Volume 4, Issue 9, September 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Performance Analysis of Fuzzy Weighted Associative Classifier Based on Positive and Negative Rules

M. Kalpana Devi
Research Scholar, Dept. of Computer Science,
Sri Padmavathi Mahila Viswavidyalayam
Tirupati, Andhra Pradesh, India

Prof. M. Usha Rani
Dept. of Computer Science,
Sri Padmavathi Mahila Viswavidyalayam
Tirupati, Andhra Pradesh, India

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

Keywords - ***Data Mining, Associative Classification, Fuzzy Weighted Associative Classifier, Fuzzificatio, Negative Rules.***

I. INTRODUCTION

Mosquito borne diseases are prevalent in more than 100 countries of the world, infecting 300-500 million people and causing about 1 million deaths every year. In India, more than 40 million people suffer from mosquito borne diseases annually. Even though we have enough past experience about epidemic diseases, but the problems are not handled in a proper way. The perusal of aforesaid statistics suggests that there should be a proper planning and implementation of control measures which require adequate exploitation of the available data for disease management and there is a need to develop prediction methods to augment existing control strategies. Therefore **Mosquito Borne Disease incidence prediction** is chosen as problem domain with the objective of developing Mosquito Borne Disease Incidence prediction system. We have developed a prediction model, which predicts mosquito borne disease incidence by identifying associations between climatic variables and disease incidence using Fuzzy Weighted Associative Classification technique.

Associative Classification (AC) is a branch of data mining, a larger area of scientific study. To build a model for the purpose of prediction, AC is a suitable prediction technique, which integrates two data mining tasks, association rule mining and classification. The main aim of classification is the prediction of class labels, while association rule discovery describes relationship between items in a transactional database. Of late, Associative Classifier is having better accuracy as compared to that of traditional classifiers.

There are many associative classification methods proposed by various authors [1], [2], [3], [4], 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 [5]. The importance of fuzz logic is discussed in section 2.

The limitation of the traditional Association Rule Mining model has been identified by researchers and its incapability for treating units differently and they have 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. [6]. He has considered cost of the item for assigning weights and a new measure was introduced, *weighted support*. The improvement in the accuracy of our system, when using attribute weightage concept is discussed in section 3.

Mining of negative association rules have been gaining much attention among researchers in recent times because, negative rules are as important as positive rules. This method has been recently introduced in the literature and these rules can provide valuable information in the prediction model. In section 4 we discussed how this negative rule generation will enhance the accuracy of the positive rule generation model.

II. WHY WE NEED FUZZIFICATION

From the point of view of a human being's intuition, no absolute crisp variable is existed in the real world. Any physical variable may contain some other components. For instance, if someone says: the temperature here is high. This high temperature contains some middle and even low temperature components. From this point of view, fuzzy control uses universal or global components, not just a limited range of components as the classical variables did.

Fuzzy ideas and fuzzy logic are so often utilized in our routine life that even none pays attention towards them. For instance, to answer some questions in certain surveys, most time one could answer with 'Not Very Satisfied' or 'Quite Satisfied', which are also fuzzy or ambiguous answers. Exactly to what degree is one satisfied or dissatisfied with some service or product for those surveys? These vague answers can only be created and implemented by human beings, but not machines. Computers can only understand either '0' or '1', and 'HIGH' or 'LOW'. Those data are called crisp or classic data and can be processed by all machines. It is possible to allow computers to handle those ambiguous data with the help of a human being, for that we need some fuzzy logic techniques and knowledge of fuzzy inference system.

In order to obtain fuzzy rules, first partition numerical attributes, usually by means of linguistic terms. Instead of assigning a class label to each record, a soft class label with degree of membership in each class is attached to each record; membership value of the class label varies from 0 to1. The use of fuzzy association rules can restrain the sharp boundary effect between intervals. The important characteristic of fuzzy logic is that a numerical value does not have to be fuzzified using only one membership function. In other words, a value can belong to multiple sets at the same time. For example, in our prediction model, a temperature value was considered as “Low", “Moderate" and “High” at the same time, with different degree of memberships as shown in the following figure.

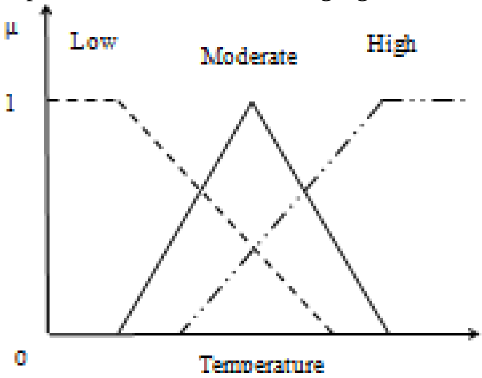

Fig.1 Fuzzy sets Low, Moderate and High

III. FUZZY WEIGHTED ASSOCIATIVE CLASSIFIER

Fuzzy Weighted Association Classification with Fuzzy Weighted Support and Confidence Framework is proposed by Sunita et al. in [7]. This method was applied in our mosquito borne disease incidence prediction system [8] and we have also enhanced Fuzzy Weighted Associative Classifier model by proposing the generation of Positive and Negative rules [9]. In our prediction system, we used 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 have 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 ‘Rainfall’ 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 two trapezoidal and one triangular membership function (Fig. 1)).

In several problem domains it does not make sense to assign equal importance to all the items, particularly in predictive modeling system where attributes have different prediction capability. The concept of weighted association rule mining is used to deal with the case where items are assigned a weight to reflect their importance and hence rules concerning them are of greater value and some authors have proposed a new Weighted Associative Classifier (WAC) that generates classification rules using weighted support and weighted confidence framework [6].

For example in the supermarket domain, weight is assigned to each of the items as per the profit it generates to the store, rather than simply calculating the percentage of transactions. A weight of an item is a non-negative real number which shows the importance of each item. A pair (x, w) is called a weighted item where $x \in I$ is an item and $w \in W$ is the weight associated with x. A transaction is a set of weighted items, each of which may appear in multiple transactions with different weights.

We have used the attribute weightage concept to incorporate attribute significance. In our proposed domain, some of the attributes have much impact to predict disease incidence. For example in the prediction of mosquito borne disease incidence, the attribute “Temperature” has more impact than the attribute “Humidity”. In our prediction model, weights assigned to each attribute were estimated based on the correlation between the attribute and the disease incidence.

A. Major steps of the proposed model

1) The monthly averages of climatic data and malaria cases of the following months were placed in a database in order to make it suitable for the mining process.
2) Each attribute is partitioned into fuzzy sets.
3) Transform original database into binary by comparing with the fuzzy ranges.

4) Select appropriate membership function and transform original database into fuzzy membership values.
5) Each fuzzy attribute is assigned a weight ranging from 0 to 1 to reflect their importance in this prediction model. Attributes that have more impact will be assigned a high weightage and attributes having less impact are assigned low weightage.
6) Generate Classification Association Rules using frequent itemset generation and is represented as X → C where X is subset of climatic variables and C is a class label such as malaria cases. Examples of such rules are {(Temperature, 'Moderate'), (Rainfall, 'High')} → (Malaria Incidence, 'High') and {(Humidity, 'Low'), (Rainfall, 'Moderate')} → (Malaria Incidence, 'Moderate').
7) Compute Fuzzy Weighted Support (FWS) and Fuzzy Weighted Confidence (FWC) for each rule.
8) Select rules that have a FWS and FWC larger than a given threshold values. These rules will be stored in Rule Base**.** Then these rules are sorted in an order of higher precedence.
9) Whenever a new record is provided, the rules from the rule base are used to predict the class label.

Performance of the system with and without weighted settings was tested; it was plotted and is showed in the following figures. Number of rules produced with weighted attributes w.r.t changing minimum confidence (and minimum support is 0.1) is shown in Fig. 2. Correctly classified instances with weighted settings are more when compared with Fuzzy Associative Classifier (Fig. 3), i.e. uninteresting rules have been removed from the final rule set in FWAC. Accuracy of the prediction system is also high with weighted settings (Fig. 8).

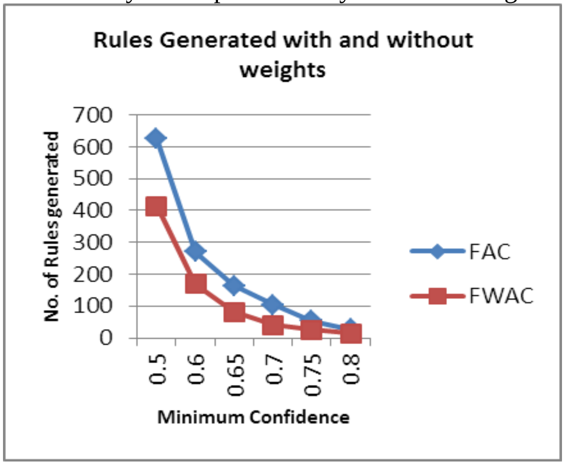

Fig. 2 Rules generated with and without weights

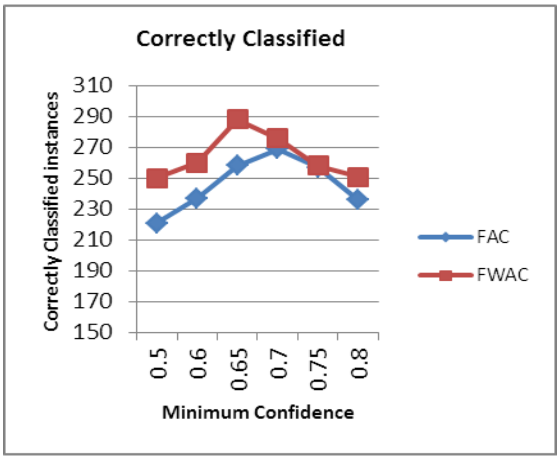

Fig. 3 Correctly classified instances with and without weights

IV. FUZZY WEIGHTED ASSOCIATIVE CLASSIFIER BASED ON POSITIVE AND NEGATIVE RULES

We enhanced Fuzzy Weighted Associative Classifier algorithm by introducing positive and negative rules. This method was tested and proved as more accurate with little extra cost. 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 [10]. Our algorithm generates all possible combinations of itemsets, and computes their correlations with class attribute. If the correlation is positive, a positive association rule with the classical support confidence idea is generated (X→Y). 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.

A. Algorithm

Traditionally the process of mining association rule has two phases, first being mining of frequent itemsets, and second being pruning for strong association rules. In our algorithm we combined two phases and generated strong association rules by analyzing correlations within each candidate itemset. We computed all possible combinations of items to analyze their correlations and keep only those rules generated from item combinations with strong correlation. The strength of the correlation is indicated by a correlation threshold. If the correlation between item combinations X and Y (X is an itemset and Y the class attribute) is negative, negative rules are generated, when their confidence is greater than minimum confidence. The produced rules are of the type *Confined negative association rules*, which will have either the antecedent or the consequent negated: (¬X → Y and X →¬Y) and both antecedent consequent negated: (¬X → ¬Y). If the correlation is positive, a positive association rule with the classical support-confidence idea is generated (X→Y). Proposed algorithm is as follows.

Algorithm: Fuzzy Weighted Associative Classifier based on Positive and Negative rules.
Input: Fuzzy Database (FD), Minimum Correlation (mincorr), Minimum Support (minsup), Minimum Confidence (minconf).
Output: Positive and Negative Classification Association Rules (PNCAR).

(1) F_1 = {large 1-itemsets in FD};

```
C = {all classes in FD};
for each i ϵ F1
    PNCAR ← GeneratePNrules ( i, C );
for (k=2; Fk-1≠φ; k++) {
    Ck= Join (Fk-1, F1);
    for each i ϵ Ck {
        supp= FWS( i, FD); //Compute fuzzy weighted support of an item i.
        if supp >= minsupp
            Fk ←Fk ∪{i}
        PNCAR ← GeneratePNrules ( i, C ); //PCAR & NCAR
    }
}
```

Fig .4 Classification Association Rule generation algorithm

Algorithm: GeneratePNrules (Generate Positive and Negative Classification rules).
Input: Itemset (i), Set of classes (C), Minimum Correlation Threshold (mincorr), Minimum Support Threshold (minsup), Minimum Confidence Threshold (minconf).
Output: Positive Classification Association Rules (PCAR), Negative Classification Association Rules (NCAR).

```
for each c ϵ C {
  r = corr (i,c);   // Compute correlation between i and c.
  if r > mincorr {
      pr = {i→c}; // Generate positive rule
      if FWC (i→c) >= minconf
          PCAR←PCAR ∪ {i→c};
   }
   else if r < - mincorr {
          nr1 = {¬i→c}; // Generate antecedent negative rule 1.
          if FWC (¬i→c ) >= minconf
               NCAR←NCAR ∪ {¬i→c};
          nr2 = {i→¬c};   //Generate consequent negative rule 2.
          if FWC (i→¬c ) >= minconf
               NCAR←NCAR ∪ {i→¬c};
          nr3 = {¬i→¬c};   //Generate antecedent consequent negative rule 3.
          if FWC (¬i→¬c ) >= minconf
            NCAR←NCAR ∪ {¬i→¬c};   }
  }
Return PCAR & NCAR
```

Fig .5 Positive and Negative Rule generation algorithm

Performance of the system with and without negative rule generation also tested. Number of rules generated by the prediction system w.r.t changing minimum confidence (and minimum support is 0.1) was high with the negative rule generation (FWPNAC) (Fig. 6) and correctly classified instances are also more with negative rule generation (Fig. 7). Accuracy is high in FWPNAC when compared with that of fuzzy weighted associative classification (FWAC), which is shown in the Fig. 8.

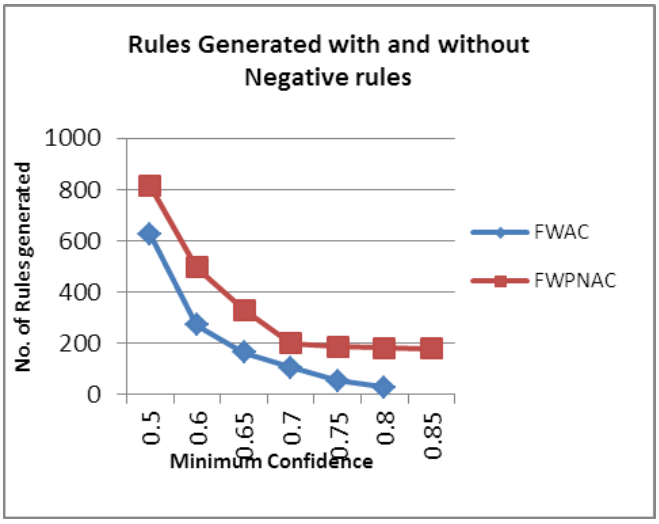

Fig. 6 Rules Generated with and without Negative rules

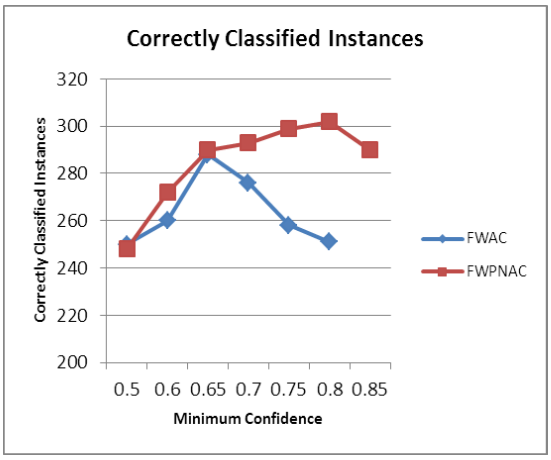

Fig. 7 Correctly classified instances with and without

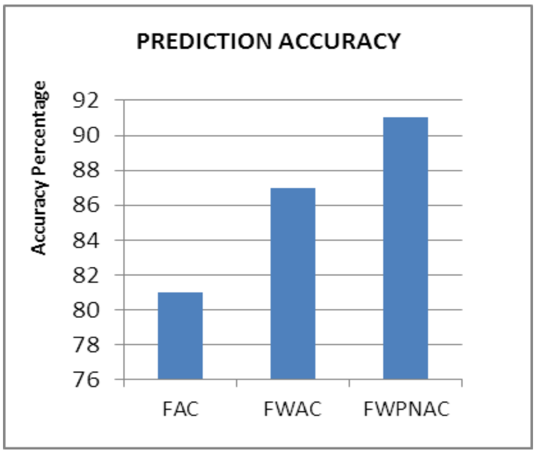

Fig. 8 Prediction accuracy negative rules

V. CONCLUSION

In this paper the performance of Fuzzy Associative Classification (FAC), Fuzzy Weighted Associative Classification (FWAC) and Fuzzy Weighted Positive Negative Associative Classification (FWPNAC) algorithms were compared. We have conducted many experiments with different minimum support and minimum confidence values. The experiment was done by collecting the real time data from departments concerned and the performance of the system is analyzed. From the results we obtained, it can be proved that attribute weightage concept increases the prediction accuracy (81% to 87%) by removing the uninterested rules and, negative rule generation will replace the week positive rules with strong negative rules, which results in higher accuracy (87% to 91%). In order to improve the accuracy of the proposed prediction model, an investigation was carried out with several approaches and proposed a model with high accuracy.

REFERENCES

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

[2] 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.

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

[4] 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.

[5] 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.

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

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

[8] 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.

[9] M. Kalpana Devi, Dr. M. Usha Rani "Fuzzy Weighted Associative Classifier based on Positive and Negative Rules" International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (4) , 2014, 5400 - 5403, ISSN: 0975 - 9646.

[10] 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

