ISSN:0975-9646

M. Kalpana Devi et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5400-5403

Fuzzy Weighted Associative Classifier based on Positive and Negative Rules

M. Kalpana Devi#, M. Usha Rani*

#*Research Scholar, Dept. of Computer Science,*
Sri Padmavathi Mahila University, Tirupati
Andhra Pradesh, India

**Professor, Dept. of Computer Science,*
Sri Padmavathi Mahila University, Tirupati
Andhra Pradesh, India

***Abstract* - Construction of effective and accurate classifier is one of the challenges facing by the researchers. Many experiments have shown that Associative Classifier is significantly more accurate than the traditional classifiers. To classify the quantitative data, Fuzzy Associative Classifier was introduced and is also proved as an effective prediction model. Mining of negative association rules have gaining much attention among researchers recently, because, negative rules are as important as positive rules. This method has been recently introduced in the literature. In this paper we propose a new framework, Fuzzy Weighted Associative Classifier based on positive and negative rules. The main aim of this paper is to improve prediction accuracy.**

***Keywords* – Associative Classifier, Fuzzy Weighted Associative Classifier, Negative association rules,**

I. INTRODUCTION

Associative rule mining is one of the most important tasks in the Data mining, introduced by Agrawal et al [1]. Association rules explore interesting and confident associations among multiple attributes. Support and Confidence are used as interesting measures of association rule discovery, where Support represents the statistical significance and is used to reduce the search space. Confidence measures the conditional probability of events associated with a particular rule [2].

Associative Classification is an excellent classification method which combines both associative rule mining and classification. It has been proved by many authors that it gives more accuracy when compared with traditional classifiers. In the associative classification model classification is being done using the class association rules (CAR). Classification Association rules are nothing but a special subset of association rules whose right-hand-side is restricted to the classification class attribute. Rule based classifiers play an important role in the classification problems in many application domains. Because classification with the real time database is a complicated task that needs to be executed accurately and efficiently. There are many Associative classification methods proposed by various authors [3], [4], [5], [6], but these methods cannot be used with quantitative data and there exists sharp boundary problem. These types of problems can be approached with fuzzy representation of data. A fuzzy approach is widely exploited among the intelligent systems, since it is very simple and similar to the human way of thinking. It is used to transform quantitative data into fuzzy data through the identification of the appropriate membership function [7].

The limitation of the traditional Association Rule Mining model has been identified and its incapability for treating units differently and proposed that weight can be integrated in the mining process to solve this problem. Giving weightage to each attribute based on its importance will produce more significant rules in the mining process. The concept of assigning weights to each attribute is first introduced by Ramkumar et al. [8]. He has considered cost of the item for assigning weights and a new measure was introduced, *weighted support*.

Traditional association rule mining provides only positive association rules, but some researchers have found the importance of negative association rules. Negative association rule extraction is a new concept and these rules can provide valuable information in the prediction model. In this paper we have enhanced Fuzzy Weighted Associative Classifier algorithm with positive and negative rule generation.

II. RELATED WORK

A. Positive and Negative Association Rules

A negative association is referred to as a negative relationship between two itemsets. In the associative classification model, itemsets in the antecedent of the rule is positively or negatively associated with the consequent class attribute. Here we add negative rules of forms $X \rightarrow \neg Y$, $\neg X \rightarrow Y$, and $\neg X \rightarrow \neg Y$ to traditional positive rule $X \rightarrow Y$. The negation denotes the absence of itemsets.

According to John Tsiligaridis [9], there are two types of negative associative rule, the Confined negative association rule and the Generalized negative association rule. *The Confined negative association rule* is one of the following: $\neg X \rightarrow Y$, $X \rightarrow \neg Y$ or $\neg X \rightarrow \neg Y$ where the entire antecedent or consequent must be a conjunction of negated or non-negated attributes. *Generalized negative association rule* is a rule that contains a negation of an item a rule for which its antecedent or its consequent can be formed by a conjunction of presence or absence of terms. An example for such association would be: A∧¬B∧¬C∧D→ Y. It would be necessary to consider just not all items within a transaction, but also all possible items absent from the transaction. The author developed BTRC (Binary Tree Rules Construction) method. The BTRC produces nested sub trees in order to find the negative association rules. BTRC is based on successive partitioning of the events of observing a sequence with a certain number of positive and negative items.

Maria-Luiza Antonie et al. [10] introduced a new associative classifier that takes advantage of both positive and negative rules. Moreover, they have presented a new way to prune irrelevant classification rules using a correlation coefficient without reducing the accuracy of the associative classifier model. Authors introduced automatic progressive thresholding process for correlation measure; it eliminates the need for manually adjusted thresholds.

Wu *et al.* [11] derived a new algorithm for generating both positive and negative association rules. They add on top of the support-confidence framework another measure called *mininterest* (the argument is that a rule $A \rightarrow B$ is of interest only if $supp(A \cup B) - supp(A)supp(B) \geq mininterest$). Although they introduce the "mininterest" parameter, the authors do not discuss how to set it and what would be the impact on the results when changing this parameter.

Ramasubbareddy et al. [12] simplified the generation and validation of negative association rules with the existing support and confidence framework. Support and confidence of negative rules can compute through those of positive rules.

Rashmi Shikhariya et al. [13] proposed a hybrid approach to deal with large size data. Proposed system is the enhancement of Frequent Pattern (FP) technique of association rule mining with positive and negative rule generation.

Nikky Rai et al. [14] proposed a new algorithm (MIPNAR_GA) for mining interesting positive and negative rule from frequent and infrequent pattern sets and the process of rule optimization is performed by Genetic Algorithm.

Many algorithms regarding mining negative association rules have been proposed recently and some of them have been quoted above. To the best of our knowledge there is no other fuzzy weighted associative classification system that generates both positive and negative rules.

III. PROBLEM DESCRIPTION

A. Fuzzy Weighted Associative Classifier

Fuzzy Weighted Association Rule Mining with Weighted Support and Confidence Framework is proposed by Sunita et al. in [15]. And this method was also applied in predicting mosquito borne disease incidence [16].

In this paper we extend the problem of classification using Fuzzy Weighted Associative Rule Mining by proposing the generation of Positive and Negative rules with Fuzzy Weighted Associative Classifier.

In the present problem, we use climatic variables (such as temperature, humidity, rainfall, evaporation, sunshine and wind velocity) as predictors and mosquito borne disease incidence as predictive variable. As climatic conditions differ from place to place, data of a particular area alone has to be considered for the prediction.

Here Fuzzy logic is incorporated to split the domain of quantitative attribute into intervals, and to define a set of meaningful linguistic labels represented by fuzzy sets and use them as a new domain. For example Temperature attribute has been divided into set of three fuzzy items such as Low, Moderate and High. The original database has to be transformed into binary using fuzzy ranges and then to membership values (using trapezoidal membership function).

The most common framework in the association rule mining is the 'support-confidence'. Although these two parameters allow the pruning of many associations that are discovered from the database, still there are cases with many uninteresting rules produced. To minimize the uninteresting rules, we need another measure called correlation coefficient.

Correlation coefficient measures the strength of the linear relationship between a pair of two variables. The importance of this measure was discussed by Antonie et al. [10]. For two variables X and Y, the correlation coefficient is given by the following formula:

$$\rho = \frac{\mathrm{Cov(X, Y)}}{\sigma x \sigma y}$$

Cov (X, Y) represents the covariance of the two variables and σ_X stands for the standard deviation. The range of values for ρ is between -1 and +1. If the two variables are independent then ρ equals 0. When ρ = +1 the variables considered are perfectly positive correlated. Similarly, When ρ= -1 the variables considered are perfectly negative correlated.

B. Algorithm

In our algorithm we generate all possible combinations of items and then compute their correlations. If the correlation between item combinations X and Y (X is an itemset and Y the class attribute) is negative, negative association rules are generated when their confidence is greater than minimum confidence. X is an itemset and the negation of it is ¬X, the absence of itemset. The produced rules have either the antecedent or the consequent negated: (¬X → Y and X →¬Y), even if the support is not higher than the support threshold. If the correlation is positive, a positive association rule with the classical support-confidence idea is generated (X→Y). In our algorithm we calculate Fuzzy Weighted Support (FWS) and Fuzzy Weighted Confidence (FWC) measures for positive rules those measures can be calculated as given in [16] as follows:

$$FWS(X \rightarrow Y) = \frac{\sum_{For\ which\ Class\ label=1} \prod_{i=1}^{|X|} (\mu(\mathrm{I\,i, lj}) * \mathrm{W(Ii, lj)})}{\mathrm{n}}$$

Where μ (I_i, l_j) is a degree of membership for fuzzy attribute, W (I_i, l_j) is weight of a fuzzy item and 'n' the number of records in D.

$$FWC(X \rightarrow Y) = \frac{Fuzzy\ Weighted\ Support(XUY)}{\text{Fuzzy Weghted Support(X)}}$$

The Support and Confidences of negative rules can be calculated through positive rules as follows:

1. Consequent Negative Rule (CNR):

$$\text{FWS } (X \rightarrow \neg Y) = \text{FWS } (X) - \text{FWS } (XUY)$$

$$\text{FWC}(X \rightarrow \neg Y) = \frac{\text{FWS } (X) - \text{FWS (XUY)}}{\text{FWS (X)}}$$

2. Antecedent Negative Rule (ANR):

$$\text{FWS } (\neg X \rightarrow Y) = \text{FWS } (X) - \text{FWS } (XUY)$$

$$\text{FWC}(\neg X \rightarrow Y) = \frac{\text{FWS } (Y) - \text{FWS (XUY)}}{1 - \text{FWS (X)}}$$

Algorithm: Fuzzy Weighted Associative Classifier based on Positive and Negative rules.

Input: Database (D), Membership function (mf()), Minimum Correlation (mincorr), Minimum Support (minsup), Minimum Confidence (minconf)

Output: Positive Classification Association Rules (PCAR), Negative Classification Association Rules (NCAR)

(1) Transform original database (D) into binary by comparing with the fuzzy ranges.
(2) Select appropriate membership function mf() and transform original database into fuzzy membership values.
(3) Each fuzzy attribute is assigned a weight ranging from 0 to 1 to reflect their importance in the prediction model. Attributes that have more impact will be assigned a high weightage and attributes having less impact are assigned low weightage.
(4) Generate Classification Association Rules using frequent itemset generation and is represented as X → c, where X is an itemset and c is a class label.
Examples of such rules are {(Temperature, ‘Moderate’), (Rainfall, ‘High’)} → (Malaria Incidence, ‘High’) and {(Humidity, ‘Low’), (Rainfall, ‘Moderate’)} → (Malaria Incidence, ‘Moderate’).
(5) Compute Correlation (corr) between fuzzy itemset and class attribute.
(6)If corr >= mincorr /*here we have taken mincorr=0.3*/
(7) Compute Fuzzy Weighted Support (FWS) and Fuzzy Weighted Confidence (FWC) for each rule.
(8) If FWS >= minsup && FWC>=minconf
(9) Store these rules into Positive rule base
(10) If corr < - mincorr
(11) Generate Negative rules such as {X→¬c}, {¬X→c}
(12) Compute FWS and FWC for each rule.
(13) If FWS>minsup && FWC>=minconf
(14) Store these rules into Negative rule base.
(15) Repeat step (4) to (14) until maximum candidate itemset length is obtained.
(16) Keep all the Positive and Negative rules into main rule base
(17) Sort main Rule base in the order of higher precedence.
(18) Whenever a new record is provided, the rules from the rule base are used to predict the class label.

C. Pruning

Apart from generating Positive and Negative rules, there is another issue which has to be given more importance is, Pruning. As shown in the Fig.1 the number of rules generated is more when we append negative rules to the positive rules. To avoid uninteresting rules we should have an effective pruning strategy. The small number of strong rules leads to classification phase faster. In our problem domain class attribute is having three classes such as (Malaria, Low), (Malaria, Moderate) and (Malaria, High), for each positive rule there exists two negative rules. The resulted negative rules will be three times more. For example a positive rule {(Temp, Low) → (Malaria, Moderate)} will have two negative rules such as {¬ (Temp, Low) → (Malaria, Moderate)} and {(Temp, Low) →¬ (Malaria, Moderate)}. {(Temp, Low) → (Malaria, Low)} will have two negative rules such as {¬ (Temp, Low) → (Malaria, Low)} and {(Temp, Low) →¬ (Malaria, Low)} and same as for (Malaria, High). In our system correlation coefficient has been used to prune uninterested positive and negative rules.

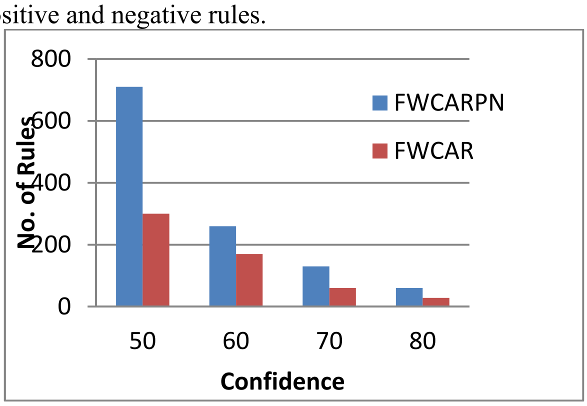

Fig.1 No. of rules produced

IV. RESULTS AND DISCUSSION

We enhanced Fuzzy Weighted Associative Classifier algorithm by introducing positive and negative rules. This method was tested and proved and it gives more accurate results at little extra cost. We developed a prediction system called Mosquito Borne Disease Incidence Prediction System using Fuzzy Weighted Associative Classifier [16]. Our model forecasts, mosquito borne disease incidence for the following month using monthly averages of climatic variables of present month. We have collected malaria cases reported month wise from Medical and Health department, Chittoor district and monthly averages of climatic data such as Temperature, Total rainfall, Relative humidity, Wind speed, Sunshine and Evaporation were collected from Regional Agricultural Research Station, Agricultural Meteorology Division, India meteorological department, Ministry of earth sciences, Tirupati.

From the above model, the accuracy we got is 87%. By enhancing with Positive and Negative rule, the system accuracy has been increased to 91%. This has proved that negative association rules are also important in the prediction model. To implement this work, we have used MATLAB 7.6.0 (R2008a) version. Some of the screen shots of our classifier model have shown in the following figures.

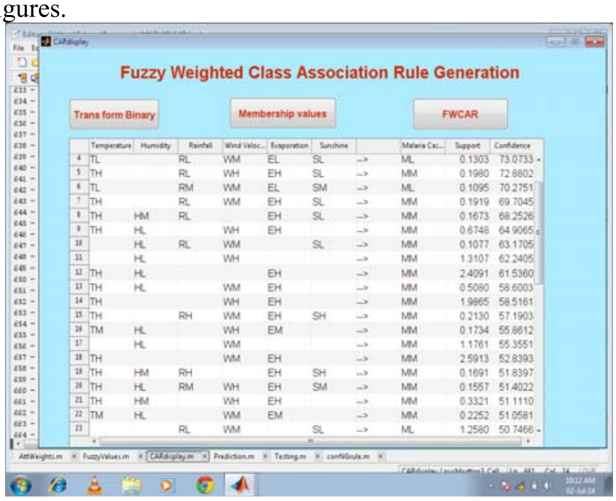

Fig.2 GUI for CAR display

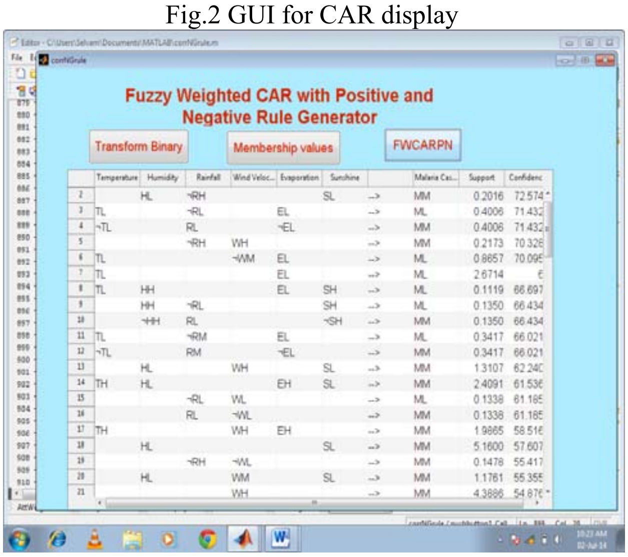

Fig.3 GUI for Positive & Negative rule display

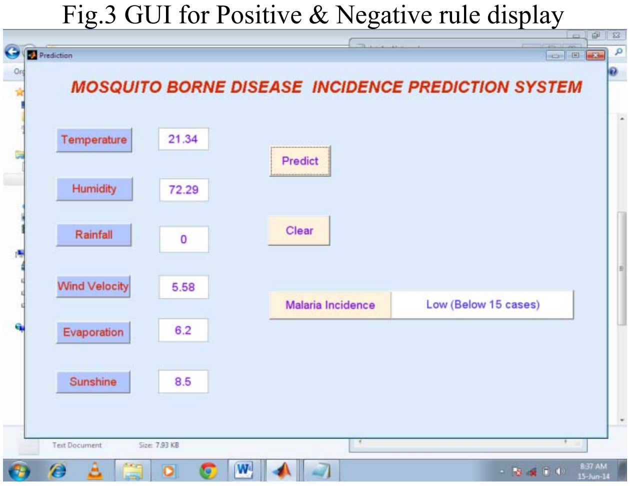

Fig.4 GUI for Disease Incidence Prediction System

V. CONCLUSION

In this paper we enhanced the algorithm Fuzzy Weighted Associative Classifier by introducing Positive and Negative rule generation. The interesting measure we used to generate strong positive and negative rules is correlation coefficient. We got better accuracy with the enhanced model and it's proved that negative rules are also as important as positive rules. If we add negative rules to already existing positive rules, we can use strong negative rules in place of weak positive rules and will get more accuracy. In future, we wish to improve accuracy of our algorithm by introducing better pruning strategy. Pruning strategy should be well organized, so that the prediction system gives more accurate results.

REFERENCES

[1] R. Agrawal, T. Imielinski, and A. Swami, 'Mining Association Rules between Sets of Items in Large Datasets' Proc. ACM SIGMOD '93, pp. 207-216, 1993.

[2] Liqiang geng and howard j. Hamilton , 'Interestingness Measures for Data Mining: A Survey'ACM Computing Surveys, Vol. 38, No. 3, Article 9, Publication date: September 2006

[3] B. Liu, W. Hsu, and Y. Ma. 'Integrating classification and association rule mining', In *KDD'98*, pp. 80-86, New York, NY, Aug. 1998.

[4] W. Li, J. Han& J. Pei, "*CMAR: Accurate and efficient classification based on multiple-class association rule*", Proceedings of the ICDM'01, 2001, pp. 369-376.

[5] X. Lin and J. Han, "*CPAR: Classification based on predictive association rule*", In SDM2003, San Francisco, CA, 2003.

[6] F. Thabtah., P. Cowling. & Peng Y, "*MCAR: Multi-class Classification based on Association Rule*", IEEE International Conference on Computer Systems and Applications ,2005, pp. 127-133.

[7] Z. Chen and G. Chen, An approach to classification based on fuzzy association rules, in T. Li, Y. Xu, D. Ruan (Eds.), *Proc. Int. Conf. Intelligent Systems and Knowledge Engineering*, 2007.

[8] G.D. Ramkumar, S. Ranka, and S. Tsur, "Weighted Association Rules: Model and Algorithm" Proc. ACM SIGKDD, 1998.

[9] John Tsiligaridis , 'Mining Positive and Negative Association Rules: An Approach for Binary Trees', (IJNCAA) 3(3): 81-87 The Society of Digital Information and Wireless Communications, 2013

[10] Maria-Luiza Antonie, Osmar R. Za¨ıane, 'An Associative Classifier based on Positive and Negative Rules', *DMKD* '04 June 13, 2004, Paris, France.

[11] X. Wu, C. Zhang, and S. Zhang. 'Mining both positive and negative association rules'. i*n Proceedings of 19th International Conference on Machine Learning*, pages 658–665, Sydney, Australia, 2002.

[12] B.Ramasubbareddy, A.Govardhan & A.Ramamohanreddy,' Classification Based on Positive and Negative Association Rules', International Journal of Data Engineering, (IJDE), Volume (2): Issue (2) : 2011

[13] Rashmi Shikhariya, Nitin Shukla2, 'An Improved Association Rule Mining with Fp Tree using Positive And Negative Integration', *Journal of Global Research in Computer Science,* Volume 3, No. 10, October 2012.

[14] Nikky Rai, Susheel Jain, Anurag Jain, 'Mining Interesting Positive and Negative Association Rule Based on Improved Genetic Algorithm (MIPNAR_GA)', Network and Complex Systems, IISTE, Vol.3, No.10, 2013.

[15] Sunita Soni and O.P.Vyas 'Fuzzy Weighted Associative Classifier: A Predictive Technique For Health Care DataMining', IJCSEI, Vol.2, No.1, February 2012.

[16] M. Kalpana Devi, M. Usha Rani, 'Mosquito Borne Disease Incidence Prediction System using Fuzzy Weighted Associative Classification', IJCA(0975-8887), Vol.91, No.13 ,April 2014.