

A Review of Fuzzy Association Rule Mining Algorithms

Rajkamal Sarma

Department of Information Technology

Rangia College, Rangia, Assam

E-mail: rajkamal_sarma@rediffmail.com

Abstract

Association Rule Mining (ARM) holds a pivotal role in the Data Mining process, generating meaningful associations in the form of rules based on predefined interestingness measures. The integration of the Fuzzy concept enhances usefulness of classical ARM technique. Fuzzy Association Rule Mining (FARM) techniques leverage membership degrees and linguistic terms, ensuring greater accuracy and natural representation of data. This paper delves into the evolution of FARM algorithms and their applications across diverse fields spanning three decades. Through a review of prominent algorithms, essential features are identified, providing valuable insights for further research in this domain.

Keywords: Fuzzy Association Rule, Data Mining, Membership degree, Linguistic Term.

Introduction

With rapid increase in the growth of data in every application, converting data into proper knowledge has already become a challenging issue. Knowledge Discovery in Database (KDD) plays an important role in conversion of data into required information. According to Fayyad et al., KDD is a nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data (U.M. Fayyad, et.al., 1996) Data mining is a key step of KDD process. Data mining consists of different functions, namely, classification, regression, clustering, summarization, image retrieval, discovering association rules, functional dependencies and rule extraction, etc. (Mitra, S. & Acharya, T., 2005). Rule mining is one of the key functions in data mining practices where rule can be mined or generated from data for the discovery of the relationship among the attributes of a dataset of transactions. Thus, the discovery of association rules is an important area of data mining research in which interesting association relationships among different attributes are described (Agrawal, R., et.al., 1993). In Market-Basket analysis, by applying techniques of association rule mining, buying habits or patterns of customers are analyzed. For this, associations between items bought by customers are found from the sales transactions. To find patterns using association rule mining it needs to detect rules of the form $A \rightarrow B$ in data containing transactions. Such a rule indicates an item containing an attribute A is likely to contain attribute B also, for example, **bread** \rightarrow **milk**. However, to generate association rules, two important quality measures viz. support and confidence are required. Depending on data representation, association rules can be classified into different types like Boolean,

generalized, quantitative etc. A Boolean association involves binary attributes, a generalized association involves attributes that are hierarchically related and a quantitative association involves attributes that can take on quantitative or categorical values. But all these types of association rules have some limitations to discover nontrivial knowledge. When a database contains values between 0 and 1, it is also possible to extend the classical mining algorithm using Fuzzy set theoretical operations to obtain Fuzzy association rules. By using Fuzzy sets imprecise terms and relations commonly employed by humans in communication and understanding can be optimally modelled (Delgado, M. et.al.,2003). Use of Fuzzy technique has been considered as one of the key components of data mining system because of their affinity with human knowledge representation (Maeda, A. et.al.,1995).

Fuzzy sets allow a flexible assignment of membership of elements to a set. While in crisp set theory, an element may or may not belong to a set, in fuzzy set theory many degrees of membership (between 0 and 1) are allowed. For this, a membership function $\mu_A(x)$ is associated with a fuzzy set A, such that the function maps each element of the universe of discourse x (or the reference set) to the interval [0, 1]. For example, a group of students is classified on the basis of the percentage of marks into poor, medium and good in a test. The interval of percentage is fixed as different category like (30%-45%) as poor, (45%-60%) as medium and (60% and above) as good. Now if a student secures 59.5%, he will be categorized as medium quality by using crisp set, though he or she is just 0.5% away from next upper category. Thus, sometimes it may not be a proper and accurate justification with a sharp boundary i.e. only 0 or 1 in a crisp set. To solve this problem and treat data with more accuracy, Fuzzy sets can be used so that data can be converted into grades between 0 and 1 with some membership function. Fuzzy set theory was introduced by L.A. Zadeh (Zadeh, L. A., 1965) to provide an approximate and yet effective way for describing the characteristics of a complex system to admit precise mathematical analysis (Zadeh, L. A. 1975). Fuzzy sets can be viewed as an extension of the classical crisp sets.

Fuzzy set theory is an expansion of traditional crisp set theory, focusing on quantifying and reasoning with natural language. It defines the process of determining whether elements from a universal set are members or non-members of a crisp set using a characteristic or discrimination function. A membership function, denoted by $\mu(x)$, represents the membership grades of elements in the set, which can be used to define fuzzy set A. Formally, given a set X of elements $x \in X$, any fuzzy subset A of X is defined as

$$A = \{x, \mu_A(x) | x \in X, \mu_A(x) \in [0,1]\} \quad (1)$$

Fuzzy set theory involves membership functions (Deng, J., & Deng, Y. 2021) that determine the degree membership of an element. These functions directly influence the fuzzification process, converting crisp data into fuzzy data. Common types of membership functions include triangular, trapezoidal, gaussian, and sigmoidal. The paper (Jain, A., & Sharma, A. 2020) analyzes techniques for designing optimal fuzzy logic systems using triangular, trapezoidal, and Gaussian functions. Results show that Trimf and Trampf provide the highest accuracy, while Gaussmf shows the most stable accuracy. Fuzzy logic is widely used in

industries like automobiles, consumer electronics, image processing, machine learning systems.

Membership degree in fuzzy set theory is a fundamental concept, indicating the degree of an element's membership in a fuzzy set. This degree, unlike binary membership, allows partial membership values ranging from 0 to 1, a feature that is crucial for handling vagueness and imprecision in real-world scenarios. The degree is determined using a membership function.

Linguistic terms are descriptive labels used in fuzzy set theory and fuzzy logic to represent qualitative or subjective concepts in human-readable language. They are often connected with fuzzy sets and defined by membership functions. Linguistic terms are crucial in fuzzy control systems, decision-making processes, and expert systems to model uncertainty and vagueness, bridging the gap between human intuition and machine logic.

Performance measures in Association Rule Mining (ARM) assess the quality and relevance of rules, focusing on key factors like support, confidence, and lift. These measures determine the frequency, strength, and co-relation of associations between items, with most algorithms evaluated using the support-confidence framework (Gyenesei A. & Teuhola J., 2001). Support indicates frequency, confidence measure reliability, and lift evaluates independence.

Some key performance measures are defined below:

Support: The support of an item set is the proportion of transactions in the dataset in which the item set appears. It indicates the frequency of occurrence of the item set in the dataset.

$$Support(A) = \frac{\text{Number of transactions containing } A}{\text{Total number of transactions}} \quad (2)$$

Confidence: The confidence of a rule $A \rightarrow B$ is the proportion of transactions containing A that also contain B . It measures how often items in B appear in transactions that contain A .

$$Confidence(A \rightarrow B) = \frac{Support(A \rightarrow B)}{Support(A)} \quad (3)$$

Related Works

Rule mining is a key function in data mining practices, where rules are mined or generated from data to discover relationships among attributes of a transaction dataset. Association rule generation is significant in data mining and research activity, as it helps explain interesting relationships among various attributes (Agrawal, R. et.al., 1993). In Market-Basket analysis, associations between items bought by customers are found from sales transactions. Association rules can be classified into different types like Boolean, generalized, and quantitative. However, these types have limitations in discovering nontrivial knowledge. Fuzzy set theory can be used to extend classical mining algorithms for databases containing values between 0 and 1, allowing for optimal representation of imprecise terms and relations (Delgado, M., et.al., 2003). The simplicity of knowledge representation has led to the recognition of fuzzy-based techniques as an important component of data mining systems (Maeda, A., et.al., 1995). IFWARM, an improved version of the standard Fuzzy weighted association rule mining algorithm, provides more stability and effectiveness using the Weighted DCP (Yang, T., & Li, C., 2015). IFWIAR, an algorithm (Sumathi, G., & Akilandeswari, J., 2020) executed based on recommendation models, improves the quality of

diagnostic and services in medicine. Fuzzy Expert system (FES) has been applied in diagnosis using different linguistic terms to classify parameters (Allahverdi, N., 2019). An attribute-oriented approach to Knowledge Discovery (KDD) was investigated and applied in generalization to reduce computational complexity in database learning processes (Han, J., et.al., 1992).

Fuzzy Association Rule Mining

The Fuzzy concept in data mining is a method that extracts association rules from quantitative databases. It addresses the boundary problem in classification of attributes based on an assumed range of values, such as crisp sets where nearby values are often ignored or overemphasized. Fuzzy sets grade membership to multiple sets, addressing this issue in categorical data. Fuzzy sets also help address the partial membership of an attribute in real-world situations by using proper linguistic terms. However, F-PNWAR (Mangayarkkarasi, K., & Chidambaram, M., 2017) mining algorithm emphasizes both positive and negative association rule mining, a new addition to rule mining techniques. E-FWARM (Mangayarkkarasi, K., & Chidambaram, M., 2018) is another modified and enhanced fuzzy-based approach that is useful in weighted association rule mining, assigning proper weight of itemset based on significance, leading to more meaningful rule generation from the database. The traditional FARM approach has an interestingness measure framework, but Fuguang Bao and et.al. (Bao, F., et.al., 2021) have analysed the merits and demerits of traditional measures used in classical ARM and proposed new parameters that are found to be more effective. In these situations, Fuzzy technique or fuzzy concept may be more effective, making FARM a mining technique based on fuzzy set theory and fuzzy logic.

Fuzzification is a method that converts crisp values into fuzzy ones to solve the sharp boundary problem between attributes. Fuzzy set theory uses membership functions to map intervals between [0,1] into different linguistic terms. However, using a sharp boundary in a crisp set may not provide accurate justification. Fuzzy set theory was proposed by L.A. Zadeh (Zadeh, L. A., 1965) to provide an approximate but effective way for describing complex system characteristics for precise mathematical analysis (Zadeh, L. A., 1975). It has been applied in various aspects of information retrieval (Pasi, G., 2008) including E-Commerce application and development (Lu, J., Ruan, D., & Zhang, G., 2008). Membership Functions are a useful method for fuzzifying values and representing them in different intervals. Fuzziness, a solution to complex fuzzy systems, balances information and uncertainty, offering flexibility and simplicity. This approach connects human reasoning to intelligent systems, simplifying the complexity of fuzzy systems.

Leading FARM Algorithms

Discovering proper association among attributes in transactional data acquired a prime attention of the researchers since the formulation of the problem, often called as “the market-basket problem.” Several algorithms are designed to discover the association rules applying different strategies. Algorithms like Apriori, Apriori TID (Agrawal, R., & Srikant, R., 1994) etc. were developed to improve the previous approaches. However, incorporation of fuzzy set has brought about a substantial transformation in rule mining technique and their

corresponding algorithms. Introduction of fuzzy set theory in association rule mining algorithms mostly involves the following steps:

- Pre-processing of raw data as sample dataset and presented as transaction dataset
- Conversion of the transaction dataset into fuzzy dataset using membership function.
- Classification of fuzzified value into linguistic terms.
- Frequent itemset generation based on the measures of interestingness.
- Generating Fuzzy Association Rules with natural and meaningful representation.

Here, some leading algorithms related to Fuzzy ARM techniques are explained.

F-APACS

Fuzzy Automatic Pattern Analysis and Classification System, commonly known as F-APACS (Chan, K. C., & Au, W. H. 1997) is an algorithm designed to extract Fuzzy Association Rules from quantitative databases. It uses linguistic terms and adjusted difference concepts, which are crucial in data mining as they simplify human knowledge representation. The algorithm uses user-defined interestingness measures like support and confidence to generate association rules, but this can lead to unnecessary noisy rules and irrelevant relationships. To address this, F-APACS introduces new measures like adjusted difference analysis and weight of evidence. The algorithm allows users to extract rules based on positive and negative association, which is a key feature of F-APACS. Most algorithms discover only positive association rules, where an attribute value is represented as a linguistic term. F-APACS is useful for mining linguistic data, as it is easy to understand. The concept of linguistic term is defined based on fuzzy set theory, making F-APACS a unique FARM approach.

FTDA

Fuzzy Transaction Data-mining Algorithm (FTDA) is an important and simple technique in data mining proposed by Hong and his co-researchers. (Hong, T. P., et.al, 1999,2008). This method operates in two main phases: firstly, the quantitative dataset is transformed into a fuzzy dataset using fuzzy membership functions, categorizing values into linguistic terms such as low, medium, or high. Subsequently, the most frequent items from each attribute, determined by their highest scalar cardinality, are chosen for subsequent iterations. An Apriori-based algorithm is then applied to uncover associations among different items, presenting them in the form of rules. To refine the results, various interestingness measures like support, confidence, and lift are utilized, aiding in the extraction of meaningful rules for further recommendation and analysis

FQARM

The Fuzzy Quantitative Association Rule Mining Algorithm is proposed by A. Gyenesei (Gyenesei, A., 2000) applied on quantitative dataset. The method of extracting Fuzzy Association rules involves finding frequent item sets by counting fuzzy support and comparing them with user-specified minimum support. The technique is used to discover association rules in datasets with both quantitative and fuzzy data. It incorporates fuzzy logic into the mining process, allowing for handling imprecise or uncertain quantitative information. The technique considers weighted associations, adding complexity for rules with varying significance levels. The findings offer insights into the integration of fuzzy logic and

quantitative data mining, providing a more nuanced perspective on association rule discovery.

CFARM

The CFARM algorithm (Khan, M. S., et.al. 2008) (Sarma, R., & Sarma, P. K. D.,2020). stands out for its ability to handle composite items, where each item comprises several attributes. Composite data item can be defined as combination of several data. The CFARM technique transforms numerical values into fuzzy sets, representing imprecision and uncertainty. It uses a comprehensive approach considering individual attributes and combinations to discover intricate fuzzy association rules. The algorithm introduces a new technique for evaluating rule interestingness, ensuring meaningful patterns extraction. It leverages fuzzy logic principles to provide a robust framework for mining fuzzy association rules from composite items

FWARM

The FWARM Algorithm (Muyeba, M., et.al., 2009) is a modified version of the Weighted ARM Algorithm (Khan, M. S., et.al., 2008), which introduces a new concept of assigning weight to items in a transactional dataset. This approach addresses the issue of some items being less important in generating interesting rules. The FWARM Algorithm also addresses the Downward Closure Property, which was not addressed by other quantitative or Boolean algorithms. The property assumes that all subsets are large if the itemset is large, but the FWARM algorithm does not hold this assumption, as each item is assigned by weight, allowing for larger itemset. The algorithm follows the breadth first search traversal method, based on tree data structures and follows the same execution process as the Apriori Algorithm. It also avoids pre-processing and post-processing steps to eliminate additional steps.

Table.19. Summary of leading Fuzzy ARM Algorithms

Sl No	Name of the Algorithms	Authors	Nature	Data Type	Features /observations
1	Fuzzy Automatic Pattern Analysis and Classification System (F-APACS)	Keith C.C.Chan, Wai-Ho Au	FCM based	Categorical, Quantitative	Linguistic terms are applied instead of discretizing the domains of quantitative attributes. Adjusted difference is utilized in place of user-defined measures. The approach enables the discovery of both positive and negative association rules.
2	Fuzzy Transaction Data-mining Algorithm (FTDA)	Tzung-Pei Hong, Chan-Sheng Kuo, Sheng-Chai Chi	Apriori TID-based	Quantitative, Categorical	A refined or degraded membership function can be applied to effectively address conventional data-related challenges. Compared to traditional crisp set mining methods for quantitative data, smoother mining outcomes with improved time complexity can be achieved.
3	Fuzzy Quantitative Association Rule Mining (FQARM)	Attila Gyensei	Apriori-based	Quantitative, Binary	Fuzzy sets can be used to overcome sharp boundary problems. The fuzzy normalization process addresses issues arising from the fuzzy partitioning of quantitative data. Incorporating correlation measures

Edited by: Dr. Sisir Kumar Rajbongshi | Dr. Kshirod Sarmah |
Dr. Rajdeep Ghosh | Dharmeswar Tarang
Department of Computer Science, Pandit Deendayal Upadhyaya
Adarsha Mahavidyalaya, Amjonga, Goalpara, Assam, India

					enhances the accuracy of the generated rules.
4	Composite Fuzzy Association Rule Mining (CFARM)	Maybin Mueyba, M. Sulaiman khan, Frans Coenen	Apriori-based	Composite data	The novel concept of <i>composite items</i> is introduced, enabling the linkage of related dataset properties and the discovery of associations among those attributes to find rules. The application of the <i>certainty factor</i> offers an effective means of identifying correlations among rules.
5	Fuzzy Weighted Association Rule Mining (FWARM)	Maybin Mueyba, M. Sulaiman khan, Frans Coenen	Weighted-based	Quantitative	Based on importance or significance of data items, some value was assigned in weighted ARM and it leads an issue Downward Closure Property (DCP). The issue of invalidation of DCP is addressed within the support-confidence framework, applicable to both classical and fuzzy associations.

Conclusions

This paper provides a survey of various Fuzzy Association Rule Mining (FARM) techniques, presenting algorithms proposed by different researchers over the past three decades. The survey reveals significant opportunities for future research aimed at enhancing the efficiency and performance of these algorithms. With the rapid and continuous growth of data, researchers increasingly require innovative methods to effectively process and analyze large datasets. From this study, several critical issues have been identified, including the reliance on predefined interestingness measures, challenges in dataset representation, the fuzzification process with appropriate membership functions, and limitations of the classical Apriori algorithm. Addressing these challenges through algorithmic improvements and novel Approaches Offers Promising Directions For Future Work.

References

- [1] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37-37.
- [2] Mitra, S., & Acharya, T. (2005). *Data mining: multimedia, soft computing, and bioinformatics*. John Wiley & Sons.
- [3] Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data* (pp. 207-216).
- [4] Delgado, M., Marín, N., Sánchez, D., & Vila, M. A. (2003). Fuzzy association rules: general model and applications. *IEEE transactions on Fuzzy Systems*, 11(2), 214-225.
- [5] Maeda, A., Ashida, H., Taniguchi, Y., & Takahashi, Y. (1995, March). Data mining system using fuzzy rule induction. In *Proceedings of 1995 IEEE International Conference on Fuzzy Systems*. (Vol. 5, pp. 45-46). IEEE.
- [6] Yang, T., & Li, C. (2015, September). A Study of Fuzzy Quantitative Items Based on Weighted Association Rules Mining. In *2nd International Conference on Intelligent Computing and Cognitive Informatics (ICICCI 2015)* (pp. 42-46). Atlantis Press.

- [7] Sumathi, G., & Akilandeswari, J. (2020). Improved fuzzy weighted-iterative association rule-based ontology postprocessing in data mining for query recommendation applications. *Computational Intelligence*, 36(2), 773-782.
- [8] Allahverdi, N. (2019, October). Applications of fuzzy approach in medicine. problems and perspectives. In 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-8). IEEE.
- [9] Han, J., Cai, Y., & Cercone, N. (1992, August). Knowledge discovery in databases: An attribute-oriented approach. In *VLDB* (Vol. 18, pp. 574-559).
- [10] Mangayarkkarasi, K., & Chidambaram, M. (2017). F-PNWAR: fuzzy-based positive and negative weighted association rule mining algorithm. *International Journal of Engineering and Technology*, 9(6), 4250-4257.
- [11] Mangayarkkarasi, K., & Chidambaram, M. (2018). E-FWARM: ENHANCED FUZZY-BASED WEIGHTED ASSOCIATION RULE MINING ALGORITHM. *Journal of Theoretical & Applied Information Technology*, 96(2).
- [12] Bao, F., Mao, L., Zhu, Y., Xiao, C., & Xu, C. (2021). An improved evaluation methodology for mining association rules. *Axioms*, 11(1), 17.
- [13] Zadeh, L. A. (1965). "Fuzzy sets." *Information and control* 8.3, 338-353.
- [14] Zadeh, L. A. (1975). "The concept of a linguistic variable and its application to approximate reasoning-III." *Information sciences* 9.1, 43-80.
- [15] Pasi, G. (2008). Fuzzy sets in information retrieval: State of the art and research trends. *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, 517-535.
- [16] Lu, J., Ruan, D., & Zhang, G. (2008). Fuzzy set techniques in e-Service applications. In *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models* (pp. 553-566). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [17] Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB* (Vol. 1215, pp. 487-499).
- [18] Chan, K. C., & Au, W. H. (1997). "Mining fuzzy association rules." *Proceedings of the sixth international conference on information and knowledge management*.
- [19] Chan, K. C., & Au, W. H. (1997, April). An effective algorithm for mining interesting quantitative association rules. In *Proceedings of the 1997 ACM symposium on Applied computing* (pp. 88-90).
- [20] Chan, K. C., & Au, W. H. (1990). "APACS: A system for the automatic analysis and classification of conceptual patterns." *Computational Intelligence* 6.3, 119-131.
- [21] Chan, K. C. (1991). A statistical technique for extracting classificatory knowledge from databases. *Knowledge discovery in databases*.
- [22] Hong, T. P., Kuo, C. S., & Chi, S. C. (1999). Mining association rules from quantitative data. *Intelligent data analysis*, 3(5), 363-376.
- [23] Hong, T. P., & Lee, Y. C. (2008). An overview of mining fuzzy association rules. *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, 397-410.
- [24] Gyenesei, A. (2000, September). Mining weighted association rules for fuzzy quantitative items. In *European Conference on Principles of Data Mining and Knowledge Discovery* (pp. 416-423). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [25] Khan, M. S., Muyeba, M., & Coenen, F. (2008). A Framework for Mining Fuzzy Association Rules from Composite Items. *ALSIP (PAKDD)*.
- [26] Sarma, R., & Sarma, P. K. D. (2020). Mining Composite Fuzzy Association Rules Among Nutrients in Food Recipe. In *Machine Learning, Image Processing, Network Security and*

Data Sciences: Second International Conference, MIND 2020, Silchar, India, July 30-31, 2020, Proceedings, Part II 2 (pp. 1-10). Springer Singapore.

- [27] Muyeba, M., Khan, M. S., & Coenen, F. (2009). Fuzzy weighted association rule mining with weighted support and confidence framework. In *New Frontiers in Applied Data Mining: PAKDD 2008 International Workshops*, Osaka, Japan, May 20-23, 2008. Revised Selected Papers 12 (pp. 49-61). Springer Berlin Heidelberg.
- [28] Khan, M. S., Muyeba, M., & Coenen, F. (2008). Weighted association rule mining from binary and fuzzy data. In *Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects: 8th Industrial Conference, ICDM 2008 Leipzig, Germany, July 16-18, 2008 Proceedings 8* (pp. 200-212). Springer Berlin Heidelberg.
