Multi-Classifier Combination for Banks Credit Risk Assessment

*Qifeng ZHOU[1], Chengde LIN[1], Wei YANG[2]

[1] Department of Automation,
Xiamen University,
Xiamen, China 361005.
[2] College of Life science&Biotechnology
Shanghai JiaoTong Unsiversity,
Shanghai,China 200030.
Email: zhouqf66@126.com

Abstract

Credit risk assessment problem belongs essentially to a classification problem. In this paper, a Multi-classifier Combination algorithm has been developed for banks credit risk assessment. We adopt Back-Propagation (BP) algorithm as the meta-learning algorithm and compared the methods of Bagging and Boosting to construct the Multi-classifier System (MCS). Experimental results on real client's data illustrate the effectiveness of the proposed method.

1 Introduction

Credit risk assessment is an important foundational work in banks credit risk management. The core problem is credit ranks, that is, to construct an available evaluating model and then classify companies according to their credit degree.

The currently wide-used credit risk assessment models are mainly Statistics Models and Neural Network Models. Statistics Models were presented after Fisher's research on heuristic knowledge [1]. Such as Z-score Model based on Multiple Discriminant Analysis (MDA), Linear Probability Model (LPM), Logit Model, Probit Model etc. Compared with traditional proportion analysis, Statistics Models are stronger in synthesis analysis and quantitative analysis. The weakness is they are drawn from asymptotic theory with large scale number of samples (infinite), but what we studied is small sample problem; their application are under some special hypothesis, in fact, our data can hardly satisfied these hypothesis.

In 1980's, some Artificial Intelligence methods such as Expert System (ES), Neural Network (NN), and Genetic Algorithm (GA) etc were introduced to these fields [2]. NN model overcomes the weakness of Statistics Model. But it is instable; perturbing the learning set can cause significant changes in their structure and accuracy. The instability affects the reliability of NN Model.

Focus on these problems, Multi-classifier Combination methods are introduced in this paper for banks credit risk assessment. We select BP algorithm as the meta-learning algorithm of MCS and use Bagging and Boosting methods to produce component classifier. The result of experiments shows that the performance of models based on Bagging is superior, and the Bagging technique can well be applied to credit risk assessment.

This paper is organized as following: in section 2, the methods and theory about combing multi-classifier are introduced; in section 3, data pre-processing and features select for credit risk assessment are presented; in section 4, credit risk assessment model based on multi-classifier combination are constructed; finally, several modeling methods are compared and conclusion is drawn in section 5.

2 A multiple classifier system approach

An ensemble of classifiers is a collection of several classifiers whose individual decisions are combined in some way to classify the test examples [3]. It is known that an ensemble often gives a much better performance than the individual classifiers that compose it. Hansen et. al. [4] shows why the ensemble gives a better performance than individual classifiers as follows. Assume that there are an ensemble of n classifiers $\{f1, f2\ldots fn\}$ and consider a test data x. if all the classifiers are identical, they are wrong at the same data, where an ensemble will show the same performance as individual classifier. However, if the classifiers are different and their errors are uncorrelated, then when $fi(x)$ is wrong, most other classifiers, except for $fi(x)$, may be correct.

2.1 Combining rule of multi-classifier

There are two commonly combining rules for multi-classifier one is vote method another is average method. If the output of each classifier is label we usually use vote method, including simple majority vote and weighted vote; if the output is the probability or others we use average method, including simple average and weighted average.

2.1.1 Vote method

Let $fi(x)$ be the decision function of the Kth classifier and Ci (i=1,2,..m) denote the label of the Jth class, here

x is the input vector. The classifier result of simple majority vote is

$$h_f(x) = \arg\max_{l\in\{1,2,\dots,L\}} \sum_{i=1}^{m} 1_{\{f_i(x)=l\}} \tag{1}$$

Where instructor function $1_{\{f_i(x)=l\}}$ equals 1 or 0 depend on $f_i(x)=l$ is true or false. Given β_i be the weight of *Ci*, the classing result of weighted vote is

$$h_f(x) = \arg\max_{l\in\{1,2,\dots,L\}} \sum_{i=1}^{m} \beta_i * 1_{\{f_i(x)=l\}} \tag{2}$$

2.1.2 Average method

Let $f_i^k(x)$ be the output of classifier *Ci* (i=1, 2, m), here $f_i^k(x)$ denote the posterior probability of x belongs to k-class, k=1, 2…L, the output of MCS is

$$h_f^k(x) = \sum_{i=1}^{m} \omega_i f_i^k(x), k = 1,2\dots,L \tag{3}$$

Where w_i is the weight of classifier C_i with $\sum_{i=1}^{m} \omega_i = 1, \omega_i \geq 0$. When $a_i = 1/m$, the formula presented above becomes the output of simple average method. The class label of x is

$$\arg\max_{k\in\{1,2..L\}} h_f^k(x). \tag{4}$$

2.2 Methods for constructing the ensemble

Many methods for constructing an ensemble of classifiers have been developed. The most important consideration in construction the component classifier ensemble is that each individual component classifier becomes different from another as much as possible [5]. This requirement can be met by using different training sets for different component classifier. We focus on representative methods, such as bagging and boosting.

2.2.1 Bagging

Bagging is a powerful and well-studied method of finding a highly accurate hypothesis (classification rule) by combining many "weak" hypotheses generated by a base-learning algorithm, each of which is only moderately accurate [6].

In bagging, several component classifiers are trained independently via a boostrap method and then they are aggregated via an appropriate combination technique. Usually, we have a single training set TR= $\{x_i,y_i\}$|i=1,2…l}. We build K replicate training data sets {Trk|k=1, 2…K} by randomly re-sampling, but with replacement, from the given training data set TR repeatedly. Each example x_i in the given training set TR may appear repeated times or not at all in any particular replicate training data set. Each replicate training set will be used to train a certain component classifier.

2.2.2 Boosting

The representative Boosting algorithm is the AdaBoost algorithm [7]. Like Bagging, each NN is also trained using a different training set. However, the selection scheme of training sample in the AdaBoost method is quite different from the bagging method.

Boosting algorithms operate in several rounds. During each round they reweigh examples in the training set and rerun the base-learning algorithm on these reweighed examples. Effectively, boosting forces the weak learning algorithm to concentrate on the "hardest examples" (the ones misclassified so far). One can achieve a sufficiently high accuracy on the training examples by running a large number of rounds of boosting.

There are also some other techniques to construct the ensemble, according to the trait of credit risk assessment, we modeling by the two methods mentioned above and compared their performance in section 4.

3 Credit risk assessment based on classifier combination

3.1 Problem formulation

Credit risk assessment is a synthesis evaluates to a company, finance organization etc according to theirs credit-ability. Usually we use the companies finance states denote their credit-ability. The methods can be described simply as studying the history sample, finding the rule from the sample, constructing a classification model and applying it to predict new sample's credit rank. It belongs to a classification problem [8]. Let $\boldsymbol{x}_i \in \boldsymbol{R}^n$, $i = 1,\dots,l$, be the input vectors drawn from some unknown probability distribution $F(x)$; let $\boldsymbol{y}_i \in \boldsymbol{Y}$ be the output of the prediction process according to a unknown conditional distribution function $F(y\,|\,x)$. The training set, on which the selection of the best predictor would be made, consists of (x_i, y_i) independent and identically distributed observations drawn from the joint distribution $F(x,y) = F(x)F(y\,|\,x)$, the learning task is to select a prediction function $f(x)$ from a family of possible functions F that minimizes the expected loss over the training set weighted by the joint distribution $F(x,y)$.

In a standard classification problem the input vector are associated with one of k classes, thus $y_i \in Y = \{1,\dots,k\}$ belongs to a set of labels denoting the class membership. In credit risk assessment problem Y is a finite set such as {AAA, AA, A, B, C} which denotes the credit degree is "excellent" "very good",

"good" "common" and "bad". We use companies' financial standards to evaluate their credit risk.

3.2 Selection of meta-learning algorithm

One of the key factors affecting the evaluate accuracy is meta-learning algorithm of MCS. Here we choose BP algorithm as the meta-learning algorithm. The main reason as follows: What we studied is a high non-linear mapping from financial ratios to enterprise's credit degree, input vectors is drawn from some unknown probability distribution, and the construct of covariance is unequal, these are obstacle for some learning algorithm, but NN can give a preferable classifying accuracy. NN don't depend on the hypothesis of variable is linear correlation or independent. The weakness of easily running into local minimum of single NN can be used to enhance the difference of each classifier. NN had already has some successful application in business and financial fields, such as [9, 10, 11]. In order to avoid unsuccessful learning, the NN classifier with validates accuracy great than 0.6 will be selected.

3.3 Data pre-processing and feature select

It is worth considering that different industries have different financial standards in credit risk assessment. So we divide the companies into several patterns by their industry. According to common-rules and expert's suggestion, we assess a company's financial status from three aspects: solvent ability, payoff ability and management ability [12, 13].

Feature selection is often an essential data processing step prior to applying a learning algorithm. The removal of irrelevant and redundant information often improves the performance of machine learning algorithms. There are two common approaches: a wrapper uses the intended learning algorithm itself to evaluate the usefulness of features while a filter evaluates features according to heuristics based on general characteristics of the data. The wrapper approach is generally considered to produce better feature subsets but runs more slowly than a filter [14]. Since the meta-learning method is BP algorithm and the size of feature is not very large, so we choose wrapper approach to extract the suitable feature for every pattern (See Figure 1).

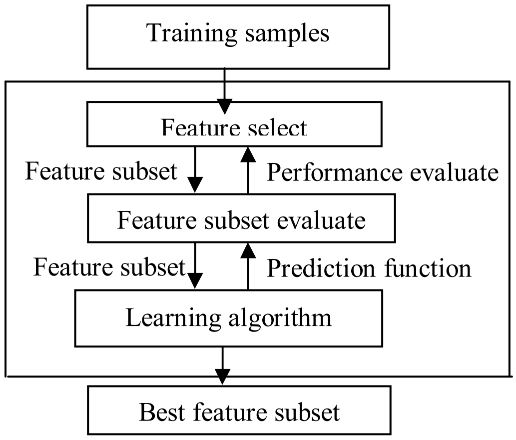

Figure 1 wrapper approach flow chart

4 Experiment result

Experiments were carried out on a subset of the database obtained from some business bank's client information. After pre-processed, we selected about 2460 companies, including 45.50% "AAA" companies, 45.09% "AA" companies, 8.43% "A" companies, and 0.97% "B" companies. Because the class "B" companies are few so we combined it with class "A" companies. Thus the total ranks are k=3, Y= {1,2,3}, 1, 2, 3 corresponding to A &B rank, AA rank and AAA rank respectively.

In this paper, we choose three levels BP algorithm as the meta-learning algorithm; the output level has one neuron. Since activate function of BP algorithm is Sigmoid function, its output between 0 and 1, so we have to change the credit ranks. Here we enactment the credit mark is 0.7, 0.3, 0.1, then the assessment result of BP algorithm classifier is

$$f(x)=\begin{cases} AAA & if \quad net(x)\geq 0.7 \\ AA & if \quad 0.3\leq net(x)<0.7 \\ A\,\&\,B & if \quad net(x)<0.3 \end{cases} \qquad (5)$$

To light industry and business industry patterns, the number of hidden neuron is 14, learning epochs is 10, to heave industry pattern the number of hidden neuron is 6, learning epochs is 8. We use AdaBoost and Bagging methods to construct assessment models respectively, the size of classifiers is 100, and use 5 folds cross validate evaluate the assessment model's accuracy. Table 1~3 show the average accuracy in every pattern.

Modeling algorithm	Best single NN	AdaBoost	Bagging
Training accuracy	88.74%	100%	87.93%
Testing accuracy	78.03%	80.39%	83.37%

Table 1 Light industry pattern assessment accuracy

Modeling algorithm	Best single NN	AdaBoost	Bagging
Training accuracy	86.54%%	100%	93.50%
Testing accuracy	77.49%	80.61%	82.18%

Table 2 Business industry pattern assessment accuracy

Modeling algorithm	Best single NN	AdaBoost	Bagging
Training accuracy	85.63%	100%	92.07%
Testing accuracy	77.08%	80.00%	80.75%

Table 3 Heavy industry pattern assessment accuracy

From table 1~3 we can see the evaluation accuracy of multi-classifier model (including Bagging and AdaBoost) is higher than the single NN. The performance of evaluate models based on Bagging is superior. In practice, we choose Bagging technique to construct assessment model. It is more stable than single NN and can avoid the weakness of over learning.

We also compared the stability and training time between Bagging and Boosting, result show that Bagging is more stable than Boosting and need less training time.

5 Conclusions

In this paper, we have proposed a multi-classifier approach based on Bagging, and experimented in several different patterns. The results showed that the MCS approach provides a better performance in generalization abilities and assessment accuracy than that provided by an individual classifier. Among the combined rules evaluated, the multi-classifier model based on Bagging technique provided the best results.

Credit risk assessment is important for banks. In this paper, we present a quantitative evaluate model, in practical application, we can combine it with qualitative analysis such as expert system to evaluate the client credit risk more carefully and provide banks a more strength assistant decision tool.

References

[1] R.A Fisher, "The Use of Multiple Measurement in Taxonomic problems," *Annals of Eugenics*, Vol. 7, pp. 179-188, 1936.

[2] E.I.Altman, "Financial ratios, discriminant analysis, and the predication of corporate bankruptcy [J]." *Journal of Finance*,Vol.23(9),pp.589-609,1968.

[3] T.G. Dietterich, "Machine learning research: Four current directions," *The AI Magazine*, Vol.18 (4), pp.97-136, 1998.

[4] P.S.L. Hansen, "Neural network ensembles," *IEEE transactions on Pattern Analysis and Machine Intelligence*, Vol.12, pp. 993-1001,1990.

[5] L.I.Kuncheva, "Relationships between combination methods and Measures of diversity in combing classifiers", *Information Fusion,* Vol.3, pp.135-148, 2002.

[6] L.Breiman, "Bagging Predictors," *Machine Learning [J]*, Vol.24(2),pp. 123-140,1996.

[7] R.E.Schapire, "the strength of weak learn ability," *Machine Learning [J]*, Vol.5(2),pp. 197-227,1990.

[8] W.X.Wei, "The Comprehensive Evaluation Method of the Credit Rating of Enterprise and Its Application," *System Engineering—Theory & Practice*, Vol.5 (2), pp26-31, 1998.

[9] G.Zhang,M.Hu,and B.Patuwo,"Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis, "*Europen J.Oper.Res.*,Vol.116,pp.6-32,1999

[10] B.A.Yang,"A Study of Commercial Bank Loans Risk Early Warning Based on BP Neural Network," *System Engineering—Theory &Practice*, Vol.9, pp.70-74, 2000.

[11] X.H.Cheng," Credit risk assessment of enterprise basing on neural network," *Journal of Systems Engineering*.Vol.17,pp.570-575,2002.

[12] Y.F.He;D.L.Shi,"Banks client credit assessment," Business Publishing House ,BeiJing, China 2002.

[13] "Enterprise Performance Interlocution," Statistic evaluate Dept, Ministry of finance, China. Economic Publishing House, Beijing, China.1999.

[14] R. Kohavi,G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, Vol.97, pp.273－324, 1997.