# Fuzzy Association Rules: new method and implementation

**Amir Ebrahimzadeh**
Sama technical and vocational training college, Islamic Azad University, Mashhad branch, Mashhad, Iran

**Abstract**
due to increasing use of huge databases , mining practical information and useful knowledge from transactions is evolving into an important area. Most data mining methods focuses on relationship among transactions. Many algorithms have been proposed to find association rules in databases with either binary or quantitative attributes. One of these approaches is fuzzy association rules mining. Fuzzy Apriori and its different variations are the popular fuzzy association rule mining (ARM) algorithms available today. Like the crisp version of Apriori, fuzzy Apriori is a very slow and inefficient algorithm for very large datasets. Hence ,in this paper, we introduce a novel technique , called FCT, for mining fuzzy association rules . Existing method discovers fuzzy association rules by scaning the database once, and performing three tasks simultaneously .First , compute the fuzzy supports of candidate 1-itemsets and then generate large 1-itemsets.Second , divide database into multiple cluster tables ,such that transaction with length k , fall into cluster table k.

Third ,builds new structure called CDi , for each cluster table i, such that CDi[A,x]=∑µA(x) ,where x is an item and A denotes linguistic term. Then fuzzy large item sets are generated according to the cluster tables, instead of scanning whole the database. In addition , if CDi[A,x]=0 for cluster i and item i, then for computing the fuzzy support of each candidate item set containing A(x) , scanning this cluster can be ignored. Consequently , we reduce incredible amount of scanning data and therefore the running time of mining algorithm is reduced greatly. Experimental results show our algorithm is many times faster than fuzzy Apriori for the very large real life dataset .
**Keywords:** Fuzzy Association Rulse , Cluster table

## Introduction

Data mining (DM) is the process used for the automatic discovery of high-level knowledge from real-world, large and complex datasets. Discovering association rules is one of several data-mining techniques described in the literature [1]. Association rules are used to represent and identify dependencies between items in a database [2]. An association rule is an expression $X \rightarrow Y$, where X and Y are sets of items and $X \cap Y = \emptyset$. It means that if all the items in X exist in a transaction then all the items in Y are also in the transaction with a high probability, and X and Y should not have a common item [3,4].

Research in this field has mainly concentrated on Boolean and quantitative association rules [4,5,6,7,8]. However, in recent years many researchers have proposed methods to mine fuzzy association rules from quantitative data in order to solve some of the problems introduced by quantitative attributes [9,10,11,12]. Recently , the fuzzy theory has been used more and more frequently in intelligent systems because of its simplicity to human reasoning. Several fuzzy mining algorithms for inducing rules from given sets of data have been designed and used to good effect with specific domains [13,14,15,16,17].

However performance is dramatically decreased in the process of many fuzzy association rules algorithms. This is due to the fact that a database is repeatedly scanned to contract each candidate itemset with the whole database level by level in process of mining fuzzy association rules. Thus, we propose an efficient method for discovering the fuzzy large item sets.

In this paper , an effective algorithm named fuzzy cluster test (FCT) algorithm is proposed. Our algorithms consist of three parts:

(1) Quantitative attributes are partitioned into several fuzzy sets by the fuzzy c-means (FCM) algorithm[18];

(2) Discovering frequent fuzzy attributes;

(3) Generating fuzzy association rules with at least a minimum confidence from frequent fuzzy attributes.

The remaining parts of this paper are organized as follows. Partitioning fuzzy set is reviewed in section 2. A new data mining algorithm for fuzzy association rules is proposed in section 3. Experimental results are stated in section 4. Conclusion are finally given in section 5.

**Partitioning fuzzy set**

Fuzzy set was proposed by Zadeh , and the division of the features into various linguistic values has been widely used in pattern recognition and fuzzy inference. From this, various results have been proposed, such as application to pattern classification by Ishibuchi et al [19] , the fuzzy rules generated by wang and Mendel [20] , and methods for partitioning feature space were also discussed by many researchers. In this paper, we view each attribute as a linguistic variable, and the variable are divided into various linguistic values. A linguistic variable is a variable whose values are linguistic words or sentences in a natural language. For example, the values of the linguistic variable 'Age' may be 'close to 30' or ' very close to 50' and referred to as linguistic values. In FCB algorithm, quantitative attributes are partitioned into several fuzzy sets by the FCM algorithm[18 ].

**FCT algorithm**

**The performance is dramatically decreased in the process of many fuzzy association rules algorithms. This is due to the fact that a database is repeatedly scanned to contract each candidate item set with the whole database level by level in process of mining fuzzy association rules. Thus, we propose an efficient method for discovering the fuzzy large it emsets.**

**Fuzzy Cluster-Test Algorithm**

After quantitative attributes are partitioned into several fuzzy sets by FCM algorithm, The algorithm employs some efficient cluster tables to represent database D by a single scan of the database, following by contrasts with the partial cluster tables. A new structure named CDi related to each cluster i , is created simultaneously with creation of each cluster table for each item x and linguistic term A.

B=Before reading each cluster table to compute the supports of fuzzy candidate itemsets , we first check if CDi[A,x]=0 ,then for each fuzzy candidate itemset containing A(x) ,we can ignore scanning the cluster table i.

Fig.1 is the algorithmic form of the algorithm , which, for ease of presentation, is divided into three parts .

Part 1 gets a set of large 1-itemsets and creates M cluster tables ,scan the database once and cluster the transaction data. If the length of transaction record is K , transaction record will be stored in the table, named cluster_tabel(k), $1 \leq k \leq M$ , where M is the length of the longest transaction record in database. Meanwhile, the set of large 1-itemsets, L1,is generated. Moreover , CDi[A,x] is created simultaneously with creation of each cluster table for each item x and linguistic term A.

CDi[A,x] represent the sum of $\mu A(x)$ ,for all values of items x in all transactions of ith cluster table.

Part 2 generates the set of fuzzy candidate k_itemsets Ck . the procedure is similar to the candidate generation of Apriori algorithm [3] .

Part 3 determines the set of fuzzy large k-itemsets Lk , as shown in Fig.3. when the length of candidate itemset is k , the support is calculated with reference to the cluster_tabel(k) . Then it is contacted with the Cluster_Table (k+1),(k+2),… .

Then before scanning each cluster table i ,for calculating the supports of candidate itemsets, structure CDi[A,x] is analyzed first(for all items x and each linguistic term).

If for item x and linguistic term A , CDi[A,x]=0 then scanning the cluster table i can be ignored for computing the support of each fuzzy candidate itemsets containg A(x).because if CDi[A,x]=0 , then the fuzzy support of each candidate itemset containing A(x) will be zero in this cluster

## Experimental results

In this section, we present some experiments that have been carried out to test the efficiency of the proposed approach.
In order to evaluate the effectiveness of FTC, we applied it to a real-life transactional along with fuzzy Apriori_like algorithm, Using Microsoft visual C# on a Pentium III 600 MHz PC with 256MB of available physical memory.

In this experiment, we compared the efficiency of the FCT algorithm to the Apriori_like algorithm. We assumed that there are 3 linguistic value for each attribute.
(1) we randomly sampled 60000 transaction records of experimental data from the real-life Database. The test database contains 10 items, in which the longest transaction record contains 7 items.
We compared The performance of FCT algorithm to Apriori –like algorithm under various users specified minimum support (MinSup), such that 0.50%, 0.40%, 0.30%, and 0.20%. Figure3 shows the execution time required by each of the two algorithms. It is obvious that whenever the min support decreases, the gap between algorithms becomes more evident.

(2) we sampled randomly 60000,70000,80000 and 90000 records of experimental data from real-life database . Again , The number of attribute is 10. We compared The performance of FCT algorithm to apriori-like algorithm. For this experiment, the minimum support was set to 30%. (Fig.4).
We can see When the number of transaction increases, again the gap between algorithms increases too.

## Conclusions

In this paper we have presented an efficient technique to handle quantitive attributes. The FCT algorithm employs some efficient cluster tables to represent database D by a single scan of the database, following by contrasts with the partial cluster tables. A new structure named CDi related to each cluster i , is created simultaneously with creation of each cluster table for each item x and linguistic term A.

Then ,before reading each cluster table to compute the supports of fuzzy candidate itemsets , we first check if CDi[A,x]=0 ,then for each fuzzy candidate itemset containing A(x) ,we can ignore scanning the cluster table i.

So we reduce great amount of scanning data and therefore the running time of mining algorithm is reduced greatly. The experiments done on the real life database show that the proposed approach has a good performance.

**Algorithms (D,Minsup)**
**<u>Input:</u> D, Minsup**
**<u>Output:</u> Answer( Answer = U $L_k$ , for 1≤k≤M )**
**Begin**
**1) initialize();** /*create cluster table , structures CD for each clusters ,compute L1 */
**2) for (k=2; $L_{k-1} \neq ø$ ; k++) do{**
**3) $C_k$= Candidate_itemset_Gen($L_{k-1}$);**
**4) $L_k$=Large_itemset_Gen($C_k$);**
**5) }**
**6) Answer=U $L_k$ ;**
**End**

Figure 1. Main program for the FCT algorithm

**Procedure Large_Itemset_Gen($C_k$)**
**<u>Input:</u> $C_k$**
**<u>Output:</u> $L_k$**
**Begin**
**1)While($C_k \neq ø$)do{**
**2) pick c from $C_k$;**
**3) support(c)=0;**
**4) for (i=k; i ≤ max_length; i++) do{**
**5) if $CD_i$[A,x] =0 and A(x) is subset of c then**
**continue;/***go to next step o of loop */
**6) temp = the fuzzy support of c**
**appearance in the Cluster_Table(*i*);**
**7) support(c) = support(c) + temp ;**
**/*compute support of fuzzy itemset c*/**
**8) }**
**9) support (c) = support ( c) / | D| ;**
**10) if (support (c) ≥ Minsup ) then{**
**11) put c into $L_k$;**
**12) }**
**13) }End.**

Figure 2.Procedure of fuzzy large k-itemsets Generation for FCT

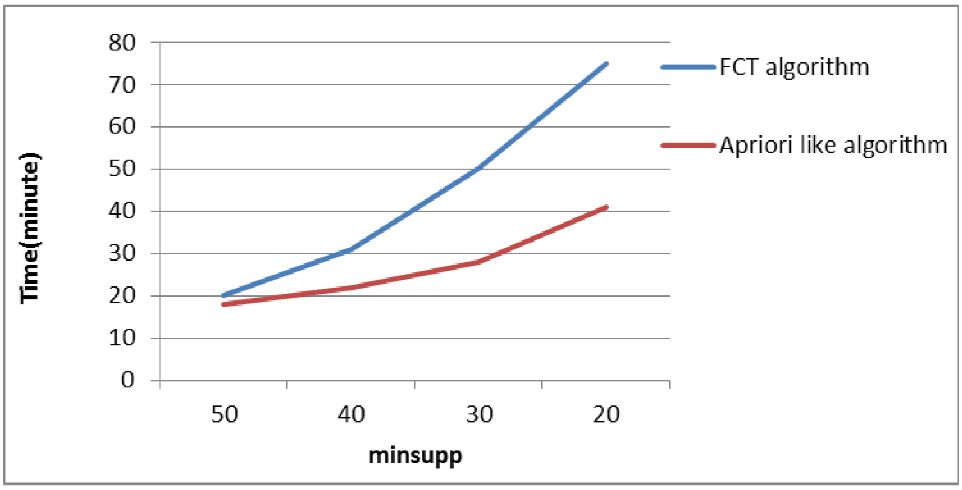
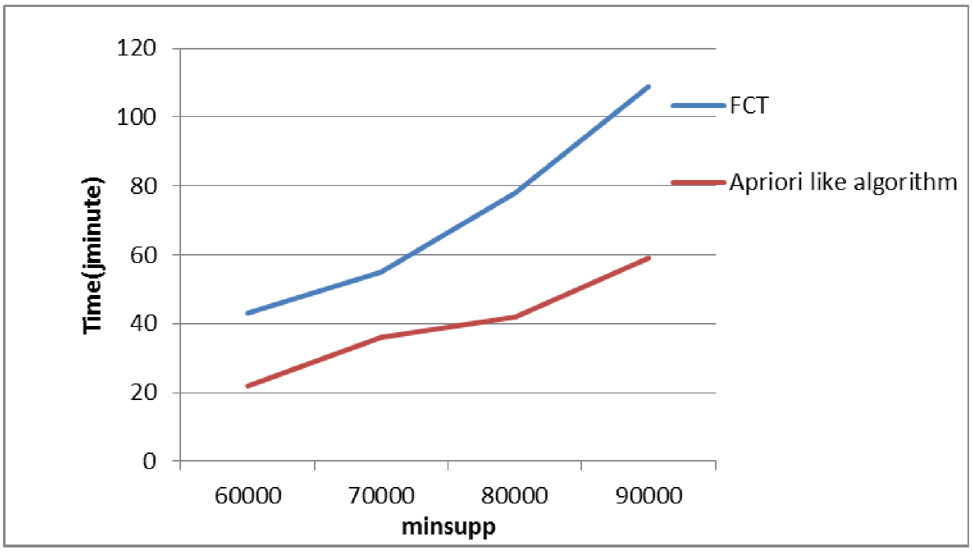



Figure 3.Performance of FCT and Apriori-like algorithm on 60000 records




. Figure 4. Performance of FCT and Apriori-like algorithm at minsupport 0.30%.

**Acknowledgement:** The research is financed *by* Sama technical and vocational training college, Islamic Azad University, Mashhad branch,

## References

1. Han J, Kamber M (2006) Data mining: concepts and techniques, 2nd edn. Morgan Kaufmann, San Fransisco , 2006 .
2. Zhang C, Zhang S (2002) Association rule mining: models and algorithms. Springer, Berlin ,2002
3. Agrawal R, Imielinski T, Swami A (1993) Mining association rules between sets of items in large databases. In: SIGMOD, Washington, D.C., USA, pp 207–216
4. Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In: International conference on very large data bases, Santiago de Chile, pp 487–499

5. Alatas B, Akin E (2006) An efficient genetic algorithm for automated mining of both positive and negative quantitative association rules. Soft Comput Fusion Found Methodol Appl 10(3):230–237
6. Alcala-Fdez J, Flugy-Pape N, Bonarini A, Herrera F (2010) Analysis of the effectiveness of the genetic algorithms based on extraction of association rules. Fundamenta Informaticae 98(1):1–14
7. Han J, Pei J, Yin Y (2004) Mining frequent patterns without candidate generation: a frequent-pattern tree approach. Data Min Knowl Dis 8(1):53–87
8. Sun K, Fengshan B (2008) Mining weighted association rules without preassigned weights. IEEE Trans Knowl Data Eng 20(4):489–495
9. Chen C, Hong T, Tseng V (2011) Genetic-fuzzy mining with multiple minimum supports based on fuzzy clustering. Soft Comput Fusion Found Methodol Appl. doi:10.1007/s00500-010-0664-1
10. Hong T, Kuo C, Chi S (2001) Trade-off between time complexity and number of rules for fuzzy mining from quantitative data. Int J Uncertain Fuzziness Knowl Based Syst 9(5):587–604
11. Hong T, Lee Y (2008) An overview of mining fuzzy association rules. In: Bustince H, Herrera F, Montero J (eds) Studies in fuzziness and soft computing, vol 220. Springer, Berlin, pp 397–410
12. Kaya M (2006) Multi-objective genetic algorithm based approaches for mining optimized fuzzy association rules. Soft Comput Fusion Found Methodol Appl 10(7):578–586
13. Delgado M, Marín N, Sánchez D, Vila M (2003) Fuzzy association rules: general model and applications. IEEE Trans Fuzzy Syst 11(2):214–225
14. Dubois D, Prade H, Sudamp T (2005) On the representation, measurement, and discovery of fuzzy associations. IEEE Trans Fuzzy Syst 13(2):250–262
15. Dubois D, Hullermeier E, Prade H (2006) A systematic approach to the assessment of fuzzy association rules. Data Min Knowl Disc 13(2):167–192lications. IEEE Trans Fuzzy Syst 11(2):214–225
16. Hullermeier E, Yi Y (2007) In defense of fuzzy association analysis. IEEE Trans Syst Man Cybern Part B Cybern 37(4):1039–1043
17. Sudkamp T (2005) Examples, counterexamples, and measuring fuzzy associations. Fuzzy Sets Syst 149(1):57–71
18. Lu Jianjiang, Song Zilin, Qian Zouping, "Mininglinguistic valued association rules", Journal of Software, 2001,12(4), pp.607-611 (in Chinese).
19. H.Ishibuchi,K.Nozaki,N.Yamamoto,H.Tanaka, Selection fuzzy if-then rules for classification problems using genetic algorithms, IEEE Transaction on fuzzy Systems 3 (3) (1995) 206-270
20. L.X.Wang , J.M.Mendel ,Generating fuzzy rules by learning from examples , IEEE transaction on systems, Man , and Cybernetics 22(6) (1992) 1414-1427