*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver. 5 (Jan. 2014), PP 83-87*
*www.iosrjournals.org*

# A Survey of Fuzzy Based Association Rule Mining to Find Co-Occurrence Relationships

Anubha Sharma[1], Asst. Prof. Nirupama Tiwari[2]
[1]*Department of Computer Science & Engineering/Shriram College of engineering & Management [SRCEM] Banmore,Gwalior (MP)/India, 474003*
[2] *Department of Computer Science & Engineering/Shriram College of engineering & Management [SRCEM] Banmore, Gwalior (MP), India, 476444*

***Abstract:*** *Data mining is the analysis step of the "Knowledge Discovery in Databases" process, or KDD. It is the process that results in the discovery of new patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract knowledge from an existing data set and transform it into a human-understandable structure.In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. Association rules are usually required to satisfy a user-specified minimum support and a user-specified minimum confidence at the same time. A fuzzy association rule mining (firstly expressed as quantitative association rule mining) has been proposed using fuzzy sets such that quantitative and categorical attributes can be handled. A fuzzy rule represents each item as <item, value> pair. Fuzzy logic softens the effect of sharp boundary intervals and solves the problem of uncertainty present in data relationships.In this paper we represent a survey of Association Rule Mining Using Fuzzy Algorithm. The techniques are categorized based upon different approaches. This paper provides the major advancement in the approaches for association rule mining using fuzzy algorithms.*
***Keywords****: Data Mining, Association Rule, Fuzzy Association Rule Mining.*

## I. Introduction

Data mining techniques operate on structured data such as corporate databases; this has been an active area of research for many years. The main tasks of Data mining are generally divided in two categories:

Predictive and Descriptive. The objective of the predictive tasks is to predict the value of a particular attribute based on the values of other attributes, while for the descriptive ones, is to extract previously unknown and useful information such as patterns, associations, changes, anomalies and significant structures, from large databases. There are several techniques satisfying these objectives of data mining. Mining Associations is one of the techniques involved in the process mentioned above and among the data mining problems it might be the most studied ones. Discovering association rules is at the heart of data mining. Mining for association rules between items in large database of sales transactions has been recognized as an important area of database research. These rules can be effectively used to uncover unknown relationships, producing results that can provide a basis for forecasting and decision making. The original problem addressed by association rule mining was to find a correlation among sales of different products from the analysis of a large set of data.

## II. Background

### 2.1Association Rule:-

The task of mining association rules over market basket data is considered a core knowledge discovery activity.

Association rule mining provides a useful mechanism for discovering correlations among items belonging to customer transactions in a market basket database. Let D be the database of transactions and J = {J1, ...,Jn} be the set of items. A transaction T includes one or more items in J (i.e., $T \subseteq J$). An association rule has the form $X \Rightarrow Y$ , where X and Y are non-empty sets of items (i.e. $X \subseteq J$, $Y \subseteq J$) such that $X \cap Y = \emptyset$. A set of items is called an itemset, while X is called the antecedent. The support sprtD(x) of an item (or itemset) x is the percentage of transactions from D in which that item or itemset occurs in the database. In other words, the support sprt () of an association rule $X \Rightarrow Y$ is the percentage of transactions T in a database where $X \cup Y \subseteq T$. The confidence or strength c for an association rule $X \Rightarrow Y$ is the ratio of the number of transactions that contain $X \cup Y$ to the number of transactions that contain X. An itemset $X \subseteq J$ is frequent if at least a fraction sprt() of the transaction in a database contains X. Frequent itemsets are important because they are the building blocks to obtain association rules with a given confidence and support.[1]

- **Support**

The rule X ⇒ Y holds with support s if s% of transactions in D contains X ∪ Y. Rules that have a s greater than a user-specified support is said to have minimum support.

- **Confidence**

The rule X ⇒ Y holds with confidence c if c% of the transactions in D that contain X also contain Y. Rules that have a c greater than a user-specified confidence is said to have minimum confidence.

Almost all association algorithms are objective and use some form of statistical analysis to determine the usefulness of a rule. Thus the set of all transactions used in the analysis must be sufficiently large in order for association rules to be concluded from it. Therefore, for the rest of this article, the term large will be used to describe a set of data with enough transactions to obtain association rules.

There are a few commonly used terms that must be defined:

→**Itemset:** An itemset is a set of items. A k-itemset is an itemset that contains k number of items.

→**Frequent itemset*:*** This is an itemset that has minimum support.

→ **Candidate set*:*** This is the name given to a set of itemsets that require testing to see if they fit a certain requirement [1] and [5].

### 2.2 Algorithms For Mining Association Rules

The 'classical' associations rule problem deals with the generation of association rules by defining a minimum level of confidence and support that the generated rules should meet. This is the case since first of all we want the rule to contain items that are purchased often - if we know that whoever buys product A byes product B as well but product A occurs only in two out of ten million transactions then it is of low interest unless the profit margin is high. Furthermore, within the set of transactions that contain item A, we want to know how often they contain product B as well; this is the role of rule's confidence.

If we introduce the term *frequent* for an itemset X that meets the criterion that its support is greater than the minimum value set- for example, we might say that we want all items or set of items that were bought by more than 70% of our customers- then our problem is restricted to finding all frequent itemsets from the database. If we know these, then we can derive all association rules by following a simple strategy. This strategy involves the calculation for every frequent itemset X and very subset Y of it- which is neither the null neither set nor X - of the confidence level of all rules of the form X\Y=>Y; the latter is essential so that we will comply with the part of the definition that demands the intersection set to be the null. As an example, if they consider an itemset that consists of pork steaks, coke and beer, the previous definition suggests that we should look for the confidence level of, say, the rule 'pork steak => coke and beer'. They, then, drop those that do not meet the minimum confidence level criterion. The problem, however, is that a small number of items is able of generating a large search space. This space, on the other hand, has a very interesting property that facilitates our work; there is a border that separates the frequent itemsets from the infrequent ones- thus, the problem is restricted on finding that border.

This is done through a mapping procedure between the set items and the set of natural numbers and the use of special classes. Each algorithm presented below will be characterized by how it looks for the border between frequent and infrequent itemsets and how it calculates the support value for each of them. The first can be done either by using the breadth-first search (BFS) or the depth-first search (DFS). In the breadth-first algorithm given a tree and a goal state they try initially all paths- ways of reaching our goal state - of length one, then all paths of length two and so on till they reach the goal state. In the depth-first search they try a path first till they get to a dead-end; then they return to the top and look for alternatives. As far as the calculation of support value is concerned, it can be done using either the number of the subset's occurrences in the database or by using set intersections [5-6].

## III. Survey Of Association Rule Mining Using Fuzzy Approach

### 3.1 Mining Fuzzy Frequent itemset using Compact Frequent Pattern(CFP) tree Algorithm

K.SuriyaPrabha, R.Lawrance,et. al. designed a novel method for generation of strong rule. The proposed construction algorithm for building a Fuzzy CFP tree from a quantitative database is described in this section. The proposed approach integrates the fuzzy-set concepts and the variation of the classic FP-tree-like approach to efficiently find the fuzzy frequent itemsets from the quantitative transactions. The Fuzzy FP-tree construction algorithm is first designed to build the tree structure for the fuzzy frequent 1-itemsets. Each node in the tree structure keeps a fuzzy frequent 1-itemset, its membership value, and the membership values of its super-itemsets in the path according to the intersection operator, which is the minimum operator here.In this algorithm they provide the following input:

**INPUT:** A quantitative database consisting of n transactions and m items, a set of membership functions and a predefined minimum support threshold s.

**OUTPUT:** A constructed CFP tree.

this proposed work integrates the fuzzy set concepts in the newly proposed CFP-tree algorithm by constructing a compact sub-tree for a fuzzy frequent item, generating candidates in batch from the compact sub-tree and later release the current subtree from memory leaving the space for next subtree thus significantly outperforms the other algorithms on both execution times, memory usages and reducing the search space finally resulting in the discovery of fuzzy frequent itemsets.[2]

### 3.2 Fuzzy Weighted Associative Classifier: A Predictive Technique for Health Care Data Mining

Sunita Soni and O.P.Vyas et al. designed an algorithm for fuzzy weighted association classification in which they extend the problem of classification using Fuzzy Association Rule Mining and propose the concept of Fuzzy Weighted Associative Classifier(FWAC).Classification based on Association rules is considered to be effective and advantageous in many cases. Associative classifiers are especially fit to applications where the model may assist the domain experts in their decisions. Weighted Associative Classifiers that takes advantage of weighted Association Rule Mining is already being proposed. However, there is also-called "sharp boundary" problem in association rules mining with quantitative attribute domains. This paper proposes a new Fuzzy Weighted Associative Classifier(FWAC) that generates classification rules using Fuzzy Weighted Support and Confidence framework. Then aïve approach can be used to generating strong rules instead of weak irrelevant rules. Where fuzzy logic is used in partitioning the domains.The problem of Invalidation of Downward Closure property is solved and the concept of Fuzzy Weighted Support and FuzzyWeighted Confidence framework for Boolean and quantitative item with weighted setting is generalized [3].

### 3.3 Frequent Item sets from Multiple Datasets with Fuzzy data

Praveen Arora, R. K. Chauhan and Ashwani Kush et.al.proposed a Traditional approaches handles crisp and fuzzy data very well but very less published results show that databases that contain multiple tables with fuzzy data having taxonomy can be handled efficiently. The Proposed algorithm is discovered by extending these traditional algorithms and helps to find the multi level fuzzy association rules in Entity –Relationship modeled databases, which is capable to handle multiple tables. The study analyzes how the attributes of several entities appear together. The Study also analyzes the rules with respect to the relationships existing between the entities and their ancestors. If several relationships exist between two or more entities, then the fuzzy association rules between their attributes and ancestors are examined with respect to each such relationship. The discovered algorithm uses the join and entity supports in determining frequent item sets. By considering the entity support it does not eliminate from the result entity item sets that are frequent with respect to their entity table but not with respect to the relationship table and it also allows the computation of correct support and confidence for rules existing among attributes of the same entity table [4].

### 3.4 .An Improved Algorithm for Mining Association Rules in Large Databases

Farah Hanna AL-Zawaidah, YosefHasanJbara and Marwan AL-Abed Abu-Zanona et. al. present a novel association rule mining approach that can efficiently discover the association rules in large databases. The proposed approach is derived from the conventional Apriori approach with features added to improve data mining performance. They had performed extensive experiments and compared the performance of the algorithm with existing algorithms found in the literature. Experimental results show that the approach outperforms other approaches and show that approach can quickly discover frequent itemsets and effectively mine potential association rules.

This paper they attack the association rule mining by an apriori based approach specifically designed for the optimization in very large transactional databases. The developed mining approach called Feature Based Association Rule Mining Algorithm.

The developed approach adopts the philosophy of Apriori approach with some modifications in order to reduce the time execution of the algorithm. First, the idea of generating the feature of items is used and; second, the weight for each candidate itemset is calculated to be used during processing. By storing the appearing feature of each interested item as a compressed vector separately, the size of the database to be accessed can be reduced greatly.

This paper is to improve the performance of the conventional Apriori algorithm that mines association rules by presenting fast and scalable algorithm for discovering association rules in large databases. The approach to attain the desired improvement is to create a more efficient new algorithm out of the conventional one by adding new features to the Apriori approach. The proposed mining algorithm can efficiently discover the association rules between the data items in large databases. In particular, at most one scan of the whole database is needed during the run of the algorithm. Hence, the high repeated disk overhead incurred in other mining algorithms can be reduced significantly. They compared our algorithm to the previously proposed algorithms found in literature. The findings from different experiments have confirmed that the proposed approach is the most efficient among the others. It can speed up the data mining process significantly as demonstrated in the

performance comparison. Furthermore, gives long maximal large itemsets, which are better, suited to the requirements of practical applications. They demonstrated the effectiveness of the algorithm using real and synthetic datasets. They developed a visualization module to provide users the useful information regarding the database to be mined and to help the user manage and understand the association rules.

The proposed technique need to improve in the mining multidimensional association rules from relational databases and data warehouses and also in mining multilevel association rules from transaction databases [5].

**3.5 An Algorithm For Mining Fuzzy Association Rules**

Reza Sheibani , Amir Ebrahimzadeh ,Member, IAUM presents a paper , in this paper, we presentan efficient algorithm named Fuzzy Cluster-Based AssociationRules(FCBAR).The FCBAR method is to create cluster tables by scanning thedatabase once, and then clustering the transaction records tothe k_th cluster table, where the length of a record is k.Moreover, the fuzzy large itemsets are generated by contrastswith the partial cluster tables. This prunes considerableamount of data, reduces the time needed to perform data scansand requires less contrast. Experiments with the real-lifedatabase show that FCBAR outperforms fuzzy Apriori_likealgorithm , a well–known and widely used association rulesalgorithm. In this paper we proposed the efficient algorithm for mining fuzzy association rules. The FCBAR algorithm creates cluster table to aid discovery of fuzzy large itemsets. Contrasts are performed only against the partial cluster tables that were created in advance. Experiments with the real-life database show that FCBAR outperforms Apriori_like algorithm, a well-known and widely used association rule.[6]

**3.6 Efficient Parallel Pruning of Associative Rules with Optimized Search**

The main focus of this research work is to propose an improved association rule mining algorithm to minimize the number of candidate sets while generating association rules with efficient pruning time and search space optimization. The relative association with reduced candidate item set reduces the overall execution time. The scalability of this work is measured with number of itemsets used in the transaction and size of the data set. Further Fuzzy based rule mining principle is adapted in this work to obtain more informative associative rules and frequent items with increased sensitive. The requirement for sensitive items is to have a semantic connection between the components of the item value pairs.

The problem of scalability and higher memory requirements are addressed in this research work by deploying parallel pruning technique at different levels of itemssets (one itemset, two itemset,etc.,). From the recent literature we came to know that, only Apriori and its adaptations are used for generating association rules. Thus, the Fuzzy based Optimal Search Space Pruning (FOSSP)is compared with existing fuzzy Apriori and the execution time is recorded.

The objective is to minimize the number of candidate sets and enhancing the association rule mining algorithm while K.Sangheetha, Dr.P.S.Periasamy,S.Prakash et.al. creating an association rules by evaluating maximal information associated with each item that occurs in given set of transaction. Initial work starts with the evaluation of weighted association rule mining in terms of item-value relational metrics. Then then umber of item metrics is taken into account of the association rule mining with reduced candidate itemset. This may decrease not only the number of itemsets generated but also the overall execution time of the algorithm. Any valued attribute will be treated as item-value relational metrics and will be used to derive the minimal number of association rules which increased the rules informationcontent.[7]

**3.7. FPrep: Fuzzy Clustering driven Efficient Automated Pre-processing for Fuzzy Association Rule Mining**

Ashish Mangalampalli, Vikram Pudi proposed the method for preprocessing of Fuzzy Association Rule Mining. This paper describes a methodology, called FPrep, to do this pre-processing, which first involves using fuzzy clustering to generate fuzzy partitions, and then uses these partitions to get a fuzzy version (with fuzzy records) of the original dataset. Ultimately, the fuzzy data (fuzzy records) are represented in a standard manner such that they can be used as input to any kind of fuzzy ARM algorithm, irrespective of how it works and processes fuzzy data. We also show that FPrep is much faster than other such comparable transformation techniques, which in turn depend on non-fuzzy techniques,

FPrep, for ARM in a fuzzy scenario.FPrep is meant for seamlessly and holistically transforming a crisp dataset into a fuzzy dataset such that it can drive a subsequent fuzzy ARM process. It does not rely on any non-fuzzy techniques, and is thus more straightforward, fast, and consistent. It facilitates user-friendly automation of fuzzy dataset.FPrep has been compared with other such techniques, and has been found to better on the basis of speed. We also illustrate its efficacy on the basis of quality of fuzzy partitions generated and the number of itemsets mined by a fuzzy ARM algorithm which is preceded by FPrep. This pre-processing technique provides us with a standard method of fuzzy data (record) representation in a fuzzy dataset such that it is useful for any kind of fuzzy ARM algorithm, irrespective of how the algorithm works. Furthermore, this pre-processing methodology has been adequately tested with two disparate fuzzy ARM

algorithms, Fuzzy Apriori and Fuzzy ARMOR, and would also work fine with other fuzzy ARM algorithm.[8]

**After surveying different techniques on Association Rule Mining Using Fuzzy Algorithm we defined the advantages and disadvantages of the techniques in the table below:**

| Techniques | Advantages/ Merits | Disadvantages /Future Direction |
|---|---|---|
| Association Rule Mining, FuzzyAlgorithm,Compact Fuzzy Tree Structure(CFT) | The proposed optimize association rule mining using new Compact Frequent tree Generation and finding Frequent sets efficiently.Through this direction it got a better result. | The algorithm does not sufficient effective and it can't incorporate with other techniques, so it will need to improve in future work [2]. |
| Data Mining, Fuzzy Weighted Associative Classifier(FWAC) | In this paper the authors have tried to predict the value of attributes in the basis of some other attributes. And run on the real life medical database. And Generate the results by using weighted confidence. Very much applicable on real life datasets. | This technique needs major modifications to improve the complexity reduction of Association rule mining.[3] |
| Fuzzy Approach, Association Rules, join operations. | In this paper author join multiple tables by applied the star schema. And then applicable to generate multi-dimensional Association Rules with the consequent part consists of single attribute and more than one attribute. The results reported in this paper are very promising. | This technique need to minimize the complexity of the algorithm and scanning of database by applying theorem on the generated rule [4]. |
| Association rules; Frequent Patterns; Apriori | The approach to attain the desired improvement is to create a more efficient new algorithm out of the conventional one by adding new features to the Apriori approach. The proposed mining algorithm can efficiently discover the association rules between the data items in large databases. | The proposed technique need to improve in the mining multidimensional association rules from relational databases and data warehouses and also in mining multilevel association rules from transaction databases [5]. |
| Association Rule Mining, FCBAR(Fuzzy Cluster Based Association Rules) | The proposed approach dealt with a challenging clustering association rule mining problem of finding interesting association rules. The results of this paper were good since the discovered rules are of a high predictive accuracy and of a high interesting value. | This technique does not sufficient reliable for a large dataset, it need to improve for the application in large data set [6]. |
| Association Rule, Apriori, Parallel Pruning | The proposed algorithm based association rule mining algorithm for the prioritization of the rules. This approach significantly reduces the Search Space. | The technique can be extended by the incorporation of the other interesting measures in the literature to future work [7]. |
| Clustering,Association Rule Mining,Automated Preprocessing. | This approach provide very fast preprocessing using clustering ,which is very much important for fuzzy based algorithm. Thus reduces the time Complexity. | The technique can not be incorporate with other techniques. Only applied with clustering based preprocessing.[8] |

## IV. Conclusion

Traditional rule mining methods, are usually accurate, but have very hard and fragile operations. Fuzzy Based algorithms on the other hand provide a robust and efficient approach to explore large search space. In recent years numerous works have been carried out using Fuzzy algorithm for mining association rules. As many works have been carried out on mining association rules with Fuzzy algorithms this paper surveys the existing work on application of Fuzzy algorithm in mining association rules and analyzes the performance of the methodology adopted. During the survey, we also find some points that can be further improvement in advanced association rules mining with Fuzzy algorithm to achieve more efficient accuracy in result and maintain a high confidence and a good coverage of the database, also providing the user with high quality rules.

## References


[1] Han, J., Kamber, M. (2001). "Data Mining: Concepts and Techniques", Harcourt India Pvt. Ltd.

[2] K.SuriyaPrabha, R.Lawrance," Mining Fuzzy Frequent itemset using Compact Frequent Pattern(CFP) tree Algorithm" International Conference on Computing and Control Engineering (ICCCE 2012), 12 & 13 April, 2012.

[3] SunitaSoni, O.P.Vyas,"Fuzzy Weighted Associative Classifier: A Predictive Technique For Health Care Data Mining", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.1, February 2012.

[4] Praveen Arora, R. K. Chauhan and AshwaniKush ," Frequent Itemsets from Multiple Datasets with Fuzzy data", International Journal of Computer Theory and Engineering, Vol. 3, No. 2, April 2011.

[5] Farah Hanna AL-Zawaidah, YosefHasanJbara and Marwan AL-Abed Abu-Zanona, "An Improved Algorithm for Mining Association Rules in Large Databases", World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 1, No. 7, 2011, pp. 311-316.

[6] Reza Sheibani , Amir Ebrahimzadeh ,Member, IAUM," An Algorithm For Mining FuzzyAssociation Rules", Proceedings of the International MultiConference of Engineers and Computer Scientists 2008 Vol I, March, 2008,pp.486-490.

[7] K.Sangeetha ,Dr.P.S.Periasamy , S.Prakash,"Efficient Parallel Pruning of Associative Rules with Optimized Search", IOSR Journal of Computer Engineering (IOSRJCE) ,volume no.3,pp.26-30.

[8] AshishMangalampalli, VikramPudi,"FPrep: Fuzzy Clustering driven Efficient Automated Pre-processing for Fuzzy Association Rule Mining", IEEE Intl Conference on Fuzzy Systems (FUZZ-IEEE), July 2010.

[9] G Vijay Krishna,PRadhaKrishna,"A Novel Approach for Statistical and Fuzzy association Rule Mining on Quantiative Data ",Journal of scientific and industrial Reasearch ,vol no.67,jul2008,pp.512-517.