ISSN: 2395-6607, Vol. 1, No. 1 March 2015 pp. 15–27

FUZZY ASSOCIATION RULE MINING APPROACH WITH MODIFIED CLUSTERING TECHNIQUES FOR PREDICTION PERFORMANCE IN MEDICAL DATABASE

Lilly P.L.[1] and Siji P.D.[2]

[1]*Assistant Professor, Department of Computer Science, St. Josephs College Irinjalakuda, Thrissur–680121, Kerala, India*

[2]*Associate Professor, Department of Mathematics St. Josephs College Irinjalakuda, Thrissur–680121, Kerala, India*

E-mail: [1]srblessy@gmail.com, [2]sr.christy@gmail.com

Abstract—*In data mining, clustering analysis is a technique for grouping data into related component based on 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.*

Keywords*: Clustering, Fuzzy Association Rules, FCM-Apriori Algorithms, Multiple Support*

INTRODUCTION

The applications of Data mining have become increasingly common in both private and public sectors. The medical community sometimes uses data mining to help and predict the effectiveness of a procedure or medicine [45]. Data mining can be performed on different types databases and data repositories. The various data mining functionalities like classification, prediction, cluster analysis etc are used to find the different kinds of pattern, the result of data mining technique. Association rule mining has been a accepted area in data mining (DM) research, more and more attracting the attention of researchers.[1] [2][3][4][5] are important works in this area. Association rules discovery presented in [6] intends to extract the characteristics, hidden association patterns and the correlation between the items (attributes) in a large database [7],[8]. The Apriori algorithm developed by [9] 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: $X \rightarrow Y$, where X is called the antecedent and Y is called the consequence. Let $I = \{i_1, i_2, \ldots, i_n\}$ be a set of distinct items (attributes). A collection of one or more items, i.e., any set of items is called an item-set. Let $D = \{t_1, t_2, \ldots, t_m\}$ be a set of transaction IDs (TIDs). Each TID in D is formed from a set of items in I. The support count is the occurrence (frequency) of X and Y together, support (XUY), and the support value is the fraction of transactions that contains both X and Y.

An item set whose support is greater than or equal to a minsupp threshold is called a frequent item set. The confidence value measures how often items in Y appear in transactions that contain X and is the ratio of occurrence (X and Y) divided by (/) occurrence(X). Support(XUY)/Support(X) An association rule is an implication expression of the form (: $X \rightarrow Y$), where $X, Y \in I$ and $X \cap Y = \phi$ A strong association rule is that which has support and confidence greater than the user defined minsupp and minconf. The main task of the association rule discovery is to find all strong rules.

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 approaches [10]. On paper [11] introduced the problem of extracting association rules from quantitative attributes by using the partitions method for these attributes. Some of the current association rule mining approaches for quantitative data neglected the values of the interval boundaries of the partitions. This causes sharpness of the boundary intervals which does not reflect the nature of human perception, justifiably argued by [12] 13]. Instead of using partition methods for the attributes, it is better to adopt the advantage of fuzzy set theory with a smooth transition between fuzzy sets. As a whole, the fuzzy approach is used for transforming quantitative data into fuzzy data. A variety of approaches has been developed in order to extract fuzzy association rules from quantitative data sets [14], [15], [16], [17], [18], [19], [20], [21].

In this paper investigates the problem of association rules extraction from quantitative data using fuzzy clustering techniques. Fuzzy clustering is a suitable method to transform quantitative data into fuzzy data, taking the advantage of fuzzy set theory over the partition method concerning the smooth transition among fuzzy sets. Fuzzy Association Rules (FARs) mining is adopted in this paper as a solution for extracting knowledge from the quantitative database.

The association rule mining aims to discover the relationships (rules) among the data attributes (features), which depend on minsupp and minconf. 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. The papers [22],[3] proposed an integrated data envelopment analysis based method to identify the most efficient association rules by ranking them using multiple criteria. Conventional association rule mining approaches like Apriori [9] and Frequent Pattern-Growth (FP-Growth) [23] are based on a single minsupp threshold. However, it was observed that using a single minsupp causes a dilemma called the "rare item problem" [24][23].

To resolve this rare item problem, author [8] developed a multiple support model called the Multiple Support Apriori (MSApriori) algorithm. MSApriori is based on the idea of setting a Mini-mum Item Support (MIS) for each item in a database, i.e., assigning multiple minsupp for different items in the database, instead of using a single minsupp for the whole database. Hence, MSaproiri is expressed as a generalization of the Apriori algorithm. Different MIS values can be assigned to assess different frequent items to facilitate the generation of frequent itemsets of rare items and prevent the production of uninteresting frequent itemsets [22] More recently, an approach has been developed to improve MSApriori called Improved Multiple Support Apriori (IMSApriori) [8],[21].

This paper also proposes Fuzzy Association Rules (FARs) generated using Fuzzy clustering on quantitative data by adopting the multiple support approaches in order to deal with the limitations of using a single minsupp. FCM–Apriori model, is based on the integration of the Fuzzy C-Means (FCM) clustering algorithm and the Apriori approach for extracting FARs. FCM–MSApriori model, is based on FCM and a multiple support thresholds approach.

Although the adoption of the MS idea from the classical partition case, the FCM–MSApriori model in the fuzzy case remains obstructed because it uses only one membership function without considering the price–quantity relation[25]. For example, In the Business field the implication of a pattern "Color Laser Printer with Low quantity" must be distinguished from that of another pattern "Printer Toner with Low quantity" although both patterns are assigned with a same fuzzy term. Managers may specify different definitions of Low quantity for Color Laser Printers and Printer Toner. Items with different prices result in different quantity demands; therefore, different membership functions must be dispatched to calculate their fuzzy term supports.

The rest of this paper is structured as follows. 2 gives a brief idea about clustering and gives the comparison about k means and fuzzy C-means algorithm. Section 3 describes Association mining with fuzzy model with the case studies. Experimental results of analysis are presented in Section 4. Finally, the conclusion are drawn in Section 5 with the key contribution of the research

K Means Clustering

The most popular class of clustering algorithms is K means algorithm, a center based, simple, and fast algorithm, aims to partition n objects into k clusters in which each object belongs to the cluster with the nearest mean. The k means algorithm is the best method to cluster the crisp data. It tries to find a user specified number of clusters. The clusters are represented by their centroid. This centroid is typically the mean of the points in the cluster. There are two phases in the algorithm. The first one is to select K centers randomly, where the value of K is user specified. The next phase is to assign each data point to the nearest center. Euclidian distance is generally considered to determine the distance between each object and center [47]. When all the objects are assigned in some clusters recalculate the mean of the clusters. This process repeats until the criterion function becomes minimum. The objective function

$$J=\sum_{j=1}^{k}\sum_{i=1}^{n}\|Xi-cj\|^{2} \qquad (1)$$

Fuzzy c Means

Integration of fuzzy logic data mining techniques has become one of the key constituents of soft computing [46]. The central idea in fuzzy clustering is the non-unique partitioning of the data into a number of clusters. The data points are assigned membership values for each of the clusters. The fuzzy clustering algorithms allow the clusters to grow into their natural shapes. In some cases the membership value may be zero indicating that the data point fuzzy c means clustering involves two processes: the calculation of cluster centers and the assignment of points to these centers using a form of Euclidian distance. This process is repeated until the cluster centers stabilize. The algorithm is similar to k-means clustering in many ways but incorporate fuzzy set's concepts of partial membership and forms overlapping clusters to support it. It assigns membership value to the data items for the clusters within a range of 0 to 1. The algorithm needs a fuzzification parameter m in the range [1,n] which determines the degree of fuzziness in the clusters. When m reaches the value of 1 the algorithm works like a crisp partitioning algorithm and for larger values of m the overlapping of clusters is tend to be more.

NEED OF FUZZY MEANS CLUSTERING

In traditional clustering algorithm, one object is assigned in to only one cluster. This is valid till the clusters are disjoint and separate. But if the clusters are touching each other or they are overlapping, then one object can belong to more than one cluster. In this case fuzzy clustering comes in to existence[47]. In fuzzy clustering, one object can be clustered in more than one cluster according to the degree of membership function.

Let a set of objects X = $\{x_1,x_2,x_3.......\}$ has to be clustered in to C = $\{c_1,c_2,c_3.......\}\delta(x,C_i)$ denote the similarity between object x and cluster C_i.

The common criteria about the membership function of datapoints is.

$$\sum_{j=1}^{c} U_{ij}=1,\ i=\{1,2,3,....N\} \tag{2}$$

The cluster cente $C_j=\frac{\sum_{i=1}^{n}(\mu j(xi))^{m}]*xi}{[\sum_{i=1}^{n}(\mu j(xi))m]}$ (3)

Since fuzzy c-means algorithm is the most popular and widely used fuzzy clustering algorithm, many approaches have been proposed to improve the performance of the algorithm. Each of these modified methods proposes a new membership function for calculating the membership of data points in clusters.

The adaptive fuzzy clustering algorithm is a modified version of the c-means clustering and it is proposed by Krisnapuram and Keller [48].The membership values in this method are calculated using Expression (48). The adaptive fuzzy clustering algorithm is efficient in handling data with outlier points. In comparison with c-means algorithm, it gives only very low membership for outlier points.Since the sum of distances of points in all the clusters involves in membership calculation this method tends to produce very less membership values when the number of clusters and points increase and this is the main limitation of it.

$$\mu_j(x_i) = \left(\frac{\left(\frac{1}{d\,j\,i}\right)^{1/(m-1)}}{\sum_{i=1}^{n}\left(\frac{1}{dj\,i}\right)^{1/(m-1)}}\right) \tag{4}$$

wher d j I $=\|x\,i-cj\|2$

In paper[3], "A Modi_ed Fuzzy C-Means Algorithm for Natural Data Exploration " they replaced the restriction imposed by exp (46) with a liberalized expression (5). That is, the sum of memberships in a cluster center must be n/2.

$$\sum_{i=1}^{n}(\mu j(xi))=n/2 \tag{5}$$

In c-mean the membership of a data point in a cluster depends directly on the sum of distances of the point in other cluster centers (2). Many limitations of the algorithm which affect the performance arise due this method [4]. Instead, they considered the sum of distances of data members in a cluster for the calculation of memberships in that cluster, it might improve the performance of the algorithm. This leads to their second modification. The new membership function for ith data point in jth cluster is given below (6).

$$\mu_j(x_i) = \frac{n}{2} * \left(\frac{\left(\frac{1}{d\,j\,i}\right)^{1/(m-1)}}{\sum_{i=1}^{n}\left(\frac{1}{dj\,i}\right)^{1/(m-1)}}\right) \tag{6}$$

Also in paper[6] "Clustering Algorithms in Biomedical Research: A Review" by Rui Xu, and Donald C. Wunsch, II explained about the proximity measures Clustering algorithms are built on the proximity of data objects, each described by a set of features, denoted as a multidimensional vector. The features can be quantitative or qualitative, continuous or discrete, which leads to different measure mechanisms. Accordingly, the data set with data objects of features will be recorded as an data matrix, with each row denoting an object and each column representing a feature.

The data matrix is designated as two-mode because its row and column indices have different meaning. This is in contrast to the one-mode proximity matrix, which is an symmetric matrix with elements representing the similarity or distance measure for any pairs of data objects in the data set, because in a proximity matrix, both dimensions share the same meaning

ALGORITHM FOR FUZZY C MEANS ALGORITHM

Initialize P= number of clusters=2

Initialize m= fuzzification parameter;

Initialize C_j=(cluster centers)

Repeat

For i =1 to n: update $\mu_j(x_i)$ applying equation 4

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

Until Cj estimate stabilize

FCM-S Apriori

The use of a single minsupp for a whole database assumes that all items in the database have the same frequency. However, in real applications, the database contains some high frequency items, while others are of low frequency. The human expert, based on do-main knowledge, can set minsupp for a specific value in order to find the frequent item sets. In that case, if minsupp is set too high it will extract a low number of frequent item sets. Thus, the rare item problem will appear and cause a dilemma (called the rare item problem). On the other hand, if minsupp is set too low, it will extract a high number of frequent item sets, which causes combinatorial explosions, i.e., all the possible associations will be found. Some of those frequent item sets are uninteresting or insignificant [24]

To overcome the dilemma of the rare item problem, [24] proposed an algorithm called MSapriroi based on a multiple minimum support thresholds approach using MIS, where the number of generated rules depends on the control parameters used.

FCM–MSMM-Apriori Model

The proposed FCM–MSMM-Apriori model adopts this multiple minimum support concept [24] and multiple membership function [25] for different attributes depending upon the frequency This model utilizes FCM, and the MSapriori approach is used for extracting FARs of rarely and highly frequent termsets from fuzzy data sets as shown in Fig. 3.

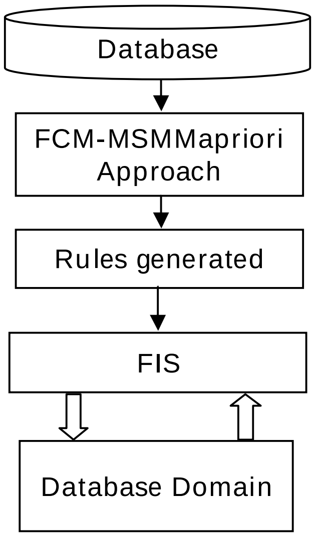

Fig. 3: FCM-MSMM-apriori Model

SOME COMMON MODIFICATIONS

Clustering

We replaced the restriction imposed by exp (5) with a little more liberalized expression (7). That is, the sum of memberships in a cluster center must be n. not n/2 that is the sum of data points in a cluster will be maximum one, taking the maximum member ship value I for every points in a cluster

$$\sum_{i=1}^{n}(\mu j(xi)) = n \tag{7}$$

Second modification

Regarding the membership value calculation of each data points, we are taking with a more accurate value by changing

$$\mu_j(x_i) = \frac{n}{p} * \left(\frac{\left(\frac{1}{d\,j\,i}\right)^{1/(m-1)}}{\sum_{i=1}^{n}\left(\frac{1}{d j\, i}\right)^{1/(m-1)}}\right) \tag{8}$$

Third Modification:

Instead of using Euclidian distance formula $J=\sum_{j=1}^{k}\sum_{i=1}^{n}\|Xi - cj\|$ 2 for finding the similar points and to form the cluster, we are using Box metric equation which is

$$d_{ji} = \max\{|\ (C_{ji} - x_{iq})| \tag{9}$$

where j denotes the no of clusters and q denotes the no of attributes in data point. we are taking the maximum value distance between the attributes of cluster centre and each data point.

Fuzzy Association Mining we are Doing the Following Modification

We list three advantages of our new knowledge discovery model as follows:

1. The FCM-Apriori is more normal and suitable in relation to human knowledge. We can easily understand linguistic terms discovered by the decision making procedure.

2. The FCM-MSApriori inherits advantages from FCM-Apriori and gives more flexibility to real-life applications. The model acquires certain patterns in which the elements include rare items with advanced profits, and excludes those that are minor with lower profits.
3. The idea of adjustable membership functions is considered to offer fuzzy quantitative information, depending on the various membership functions. That is, although any two items may have the same fuzzy term, the meanings of the quantitative natures differ.

Database Description

The thalassemia is autosomal recessive disorder which results in reduced production of one or more of the submits hemoglobin [36]. Thalassemia is a public health problem in the tribal area of India. Beta thalassemia major produces severe anemia that requires lifelong blood transfusions for survival. The molecular defects producing beta thalassemia are heterogeneous, and each ethnic group possesses its own specific set of mutations [37][38]. Treatment of Thalassemia involves lifelong treatment [39]. Management includes regular blood transfusions, iron chelation treatment, management of complications including osteoporosis, cardiac dysfunction, endocrine problems, hepatitis B and C infection, HIV infection. Life-expectancy for Thalassemia has improved significantly with modern medical treatment. [40–42] But it has been estimated that only 5–10 percentage thalassemia children born in India receive optional treatment[43] without access to regular chelation treatment and medical care, the majority of children with Thalassemia major do not reach the age of 20.

Materials and Methods: population; sample size is 61.the study was done October 2006–Jan 2008 [44]

***Source of Data*:** With the same study in our previous work,it was prospective observational study done in 61 thalassemic patients to observe the growth and sexual maturation. Data was collected from 61 children between the age group of 3 to 15 years who were diagnosed as having Beta Thalassemia major by hemoglobin electrophoresis and receiving blood transfusion from Thalassemia clinic of St.Johns Medical College Hospital Bangalore. Linear growth was assessed in all children between the age group of 9–15 years. The database has been normalised for clustering as well as for rule mining to obtain more accurate results

Table 1: Analysis on Betathalesemia Data Base

Sl. No.	Details	No
1	Data size	61
2	No of variables	6
3	Min support	2.58
4	Min confidence	0.4

Discussion and Results

In all experiments we use MATLAB software as a powerful tool to compute clusters. The fact that the number of patients with thalassemia decreases beyond 15 years could be explained by death mostly among children older than 15 years.This can be explained by the fact that if children are not transfused, they die before the age of 6 years and if they are transfused and non-chelated,they die before the age of 20.The clustering of number of patients shows that the age group between 8 and 10 years old are mostly affected by this diease.The mean age is 10=(not equal to) 5 years. Beyond 15 years,the number of cases decreases (Fig. 4 (x-axis no of patients and y axis age)).

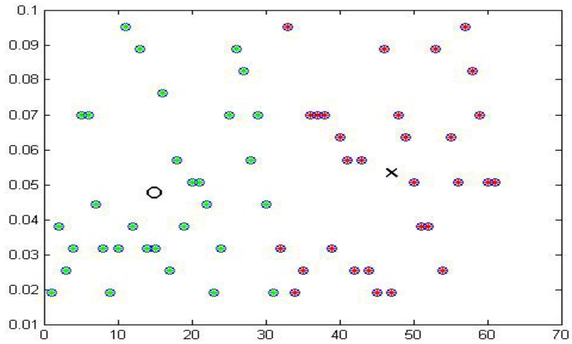

Fig. 4: The Two Clusters Formed using the Variables Patient Id with their Age using Color Discrimination

The distribution of Thalassemia patients according to sex shows male predominance. However, there is no significant difference between male and female regarding the occurrence of the disease.

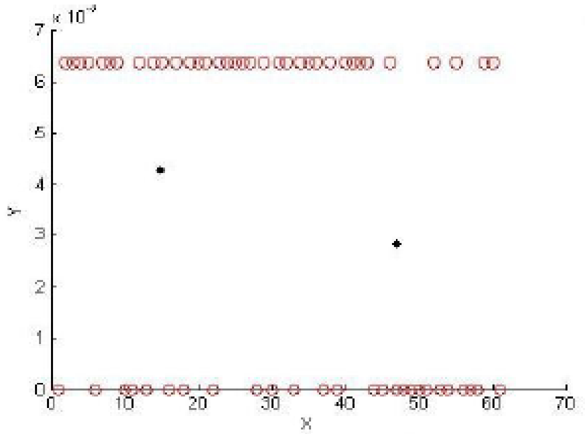

Fig. 5: The Distribution of Thalassemia Patients According to Sex

The patients issued from consanguineous marriages are affected by the disease with a rate of 57 and 43.

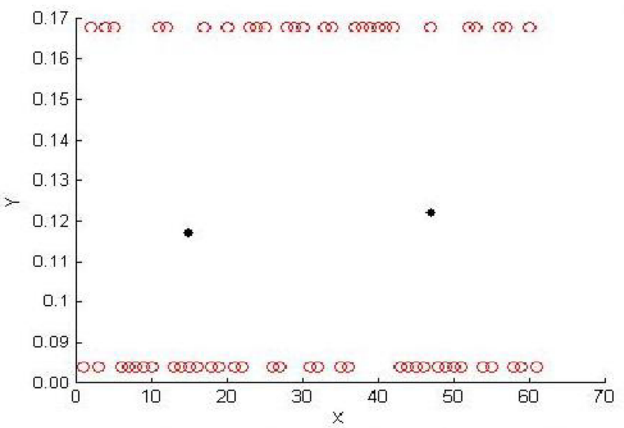

Fig. 6: The Distribution of Thalassemia Patients According To Consanguineous Marriages

The FCM–MSMM-Apriori model.

For analysis and validation purposes, Betathalesemia data set (Section 3.1) is used. Betathalesemia prediction (including age) and consanguinity has long been regarded as a critical concern for the prediction of disease [44]. The FCM–Apriori model discussed in Section 2 is implemented on the database of Betathalesemia; Furthermore, the simulations and experiments are illustrated. Subsequently, the results' analysis of the model application is discussed.

Example: how the proposed FCM–Apriori model works

This example illustrates the steps of the model applied to the Betathalesemia patients database.

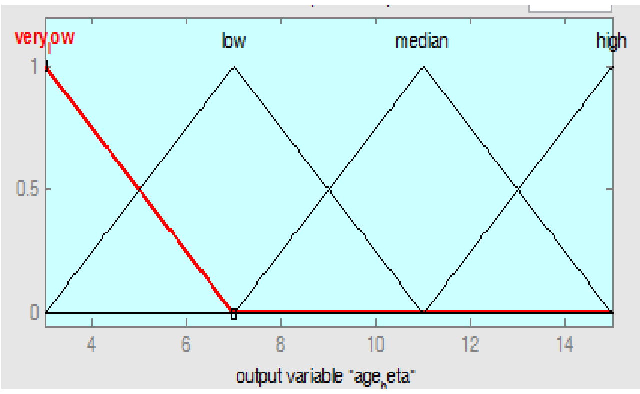

Fig. 7: An Example of the Age Field (3–15) and its Membership Functions

Figure 7 represents an example of the age field (3–15) and its membership functions. all fields have four fuzzy classes including: Very Low (VL), Low (L), Medium (M) and High(H). Here, each fuzzy class (fuzzy set) is mapped into numbers,

Figure 8 explains the analysis of min supp and minconf values on the MAPE for the betathalesemia Dataset. The graph of minconf 0.4 and minsupp 2.5 shows MAPE value 11.5 which is the Minimum

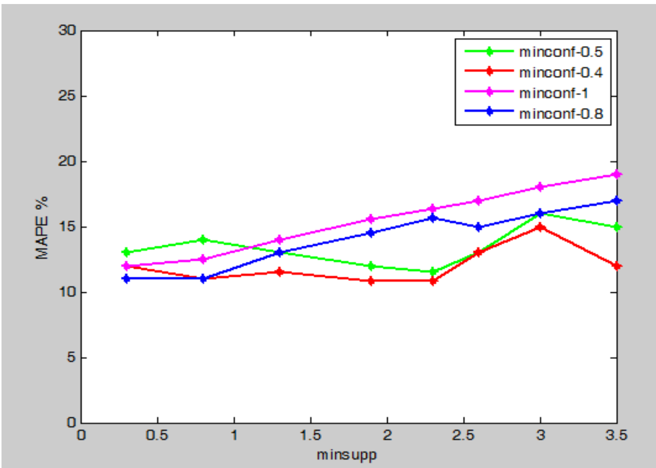

Fig. 8: The MAPE for Different Minsupp and Minconf

Again Fig. 9 shows that the minimum MAPE is produced when of minconf 0.4 and minsupport 2.58. The selection of an appropriate no of rules for accurate prediction depends on the selection of minsupp and minconf values.

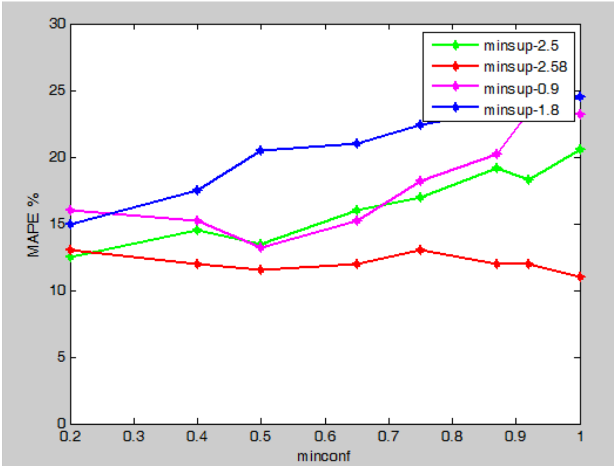

Fig. 9: The MAPE for Different Minsupp and Minconf

Figure 10 shows the performance analysis of Existing Algorithm and Proposed Algorithm. The FCM MSMMApriori gives minimum MAPE value when min support is 2.58 and minconf 0.4.

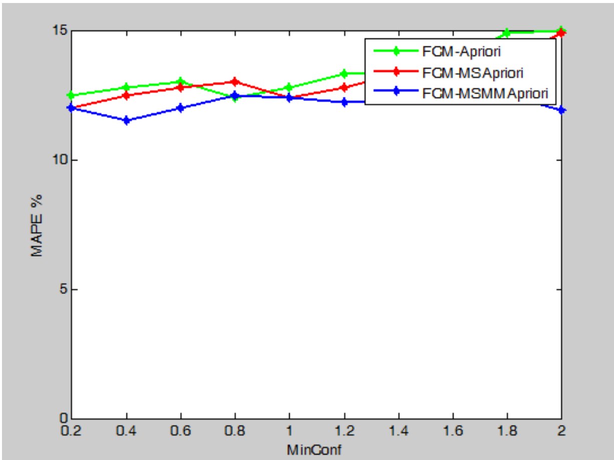

Fig. 10: Analysis of Existing Algorithm and Proposed Algorithm

Figure depicts the analysis of minsupp and minconf values on the MAPE for the data set. The minimum MAPE value of 11.4%, and it contains rules that cover most cases in the graph when the minconf 0.4 with minsupp 2.5. When minconf is less than 0.4,it will increase the MAPE; this is explained by producing a large number of rules (a decrease in minconf shows an increase in the deviated rules, which causes error for the FIS). Also it is noted that if minconf is greater than 0.4, it will lead to an increase in the MAPE. Again this is explained by producing a small number of rules, which does not give robust results for the FIS (the increase in minconf implies a decrease in the number of relevant rules).

Calculation of MAPE

FCM and Apriori	13.4
FCM and MS apriori	12.9
FCM and MSMM Apriori	11.5

CONCLUSION

The modified C means algorithm with membership calculation and the restriction regarding the objective function performs better result in terms of accuracy. Also using Box Metric equation for proximity measure instead of Euclidian distance gives betters clusters. The modifications implemented in real time dataset.

The Beta thalassemia database gives a clear idea about the percentage of linear growth by analyzing the different variables. This analysis is done using clustering techniques. The database helps to identify the percentage of linear growth of b-thalassemia patients in children by the techniques of new modified clustering FCM algorithms.

This paper has presented an enhanced prediction models using a Fuzzy association rule mining approach. The FCM–Apriori model is based on a single support value, which has been tested for data sets in a Beta thalesemia patients. It is noted from the results that the model has efficiently minimized MAPE, which is sensitive to minsupp and minconf values. The model used FCM to decide centers for each field separately from the whole field. It is noted that FCM may be a basis an overlapping problem to fuzzy sets (membership functions) for the whole data set. In addition, FCM–MSapriori approach used a multiple minsupp for the whole database, for instance, by considering and assuming the same frequency for all items (attributes) in a particular data set. The FCM–MSMMapriori model was developed based on the integration of FCM–MSapriori and the multiple membership function approach, which is able to generate dominating FARs. It is noted that the proposed model offers the best prediction performance as compared to the existing models reported in the literature. In the future, an improvement of FARs extraction can be investigated to enhance prediction accuracy and performance further.

REFERENCES

Jain, V., Benyoucef, L., & Deshmukh, S. G. (2008). A new approach for evaluating agility in supply chains using fuzzy association rules mining. Engineering Applications of Artificial Intelligence, 21(3), 367–385.

Toloo, M., & Nalchigar, S. (2011). On ranking discovered rules of data mining by data envelopment analysis: some models with wider applications. In K. Funatsu & K. Hasegawa (Eds.), New Fundamental Technologies in Data Mining (pp. 425–446). InTech publisher.

Toloo, M., Sohrabi, B., & Nalchigar, S. (2009). A new method for ranking discovered rules from data mining by DEA. Expert Systems with Applications, 36(4), 8503–8508.

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

Chiu, H. P., Tang, Y. T., & Hsieh, K. L. (2012). Applying cluster-based fuzzy association rules mining framework into EC environment. Journal of Applied Soft Computing, 12(8), 2114–2122

Agrawal, R., Imielinski, T. & Swami, A. (1993). Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD international conference on management of data, Washington, DC, United States.

Kannan, S., & Bhaskaran, R. (2009). Association rule pruning based on interestingness measures with clustering. International Journal of Computer Science Issues (IJCSI), 6(1), 35–43.

Kiran, R., & Reddy, P. (2010). Mining rare association rules in the datasets with widely varying items' frequencies. In Proceedings of 15th international conference on database systems for advanced applications (DASFAA 2010). Tsukuba, Japan: Springer.

Agrawal, R. & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In Proceeding of 20th international conference on very large databases (VLDB), Santiago, Chile.

Gedikli, F., & Jannach, D. (2010). Neighborhood-restricted mining and weighted application of association rules for recommenders. Lecture Notes in Computer Science (Vol. 6488, pp. 157–165). Springer.

Srikant, R. & Agrawal, R. (1996). Mining quantitative association rules in large relational tables. In Proceedings of the 1996 ACM SIGMOD international conference on management of data, Montreal, Quebec, Canada.

Kuok, C. M., Fu, A., & Wong, M. H. (1998). Mining fuzzy association rules in databases. ACM SIGMOD Record, 27(1), 41–46.

Kaya, M. & Alhajj, R. (2003). Facilitating fuzzy association rules mining by using multi-objective genetic algorithms for automated clustering. In Proceedings of the third IEEE international conference on data mining (ICDM'03).

Hong, T. P., Kuo, C. S., & Wang, S. L. (2004). A fuzzy Apriori Tid mining algorithm with reduced computational time. Applied Soft Computing Journal, 5(1), 1–10.

Zhang, L., Shi, Y. & Yang, X. (2005). A fuzzy mining algorithm for association-rule knowledge discovery. In Proceedings of the eleventh Americas conference on information systems, Omaha, NE, USA.

Huang, M. J., Tsou, Y. L., & Lee, S. C. (2006). Integrating fuzzy data mining and fuzzy artificial neural networks for discovering implicit knowledge. Knowledge-Based Systems, 19(6), 396–403.

Lei, Z. & Ren-hou, L. (2007). An algorithm for mining fuzzy association rules based on immune principles. In Proceedings of the 7th IEEE international conference on bioinformatics and bioengineering, Boston, MA.

Pach, F. P., Gyenesei, A., & Abonyi, J. (2008). Compact fuzzy association rule-based classifier. Expert Systems with Applications, 34(4), 2406–2416.

Chen, C. H., Tseng, V. S., & Hong, T. P. (2008). Cluster-based evaluation in fuzzy- genetic data mining. IEEE Transactions on Fuzzy Systems, 16(1), 249–262.

Ashish, M. & Vikramkumar, P. (2010). FPrep: Fuzzy clustering driven efficient automated pre-processing for fuzzy association rule mining. In Proceedings of IEEE international conference on Fuzzy systems, India (pp. 1–8).Bezdek, J. C. (1981). Pattern recognition with fuzzy objective function algorithms. ISBN 0-306-40671-3.

Palacios, A. M., Gacto, M. J., & Alcala-Fdez, J. (2010). Mining fuzzy association rules from low-quality data. Soft Computing. doi:10.1007/s00500-011-0775-3.

Hu, Y-H., & Chen, Y-L. (2006). Mining association rules with multiple minimum supports: a new mining algorithm and a support tuning mechanism. Decision Support Systems (vol. 42, pp. 1–24). Elsevier.

Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. In Proceedings of the 2000 ACM SIGMOD international conference on management of data. Dallas, Texas, United States: ACM.

Liu, B., Hsu, W., & Ma, Y. (1999). Mining association rules with multiple minimum supports. In Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99). San Diego, California, United States: AC

Tony Cheng-Kui Huang (2013) Discovery of fuzzy quantitative sequential patterns with multiple minimum supports and adjustable membership functions, Information sciences 222,126-146-Elsevier

Y.C. Lee, T.P. Hong, W.Y. Lin, Mining fuzzy association rules with multiple minimum supports using maximum constraints, The Eighth International Conference on Knowledge-based Intelligent Information and Engineering Systems 3214 (2004) 1283–1290.

Y.C. Lee, T.P. Hong, W.Y. Lin, Mining association rules with multiple minimum supports using maximum constraints, International Journal of Approximate Reasoning 40 (2005) 44–54.

Y.C. Lee, T.P. Hong, T.C. Wang, Multi-level fuzzy mining with multiple minimum supports, Expert Systems with Applications 34 (2008) 459–468.

Y.C. Hu, G.H. Tzeng, C.M. Chen, Deriving two-stage learning sequences from knowledge in fuzzy sequential pattern mining, Information Sciences 159 (2004) 69–86.

N.P. Lin, H.J. Chen, W.H. Hao, H.E. Chueh, C.I. Chang, Mining negative fuzzy sequential patterns, in: Proc. of the 7th WSEAS International Conference on Simulation, Modelling and Optimization, Beijing, China, 2007, pp. 52–57.

C. Fiot, A. Laurent, M. Teisseire, From crispness to fuzziness: three algorithms for soft sequential pattern mining, IEEE Transaction on Fuzzy Systems (6) (2007) 1263–1277.

T.C.K. Huang, Knowledge gathering of fuzzy multi-time-interval sequential patterns, Information Sciences 180 (170) (2010) 3316–3334.

J. Xue, M. Krajnak, Fuzzy expert systems for sequential pattern recognition for patient status monitoring in operating room, in: Proc. of the 28th IEEE EMBS Annual International Conference, New York City, USA, 2006, pp. 4671–4674.

Bilal Sowan, Keshav Dahal M.A. Hossain, Li Zhang, Linda Spencer, Fuzzy association rule mining approaches for enhancing prediction performance, Expert Systems with Applications 40 (2013) 6928–6937-Elsevier

Bezdek, J. C. (1981). Pattern recognition with fuzzy objective function algorithms. ISBN 0-306-40671-3.

Tajunishaand Saravanan Department of Computer Science, Sri Ramakrishna College of Arts and Sci ence (W), Department of Computer Application, Dr. N.G.P. Institute of Technology, Coimbatore, Tamil nadu, India "An ecient method to improve the clustering performance for high dimensional data by Principal Component Analysis and modified K- means", International Journal of Database Management Systems (IJDMS), Vol.3, No.1, February 2011

Jin-AiMary Anne Tan, Ping-Chin Lee,Yong- ChuiWee, Kim-Lian Tan, Noor Fadzlin Mahali, Elizabeth George, and Kek-Heng Chua1 "High Prevalence of Alpha- and Beta-Thalassemia in the Kadazan-dusuns in East Malaysia: Challenges in Providing Effective Health Care for an Indigenous Group". J Biomed Biotechnol. 2010;2010. pii: 706872. doi: 10.1155/2010/706872. Epub 2010 Sep 5., www.ncbi.nlm.nih.gov/pubmed/20871816

S. H. Orkin and H. H. Kazazian Jr., The mutation and polymorphism of the human beta-globin gene and its surrounding DNA", Annual Review of Genet- ics, vol. 18, pp. 131171, 1984

J. A. M. A. Tan, P. S. Chin, Y. C. Wong, K. L. Tan, L. L. Chan, and E. George," Characterisation and confirmation of rare beta-thalassaemia mutations in the Malay, Chinese and Indian ethnic groups in Malaysia", Pathology, vol. 38, no. 5, pp. 437441, 2006., Pathology. 2006 Oct; 38(5):437-41, http://www.ncbi.nlm.nih.gov/pubmed/17008283

Mary Petrou Haemoglobinopathy Genetics Centre, University College London, Institute of Women's Health, University College London Hospitals NHS Foundation Trust, Pathology Division 86-96 Chenies Mews London WC1E 6HX " Screening for beta- thalassaemia", Indian Journal of Human Genetics (2010) 16: 1-5, January 01, 2010

Telfer P, Coen PG, Christou S, Hadjigavriel M, Kolnakou A, Pangalou E, et al. Survival of medically treated thalassaemia patients in Cyprus. "Trends and risk factors over the period 1980 2004.",Haematologica.2006; 91:1187-92. [PubMed]

Telfer P, Constantinou G, Andreou P, Chris-tou S, Modell B, Angastiniotis M. "Quality of Life in Thalassemia"., Indian J Hum Genet. 2010 Jan-Apr; 16(1): 1–5. doi: 10.4103/0971-6866.64934,Ann N Y Acad Sci. 2005;1054:273-82.[PubMed]

Modell B, Khan M, Darlinson M. Survival in beta thalassaemia major in the UK:Data from the UK Thalassaemia Register. Lancet. 2000; 355:2051- 2. [PubMed]

Lilly P.L,Siji P.D, A New Featured Fuzzy Clustering Algorithm Based On Adaptive Clustering,VISTAS Vol.2,No.1,ISSN- ISSN: 2319-577, pp 7-11

D.Napoleon A and S.Pavalakodi Department of Computer Science,Bharathiar University,"A New Method for Dimensionality Reduction using K-Means Clustering Algorithm for High Dimensional Data Set" Volume 13 No.7, January 2011

Kiran Jyoti Department of CSE and IT GNDEC, Ludhiana, Punjab, India Dr. Satyaveer Singh Department of Mathematics JJT University, Jhunjunu, Rajasthan "Data Clustering Approach to Industrial Process Monitoring, Fault Detection and Isolation".

Binu Thomas, Raju G., and Sonam Wangmo "A Modi_ed Fuzzy C-Means Algorithm for Natural Data Exploration " E. Cox, Fuzzy Modeling And Genetic Algorithms For Data Mining And Exploration, Elsevier, 2005

Pavel Berkhin, "Survey of Clustering Data Mining Techniques", Available: http://citeseer.ist.psu.edu/berkhin02survey.html