A Comparison Study of Credit Scoring Models

Defu Zhang[1,2], Hongyi Huang[1], Qingshan Chen[1], Yi Jiang[1]

[1]*Department of Computer Science, Xiamen University, 361005, China*
[2]*Longtop Group Post-doctoral Research Center, Xiamen, 361005, China*
dfzhang@xmu.edu.cn cn

Abstract

In this paper we consider a credit scoring problem. We compare three powerful credit scoring models: genetic programming (GP), backpropagation neural networks (BP) and support vector machines (SVM) when applied to this problem, then we give a combined model. The results show that the combined model produces good classification results.

1. Introduction

With the rapid growth of the credit industry and the need of competition, credit scoring models have been extensively used for the credit admission decision [1]. In the last two decades, many credit scoring models [2] have been developed for the credit scoring problem. Generally, there are two classes of credit scoring models: Traditional mining models based on the classical statistical methods [3] and the modern mining models based on artificial intelligence techniques. The traditional models have failed to meet the needs of financial market. The modern mining models [4-10], for example, neural networks, genetic programming, fuzzy rules, have been growing in popularity and have become a hot research. The popular credit scoring models include logistic regression, neural networks, rough sets, genetic programming, support vector machines. Logistic regression model is one of the most popular classification models. In contrast with other statistical regression models, it is suitable for a variety of distribution function and more appropriate for credit scoring problems. Logistic regression has good linear classification capability. However, it performs worse for nonlinear problems. Rough Set is a mathematical tool to deal with ambiguity, uncertainty of information. Compared with fuzzy sets, it does not need to make any assumptions to handle the data. A rough set is based on an inductive method, the advantage lies in its ability to provide decision support in order to understand the rules. Although this method has been successfully applied to credit scoring, its weakness is poor forecasting, for a new input, it is possible that it cannot find any rule to match the input. Genetic programming can be understandable and has been widely used in symbolic regression. However, GP sometimes costs much time and does not find any rule for new customer. Artificial neural networks are a simple and abstract simulation of the human brain to describe the characteristics of a certain system, they have many advantages. However, they are easy to trap into local minima and overfitting. It is quite difficult to understand the learning and decision-making processes due to their black box nature. SVM can often obtain global optimal solutions. However, it is very difficult to determine what knowledge is redundancy, which knowledge is more useful and has more important role. Especially, GP, BP and SVM are sensitive to the selection of parameters. So it is very necessary to study them.

2. The credit scoring problem

Given a certain amount of customers, each customer has certain attributes or characteristics. The credit scoring problem is how to conduct a credit scoring model to implement systematic analysis of these data, and mine the behavior pattern and credit characteristics, capture the relationship between historical information and future credit performance. The model is then used to predict the future credit performance of customers and new customers.

Specifically, credit scoring problem can be described as follows: Given a customer data set $S=\{(x_1,y_1),\cdots,(x_j,y_j),\cdots,(x_n,y_n)\}$. Each customer x_j contains m attributes: $x_{j1},x_{j2},\cdots,x_{jm}$, y_j denotes the type of customer, for example, good or bad. The task of the credit scoring problem is to construct a model f, for the new x, we can predict y. Namely, $y=f(x)$.

The credit scoring problem has been tackled in various models. This paper does not directly address the general credit scoring problem. Rather, it takes a simplified problem with two classifications and investigates three models to solve it. The three models based on artificial intelligence techniques have been applied since the beginning 90's. These include genetic programming (GP), backpropagation neural networks (BP), Support vector machines (SVM). The aim of solving a simplified version of the problem is to ascertain whether or not these credit

Third International Conference on Natural Computation (ICNC 2007)
0-7695-2875-9/07 $25.00 © 2007 IEEE

scoring models offer a sensible model for solving this type of problem. We know that these models have had success in this area but the three models have not been rigorously compared on the particular problem that we are interested in [4, 6].

If these modes can show a promising result on a simplified version of the problem, this would indicate that it is worthwhile investigating the same techniques for more difficult problems, for example, multiclassification problem.

3. Three credit scoring models

3.1 Neural network

Neural networks have a strong ability to deal with complicated problems by simulating the human brain, it can be used to simulate the non-linear relationship in complicated data. The feed-forward networks are the most widely used architecture because they offer good generalization abilities and are readily to implement. The network architecture used in the paper is consists of three layers of neurons connected by weights. The input of each neuron is the weighted sum of the network inputs, and the output of the neuron is a sigmoidal function value based on its inputs. Given a finite number of pattern pairs consisting of an input pattern x_j and a target output pattern y_j , this network is trained by supervised learning. Generally, the backpropagation algorithm, which is the most popular learning algorithm, is adopted to perform steepest descent on the total mean squared error (MSE)

$$MSE = \frac{1}{2}\sum_{t=1}^{n}(\widetilde{y}_j - y_j)^2 ,$$

where n is the total number of pattern pairs.

Given an initial weights and threshold, each input pattern passes this network and gets an output pattern. Then the error between the output pattern and the target pattern is determined by MSE , and adjustment to weights and thresholds. The process is repeated with each pattern pair assigned for training network until the train error is within a prescribed tolerance level.

More detailed descriptions of neural network, the reader can be referred to [5,11].

3.2 Genetic Programming

GP is based on the Darwinian principle of natural selection and evolution. GP makes use of the idea of survival of the fittest by progressively accepting better solutions to the problem. It is inspired by biological processes of inheritance, mutation, natural selection, and the genetic crossover that occurs when parents mate to produce offspring. GA differs from conventional non-linear optimization algorithms in that it searches by maintaining a population of solutions from which better solutions are created rather than making incremental changes to a single solution to the problem. For the credit scoring problem, GP started a certain number of randomly generated population (a group of randomly generated IF-THEN rules) Then make use of selection operator, mutation operator and cross operator to operate continuously evolution (after several generations of reproduction) until they found the best individual.

In GP, we will improve the functional processes, such as LISP function. Such a procedure can be expressed as a node labeled with a limited number of trees. Internal nodes have one or more parameters to the function, or predicate action. Leaf nodes include process constant, variable, or non-action function.

In order to obtain the classification rules efficiently, discretization of continuous attributes should be first employed. Then, the maximum GP-tree depth of six is enforced to ensure for obtaining a simple rule. We make use of crossover, copy and mutation operators.

The measurement of fitness is a rather nebulous subject since it is highly dependent on problem. After a GP model is built, we can use it to classify new credit customers. More detailed descriptions of GP for credit scoring problem, the reader can be referred to [7,10,13].

3.3 Support vector machines

SVM implemented the principle of structure risk minimization by constructing an optimal separating hyper plane. The hyper plane for the problem of separating two classifications is:

$$wx + b = 0 .$$

To find the optimal hyper plane, the norm of the vector w needs to be minimized, where the optimal non-separable hyperplane should satisfy

$$y_i\left[(w \cdot x) + b\right] - 1 + \varepsilon_i \geq 0, i = 1, 2, \cdots, n$$

where ε_i denotes the nonnegative slack variable. In terms of these introduced slack variables, the problem of finding the hyperplane that provides the minimum number of training errors (i.e., to keep the constraint violation as small as possible) has the following formal expression:

$$\Phi(w, \varepsilon) = \frac{(w \cdot w)}{2} + C\sum_{i=1}^{n}\varepsilon_i ,$$

where C denotes a penalty parameter on the training error.

For nonlinear case, we map the original space into a high dimension space by nonlinear mapping, in which an optimal hyperplane can be sought. The inner product functions are replaced by the kernel function, however the computation complexity will not increase. Using a dual problem, the quadratic programming problems can be rewritten as

Third International Conference on Natural Computation (ICNC 2007)
0-7695-2875-9/07 $25.00 © 2007 IEEE

$$w(a)=\sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i,j=1}^{n} a_i a_j y_i y_j k(a_i, a_j)$$,

With the decision function

$$f(x)=\mathrm{sgn}\left\{\sum_{i=1}^{n} a_i^* y_i k(x_i \cdot x_j)+b^*\right\}.$$

Radial basis function (RBF) is a common kernel function

$$k(x_i, x_j)=\exp\left(-\gamma\|x_i - x_j\|^2\right).$$

More detailed descriptions of SVM for credit scoring problem, the reader can be referred to [6,13,14].

Artificial neural networks, genetic programming and SVM models were implemented by C++ language for two classification problems. The test examples come from the two types of credit data [15]: Germany and Australia. For Germany credit data set, we generate 8 group data from G1 to G8, each group chooses randomly 70% of the data as a training set, the remainder 30% of the data is the test data set. Similarly, we obtain 8 group data from A1 to A8 for Australia credit data set. The computational results were reported in Table 1 and 2. The performance of the credit scoring models is determined by the classification accuracy of test data set, namely

4. Experimental Results

$$\text{Accuracy}=\frac{\text{The number of matched customers in the test data set}}{\text{The number of customers in the test data set}}\times\%$$

Table1. The classification accuracy of different models for German credit data set (%)

	G1	G2	G3	G4	G5	G6	G7	G8	average
BP	81.06	77.74	79.07	80.73	78.4	**80.4**	82.06	**78.73**	79.77
GP	80.73	78.74	78.4	81.06	78.4	79.07	82.06	77.74	79.53
SVM	81.06	78.07	77.74	81.73	77.07	77.07	80.39	78.4	78.94
CM	**81.72**	**79.73**	**80.73**	**82.06**	**78.73**	80.07	**82.39**	**78.73**	80.52
Best	**81.72**	**79.73**	**80.73**	**82.06**	**78.73**	**80.4**	**82.39**	**78.73**	80.56

Table2. The classification accuracy of different models for Australia credit data set(%)

	A1	A2	A3	A4	A5	A6	A7	A8	average
BP	89.42	**89.9**	89.42	86.53	91.83	88.94	90.34	87.92	89.29
GP	89.9	88.46	89.9	87.98	**92.30**	**89.9**	**90.82**	86.47	89.47
SVM	**89.42**	88.46	89.9	88.46	89.9	87.01	88.41	86.47	88.5
CM	**89.42**	89.42	**90.38**	87.98	91.83	89.42	**90.82**	**88.89**	89.77
Best	**89.42**	**89.9**	**90.38**	**88.46**	**92.30**	**89.9**	**90.82**	**88.89**	90.01

From table 1 and 2, we can observe that all the three models perform better. In detail, for different data groups, each model obtains the different classification accuracy. For BP, GP and SVM, Credit Data of Germany reached 79.77%, 79.53% and 78.94% respectively. 89.29%, 89.47% and 88.5% can be achieved for the Australian credit data respectively. Although BP and GP are better on the average than SVM, the classification accuracy of SVM is more stable, namely, each run has the same result for the same data set. SVM is relatively simple and fast because we have no detailed study on the selection of SVM parameter. The performance of three models are not too good for the German credit data, the reason may be that German credit data set has too much good customers, reaching 70%. In addition, it is well known that three models are sensitive to the selection of parameters such as C and γ in SVM, crossover and mutation probability in GP, learning rate and momentum factor in BP et al. In this paper, we implement the self-adaption selection of parameters for BP and GP.

In order to obtain the better classification accuracy, we construct a combined model (CM) by majority voting as follows: For one customer, if there are two or three models with same classification result *A*, then the customer is classified as *A*. Otherwise, the classification result of the customer is the same as that of the model with the highest accuracy.

From table 1 and 2, we can observe that the average accuracy of CM is 80.52% for German credit data set data, 89.77% for Australia credit data set. On the average, it outperforms BP, GP and SVM. Especially, CM improves the results of G1-G5, G7, G8, A1, A3, A7 and A8.

At last, we output the best results(Best) with highest accuracy (using bold black) among four models.

Third International Conference on Natural Computation (ICNC 2007)
0-7695-2875-9/07 $25.00 © 2007 IEEE

5. Conclusions

In this study, we first compare BP, GP and SVM for the credit scoring problem. Then we develop a combined model based on the three models. The experimental results have shown that the three models can obtain good classification results for the credit scoring problem. Especially, the combined model can obtain better results than BP, GP and SVM.

Acknowledgments

This work has been supported by academician start-up fund (Grant No. X01109) and 985 information technology fund (Grant No. 0000-X07204) in Xiamen University.

References

[1] Thomas, L. C. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. International Journal of Forecasting, 2000,16 (2), 149–172.

[2] Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society, 2003,54(6), 627–635.

[3] Henley, W. E. Statistical aspects of credit scoring. Dissertation,The Open University, Milton Keynes, UK.1995.

[4] V. Desai, J. Crook, G. Overstreet, A comparison of neural networks and linear scoring models in credit union environment, European Journal of Operations Management, 1996, 95 (1) 24–37.

[5] West, D. Neural network credit scoring models. Computers and Operations Research, 2000,27 (11–12), 1131–1152.

[6] Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., & Wu, S. Credit rating analysis with support vector machines and neural networks: a market comparative study. Decision Support Systems, 2004, 37(4),543–558.

[7] Ong, C.-S., Huang, J.-J., & Tzeng, G.-H. Building credit scoring models using genetic programming. Expert Systems with Applications,2005,29(1), 41–47.

[8] Yongqiao Wang, Shouyang Wang, and K. K. Lai.A New Fuzzy Support Vector Machine to Evaluate Credit Risk, IEEE TRANSACTIONS ON FUZZY SYSTEMS, 2005, 13(6).

[9] Hsieh, N.-C. Hybrid mining approach in the design of credit scoring models. Expert Systems with Applications, 2005, 28(4), 655–665.

[10] Jih-Jeng Huang, Gwo-Hshiung Tzeng,Chorng-Shyong Ong. Two-stage genetic programming (2SGP) for the credit scoring model.Applied Mathematics and Computation, 2006,174, 1039–1053

[11] Defu Zhang, Qingshan Jiang, and Xin Li. A Hybrid Mining Model Based on Neural Network and Kernel Smoothing Technique. Lecture Notes in Computer Science.2005, 3516: 801–805.

[12] Qing-Shan Chen, De-Fu Zhang, Li-Jun Wei and Huo-Wang Chen. A Modified Genetic Programming for Behavior Scoring Problem. 2007 IEEE Symposium on Computational Intelligence and Data Mining. To be printed.

[13] Defu Zhang, Qingshan Chen, Lijun Wei. Building Behavior Scoring Model Using Genetic Algorithm and Support Vector Machines. International Conference on Computational Science 2007. To be printed.

[14] Chang, C. C., & Lin, C. J. LIBSVM: a library for support vector machines. 2001. Available from http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[15] C. L. Blake and C. J. Merz. UCI repository of machine learning databases, 1998.