Expert Systems with Applications 37 (2010) 127–133

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Least squares support vector machines ensemble models for credit scoring

Ligang Zhou [a,*], Kin Keung Lai [a], Lean Yu [b]

[a] Department of Management Sciences, City University of Hong Kong, Hong Kong
[b] Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China

ARTICLE INFO

Keywords:
Credit scoring
Support vector machines
Ensemble model

ABSTRACT

Due to recent financial crisis and regulatory concerns of Basel II, credit risk assessment is becoming one of the most important topics in the field of financial risk management. Quantitative credit scoring models are widely used tools for credit risk assessment in financial institutions. Although single support vector machines (SVM) have been demonstrated with good performance in classification, a single classifier with a fixed group of training samples and parameters setting may have some kind of inductive bias. One effective way to reduce the bias is ensemble model. In this study, several ensemble models based on least squares support vector machines (LSSVM) are brought forward for credit scoring. The models are tested on two real world datasets and the results show that ensemble strategies can help to improve the performance in some degree and are effective for building credit scoring models.

1. Introduction

Credit risk assessment has become an increasingly important area for financial institutions. In recent financial crisis, many financial institutions suffered heavy loss from a steady increase of customers' defaults on loans. In USA, general credit cards' issuers wrote off $27.19 billion of debts as losses in 1997 and this figure had increased to $31.91 billion in 2006 (HSN Consultants Inc., 2007). In addition, the recent subprime mortgage crisis in USA has caused some companies to loss billions of dollars due to customers' default. However, the financial institutions can not refuse all customers from the growing credit market solely to avoid credit risk. Therefore effective credit risk assessment has become a crucial factor for gaining competitive advantages in credit market which can help financial institutions to grant credit to credit worthy customers and reject non-creditworthy customers thus reduce loss.

There is a clear need for accurate decision support for making the credit granting decision, since an improvement in accuracy of even a fraction of a percent can translate into significant future savings for the credit industry (Thomas, Edelman, & Crook, 2002). Credit scoring is the most widely used techniques that help lenders to make credit granting decisions. Credit scoring's main idea is to estimate the probability of applicant's default in terms of the characteristics recorded in the application form or credit bureau, by a quantitative model that is built on basis of information of past applicants. Most quantitative methods from different discipline have been used for building credit scoring models, such as linear discriminant analysis, logistic regression, decision tree, bayes network from statistics; linear programming; artificial neural network, support vector machines from artificial intelligence and some other hybrid approaches.

Support vector machines method was originally proposed by Vapnik and has been successfully applied to a number of real-world problems, such as handwriting digital recognition, text categorization and speaker identification and so on. In most of these applications, SVM's generalization performance either matches or is significantly better than that of other competing methods (Burges, 1998). For credit scoring problem, Baesens (2003) studied the performance of various state-of-the-art classification algorithms on eight real-life credit scoring datasets. Among the 17 methods tested, in terms of classification accuracy (measured by percentage correctly classified, PCC), average ranking of SVM is the highest (Thomas, Oliver, & Hand, 2005). To improve the performance of PCC, Huang, Chen, and Wang (2007) constructed a hybrid SVM-based credit scoring models with some heuristic searching methods for selection of input features and model parameters. The experimental results showed that SVM classifiers, with relatively few input features, can still achieve classification accuracy as good as neural networks, genetic programming and decision tree classifiers.

Although single SVM models have good performance in classification, they are sensitive to samples and parameters setting; generally, a single SVM classifier with a fixed group of training samples and fixed parameters setting may have some kind of inductive biases. One effective way to reduce the bias is ensemble model. Ensemble models can effectively make use of diversity to reduce the variance-error and are believed to be able to produce classification performance that is better than a single classifier (Breiman,

* Corresponding author. Tel.: +852 2788 8374.
E-mail address: mszhoulg@cityu.edu.hk (L. Zhou).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.05.024

1996, 1999). The main idea of the ensemble method is to combine a set of models, each of which solves the same original problem, in order to obtain a better composite model with more accurate and reliable estimates or decisions than what can be obtained from a single model (Maimon & Rokach, 2005). Fig. 1 shows the mechanism of the ensemble method.

This paper proposes several new SVM ensemble models that aggregate least squares SVM (LSSVM) classifiers of individual members by different measures. These ensemble models can be categorized into two groups. One group is based on the reliability of the decision made by each classifier, and the other group is based on different weight assignment strategies.

Reliability-based ensemble strategies are different from the "one-member-one-vote" commonly used in ensemble method. The modification comes from the intuition that "one-member-one-vote" ensemble strategy does not make full use of the information from each classifier member. For example, an ensemble model has three SVM member classifiers with classification hyperplanes $\mathrm{H_a}$, $\mathrm{H_b}$ and $\mathrm{H_c}$, as shown in Fig. 2. For an unknown observation with features vector as at point *A*, two ensemble members $(\mathrm{H_a}, \mathrm{H_b})$ are neutral to its prediction but have a slight preference to classify it as negative class, while the third member $\mathrm{H_c}$ strongly supports its classification as positive class. Then what should be the reasonable classification results of the ensemble? In other words, decisions values of *A*, obtained from two member SVM classifiers, are negative but are very close to the classification hyperplane, while value from the third is positive, but is beyond the optimal hyperplane for positive class. Obviously, the final decision from the "one-member-one-vote" ensemble is to classify *A* as negative, since two members are for negative class and only one is for positive class, under the rule that the minority is subordinate to the majority. One intuitive method is to assign different weights to different classifiers, instead of "one-member-one-vote". The weights can be determined by the overall performance of the classifiers, and are set to be a constant after training, or are set to be a function of the performance of the classifier on the sample.

Ensembles of SVM classifiers have also been extensively studied. Kim, Pang, and Je (2003) proposed a SVM ensemble with bagging or boosting and found that it greatly outperforms a single SVM in terms of classification accuracy. Sun and Huang (2004) proposed a least square SVM ensemble model and found it more robust on noisy data. Other analyses of SVM ensemble models include (Coelho, Lima, & Von Zuben, 2003, & Ma et al., 2004).

To achieve high classification performance, there are two essential requirements to be met by ensemble members and the ensemble strategy. First, ensemble members should have some sort of diversity, i.e. classifiers must show different classification properties. In an extreme case, if all classifiers make the same decision on all the test samples, the ensemble model with such classifiers will be of no help in improving the decision making. Diversity of ensemble members can be ensured by using different training samples or sampling methods, or by using different classifier parameters for different types of classifiers. Secondly, a good ensemble strategy is also required, on a set of complementary classifiers, in order to obtain high classification performance.

The rest of this paper is organized as follows. The framework of LSSVM ensemble model and some ensemble strategies are introduced in Section 2. To verify the effectiveness of the proposed methods, empirical analysis of the models is reported in Section 3. In Section 4, a short conclusion and discussion are presented.

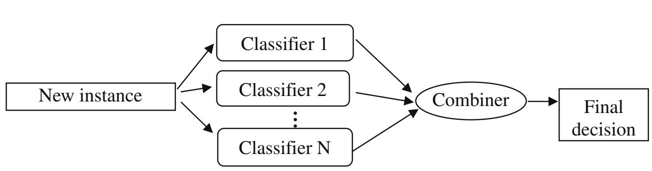

Fig. 1. The framework of ensemble model.

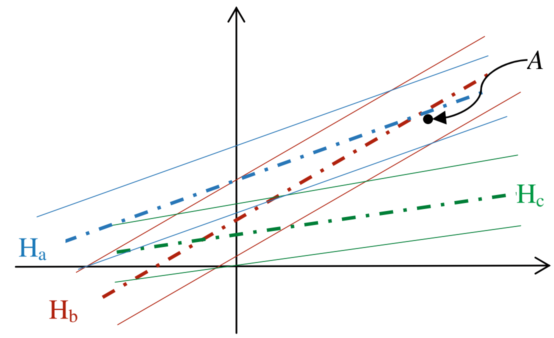

Fig. 2. Example of decision by an ensemble model.

2. Least squares SVM ensemble models

2.1. LSSVM ensemble framework

In this section, a LSSVM ensemble learning framework is proposed for credit risk evaluation. The basic idea of LSSVM ensemble model is originated from using all the valuable information hidden in LSSVM classifiers, where each can contribute to improvement of generalization. In our proposed multistage SVM ensemble learning model, to increase the diversity of the classifier, a sampling approach is first used to generate different training sets for each classifier. In terms of different training datasets, multiple individual classifiers are trained. Accordingly, classification results and performance values of each SVM classifier, on validation datasets, are obtained. Then a special measure is used to select appropriate ensemble members from multiple trained SVM classifiers. Finally, ensemble members are aggregated in terms of some criteria, and their ensemble results are obtained. The final result is called the ensemble output. The general architecture of the multistage reliability-based SVM ensemble learning model is illustrated in Fig. 3.

Usually, credit institutions may have millions of customers and each customer has at least one record in the dataset. Thus the sample size for all customers is huge. In practice, it is not possible to use all the existing samples to build the model since it may consume unacceptably high computational time. Therefore, it is necessary to select samples from the original dataset to construct the training, validation and testing subsets, for building the models. Given that the size of the original dataset *DS* is *N*, sizes of new training data *TR*, validation dataset *VS* and testing dataset *TS* are N_r, N_v, and N_t, respectively. The number of new training data subsets is *n*, which is the number of classifiers. In this study, to increase the diversity of classifiers in the ensemble model, we randomly select a fixed proportion of samples in *TR* for each training subset. The drawback of this sampling method is that some samples will be selected several times and some will never be selected.

According to the definition of effective ensemble classifiers by Hansen and Salamon (1990), 'a necessary and sufficient condition for an ensemble of classifiers to be more accurate than any of its individual members is if the classifiers are accurate and diverse.' Generally, an effective ensemble classifier consisting of diverse models with much diversity is more likely to have a good generalization performance in terms of the principle of bias-variance trade-off (Yu, Lai, Wang, & Huang, 2006). Therefore, how to generate diverse models is a crucial factor. Several methods have been investigated for generation of different ensemble members making