Enhanced Fuzzy Association Rule Mining Approaches for Prediction Performance in Betathalesemia's Patients

Dr. Lilly P. L[1], Siji P. D[2]
1. Associate Professor, Department of Mathematics, St.Josephs College Irinjalakuda, Thrissur
2. Assistant Professor, Department of Computer Science, St.Josephs College Irinjalakuda, Thrissur

***Abstract*-This paper presents an investigation into a fuzzy association rule mining model for enhancing prediction performance in a medical database. This model (the FCM–MSMMApriori model) integrates multi membership and multiple support approach for Betathalasemia disease for performance prediction. The idea of adjustable membership functions is measured to offer fuzzy quantitative information, depending on the different membership functions. Betathalasemia data which usually has a quantitative structure in nature. Traditional data mining algorithms such as association rules, when applied alone, often yield uncertain and unreliable results. The new algorithm focuses on characteristics of the variety structure of data, and the association rules of data causes can be calculated more accurately and in higher rates. The Novel application of Multi Support Multiple membership function to the attributes according to their nature gives better prediction and good performance analysis.**

Keywords: Fuzzy association rules, FCM-Apriori algorithms, Multiple Membership,

1. INTRODUCTION

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. The next section presents the Existing algorithms and prediction models in the literature. Section 3 describes the proposed model with the case studies used to demonstrate the models. Experimental results of analysis are presented in Section 4. Finally, the conclusion are drawn in Section 5 with the key contribution of the research

2. RELATED WORK

The FCM–Apriori inherits benefits from FCM and Apriori and gives more flexibility to real-life applications, especially in business cases. The model acquires certain patterns in which the elements include rare items with higher profits, and excludes those that are trivial with lower profits. In recent years, classical extensions have been proposed in various applications. The paper [24] extended the idea to develop an algorithm for mining generalized association rules with multiple minimum supports. A method to address the issues of mining association rules with multiple minimum supports and maximum constraints is proposed in [26]. For fuzzy extensions, the paper [27] proposed an approach to find large-itemsets and association rules under the maximum constraint of multiple minimum supports. Subsequently, on integrated fuzzy set concepts, data mining, and multiple level taxonomy to find fuzzy association rules in a specified transaction dataset. In the line of mining sequential patterns, the concept [29] used the minimum spanning table method to find two-stage learning sequences in fuzzy sequential pattern mining. The author [30] proposed an idea on the absences of frequent fuzzy itemsets, and developed a method for mining negative fuzzy sequential patterns. Conversely, traditional fuzzy sequential pattern mining is referred to for mining positive patterns. There are three differing approaches in [31] used for the evaluation of the support, and extracted various levels of information for mining fuzzy sequential patterns. In [32] presented a multi-time-interval approach to discover fuzzy multi-time-interval sequential patterns. Several fuzzy inference systems developed by [33] for monitoring patient status; in particular, they included recursive fuzzy inference and non-recursive with sequential patterns as inputs

the author [34] proposed FCM–Apriori model extracts fuzzy rules for building a KB from a database, and our work is based on this paper. heir model utilizes the following two methods:FCM is used to transform the quantitative data set into fuzzy sets (terms). FCM is one of the fuzzy clustering algorithms based on an objective functioning method, developed by Bezdek [35] adopting the fuzzy set theory. In other words, it assigns a data object to more than one cluster. The Apriori 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 Apriori 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.

Fig. 1shows the outline of the process: (i) getting from the database the data set, which is analyzed for consistency and any noisy data set will be removed, (ii) transforming the quantitative data set into fuzzy sets while using FCM and applying the Apriori approach to extract FARs (iii) then saving these rules in the KB, (iv) using a Fuzzy Inference System (FIS) to command the knowledge (rules) for a prediction and (v) testing the feasibility of the model in the case studies.

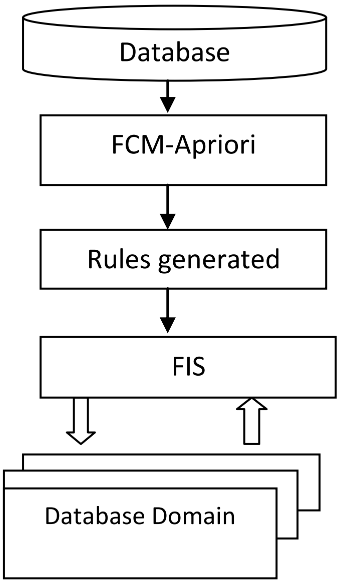


Fig. 1The FCM Apriori model

The following definitions (notations) are used in the model:
Field: Attribute (item or column) of the crisp input data.
Record (Case): Row with all fields.
Term: Fuzzy set class (fuzzy term).
x_{ij}: Value of the crisp input data.
$\mu(x)_{if}$: Fuzzy set membership value.
Sum_{jf}: Summation of each fuzzy term for all records.
Termset: A set of terms containing one term or more.
C_k: Contains candidate termsets, $1 \leq k \leq n$, where n = the maxi-mum number of fields.
L_k: Contains large termsets, $1 \leq k \leq n$, where n = maximum number of the fields.
minsupp: Minimum support threshold value (observing that minsupp = 2.45)
minconf: Minimum confidence threshold value (observing that minconf = 0.4 for this value is selected based on many experiments run to find out the appropriate ones that enable us to extract useful rules. Thus the error is minimized.).

The FCM–Apriori model shown in Fig. 2 works as follows:

```
Begin
  FCM; {clustering data set}
  Find the fuzzy sets of the quantitative data set, based on FCM
  Calculate the sum of the membership values for each fuzzy term for all records, using Eq. (1)
  IF Sum_jf ≥ minsupp Then
      Insert the fuzzy term into L_1, L_1={frequent termsets}
  For (k = 2; L_k−1 ≠ Ø; k++) do
      C_k = generate candidate from L_k−1(join  L_k−1called (p) with L_k−1 called (q));
      {
          Insert into C_k
          Select termset;  p.term_1, p.term_2 ... p.term_k−1, q.term_k−1
          From p, q
          Where p.term_1 = q.term_1 ... p.term_k−2 = q.term_k−2 , p.term_k−1 ≠ q.term_k−1
      }
     For each termset c ∈ C_k do
          Check all the sub-termsets of all termsets in C_k, and it should be a frequent termsets  in L_k−1
          For each (k − 1) subset s of c do
              IF s ∉ L_k−1 Then
                    Delete c from C_k
           End
      End
      For each termset candidate in C_k do
              Calculate the support value
              IF Sum_jf ≥ minsupp Then
                  Insert the fuzzy termset into L_k, L_k={frequent termsets}
       End
  End
  Select the frequent termsets including the target attribute (output attribute)
  Form the frequent termsets (rules) that exist in L_2 to L_n under the form "IF-Then"
  For each rule
     Calculate the confidence value for each rule
     IF CV ≥ minconf Then
        Accept the rule
  End
  Check the rules for contradiction
  Insert all the accepted rules in KB
  Infer the generated rules in KB using FIS
Stop
```

Fig-2.The FCM Apriori algorithm

(1) FCM is used to cluster the data into terms and then to deter-mine the centre of each fuzzy set and the maximum and minimum value for each field of the input data set.
(2) The data set is converted into a fuzzy data set, using one of the standard membership functions (the triangular and trapezoid membership functions [14]
(3) A support value is calculated for each term by summing the fuzzy membership values in each term for all records using $Sum_{jf}=\sum_{i=1}^{n} \mu(x)ij$ (Eq. (1), then this summation value is stored in the candidate termset C_1.
(4) Terms which are greater than or equal to minsupp are moved to L_1.
(5) Terms are joined up and combined, as (L_1 join L_{1-}) = {{c[1], c[i]},{c[1], c[i + 1]} . . . {c[1], c[n]}}, where c[1] represents the first fuzzy term, c[2] indicating the second fuzzy term with c[n] indicating the last fuzzy term. Also $c[1] \cap c[i] = \Phi$, $c[1] \cap c[i + 1] = \phi$, . . . $c[1] \cap c[n] = \phi$ (i.e., the terms for each termset do not belong to the same field).Once every termset is stored in the candidate termset C_2, the support value for each termset will be calculated using a minimum operator for the fuzzy values of the terms in the termset. Also the result of the minimum values in that termset is summed for all records. Finally, the results' summations will be stored in the candidate termsets C_2.
(6) Termsets greater than minsupp are moved to L_2.
(7) Termsets are joined up and combined again as L_2 = p join L_2 = q, where $p.term_1 = q.term_1$ . . . $p.term_{k\_2} = q.term_{k\_2}$, $p.term_{k\_1}$ – $q.term_{k\_1}$. This combination is based on every sub-termset of the candidate termset existing in C_k. The can-didate termset should be a frequent termset in the previous large termset of L_{k_1}. Also the terms for each termset in C_k do not belong to the same field.

(8) Termsets are stored in the candidate termset C_3, then the support value is calculated for each candidate termset.

(9) Termsets and their support values in C_3 greater than or equal to minsupp are moved to L_3.

(10) Termsets are joined up and combined, until L_n is empty.

(11) Termsets are pruned by selection of the termsets including the target attribute. As a consequence, termsets are phrased in IF–Then form, then the Confidence Value (CV) is calculated based on Eq. (2). The rules that are greater than or equal to minconf are accepted. Then the contradiction rules are removed, based on the CV.

(12) $$CV=\frac{\sum[(IF)\cap(Then)]}{\sum(\min(If))} \qquad \text{-eq(2)}$$

Once the extracted rules are stored in the KB, they will be used later in the FIS.

For the purpose of evaluation and validation, prediction quality is assessed using statistical evaluation metrics Mean Absolute Percentage Error (MAPE)

$$\frac{1}{N}\sum_{i=1}^{N}\left(\left|\frac{PVi-RVi}{RV_i}\right| * 100\right) \qquad \text{......eq(3)}$$

where

PV: the predicted output value

RV: the real output value

N: the total number of comparison records

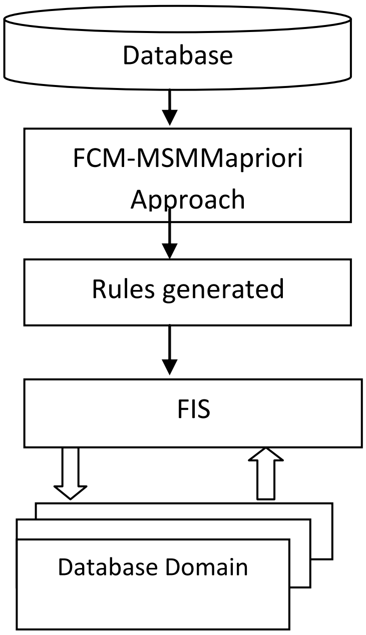
2.2 The FCM–MSapriori model

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 itemsets. In that case, if minsupp is set too high it will extract a low number of frequent itemsets. 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 itemsets, which causes combinatorial explosions, i.e., all the possible associations will be found. Some of those frequent itemsets 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.

3. The proposed 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.



Advantages of the proposed Algorithm

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

1. The FCM-Apriori is more natural and appropriate in relation to human knowledge. Managers can easily understand fuzzy linguistic terms or soft information discovered by the decision making process.
2. The FCM-MSApriori inherits benefits from FCM-Apriori and gives more flexibility to real-life applications, especially in business cases. The model acquires certain patterns in which the elements include rare items with higher profits, and excludes those that are trivial 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. We prepare various quantities of items, even if their fuzzy terms are the same.

3.2 Data sets for the case studies

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

Source of Data: The study 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

SL..NO.	DETAILS	NO
	Data size	61
	No of variables	7
	Min support	2.58
	Min confidence	0.4

Table1. analysis on Betathalesemia Data base

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 (Figure 4 (x-axis no of patients and y axis age)).

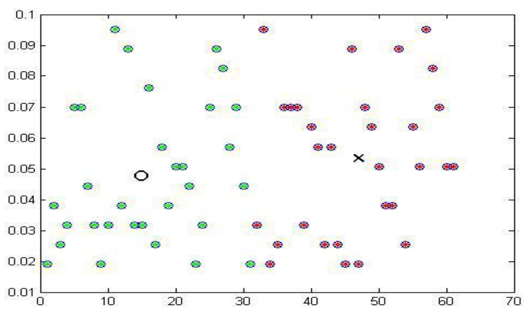


Fig.4.The two clusters formed using the variables patient id with their age using color discrimination

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.

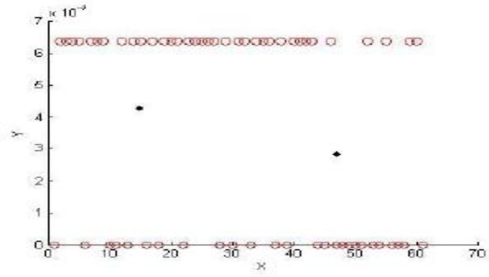


Fig.5 The distribution of Thalassemia patients according to sex

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

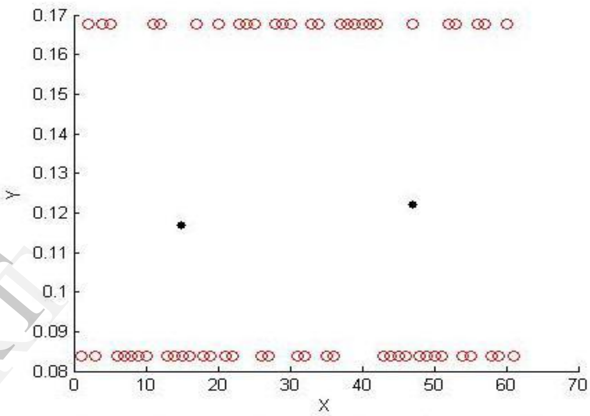


Fig 6.The distribution of Thalassemia patients according to consanguineous marriages

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

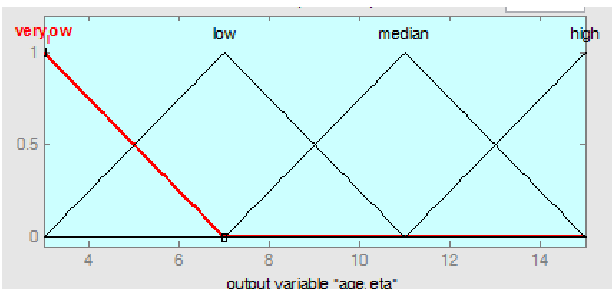


Fig. 7 represents an example of the age field (3-15) and its membership functions.

all fields have four fuzzy classes including: Very Low (VL), Low (L), Medium (M) and High(H). For abbreviation, each fuzzy class (fuzzy set) is mapped into numbers,

Fig 8 explains the analysis of min supp and minconf values on the MAPE for the betathalesemia Dataset. the graph of minconf 0.4 and minsupport 2.5 shows the Minimum MAPE value for 11.5

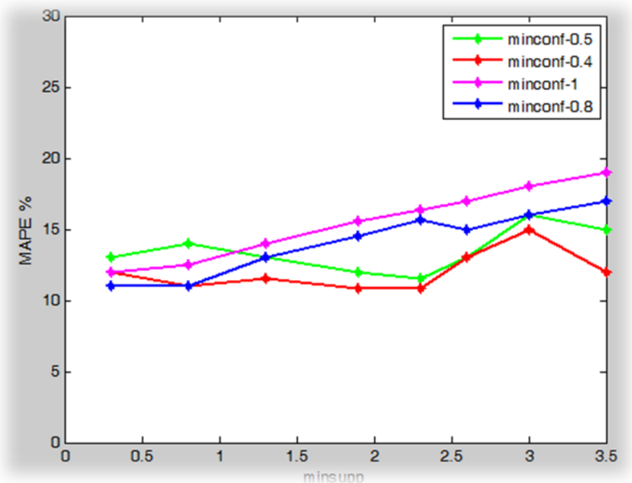


Fig 8. the MAPE for different minsupp and minconf

Fig 9 shows that the graph showing 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.

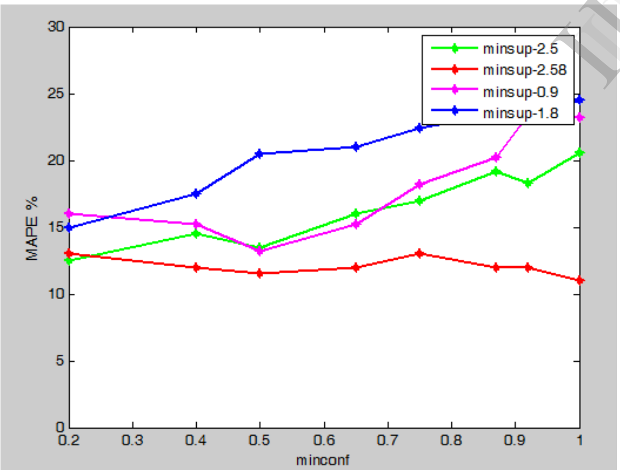


Fig 9. the MAPE for different minsupp and minconf

Fig 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.

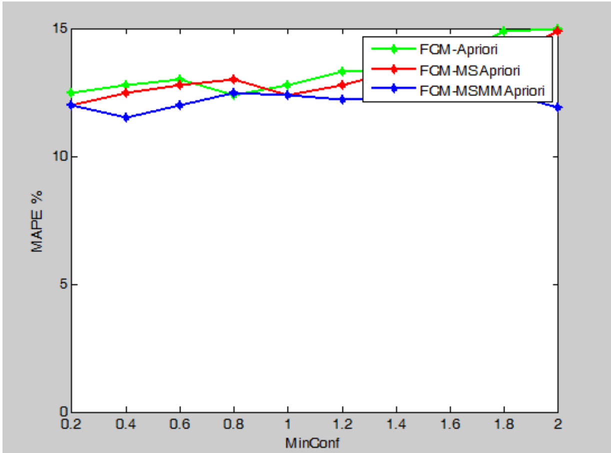


Fig 10. Analysis of Existing Algorithm and Proposed Algorithm

Fig depicts the analysis of minsupp and minconf values on the MAPE for the data set. The graph of minconf 0.4 with minsupp 2.5 shows the minimum MAPE value of 11.4%, and it contains rules that cover most cases. minconf less than 0.4 will increase the MAPE; this is explained by producing a large number of rules (a decrease in minconf implies an increase in the deviated rules, which causes noise for the FIS). It is noted that if minconf is greater than 0.4, it will also 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). The graph of minconf 1 is more highly affected by minsupp than the others, in other words, minsupp has a large influence on high
minconf values.

Calculation of MAPE

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

5. CONCLUSION

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

1. 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.
2. 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.
3. 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.
4. 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.
5. 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
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Kuok, C. M., Fu, A., & Wong, M. H. (1998). Mining fuzzy association rules in databases. ACM SIGMOD Record, 27(1), 41–46.
13. 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).
14. 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.
15. 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.
16. 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.
17. 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.
18. Pach, F. P., Gyenesei, A., & Abonyi, J. (2008). Compact fuzzy association rule-based classifier. Expert Systems with Applications, 34(4), 2406–2416.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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
25. 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
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. T.C.K. Huang, Knowledge gathering of fuzzy multi-time-interval sequential patterns, Information Sciences 180 (170) (2010) 3316–3334.
33. 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.
34. 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
35. Bezdek, J. C. (1981). Pattern recognition with fuzzy objective function algorithms. ISBN 0-306-40671-3.
36. 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
37. Jin-AiMary Anne Tan, Ping-Chin Lee,Yong- ChuiWee, Kim-Lian Tan, Noor FadzlinMahali, 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*
38. S. H. Orkin and H. H. Kazazian Jr., The mutation and polymorphismof the human beta-globin gene and its surrounding DNA", Annual Review of Genet- ics, vol. 18, pp. 131171, 1984

39. 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
40. 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
41. 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]
42. 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]
43. 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]
44. Siji P.D. and P. L. Lilly (2013), A new featured Fuzzy clustering Algorithm Based on Adaptive clustering , VISTAS , ISSN: 2319-5770, Vol. 2, No.1, 2013,127-133.