Data Mining Approach for Feature Reduction Using Fuzzy Association Rule

Siji P D[1*], M.L. Valarmathi[2]

[1*]Department of Computer science, St. Josephs College Irinjalakuda, Thrissur, India
[2]Dept. of EEE, Alagappa Chettiar College of Engineering and Technology College, College Road, Karaikudi, India

Corresponding Author: srblessy@gmail.com

Available online at: www.ijcseonline.org

Abstract*:* Data mining is an upgrading technology for knowledge extraction in many fields like medical, educational, industrial, etc. Extracting an important data from large database is most vital factor. Data extraction processwere done through many techniques like feature extraction, prediction, classification, etc. for our research analyses prediction of data mining helps a lot for accessing useful information. In this paper we focused on road traffic dataset and we used fuzzy data extraction for membership function by using FCM. For the knowledge extraction process here we implemented the correlation and coefficient algorithm for road traffic dataset and attribute reduction were done by using Genetic algorithm and finally with the help of A-Priori algorithm we generate the rule for the mining the associate object for feature reduction.

Keywords: Data Mining, Prediction, Feature Reduction, Fuzzy, Association Rule and Rule Generation.

I. INTRODUCTION

Today data mining is the major research area for every researcher. Extracting of the supportive information in a huge database is an indispensable for every person in today. Data Mining (DM) is the skillful of resolve expensive information from the large database. The purpose of the DM is to discover information and present it in an understanding that is simply understandable to the people. Knowledge detection in database is detailed method which presenting a number of functional, relevant information. Data mining, or information detection, is the procedure of excavation and investigate enormous sets of data and then eradicate the connotation of the data [1]. In DM they were many perform are used like association rule, prototype matching, feature extraction, predictive analyses etc. With the support of the DM analytical analyses it mines the possible result. Predictive analytics is used to conclude the possible outlook result of an experience or the likelihood of a conditions happening. It is the separation of data mining come inside reach of with the approximation of future probability and trends. Data Mining, on the other hand, is the extraction of tricky to comprehend or hidden predictive information from huge databases or data warehouses. A predictive analytical replica is built by data mining tools and techniques. The primary step consists of extort data by accessing enormous databases. The data thus get hold of is practice with the help of greater algorithms to find hidden patterns and predictive information. Predictive analytics is used to routinely analyze large amounts of data with disparate variables. The most commonplace and significant applications in data mining almost certainly occupy predictive modeling. Clustering refers to the

prediction of a target unreliable that is uncompromising in nature, such as deduction road safety data. Presumption, on the other hand, refers to the calculation of an objective variable that is metric (i.e., interval or ratio) in scenery, such as predicting the degree of stay or the quantity of resource utilization [2]. For predictive illustration the data mining techniques frequently used encompass traditional statistics, such as numerous distinguish analysis and logistic deterioration investigation. They also include non-traditional methods developed in the areas of artificial smartness and machine learning.

Predictive model of DM is very essential for the fitness sector to analyze the Beta Thalesemia data. The exposed knowledge can be used by the road safety administrator to progress the dominance of service. In Beta Thalesemia, data miningis attractive gradually more well-liked, if not increasingly more indispensable. Several factors have provoked the use of data mining applications in this field. The existence of therapeutic insurance fraud and abuse, for example, has led numerous healthcare insurers to challenge to diminish their wounded by using data mining tools to help them find and trail offenders. Freshly, there has been information of successful data mining submission in healthcare sectors [3]. Data mining can get better association mining by discovering relational data and trends in large amounts of complex data. Most common seen in data mining are extraction association rule from database. Association rule is mainly used for discovering the attributes between the huge dataset along with relationship among the dataset. This rule can be efficient providing results for uncovering unknown relationship based on the forecast and

decision. Based on the above statement, association rule isone of the famous areas along with application and development in data mining. By implement a technique for mining association rule based on quantitative and categorical attributes from the database. For information process, mining fuzzy rule system is an important factor and can be used for various researches. This paper proposed the method for extraction of association rule from quantitative data using fuzzy clustering technique and the Apriori approach [4]. The Apriori algorithm is a classic and popular algorithm for strong association rules (knowledge) extraction from a transaction database with high frequent itemsets using the pre-defined threshold measures. These thresholds are minimum support (minsupp) and minimum confidence (minconf). Association rules are formally written and presented in the form of ''IF–Then'' as follows: X ? Y, where Xis called the antecedent and Y is called the consequence [5].One of the advantages of association rule discovery is that it extracts explicit rules that are of practical importance for the user/human expert to understand the application domain. Therefore this can be facilitated to adjust (extend) the rules manually with further domain knowledge, which is difficult to achieve with other mining approach. Extracting the knowledge from the quantitative database based on Fuzzy dataset extraction. This paper proposed a genetic algorithm (GA) for attribute reduction from the database. Prediction ofthe database and the extraction of the Fuzzy Association Rule are based on two methods; thefirst method is the combination of the Fuzzy C-Means (FCM) and Genetic algorithm (GA).With the help of A-Priori algorithm we generate the rule for the mining the associate object for feature reduction. The result is verified with two databases .One is Vehicle database and the second one is Beta Thalassemia Patients record. In this paper Section 2 gives a brief idea of previous work in this area. Section 3 and 4 explains A- Priori Fuzzy Concepts and the proposed new Attribute reduction techniques using GA. Section 5 gives the results with the tables and figures. Section 6 gives the conclusion.

II.   LITERATURE REVIEW

In [6] Predictive analysis is an advanced division of data manufacturing which generally predicts some incidence or likelihood based on data. Predictive analytics uses data-mining techniques in arrange to make forecast about future events, and make suggestion based on these predictions. The development engrosses an analysis of momentous data and based on that examination to guess the future incidence or events. A replica can be created to guess using Predictive Analytics replica techniques. The form of these extrapolative models varies depending on the data they are using. Classification & deterioration are the two main objectives of analytical analytics. Predictive Analytics is collected of various arithmetical & analytical techniques used to enlarge models that will guess future incidence, events or likelihood.

Predictive analytics is able to not only agreement with unremitting changes, but alternating revolutionize as well. Classification, calculation, and to some degree, resemblance analysis constitute the investigative methods employed in predictive analytics.

In [7] describe that analysis technique to determine a small set of rule in the database to forms a precise classifier Association rule mining is significant. They initiate the combined advance that incorporates association rule removal and categorization rule mining. This is new categorization approach is realize by focusing on mining a particular subset of involvement rules called classification association rule, then categorization is being execute using rules. The associative classifiers are especially fit to submission were the replica may support domain experts in their conclusion There are many associative organization come within reach of that have been proposed freshly such as CBA, CMAR, CPAR and MCAR and MMAC.

In [8], analyze the data mining clustering technique for grouping data into related component basedon similarity metrics. Integration of fuzzy logic with data mining techniques has become one of the key constituents of soft computing. In traditional clustering algorithm, one object is assigned in to only one cluster. But if the clusters are touching each other or they are overlapping, fuzzy clustering comes in to existence. In this paper the membership calculation for clustering the points and its criteria is modified. The Box metric equation is applied as a proximity measure. This paper also presents an investigation into a fuzzy association rule mining model for enhancing prediction performance in a medical database. This model (the FCM–MSMM Apriori model) integrates multi membership and multiple support approach for Betathalasemia disease for performance prediction.

In [9] analyze the health care environment needs knowledge based discovery forhandling wealth of data. Extraction of the potential causes of the diseases is the most important factor for medical data mining. Fuzzy association rule mining is well-performed better than traditional classifiers but it suffers from the exponential growth of the rules produced. In the past, we have proposed an information gain based fuzzy association rule mining algorithm for extracting both association rules and membership functions of medical data to reduce the rules. It used a ranking based weight value to identify the potential attribute. When we take a large number of distinct values, the computation of information gain value is not feasible. In this paper, an enhanced approach, called gain ratio based fuzzy weighted association rule mining, is thus proposed for distinct diseases and also increase the learning time of the previous one. Experimental results show that there is a marginal improvement in the attribute selection process and also improvement in the classifier accuracy. The system has been implemented in Java platform and verified

by using benchmark data from the UCI machine learning repository.

In [10] tells that prediction is an important and interesting task in Data Mining. The good prediction model should predict a future value accurately. This can be done by building a model that generates and evaluates a set of rules for prediction. Associative Classification is a rule based model and it gives more accuracy than traditional classification methods. Mosquito Borne Disease Incidence Prediction System using Fuzzy Weighted Associative Classifier was developed which is later enhanced with negative rule generation technique by us. In this paper we discussed the performance of the fuzzy associative classifier with as well as without weighted concept and the same were also tested using negative rule generation. We also emphasized the need for fuzzification of quantitative data.

III.   MEMBERSHIP EXTRACTION USING FCM

FCM is used to cluster the data into terms and then to determine the center of each fuzzy set and the maximum and minimum value for each field of the input data set. The data set is converted into a fuzzy data set, using one of the standard membership functions. Fuzzy association rule mining models is used for enhancing predictive object. For extraction of dataset we use fuzzy model. And to predict the object genetic algorithm were used to correlate the object in the database. With the help of genetic learning of the membership function for extracting the both fuzzy association rule and membership function from quantitative transactions. It is composed of two stages, first stage to learn the membership function by genetic process, second method to mine the fuzzy association rule [11, 12]. Mining a Fuzzy association rule in a probabilistic quantitative database by formulating a fuzzy–probabilistic database is implemented using FCM. FCM is used to define the membership functions.

Algorithm: Fuzzy Clustering for data extraction

Initialize P= number of clusters=2

Initialize m= fuzzification parameter;

Initialize Cj=(cluster centers)

Repeat

For i =1 to n: update μj(xi) applying equation 4

For j=1 to p: update Ci with (4)with current μj(xi)

Until Cj estimate stabilize

Algorithm 1: Fuzzy data extraction

Three data sets have been employed for the model performance analysis. Two data sets are related to a road traffic problem and the third one is the Abalone bench benchmark data set. The road traffic data has been generated using a traffic simulation model, (called the METANET macroscopic flow model consists

- Traffic state, which is represented by:
- traffic demand in road 1 (the number of vehicles that need to use road 1)
- traffic demand in road 2
- traffic density in road 1 (the number of vehicles that are using road 1 per km)
- traffic density in road 2.

It shows information about the input of road traffic data set. Predicted Average Travel Time (ATT) (ATT is the total average time required for a vehicle to cross the traffic network).

The data set is converted into a fuzzy data set, using one ofthe standard membership functions.For the purpose of evaluation and validation, prediction quality is assessed using different statistical evaluation metrics.

IV.   ATTRIBUTE REDUCTION USING GA

Attribute reduction is an extraction of unwanted data. In the large no.of database relation extraction is a critical issue. By reducing the unwanted information we can get the needed information. For that purpose association rule mining were used to extract data and also fuzzy dataset extraction were used for large database extraction. For the extraction purpose genetic analyze of data may very well to extract the data, related of the data were extracted from the large database is done by Genetic Algorithm [13, 14]. Genetic Algorithms (GAs) are adaptive heuristic search algorithms based on the evolutionary ideas of natural selection and genetic processes. It is inspired by the mechanism of natural selection where stronger individuals are likely the winners in a competing environment. GA is successfully applied to many fields such as optimization, machine learning, neural network, fuzzy logic controllers. Idea behind the GA is genetic and evolutionary theory. Three main operators in the genetic algorithm is crossover, mutation and selection operation. The crossover operation generates offspring from two chosen individuals in the road traffic data by exchanging some bits in the two individuals. The offspring thus inherit some characteristics from each parent. The mutation operation generates offspring by randomly changing one or several bits in an individual. The offspring may thus possess different characteristics from their parents. Mutation prevents local searches of the search space and increases the probability of finding global optima.

Algorithm 2: Genetic algorithm

Input: A transaction database DB and a minimum support threshold ξ.

Output: relational item of DB.

Step 1: Scan the transaction database DB once. **Step 2:** Collect R, the set of relational items, and the support of each relational item.

Step 3: Sort R in database RList, the list of relational items.

Step 4: Create the map of an relational tree, T , and label it as "extraction". For each transaction Trans in DB do the following.

Step 5: Select the relational items in Trans and sort them according to the order of FList.

Step 6: Let the sorted relational-item list in Trans be [p | P], where p is the first element and P is the remaining list. Call insert tree([p | P], T). The relation insert tree([p | P], T) is performed as follows. If T has a child N such that N.item-name = p.item-name, then increment N's count by 1; else create a new node N, with its count initialized to 1, its parent link linked to T , and its node-link linked to the nodes with the same item-name via the node-link structure. If P is nonempty, call insert tree (P, N) recursively.

With the help of the GA feature reduction were done. Our experimental shows the reduction result in above chart.
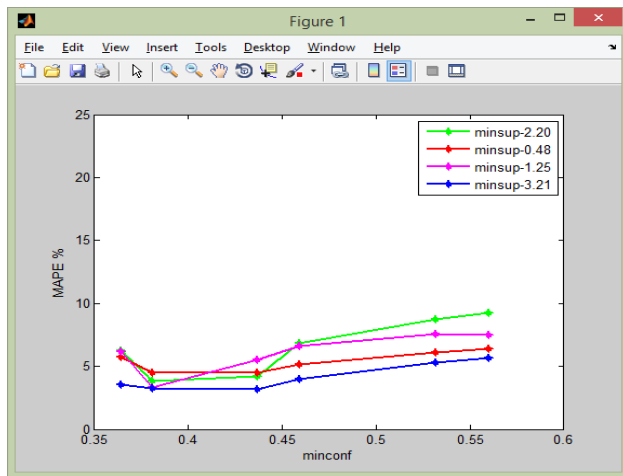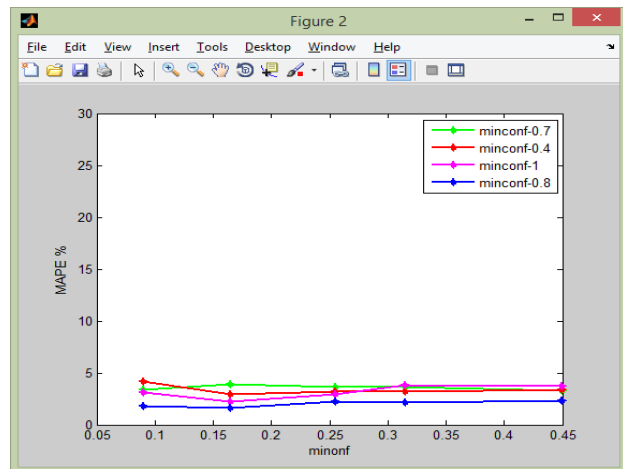

Fig. 1: Finding Minsup using GA using dataset1

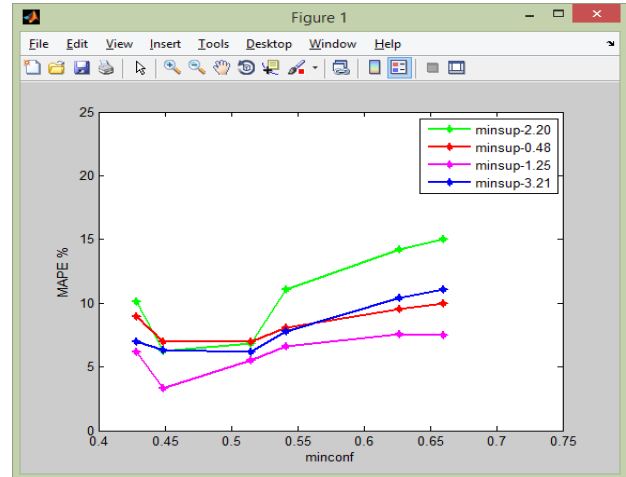**Fig. 2**: Finding Minconf using GA using dataset1

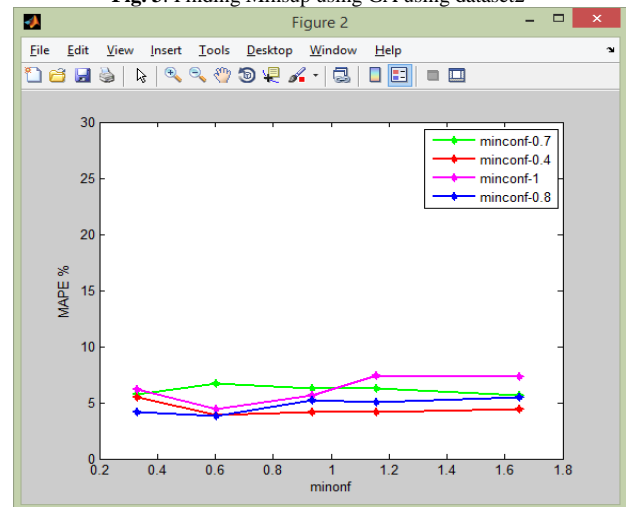**Fig. 3**: Finding Minsup using GA using dataset2

Fig. 4: Finding Minconf using GA using dataset2

V.   RULE GENERATING USING A-PRIORI

The A-priori approach is used for extracting fuzzy termsets (frequent itemsets) from a fuzzy data set based on interesting measures (minsupp and minconf). It is worth mentioning that the A-priori algorithm is adopted in order to deal with fuzzy data and therefore able to generate FARs. Throughout the rest of the paper, the term 'itemsets' corresponds to its termsets.For analysis and validation purposes, the small road traffic dataset and large road traffic data set were used. Traffic state prediction (including traffic flow(traffic density) and traffic demand) has long been regarded as critical concern for intelligent road traffic systems [15, 16]. For selection we follow the rules which must be either or one in the below rule. rules for accurate prediction depends on the selection of the minsupp and minconf values.

Rules are formally written and presented in the form of "IF– Then" as follows: X⊂Y, where X is called the antecedent and Y is called the consequence. Consequently, large numbers of rules are anticipated, particularly if minsupp is set

to be very low. Practically, a single minsupp is a vital parameter that controls the extracted number of association rules

- IF Demand 1 = High and Density 1 = Medium Then ATT = Medium; CV = 072;
- IF Demand 1 = High and Density 1 = Medium Then ATT = High; CV = 0:61:

The FCM is applied to road traffic controlmanagement; two case studies with the two different data set sizes are used for the road traffic domain for predicting the traffic. Furthermore, the simulations and experiments are illustrated. Subsequently, the results' analysis of the model application is

Table 1

1	32.43401
2	63.30693
5	32.43401
6	63.30693
9	32.43401
10	63.30693
13	32.43401

This table shows part of the candidate terms after the calculation of the support value for each term. For instance, the summation value for the term VL in Demand 1 is equal.

Table 2

1	10	11	1.184472
1	11	12	0.026363
1	12	13	0.026363
1	13	14	14.70932

This table 2 shows part of the large termset of L1

Table 3

1	1	2
1	5	6
1	9	10
1	13	14
1	17	18
1	21	22

This table 3 shows part of the candidate termsets of Term1 and candidate terms of C2 respectively.

Table 4

1	5	32.43401
1	6	14.70932
1	9	32.43401
1	10	14.70932
1	13	32.43401
1	14	14.70932
1	17	32.43401

Table 4 shows the large termset of L2,

Table 5

1	9	10	4.013889
1	10	11	0
1	11	12	0
1	12	13	0
1	13	14	4.013889

Table 5 shows part of the candidate termsets that are stored

Table 6

1	41	42	1
1	45	46	1
1	49	50	1

Table 6 shows part of the large termset of L3.

Table 7

32.43401	32.43401	1	32.43401
63.30693	14.70932	0.232349	63.30693
32.43401	32.43401	1	32.43401

Table 7 shows the frequent termsets in L2 to Ln attribute

VI. CONCLUSION

Data mining is an important research area for feature reduction to find out the useful information. For feature reduction in road side data this paper proposed a fuzzy association rule mining models for enhancing prediction object. We generate the membership function as an output by using correlation and coefficient algorithm. Reducing the feature Genetic algorithm were applied for getting correct output and we generate the rule using A-priori algorithm for feature reduction. The result proves that the attribute reduction technique using GA is more predictive than the convention A-Priori FCM method.

REFERENCES

[1] V. Palanisamy and A. Kumarkombaiya, "*Designing a Knowledge Discovery of Clustering Techniques in Pharmaceutical Compounds*", International Journal of Computer Sciences and Engineering, Vol.3, Issue.4, pp.58-63, 2015.

[2] Raghupathi W. "*Data mining in healthcare. Healthcare Informatics: Improving Efficiency through Technology*", Analytics, and Management. 2016 Apr 27:353-72..

[3] Biafore, S. (1999). Predictive solutions bring more power to decision makers. Health Management Technology, Vol.20(10), pp 12-14.

[4] Lu, J., Xu, B. & Jiang, J. (2003). *A prediction method of fuzzy association rules*. InProceedings of IEEE international conference on information reuse and integration(pp. 98–103), Nanjing, China.

[5] Siji PD, Valarmathi ML, Mohana S. *A new soft computing technique for efficient rule mining*. Journal of Theoretical and Applied Information Technology. 2016 Jan 1;83(3):434.

[6] Lei, Z. &Ren-hou, L. (2007). *An algorithm for mining fuzzy association rules basedon immune principles*. In Proceedings of the 7th IEEE international conference onbioinformatics and bioengineering, Boston, MA.

[7] Mishra N, Silakari DS. *Predictive Analytics: A Survey, Trends, Applications, Oppurtunities & Challenges*. International Journal of Computer Science and Information Technologies. 2012;3(3):4434-8.

[8] SunitaSoni and O.P.Vyas *Using Associative Classifiers for Predictive Analysis in Health Care Data Mining* ―International Journal of Computer Applications (0975 – 8887) Volume 4 – No.5, July 2010.

[9] Kumari Nirmala, R.M.Singh and Shilpi Gupta, "*Analysis for Heart Related Issues using comprehensive Approaches: A Review*", International Journal of Computer Sciences and Engineering, Vol.3, Issue.3, pp.184-187, 2015.

[10] Nithya NS, Duraiswamy K. *Gain ratio based fuzzy weighted association rule mining classifier for medical diagnostic interface*. Sadhana. 2014 Feb 1;39(1):39-52.

[11] Kuok, C. M., Fu, A., & Wong, M. H. (1998). *Mining fuzzy association rules indatabases*. ACM SIGMOD Record, 27(1),pp-41–46.

[12] M. Patel, A. Hasan , S.Kumar, "*A Survey: Preventing Discovering Association Rules for Large Data Base*", International Journal of Scientific Research in Computer Science and Engineering, Vol.1, Issue.2, pp.30-32, 2013.

[13] Kaya, M. &Alhajj, R. (2003). *Facilitating fuzzy association rules mining by usingmulti-objective genetic algorithms for automated clustering*. In Proceedings ofthe third IEEE international conference on data mining (ICDM'03).

[14] Ho, G. T. S., Ip, W. H., Wu, C. H., &Tse, Y. K. (2012). *Using a fuzzy association rulemining approach to identify the financial data association*. Expert Systems withApplications, 39(10), 9054–9063. doi: 10.1016/j.eswa.2012.02.047.

[15] Sowan B, Dahal K, Hossain MA, Zhang L, Spencer L. *Fuzzy association rule mining approaches for enhancing prediction performance*. Expert Systems with Applications. 2013 Dec 1;40(17):6928-37.

[16] Hong, T. P., Kuo, C. S., & Wang, S. L. (2004). *A fuzzy AprioriTid mining algorithmwith reduced computational time*. Applied Soft Computing Journal, 5(1), 1–10.