Extreme Learning Machine for Bank Clients Classification

Ganglong Duan
Xi'an University of Technology
Xi' an, China
E-mail: gl-duan@xaut.edu.cn

Zhiwen Huang
Xi'an University of Technology
Xi' an, China
huangzhiwen117@163.com

Jianren Wang,
Xi'an University of Technology
Xi' an, China
wjr@xaut.edu.cn

***Abstract*—In this paper, a classification mode for commercial bank clients' classification using the extreme learning machine (ELM) algorithm is proposed to study the commercial banks VIP loss. Firstly, we adopt the existing data sets of banks to train the ELM model; then, customer classification algorithm and its parameters are selected for classification purpose. Lastly, comparative analysis with existed methods are also compared, which showed that its advantages with the traditional gradient algorithm and other classification algorithm, which further indicate that ELM algorithm can not only overcome their drawbacks but also has faster learning rate, higher rate of accuracy, and better generalization.**

***Keyword* — ELM; Data Mining; Classification; Business Intelligence**

I. INTRODUCTION

After feedforward neural networks with randomly selected hidden layers were first proposed by Pao et al for multilayer perception neural networks[1],and related work on RBF networks was published earlier by Broom head and Lowe for RBF neural networks[2], a new learning algorithm called Extreme Learning Machine (ELM) was proposed recently by G.-B. Huang et al[3], unlike these traditional implementations, which randomly all the hidden nodes parameters of generalized Single-hidden Layer Feedforward Networks (SLFNs) and analytically determines the output weights of SLFNs. It is clear that the learning speed of feedforward neural networks is in general far slower than required and it has been a major bottleneck in their applications for past decades. However, all the parameters of ELMs can be analytically determined instead of being tuned. In theory, this algorithm tends to provide the good generalization performance at extremely fast learning speed.

Customer classification means that customers are divided into different groups according to the customer attributes. It is one data mining research topic and has widely and successful application in CRM. The Pareto Principle indicated that 80 percent of the profits of an enterprise are gotten from 20 percent of its customers. Thus, a bank with limited resources should identify the customers who generate the most profits by establishing a large data warehouse and analyzing comprehensively the massive data accumulated for a long time. Therefore, the customer base will be further divided through a variety of groups and then for different customer segmentation, different services strategy will be implemented. As the goal to provide service for "one-on-one" customer, in line with customer demand and the demand for psychology, banks are able to retain and maintain more quality customers so as to create more profits.

At present, the study of the commonly used data mining algorithms for bank customer classification are decision trees, neural networks, bayes, support vector machines, they all have some shortages like slow learning, or accuracy of the test, or pan-defects of poor results. What's more, these algorithms are extremely vulnerable to the error of local optimum. Therefore, more efficient and effective algorithms are strongly needed. In this paper, we adopt a new machine learning method, ELM, to establish mathematical models to address the problem of the loss classification of bank customers.

II. CLIENT CLASSIFICATION METHOD

Taking a commercial bank, the individual VIP (clients being the bank's largest customer base) as the classification background, whether the upcoming VIP customers will be the classified targets lost or not are predicted. Accordingly, the VIP customers will be classified as incoming-loss customers and stable customers.

In the Paper, the main idea to realize the classification of the VIP customer in the commercial bank is to use the existing commercial banks VIP customer attribute data, including their static information, data and access to their accounts within three months of continuous business data (as input), and the loss after three months (as a Category output). Then to design a model, take it to classify the loss of the VIP customer.

To combine with the classification of data mining methods, the loss of customers of the bank forecast VIP

978-0-7695-3876-1/09 $26.00 © 2009 IEEE
DOI 10.1109/ICIII.2009.277

classification of specific applications and The characteristics of Extreme Learning Machine (ELM) algorithm is designed to solve the loss of customers of banks VIP classification of the overall operation, specific operations related to the following Process: data preparation, the establishment of data warehouses, model building, training test model, the empirical analysis of the five steps for this research by the various steps of the research work can be summarized as shown in Figure I.

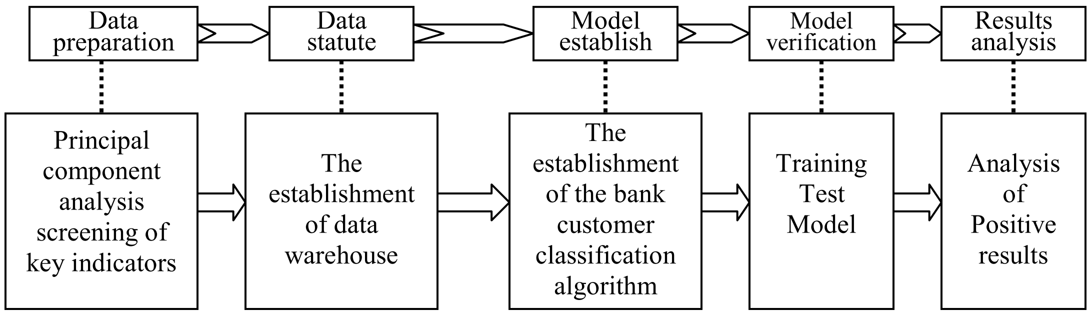

Figure I. Flow chart `for` method overall customer classification

III. ELM Customer Classification Model

A. Selecting indicators

First through research and interviews, as well as through the use of attribute-oriented induction (AOI) after the removal of some indicators, 15 more important indicators will be retained. And then we use principal component analysis from the 15 targets select some indicators which exert dominancy on the classification of the VIP customer.

According to principal component analysis (PCA) algorithm, the experiments were performed directly on MATLAB toolbox to treat Matrix a generated from the points (using five points system by experts) data of indicators for principal component analysis. By calling function princomp(A) in MATLAB, which can get the principal component matrix pc and m non-negative Eigen values of covariance matrix $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$, And calculate the contribution rate and the total contribution rate of every Eigen value, then, according to the size of contributions, select the first s $(1 \leq s \leq m)$ Eigen values, making its cumulative contribution rate $th_s \geq 90\%$, the s indicators corresponding to the s Eigen value is the major indicators filtered out, which is key indicators.

Running [pc, socre, latent, tsquare] = princomp (A) in MATLAB, in accordance with the return results of principal component analysis and the formula (contribution rate: $h_j = \lambda_j / \sum_{k=1}^{m} \lambda_k$, the total contribution rate:

$$th_j = \sum_{k=1}^{j} \lambda_k / \sum_{k=1}^{m} \lambda_k \quad (j = 1, 2, \cdots m)).$$

For calculating contribution rate and the total contribution rate to calculate the relevant data, the results are shown in the table I.

TABLE I. Eigen value, contribution rate and Cumulative contribution rate Eigen value of Covariance matrix

Indicators	Eigen value of covariance matrix	Contribution rate (%)	Cumulative contribution rate (%)
DAB	3.2128	33.77	33.77
LB	2.0688	21.74	55.51
FOW	1.1329	11.91	67.42
FOD	0.9044	9.51	76.93
DT	0.6582	6.92	83.85
CL	0.4760	5.00	88.85
CMI	0.4101	4.31	93.16
Withdrawals total	0.2413	2.54	95.70
Number of withdrawals	0.1917	2.01	97.71
Number of deposit	0.1095	1.15	98.86
Gender	0.0680	0.71	99.57
Vocational	0.0207	0.22	99.79
Age	0.0176	0.18	99.97
Qualifications	0.0022	0.03	100
Marital status	0	0	100

It can be showed from Table 4-1, the cumulative contribution rate of the first seven principal component indicators was 93.16 percent, which was higher than 90 percent, according to the main idea of the principal component analysis, in the 15 attributes, and the former 7 attributes are the most important indicator, that is, as the key variables of model. These 7 indicators are: daily average balance (DAB), last balance (LB), frequency of withdrawals (FOW), frequency of deposit (FOD), deposits total (DT) of an account in three consecutive month, and customer locations (CL) and customers monthly income (CMI).

B. ELM Model of Customer Classification

After selecting key indicators and establishing of data warehouse, ELM algorithm proposed recently was improved and used to design a mathematical model for a specific application of classification of VIP customer loss of banks. By studying the specific issues of customer classification of bank and the basic ideas and characteristics of ELM algorithm, the ELM model of bank's VIP customer loss of classified can be established as follow[3][9].

For N arbitrary distinct samples $(x_i, t_i) \in R^n \times R^m$, where $x_i \in R^n$ is inputs (x_i is a vector that includes 7 components, as that each x_i is a group of values of the 7 indicators selected before), $t_i \in R^m$ is target outputs (t_i is a numerical, '1' a loss,'0' indicates not loss). The SLFNs with L hidden neurons and activation $g(x)$ can approximate these N samples with zero error means that there are β_i, a_i, b_i to be modeled as

$$f_l(x) = \sum_{i=1}^{l} \beta_i G(a_{i,} b_{i,} x) = t_i \tag{1}$$

Where $x \in R^n, a_i \in R^n, \beta_i \in R^m$ $i=1,\dots,N$, in Formula (1), a_i is the weight vector connecting the ith hidden neurons and input neurons, b_i is the threshold of the ith hidden neurons; $\beta_i=[\beta_{i1}, \beta_{i2},\cdots, \beta_{im}]^T$ is the weight vector connecting the ith hidden neurons and output neurons; $a_i \cdot x$ denotes the inner product of a_i and x_i .There are activation function $g(x)$ to make $G(a_i, b_i, x) = g(a_i \cdot x + b_i), b \in R.$

The above N equations can be written as

$$H\beta = T \tag{2}$$

where

$H(a_1,\dots,a_N, b_1,\dots,b_N, x_1,\dots,x_N,)=$

$$\begin{bmatrix} G(a_1, b_1, x_1) & \dots & G(a_{N^-}, b_{N^-}, x_1) \\ \\ G(a_1, b_1, x_N) & \dots & G(a_{N^-}, b_{N^-}, x_N) \end{bmatrix}_{N \times N^-};$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{N^-}^T \end{bmatrix}_{N^- \times m}; \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}. \tag{3}$$

in Formula (3), H is called the hidden layer output matrix of the neural network, the ith column of H is the ith hidden node output with respect to inputs $x_1, x_2, \dots, x_N$. Formula (3) can become a linear system; therefore, the output weight β can get from Formula (4):

$$\beta = H^{-1}T. \tag{4}$$

in Formula (4), H^1 can be generated by Moore-Penrose inverse of the hidden layer output matrix H.

In the model, the parameters such as a_i, b_i, β, as can be generated through training the exited data sets repeatedly, which were accorded to calculate the output corresponded to the input data, and then, compared to the actual output to generate training and testing accuracy.

Two issues should be much noted in the model. First, the number of nodes of the hidden layer can be set arbitrarily, but requiring it is more than the number of input variables. Second, there are many activation function type to choose, and the most often used activation function are S-curve function (Sig), sine curve function (Sin), sub-function (Hardlim), and so on[3][7][8].

IV. SIMULATION EXPERIMENTAL

A. Model Verification

After establishing the ELM model of classification of bank customer loss, take tow sets of standard data which included complete input and output data and had been statute well from the warehouse. And one sets of data used for training, while the other used for testing.

In this paper, the simulation was done in the MATLAB environment, the loss of VIP classification model based on the ELM algorithm can be defined as a function:

Function[TrainingTime,TestingTime,TrainingAccuracy,TestingAccuracy,output] = VIPCC(train_data, test_data,Elm_Type,NumberofHiddenNeurons,ActivationFunction)

An example of simulation: enter orders statement in MATLAB command window as:

[TrainingTime,TestingTime,TrainingAccuracy,TestingAccuracy,output]=VIPCC(c_train,c_test,1,100,'sin')

The number of hidden layer nodes is set to 100, the activation function type is set sine curve, and the results of training and testing are: training time is 0.8590s, testing time is 0.3440s, training accuracy is 100%, testing accuracy is 92 %, and the output results of the groups of testing data sets are 0 and 1, which meet the requirements. Changing the number of hidden nodes and activation function type, train and test repeatedly the ELM model of the loss of bank customers classification, and compare with its results of training and testing, the results listed in the table II.

B. Experimental Analysis

The above simulation results show that the performance (include study rate and accuracy) of the loss of bank VIP classification ELM model is good and

TABLE II. The training and testing results of ELM model for different number of hidden nodes and activation function type

Number of hidden nodes \ Parameters		Training time (s)	Testing time (s)	Training accuracy	Testing accuracy
100	Sin	0.8590	0.3440	1	0.9200
	Sig	0.0150	0.0050	1	0.8800
	hardlim	0.6880	0.0180	0.96	0.9200
50	Sin	0.0310	0.0110	1	0.5200
	Sig	0.1600	0.0670	0.9400	0.7400
	hardlim	0.0150	0.0017	0.8800	0.8800
20	Sin	0.0140	0	0.6400	0.5600
	Sig	0.0940	0	0.7400	0.7600
	hardlim	0.0050	0	0.8400	0.8800

different hidden nodes number and different activation function are affected to the results of training and testing ELM model. At experiment in the papers, the choice of 100 nodes and hardlim as activation function can get the best results. Therefore, as long as choose the right number of hidden nodes and activation function type, it can make the testing accuracy of the ELM model is higher than 90 percent. In addition, the authors also used the other three data mining algorithm (decision trees, neural networks, Bayes) on the same training test data for the training and testing, and compared the training and testing results of them with it of ELM, the training and testing results of every algorithm listed in the table III.

TABLE III. Training and test results of ELM and other three algorithms

Algorithm \ Parameters	Training time (s)	Testing time (s)	Training accuracy	Testing accuracy
ELM	0.8590	0.3440	0.9530	0.9200
Decision Tree	3.6520	1.3450	0.8600	0.8300
Neural Networks	3.1231	1.1240	0.9460	0.8840
Bayes	2.9315	0.9134	0.9840	0.9000

V. CONCLUSION

Bank customer classification is a problem of application of data mining. In the paper, by studying the specific issues of customer classification of bank and the basic ideas and characteristics of ELM algorithm, the ELM model of bank's VIP customers' loss of classified has been designed and established. At the process of establishing model, principal component analysis was taken to selecting the key variables. And training and testing the model repeatedly in the MATLAB, while other classification algorithms were used to training and testing the same data, the simulation results are satisfactory, and by comparing with the simulation results of these algorithms, it can get the conclusions: the learning rate and accuracy of ELM algorithm must be higher than the other three algorithms, but also shows that the ELM algorithm has better generalization. The simulation results also showed that as long as choosing right number of hidden nodes and activation function, the testing accuracy of the model based ELM algorithm for the loss of bank's VIP classification may be higher than 90%, that is, it can meet the request of the practical application of the loss of bank's VIP classification.

REFERENCES

[1] Y.H. Pao, G. H. Park, and D. J. Sobajic, "Learning and generalization characteristics of random vector functional-link net, Neurocomputing, vol. 6, pp. 163180, 1994.

[2] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks, Complex Systems, no. 2, pp. 321-355, 1988.

[3] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neural computing, 2005

[4] Singh R (Singh,Rampal),Balasundaram S (Balasundaram, S.), Application of Extreme Learning Machine Method for Time Series Analysis, Proceedings of World Academy of Science Engineering and Technology ,26 ,(2007) 361-367.

[5] Nizar AH, Dong ZY, Wang Y,Power utility nontechnical loss analysis with extreme learning machine method, IEEE Transactions on power systems,23/3(2008)946-955.

[6] P.L. Bartlett, The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network, IEEE Trans. Inform. Theory 44 (2) (1998) 525–536.

[7] Kim JW, Shin HS, Lee YW, et al, Algorithm for classifying arrhythmia using extreme learning machine and principal component analysis, proceedings of annual international conference of the IEEE engineering in medicine and biology society,3257-3260,2007.

[8] M.-B. Li, G.-B.Huang, P.Saratchandran, N. Sundararajan, Fully complex extreme learning machine, Neural computing, 2005, to appear.

[9] Hai-Jun Rong, Yew-Soon Ong, Ah-Hwee Tan, Zexuan Zhu. A fast pruned-extreme learning machine for classification problem. Neurocomputing, Volume 72, Issues 1-3, December 2008, Pages 359-366