World Applied Sciences Journal 17 (12): 1567-1570, 2012
ISSN 1818-4952

Using Support Vector Machine in Fuzzy Association Rule Mining

[1]Morva Ebrahimpour and [2]Hamid Mahmoodian

[1]Department of Electrical Engineering, Science and Research branch, Islamic Azad University, Fars, Iran
[2]Faculty of Electrical Engineering, Najafabad Branch, Islamic Azad University, Iran

Abstract: Fuzzy rule based classification systems is one of the most popular in pattern classification problems. The rules in the fuzzy models can be weighted to show the importance of generated rules where all attributes in the antecedent part of the rules have been usually weighted equally. Whereas the contributed attributes in a fuzzy model may have different influences on the decision making, a new method based on support vector machine-recursive feature elimination (SVM-RFE) has been proposed in this study to show the effects of attributes by weighting factors. Apriori algorithm and fuzzy association rule mining (FARM) have been used to generate the suitable rules which are weighted by fuzzy support value. The combination of the proposed method for attribute weighting and fuzzy support value for weighting the generated rules have been used to discriminate the samples of two different well known datasets iris and wine. The results show that this simple method can increase the rate of accuracy and reduce the dependency of model to fuzzy support value in Apriori algorithm and the number of rules.

Key words: Support vector machine . fuzzy association rule . data mining . classification . feature weighting

INTRODUCTION

One of the most addressed knowledge discovery tasks is to find a small set of rules from large databases. The larger the amount of stored data, the important the demand of discovering information which usually help researchers to make better decision in different aspects such health care services or business.

Extracted rules from a large scale database can used in making a rule base which in turn make a model for classification problems. Rule-based techniques such as RIPPER [1] or FOIL [2] have been presented earlier which are fast in classification but not efficient in classification accuracy in most cases.

In 1993, mining association rule discovery has been presented by Agrawal [3]. The proposed method in [3] tries to discover meaningful associations among different sets of attributes. Association rule discovery originates from market basket analysis and aims to find relationships within the items purchased in a basket. The general form of association rules are such $X \Rightarrow Y$ where X and Y are antecedent and consequent part respectively. In classification problems the antecedent part contains some of the items or attributes and the consequent part is usually a class label.

Association rules discovery is usually divided into two categories, crisp and fuzzy. In crisp category, which is usually used for attributes with binary value, 1 and 0 show the existence or not existence of an item in a transaction respectively. This method has been developed in other studies by fuzzy concepts (termed fuzzy association rule-FAR) to use linguistic variables to have higher interpretability and more comprehensively [4-7]. The model of classifier which is made based on FAR are closer to human language concepts. Association rule mining and fuzzy association rule mining (FARM) have been used in many studies that show their ability in rule extraction of high dimensional datasets [8-11].

One of the first attempts to mine association rules from a large dataset has been done by Apriori algorithm [3]. The algorithm can be used for both, finding frequent patterns and also deriving association rules from them. This algorithm can be used for both crisp and FAR and two variables Support value and Confidence value are used to show the reliability and accuracy of a mined rule [6, 7].

Actually, the importance of the selected attributes in a classification problem might be difference and most of proposed classification methods such as K-nearest neighborhood, support vector machine, logistic regression ignore it. In addition, It is shown that the procedures where consider a weighting factor for attributes in classification (such as fuzzy classifier) or in feature selection procedure (such as SVM-RFE) can improve the performance of the model.

Corresponding Author: Morva Ebrahimpour, Department of Electrical Engineering, Science and Research branch, Islamic Azad University, Fars, Iran

In this study, a new method in feature weighting based on SVM-RFE has been proposed to weight the attributes which are contributed in a fuzzy classifier structured by FARM. In this procedure, both attributes and fuzzy rules have been weighted by using SVM-RFE and fuzzy support values respectively simultaneously to consider the influences of attributes in final decision made by classifier. External 10-fold cross validation has been used to validate the procedure and show that this weighting method can increase the accuracy of classification as well as reduce the number of selected attributes if final dataset.

MATERIAL FRAMEWORKS

Fuzzy association rule mining: In this section the basic definition of FAR is introduced briefly. We will define weighted attribute-fuzzy association rule (WA-FAR) which also includes original FAR when all weighting factors are equal one.

Given a dataset matrix $G^{n\times m}$ where $g_{ij}(i=1,\ldots,n \text{ and } j=1,\ldots,m)$ is an element of matrix G which shows the amount of attribute j in i^{th} instance or samples and vector $C^{n\times 1}$ is the class label. Fuzzy membership values of jth attribute in i^{th} samples can be presented by

$$f_{i,j} = [A_{j,1}(g_{i,j}), A_{j,2}(g_{i,j}), \ldots, A_{j,s}(g_{i,j})] \quad (1)$$

where $A_{j,t}(\cdot)$ is $t^{tn}(t = 1,\ldots,s)$ membership function for j^{th} attribute. The number of s (or the number of membership functions) can change in different attributes but without losing generality, we suppose that s is same for all attributes. Let's the vector $W = [w_1, w_2, \ldots w_m]$ is weighting vector for attributes. The ith fuzzy sample can be represented in vector h_i where:

$$h_i = [w_1A_{1,1}(g_{i,1}), \ldots, w_1A_{1,s}(g_{i,1}), w_2A_{2,1}(g_{i,2}), \ldots, w_2A_{2,s}(g_{i,2}), w_mA_{m,1}(g_{i,m}), \ldots, w_mA_{m,s}(g_{i,m}), c] \quad (2)$$

and c is the class label.If we denote the attributes with r_1 to r_m, then a fuzzy item set can be presented by :

$$< R : A >= [<r_1 : A_{1tos}> \bigcup <r_2 : A_{1tos}> \bigcup \ldots \bigcup <\eta : A_{1toss}>] \text{ where } \leq m+1 \quad (3)$$

For example, when attributes are R = {Temperature, Pressure}, a fuzzy item set can be such as <Temperature: High>∪<Pressure: Medium>. Fuzzy support (FS) is a well-known evaluation measure which shows the usefulness of a fuzzy item set. If FS of a fuzzy item set be larger than a predefined value (λ), it is called a frequent fuzzy item set. FS for a fuzzy item set such as <R:A> is measured by:

$$\frac{\sum_{k=1}^{n}\prod < r_i : A_{i,j} > \varepsilon < R : A >^h k(r_i)}{n} \quad (4)$$

For example, if fuzzy 2-item set (X:A) = [<Temperature: High>∪<Pressure: Medium>] for three samples be:

Samp1={(Temperature:High/0.8),(Pressure: Med/0.3)},
Samp2={(Temperature:High/0.6),(Pressure: Med/0.8)},
Samp3={(Temperature:High/0.7),(Pressure: Med/0.5)},

Then:

$$FS(X:A) = \frac{0.8\times 0.3 + 0.6\times 0.8 + 0.7\times 0.5}{3} = 0.356$$

A fuzzy association rule has a form like (X:A)⇒(Y:B) where there is no intersection between (X:A) and (Y:B). To evaluate the accuracy of a rule, another well-known value, fuzzy confidence (FC) is defined by:

$$FC(X:A > \rightarrow < Y:B >) = \frac{FS(< X:A > \bigcup < Y:B >)}{FS(X:A)} \quad (5)$$

Considered rules should have FC larger than a predefined value (γ).

Assigning a class label: Suppose that V is set of all rules and v_c is the set of rules which their consequent part is the class with label c. Let denote $\mu^d_{A(.)}$ be the membership value of the new sample ϑ_{new} for fuzzy item sets in d^{th} rule which has FC_d confidence value. In this case, the new sample (ϑ_{new}) is assigned to be in the class which has maximum cl_c where:

$$cl_c = \frac{\sum d\varepsilon V_c(\prod A_{(.)}\,\varepsilon\, d\mu^d_{A(.)}) \times FC_d}{\sum d\varepsilon V(\prod A_{(.)}\,\varepsilon\, d\mu^d_{A(.)})} \quad (6)$$

Assigning the weighting factors: Support vector machine (SVM) has been known as a powerful classifier which has been widely used in many different aspects of data classification such as bankruptcy prediction [12], bioinformatic [13], drought prediction [14] or biometric [15]. According to this method, Guyon [16] has shown that in a linear discriminant function like:

$$D(x) = W^u.x + b \quad (7)$$

which W^u is weighting vector and b is bias value, the features (in this case, attributes) can be ranked by amount of $\|w_j^u\|^2$ where w_j^u are elements of vector W^u. In (7), $x = g_i^T (g_i^T)$ is the transpose of ith row of matrix G) and W^u can be measured based on proposed method in support vector machine classification problem [17] and it is equal:

$$W^u = \sum_{i=1}^{n} y_i \alpha_i g_i \tag{8}$$

where, y_i and α_i can be calculated by following optimization problem:

$$\max_\alpha \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{f=1}^{n} y_i y_f \alpha_i \alpha_f g_i^T g_f \tag{9}$$

Subject to $\sum_{i=1}^{n} y_i \alpha_i = 0$ and $0 \le \alpha_i \le C$(C is a constant which shows the rate of penalize misclassification and margin errors). This problem can extend to nonlinear problem which is not used in this paper. To use W^u as weighting factor for attributes, they can be normalized in [0, 1] by:

$$W = \{w_j \mid w_j = \frac{|w_j^u|}{\max_j(|w_j^u|)}\} \text{ and } j = 1, \ldots, m \tag{10}$$

EXPERIMENTAL RESULTS

To analyze the effect of weighted attributes in fuzzy association rule classifier, we used two well-known datasets (iris and wine). Iris dataset contains 150 samples of three types of flowers with four attributes and wine dataset includes 178 samples with 13 attributes which are divided in three classes. We consider 10-fold cross validation in all steps of experiments and used One Against All (OAA) method to train a linear support vector classifier and find three normalized weighting vector for each of datasets. The value of C in optimization problem (9) was considered equal infinity.

Linguistic variables, for both datasets, are defined small, medium and large with membership functions trapezoid, triangular and trapezoid respectively. K-means clustering method was used to divide the records of each attribute in 3 clusters. Center of each cluster (mean of amounts of the records) is considered to be vertex of membership functions which are designed such that sum of membership value of each attribute be equal one.

For rule extraction, The Apriori algorithm has been used separately for each class and the rules with the same class label in consequent part were considered to be candidate for rule base before pruning. Membership function for a class label was considered constant and equal 1 and so, the fuzzy confidence value (γ) for rules were equal 1. To prun the rules, if antecedent part of a rule was completely common with a larger rule, the smaller rule is pruned.

The procedure has been repeated 100 times and the mean of the number of rules and mean of accuracy (%) with different value for fuzzy support (λ) with weighting factor and without them ($w_j = 1$) have been calculated.

The results are shown in Table 1 and 2. Note that in most cases, the decreasing the number of rules causes that the size of rules be larger. (In tables, 'W' and 'NW' show the results with weighted attributes, non weighted attributes respectively where 'NR' means "no rule extracted").

The tables show that the attributes which are weighted based on their weighting factor, increase the rate of the classification accuracy comparing with non weighted case (except in wine dataset with FS=0.3). On the other hand, the tables show that the rates of accuracy and number of rules in non-weighted attributes have more dependency to the fuzzy support value than weighted ones. Weighted attributes have caused a decreasing in the number of rules which in turn, reduce the calculation time and increase the interpretability of the model (e.g. in wine dataset we have 87.08% accuracy with almost 171 Rules with FS=0.3 for non weighted model vs. 90.45 accuracy with almost 5 rules with FS=0.4 for weighted model).

CONCLUSION

The attributes in a dataset usually have different influences which should be considered in classification. Since SVM-RFE and fuzzy classifier have been known as two powerful methods for feature selection and classification respectively, a hybrid algorithm based on these two methods have been proposed to weight the attributes contributed in the classification procedure. The results present that the model with weighted attributes have higher performance with respect to the model with no weighted features. In addition the number of fuzzy rules in former model have been dramatically reduced which causes to reduce the computational cost as well as complexity in the model which consequently increase the reliability of the model. The tables have also shown that the effect of fuzzy support value in the rate of accuracy of the no weighted model is more than model with weighted attributes which shows that the improvement in robustness in proposed model.

Table 1: Iris dataset

	FS=0.3		FS=0.4		FS=0.5		FS=0.6		FS=0.7		FS=0.8	
	W	NW	W	NW	W	NW	W	NW	W	NW	W	NW
Rate of accuracy (%)	92.6	70	93.3	64.8	92.8	71.5	94.7	75.7	89.4	82.46	92.6	91.6
Number of rules (mean)	3	6.03	3	7.56	3.12	7.49	3.2	6.13	4.75	5.82	4	4.1

Table 2: Wine dataset

	FS=0.3		FS=0.4		FS=0.5		FS=0.6		FS=0.7		FS=0.8	
	W	NW	W	NW	W	NW	W	NW	W	NW	W	NW
Rate of accuracy (%)	80.9	87.08	90.45	87.1	85	83.8	83.6	81.9	---	90.02	---	---
Number of rules (mean)	13.28	171.2	5.2	81.2	4.5	36.4	4.44	17.36	NR	12.46	NR	NR

REFERENCES

1. Cohen, W.W., 1995. Fast effective rule induction. In ICML, pp: 115-123.
2. Quinlan, J.R., 1993. Cameron-Jones. RM.: Foil: A Midterm Report, pp: 3-20.
3. Agrawal, R, T. Imielinski and A. Swami, 1993. Mining association rules between sets of items in large databases: InProc. of the 1993 ACM SIGMOD Conference, pp: 207-216.
4. Mannila, H., H. Toivonen and I. Verkamo, 1994. Efficient algorithms for discovering association rules. In: Procd AAAIWorkshop on Knowledge Discovery in Databases, pp: 181-192.
5. Au, W.H. and K.C.C. Chan, 1997. Mining fuzzy association rules. In Proc. of 6th Int. Conf. On Information and Knowledge Management, Las Vegas, NV, USA, pp: 209-215.
6. Chen, G., Q. Wei and E. Kerre, 2000. Fuzzy data mining: Discovery of fuzzy generalized association rules. RecentIssues on Fuzzy Databases, Bordogna, G. and G. Pasi, (Eds.). Studies in Fuzziness and Soft Computing. Series Physica-Verlag.
7. Guoqing Chen and Quiang Wei, 2002. Fuzzy Association Rules and the extended mining algorithms. Information Sciences, 147: 201-228.
8. Sheikhan, M. and M. Sharifi Rad, 2011. Intrusion Detection Improvement Using GA-Optimized Fuzzy Grids-Based Rule Mining Feature Selector and Fuzzy ARTMAP Neural Network. World Applied Sciences Journal, 14 (5): 772-781.
9. Sheikhan, M. and M. Sharifi Rad, 2010. Misuse Detection Based on Feature Selection by Fuzzy Association Rule Mining. World Applied Sciences Journal 7 (Special Issue of Computer & Electrical Engineering): 32-40.
10. Sheikhan, M. and Z. Jadidi, 2009. Misuse Detection Using Hybrid of Association Rule Mining and Connectionist Modeling. World Applied Sciences Journal 7 (Special Issue of Computer & IT): 31-37.
11. Nikbakht Shahbazi, A., B. Zahraie, H. Sadeghi, M. Manshouri and M. Nasseri, Seasonal Meteorological Drought Prediction Using Support Vector Machine. World Applied Sciences Journal, 13 (6): 1387-1397.
12. Moradi, M., M. Shafiee Sardasht and M. Ebrahimpour, 2012. An Application of Support Vector Machines in Bankruptcy Prediction. Evidence from Iran. World Applied Sciences Journal, 17 (6): 710-717.
13. Dehghani, S. and M. Abbasi Dezfooli, 2011. Breast Cancer Diagnosis System Based Contourlet Analysis and Support Vector Machine. World Applied Sciences Journal, 13 (5): 1067-1076.
14. Nikbakht Shahbazi, A., B. Zahraie, H. Sadeghi, M. Manshouri and M. Nasseri, 2011. Seasonal Meteorological Drought Prediction Using Support Vector Machine. World Applied Sciences Journal, 13 (6): 1387-1397.
15. Radmehr, M., M. Anisheh, M. Nikpour and A. Yaseri, 2011. Designing an Offline Method for Signature Recognition. World Applied Sciences Journal, 13 (3): 438-443.
16. Guyon, I., J. Weston and S. Barnhill, 2002. Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning, 46: 389-422.
17. Cortes, C. and V. Vapnik, 1995. Support vector networks. Machine Learning, 20: 273-297.