Class Strength Prediction Method for Associative Classification

Suzan Ayyat
Department of Informatics
Huddersfield University
Huddersfield, UK
U1277021@hud.ac.uk

Joan Lu
Department of Informatics
Huddersfield University
Huddersfield, UK
j.lu@hud.ac.uk

Fadi Thabtah
Ebusiness Department
Canadian University of Dubai
Dubai, UAE
fadi@cud.ac.ae

***Abstract*—Test data prediction is about assigning the most suitable class for each test case during classification. In Associative Classification (AC) data mining, this step is considered crucial since the overall performance of the classifier is heavily dependent on the class assigned to each test case. This paper investigates the classification (prediction) step in AC in an attempt to come up with a novel generic prediction method that assures the best class assignment for each test case. The outcome is a new prediction method that takes into account all applicable rules ranking position in the classifier beside the class number of rules. Experimental results using different data sets from the University of California Irvine (UCI) repository and two common AC prediction methods reveal that the proposed method is more accurate for the majority of the data sets. Further, the proposed method can be plugged and used successfully by any AC algorithm.**

Keywords-associative classification; data mining; prediction phase.

I. INTRODUCTION

Associative Classification (AC) is an emerging classification research topic which employs association rules to solve classification problems in data mining [1]. The goal of the AC method is to learn a classification model from input classified data (historical data) that in turn is used to assign the right target class in new data (test data). For example, in text categorization the target class is the document's category. This type of application can be seen as supervised learning because learning is focused on a special attribute in the training data set called the target class.

Recent experimental research [2][3] indicated that AC methods usually devise good classification models in terms of predictive accuracy when contrasted to other classification methods such as statistical, covering, and decision trees. For instance, in a recent research study [4], and using 20 University of California Irvine (UCI) data sets [5] the accuracy of an AC algorithm called MAC [4] is 1.86%, 3.12 % and 3.11% higher than PART[6], RIPPER [7], and C4.5 [8] algorithms, respectively. This evidence, if limited, reveals the predictive power of AC when building classification models which increase the usage of this type of classification models in applications. The main reason for learning high accurate classification models by AC approach is the new rules (knowledge) induced during the learning step where the majority of the items and the target class combinations in the training data are evaluated for possible positive correlations [9]. Nevertheless, the number of rules could be large [10].

Recently, a number of AC algorithms have been developed in the research literature like, CBA [1], CPAR [11], LC [9], MAC [4] and others. These methods utilise various methodologies to induce rules, store rules, prune rules and predict the class of test data. This paper concentrates on the class prediction step in AC. Predicting the right class of a test data by the classifier is considered the most important step in the AC algorithm's lifecycle. In this step, the AC algorithm uses the rules learnt to predict the class labels of the test data and coming up with the right class for each test data is crucial because the overall predictive performance of the classifier depends on this decision. In addition, choosing the right rules in the classifier to assign the predicted class is a challenging tasks [12][13]. This is since there could be multiple rules applicable to the test data yet associated with different class labels.

To deal with the class prediction step in AC, we develop a novel class prediction method that considers all possible rules applicable to the test data. This is unlike most existing methods that:

a) Either consider the first rule in the classification model that is similar to the test data items [14].
b) Or computes rules' weights based on complex mathematical formula [11][15].

The main problem that this paper addresses is the inability of existing AC prediction methods of making use of all class labels in the classification model in cases when there are more rules similar to the test data items. For example, suppose we have a test data (a,b,c) that requires classification, and we have in the classification model 3 rules R1: (a^b, class2), R2:(b^c, class1), and R3:(b, class1). Assume that R1>R2>R3 in the classification model. Now, most existing AC methods like CBA, MCAR and MAC allocate class2 to the test data dismissing rules (R2 and R3) which indeed jeopardizes the prediction decision. On the other hand, few AC algorithms, like CMAR, groups rules applicable to the test data, with respect to their class labels, and then computes each group's rules support and confidence. This is problematic especially when we have a large number of rules similar to the test data or the number of test data to be classified is huge. So we intend to use all classes of the rules that are similar to the test data items for prediction. Then, when computing the accuracy of the classification model, the fired rule's assigned class to the

 ISBN: 978-1-61208-364-3

test data is counted which makes the decision more reliable and will possibly enhance the accuracy of the model.

Our new class prediction method considers the class labels of rules similar to the test data and gives the test data the class belonging to the highest score. Later in Section 3, we show how the class score is computed. This prediction is more realistic than one existing rule prediction methods simply because none of the applicable rules that are similar to the test data is ignored. The research question that the article is trying to solve is:

- Can we come up with a prediction method that takes into account both the rules position in the classifier and the class representation in the context of number of rules in order to have a fair and accurate prediction decision?

This paper is structured as follows: The literature review is given in Section 2. Our prediction method is discussed in Section 3 along with a detailed example. Section 4 is devoted to present the comparison results between the proposed methods and other classification methods in AC. Finally, the conclusions are given in Section 5.

II. LITERATURE REVIEW

Generally, there are two main methods in predicting the class of test data in AC. The first method is a group-based method that assigns the class that belongs to a group of rules to the test data during the classification step. The second method takes on only a single rule class often the class of the first rule (highest position one) similar to the test data items. This is the class that these methods assign to the test data to determine whether the test data are a hit or a miss when computing the model's accuracy. Typical algorithms that employ this kind of prediction are MAC, MCAR, CBA and many others. This prediction method assumes:

1) The rules in the classification model are sorted based on certain criteria
2) Only one rule is used for prediction

The second condition above has been criticized by several researchers [4][15], due to the following facts:

1) There could be more than one rule similar to the test data
2) These matching rules may have close ranking position in the classification model

Therefore, using a single rule is seen to be biased and an unfair decision. Nevertheless, this approach is simple, especially in circumstances when there is only one rule similar to the test data.

In circumstances when multiple rules are similar to the test data during the prediction step, the decision to only fire one rule becomes debatable. Consequently, a more fair decision is to use the information provided by all matching rules for the class prediction decision. In 2011, two prediction methods were proposed by Thabtah F. et al. [12] on using the classifier's rules confidence values matching the test data to make the prediction decision. The first prediction method groups all rules matching the test data into collections based on their classes and then the average confidence for each collection is computed. This method assigns the class of the group with the largest average confidence. The other method described by Thabtah F. et al. [12] is similar to the method described above but it does not require that all items of the rules be identical to the test data items by allowing partial similarity than full similarity aiming to have larger number of rules within the collections.

Veloso A. et al. [16] developed a prediction method in AC that utilises all rules applicable to the test data after dividing them into groups with respect to their class labels. Then, a group score consisting of the confidence and support value of the rule(s) is computed and the group class with the largest score is given to the test data. A few years ago, a greedy AC called CPAR used a multiple rules prediction method based on the Laplace expected accuracy. This method works as follows: For a test data (t) that is about to be predicted, the method groups all rules in the classifier contained in t in groups based on the rules class labels. Then the group expected accuracy average is calculated and t is allocated the largest group's expected accuracy class. The group's Laplace expected accuracy is computed according to the equation below:

$$\text{Laplace (Cluster)} = \frac{(D_c(r\_group)+1)}{(D_{tot}(r\_group)+D)} \quad (1)$$

where

D_c (r_group) is the number of training instances covered by the group's rule (head and tail).

D_{tot}(*r_group*) is the number of training instances similar to the group's rule body.

D is the number of class labels in the training data.

III. THE PROPOSED PREDICTION METHOD

In this section, we discuss the main contribution which is the development of a novel AC prediction method that will enhance the predictive power of any AC algorithm in forecasting test data. Our method falls under the category of a group-based method to come up with the most accurate class to assign to the test data during the classification step. The method assumes the following before it gets invoked:

1) All rules are generated and the classifier is built.
2) All rules within the classifier are sorted according to any sorting procedure.

So, when a test case is demanding a class during the prediction step, our method (Figure 1) works as follows:

It scans the classifier and marks any rule that is similar to the test data items. Here, we have several situations:

a) When there is only a single rule matching the test data items the situation is simple and we assign the class of that rule to the test data.
b) When more than one rule is similar to the test data and all of them are connected with a similar class, our

method assigns that class to the test data in a straightforward manner.

c) When multiple rules are similar to test data and these rules are connected with different class labels the situation becomes challenging and this is where the novelty of our method applies. Firstly, our method clusters the applicable rules into groups with respect to their classes. Then, based on both the rules rank in each cluster and the cluster size the decision of which class to assign to the test data is decided. We have combined both “the rules rank per group” and the “size of the group” into a ranking formula that we name the Class_Strength as shown in the equation below.

$$\text{Class\_Strength Ci=Score Ci + Ci Number of Rules} \quad (1)$$

$$\text{Score Ci} = \sum_{\mathrm{i}=1}^{\mathrm{j}} n - (Xci - 1) \quad (2)$$

Where

Xi is the number of rules matching the test data for

class ci

n is the total number of rules matching the test data

So, for each group’s class, its strength will be calculated and the class belonging to the group that has the largest strength gets assigned to the test data. In the case that more than one group has the same number of rules; the choice will be based on the class representation (number of rules per group).

This method takes advantage of two previous group-of rule prediction methods in AC; the one that considers the rules rank are the primary criteria (confidence) and the other

```
Input: test data set (T), Classifier (C)
Output: Error rate E

Iterate over Ts
Iterate over C
      locate rules that are similar to the current test data
      cluster the rules per class label
         compute the class strength per cluster according to
Equation (1)
        assign the class with the largest strength to the current
test data
end
      end if
   else assign the default rule to the current test data
   end if
end
end
compute the number of errors of T
```

Figure 1 The new prediction method

that considers the class representation per rule (number of rules). We have combined both procedures into a novel measure called the Class_Strength for a more legitimate prediction decision. The class assignment of test data has improved when contrasted with classification procedures such as that of CBA and its successors that take the class of the first ranked rule in the classifier matching the test data to make the prediction decision. Furthermore, it also overcome multiple rule prediction methods in AC like CPAR and CMAR that employ mathematical based attribute assessment formulas, e.g., confidence, support, weighted Chi-Square. Now, instead of favouring rules with high confidence (ranking position) or rules belonging to the most representative class to make the classifications decision, the new measure takes advantage of both approaches which give the decision of assigning class to test data legitimacy and accuracy. Finally, when no rules in the classifier are applicable to the test case, the default class rule will be assigned to that case.

Example

Consider the test data shown in Table I to be predicted, Table II shows all relevant rules from the classifier that are similar to the test data. The similarity has been based on the rule’s body and the test data items. Now a typical AC like CBA or MCAR will take on rule rank # 1 and will allocate its class, i.e. (c3), to the test data simply because this is the best ranked rule matching the test data. On the other hand, a class representation based prediction method, like MAC, assigns class (c1) to the test data since this class has the most number of rules matching the test data. For our prediction method, we first divide the rules in Table II into groups based on their class labels as shown in Table III. We then compute the new rule score based on the Equation (1), i.e $n - (Xi - 1)$ and the rules score are shown below in Table IV. Finally, we sum up each class score with the number of rules belonging to it to derive the class strength, as shown in Table V. In this example, we have a tie score between class c_3 and c_1. Nevertheless, we assign class c_1 to the test data since it has a larger number of rules.

TABLE I. TEST DATA

Attribute1	**Attribute2**	**Attribute3**	**Attribute4**	**Class**
a_1	b_1	l_1	g_5	?

TABLE II. RULES MATCHING THE TEST DATA OF TABLE I

Rank	**Rules**
1	g_5 ^ $\rightarrow c_3$
2	l_1^b_1 $\rightarrow c_2$
3	a_1 ^ $\rightarrow c_1$
4	a_1 ^b_1 $\rightarrow c_1$

 ISBN: 978-1-61208-364-3

TABLE III. NUMBER OF RULES PER CLASS

Class	# of Rules
c_3	1
c_2	1
c_1	2

TABLE IV. RULES SCORE CALCULATIONS

Rank	Rule	Score Calculation	Rule weight
1	g_5 ^ $\rightarrow c_3$	4 - (1-1)	4
2	l_1^b_1 $\rightarrow c_2$	4 - (2-1)	3
3	a_1 ^ $\rightarrow c_1$	4 - (3-1)	2
4	a_1 ^b_1 $\rightarrow c_1$	4 - (4-1)	1

TABLE V. CLASS SCORES

Class	Class Score Eq. 1	Class Strength Score+#of Rules
c_3	4	4+1=5
c_2	3	3+1=4
c_1	3	3+2=5

IV. EXPERIMENTAL RESULTS

A. Settings

Different data collections from the UCI repository [5] have been utilised to measure the impact of the new prediction method on the classification accuracy of the classifiers resulting from the experiments. We have used 12 data sets that have different size and attribute types. Tenfold cross validation testing method has been used to run the experiments. This method is used in data mining research to derive accurate and fair results. In particular, it divides the input data set into 10 folds randomly and the classifier is trained on 9 folds and then tested on the hold out fold to derive its error rate. The same process is repeated 10 times and the accumulated results are then averaged.

We have selected two main prediction methods in AC to compare our results with mainly because they use different prediction methods for assigning the class labels to test data. The main measure used for comparing these methods and ours is the error rate since we would like to answer the question "whether combing rules rank with class number of rules enhance the predictive power of AC algorithms?". The first methodologies used are based on CBA and MCAR and consider the highest rank rule prediction [1][14]. The second method was recently developed and uses a group-based method that considers the class associated with the maximum number of rules [4]. All these methods and ours have been implemented in Java in MCAR algorithm.

In the experiments, the AC main parameters, which are minimum support (*minsupp*) and minimum confidence (*minconf*), have been set to 2% and 50%, respectively. The reason for setting the *minsupp* to 2% is because it has been used previously by many research studies and proved to be fair in compromising between the number of rules extracted and the accuracy rate. The *minconf* has a limited effect on the performance of the AC algorithm so we have set it to 50%. Lastly, the experiments have been performed on I3 PC with 4.0 GB RAM and 2.7 GH processor.

B. Results Analysis

We have generated the error rate of the considered prediction methods on 12 UCI data sets, as shown in Figure 2. The figure clearly demonstrated that the proposed prediction method has enhanced the predictive rate of the classifiers devised on the data sets. In particular, our method achieved a decrease in the error rate on average by 1.18% and 1.12% on the 12 data sets we consider when contrasted with MAC and CBA algorithms respectively. A possible reason for the decrease in the error rate is mainly due to that new prediction methodology that allocates the test data the most appropriate class based on the class strength that we compute during the prediction step and for each test case. The fact that we consider both the rules rank and the class of rules for each class cluster gives a legitimate and accurate decision of which class to assign. This is since we have accounted for multiple rules and considered these rules rank in a new formula that assures a score for each class. In other words, we allocate the class of the cluster having the largest score (strength) to the test data based on both number of rules applicable to the test data and these rules

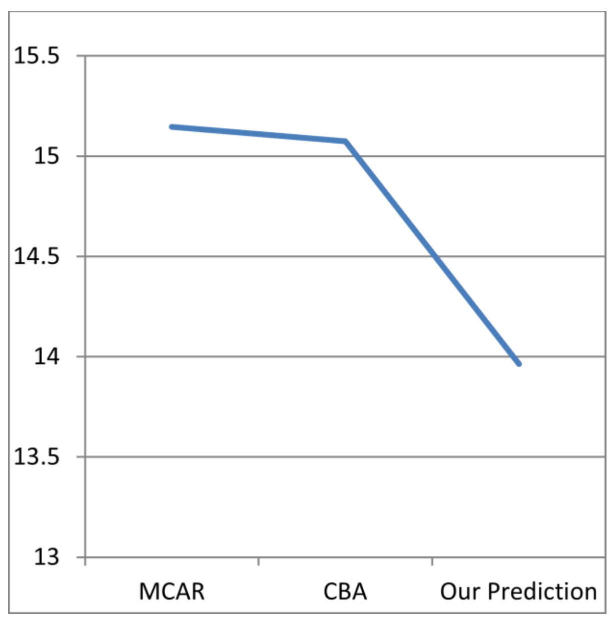

Figure 2 Average error rate produced from the UCI data sets by the prediction methods

rank in the classifier. This, surely, should minimize the error rate of the classifier, as shown in figure 2.

We have looked into more detailed results and for each UCI data sets, as depicted in Table VI. The figures in the table reveal consistency in the error rate results between the prediction methods we consider in this article. This means there are no large significant differences in the error rate results for most of the data sets for both CBA and MAC, except the fact that we have improved the predictive power of the classifiers for most of the data sets. Precisely, our prediction method outperforms both MAC and CBA prediction methods on most of the data sets and the won-tie-lost records are 10-2-2 and 9-3-0, respectively. The fact that our method investigates both the rank of the rules and the class representation per rules has a definite advantage and the most suitable class gets allocated to the test data.

Table VII displays the runtime for the prediction phase in seconds computed from two implementations (MAC single rule and our multi-rule prediction methods) for a sample of the data sets. It is obvious from the figures in the table that our prediction method normally takes longer to forecast test data than single rule based methods such as MAC. Nevertheless, the proposed prediction method has enhanced the predictive performance of the final classifiers if compared to those of MAC. In addition, the time spent in assigning test cases the right class labels is not excessive according to Table VII. There should be a tradeoff between precision and test data prediction time where longer time can be tolerated in exchange for higher level of predictive accuracy.

TABLE VI. THE ERROR RATE OF THE CONSIDERED PREDICTION METHODS ON 12 UCI DATA SETS

Data set	Size	MAC	CBA Prediction	Our Prediction
Breast	699	5.36	6.76	5.42
Cleve	303	18.54	16.9	18.36
Glass	214	24.76	23.47	22.58
Heart	294	18.8	18.13	18.1
Hybothroid	3772	6.3	7.68	6.3
Iris	150	7.06	6.69	5.74
Labor	57	16.49	13.67	14.04
Led	3200	28.1	30.53	25.2
Lymph	148	26.08	25.57	23.1
Pima	768	24.44	25.42	24.44
Tic-tac	958	1.02	1.04	0.18
Wine	178	4.8	5.04	4.1

TABLE VII. THE RUNTIME FOR PREDICTION IN SECONDS

Data set	MAC	Our Prediction
Cleve	0.06	0.17
Breast	0.25	0.38
Glass	0.03	0.16
Iris	0.02	0.09
Pima	0.08	0.19
Tic-Tac	0.14	0.22
Led	0.19	0.38
Heart	0.015	0.12

V. CONCLUSIONS AND FUTURE WORKS

Predicting test data in AC is an interesting research problem that requires careful consideration due to the fact that more than one rule could be similar to the test data and that makes the prediction decision a hard task. This paper presented a prediction method based on two main criteria:

- The class representation in the context of the numbers of rules
- The rules rank

The outcome is a novel method which considers all rules that are similar to the test data during the classification step and computes the class strength per class assigning the class that has the largest strength. The class strength is based on the rules ranking position as well as the number of rules per class. Experimentations using 12 data sets from the UCI data repository and two common AC prediction methods have been conducted to measure the success and failure of our method. The results with respect to one-error rate reveal that the new prediction method has enhanced the predictive power of the resulting classifiers and on most data sets we used. In the near future, we would like to use our prediction method on unstructured textual data in the domain of text mining.

REFERENCES

[1] Liu B., Hsu W., and Ma Y. “Integrating classification and association rule mining”. Proceedings of the KDD. New York, NY 1998, pp. 80-86.

[2] Jabbar M. A., Deekshatulu B. L., and Chandra P. “Knowledge Discovery Using Associative Classification for Heart Disease Prediction”. Advances in Intelligent Systems and Computing, Volume 182, 2013, pp. 29-39.

[3] HooshSadat M. and Zaiane O. “An Associative Classifiers for Uncertain Datasets”. Proceedings of the 16th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKFF 2012), Malaysia, 2012, pp. 342-353.

[4] Abdelhamid N., Ayesh A., Thabtah F., Ahmadi S., and Hadi W. “MAC: A multiclass associative classification algorithm”. Journal of Information and Knowledge Management, Volume 11, issue 2, 2012.

[5] Merz C. and Murphy P. “UCI repository of machine learning databases”. Irvine, CA, University of California, Department of Information and Computer Science, 1996.

[6] Frank, E. and Witten, I. “Generating accurate rule sets without global optimization”. In: Proceedings of the Fifteenth International Conference on Machine Learning, Morgan Kaufmann Madison, Wisconsin, pp. 144–151,1998.

[7] Choen W. and Singer Y. "Context-Sensitive Learning Methods for Text Categorization". ACM Transactions on Information Systems, Volume 17, No. 2, pp. 141–173, 1999.

[8] Quinlan, J. “C4.5: Programs for machine learning”. *San Mateo, CA: Morgan Kaufmann the KDD. New York, NY.*,1998 pp. 80-86.

[9] Thabtah F., Mahmood Q., McCluskey L., and Abdel-jaber H . “A new Classification based on Association Algorithm”. Journal of Information and Knowledge Management, World Scientific, Volume 9, No. 1, pp. 55-64, 2010.

[10] Wang X., Yue K., Niu W., and Shi Z. “An approach for adaptive associative classification”. Expert Systems with Applications: An International Journal, Volume 38 Issue 9, pp. 11873-11883, 2011.

[11] Yin, X. and Han, J. “CPAR: Classification based on predictive association rule”. Proceedings of the –the SIAM International Conference on Data Mining -SDM, pp. 369-376, 2003.

[12] Thabtah F., Hadi W., Abdelhamid N., and Issa A. “ Prediction Phase in Associative Classification”. Journal of Knowledge Engineering and Software Engineering. WorldScinet ,Volume: 21, Issue: 6, pp. 855-876, 2011.

[13] Abdelhamid N., Thabtah F., and Ayesh A. “ Phishing detection based associative classification data mining”. *Expert systems with Applications Journal*. 41 pp.5948–5959, 2014.

[14] Thabtah, F., Cowling, P., and Peng, Y. “MCAR: Multi-class classification based on association rule approach”. Proceeding of the 3rd IEEE International Conference on Computer Systems and Applications. Cairo, Egypt 2005, pp. 1-7.

[15] Li, W., Han, J., and Pei J. “CMAR: Accurate and efficient classification based on multiple-class association rule”. Proceedings of the IEEE International Conference on Data Mining –ICDM, pp. 369-376, 2001.

[16] Veloso A., Meira W., Gonçalves M., and Zaki. M . “Multi-label Lazy Associative Classification”. Proceedings of the Principles of Data Mining and Knowledge Discovery - PKDD, pp. 605-612, 2007.