm with reduced computational time. *Applied Soft Computing,* 5**,** 1-10.

[18] HU, Y.-H. & CHEN, Y.-L. 2006. Mining association rules with multiple minimum supports: a new mining algorithm and a support tuning mechanism. *Decision Support Systems,* 42**,** 1-24.

[19] JAFARZADEH, H., TORKASHVAND, R. R., ASGARI, C. & AMIRY, A. 2015. Provide a new approach for mining fuzzy association rules using apriori algorithm. *Indian Journal of Science and Technology,* 8**,** 127-134.

[20] JAIN, V., BENYOUCEF, L. & DESHMUKH, S. 2008. A new approach for evaluating agility in supply chains using fuzzy association rules mining. *Engineering Applications of Artificial Intelligence,* 21**,** 367-385.

[21] KALIA, H., DEHURI, S. & GHOSH, A. 2013. A survey on fuzzy association rule mining. *International Journal of Data Warehousing and Mining (IJDWM),* 9**,** 1-27.

[22] KIRAN, R. U. & REDDY, P. K. Mining rare association rules in the datasets with widely varying items' frequencies. International Conference on Database Systems for Advanced Applications, 2010. Springer, 49-62.

[23] LIU, B., HSU, W. & MA, Y. Mining association rules with multiple minimum supports. Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999. ACM, 337-341.

[24] MALLIK, S., MUKHOPADHYAY, A. & MAULIK, U. 2015. RANWAR: rank-based weighted association rule mining from gene expression and methylation data. *IEEE transactions on nanobioscience,* 14**,** 59-66.

[25] MANGAYARKKARASI, K. & CHIDAMBARAM, M. Rank-Based Weighted Association Rule Mining Technique for Secure Cloud Computing Environment *Global Journal of Engineering Science and Research Management,* 4.

[26] MANSHA, S., BABAR, Z., KAMIRAN, F. & KARIM, A. Neural Network Based Association Rule Mining from Uncertain Data. International Conference on Neural

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Information Processing, 2016. Springer, 129-136.

[27] MARAFI, S. 2014. *Market Basket Analysis with R* [Online]. Available: http://www.salemmarafi.com/code/market-basket-analysis-with-r/.

[28] MUYEBA, M., KHAN, M. S. & COENEN, F. Fuzzy weighted association rule mining with weighted support and confidence framework. Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2008. Springer, 49-61.

[29] NITHYA, N. & DURAISWAMY, K. 2015. Correlated gain ratio based fuzzy weighted association rule mining classifier for diagnosis health care data. *Journal of Intelligent & Fuzzy Systems,* 29**,** 1453-1464.

[30] PAUL, A. K., SHILL, P. C., RABIN, M. R. I. & MURASE, K. 2017. Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease. *Applied Intelligence***,** 1-18.

[31] PEARS, R., KOH, Y. S., DOBBIE, G. & YEAP, W. 2013. Weighted association rule mining via a graph based connectivity model. *Information Sciences,* 218**,** 61-84.

[32] RUBIA, K. & SASIKALA, S. 2014. A comparative study of fuzzy logic and weighted Association rule mining in frequent datasets. *International Journal of Computer Science & Applications (TIJQSA) 3,* 1.

[33] SARKAR, S., LOHANI, A. & MAITI, J. Genetic Algorithm-Based Association Rule Mining Approach Towards Rule Generation of Occupational Accidents. International Conference on Computational Intelligence, Communications, and Business Analytics, 2017. Springer, 517-530.

[34] SHENG, G., HOU, H., JIANG, X. & CHEN, Y. 2016. A novel association rule mining method of big data for power transformers state parameters based on probabilistic graph model. *IEEE Transactions on Smart Grid*.

[35] SIJI, P. & VALARMATHI, M. Enhanced Fuzzy Association Rule Mining Techniques For Prediction Analysis in Betathalesemia's Patients.

[36] SOWAN, B., DAHAL, K., HOSSAIN, M. A., ZHANG, L. & SPENCER, L. 2013. Fuzzy association rule mining approaches for enhancing prediction performance. *Expert Systems with Applications,* 40**,** 6928-6937.

[37] TOLOO, M. & NALCHIGAR, S. 2011. On ranking discovered rules of data mining by data envelopment analysis: some models with wider applications. *New Fundamental Technologies in Data Mining.* InTech.

[38] TOLOO, M., SOHRABI, B. & NALCHIGAR, S. 2009. A new method for ranking discovered rules from data mining by DEA. *Expert Systems with Applications,* 36**,** 8503-8508.

[39] VIDYA, V. 2006. AN EFFICIENT FUZZY WEIGHTED ASSOCIATION RULE MINING WITH ENHANCED HITS ALGORITHM.

[40] VIDYA, V. 2014. A Hs-Hybrid Genetic Improved Fuzzy Weighted Association Rule Mining Using Enhanced Hits Algorithm. *Journal of Agricultural & Biological Science,* 9.

[41] VO, B., COENEN, F. & LE, B. 2013. A new method for mining Frequent Weighted Itemsets based on WIT-trees. *Expert Systems with Applications,* 40**,** 1256-1264.