International Journal of Soft Computing 1 (1) : 30-35, 2006

Detecting Credit Card Fraud by Using Support Vector Machines and Neural Networks

[1]Rong-Chang Chen, [2]Luo Shu-Ting and [1]Li Shiue-Shiun
[1]Department of Logistics Engineering and Management
[2]Graduate Institute of Business and Management National Taichung Institute of Technology, No. 129, Sec. 3, Sanmin Road., Taichung, Taiwan 404, R.O.C.

Abstract: Conventionally, historical actual transaction data are used to set up a model for detecting credit card fraud. Instead of using traditional approaches, a new personalized approach has recently been presented to prevent fraud. The personalized approach proposes to prevent credit card fraud before initial use of a new card, even users without any real transaction data. Though this approach is promising, there are some problems waiting to be improved. A main issue of the personalized approach is how to predict accurately with only few training data, since it collects quasi-real transaction data by using an online questionnaire system and users are generally not willing to spend too much time to answer questionnaires. This study employs Support Vector Machines (SVM) and Artificial Neural Networks (ANN) to investigate the problem of fraud detection of credit cards. The type of ANN models we use in this study is the Back Propagation Networks (BPN). The performance of neural networks is compared with that from SVM. Experimental results from this study show that both BPN and SVM can offer good solutions. When the data number is small, SVM can have better prediction performance than BPN in predicting the future data. Besides, the average prediction accuracy reaches a maximum when the training data ratio arrives at 0.8.

Key words: Fraud detection, credit card fraud, questionnaire-responded transaction, SVM, neural network, back propagation networks

INTRODUCTION

Payments using credit cards have become more and more popular in recent years. However, the popular use of credit card is accompanied by a large number of deceptive transactions, which cost hundreds of millions of dollars each year. It is, therefore, very essential to use an effective method to solve this problem and reduce fraudulent losses.

In dealing with credit card fraud, typically, real transaction data are used to produce models for detecting a new case [1-4]. This approach is a good solution in some situations. However, the most common-see fraud for credit cards happens when a thief uses personal information to open a credit card account in one's name or a new card is stolen before the use of an applicant. Consequently, it is very important for users to protect themselves from frauds before using a new card.

Rather than detecting credit card fraud by past transaction data, Chen *et al.*,[5-6] proposed a novel approach to solve the fraud problem. They suggested building up a personalized Questionnaire-Responded Transaction (QRT) model based on personal data collected by an online questionnaire system. Since the unauthorized user's consumer behavior is usually different from that of the cardholder, the fraud can be avoided from the initial use of a credit card.

This study employs the powerful learning tools, Support Vector Machines (SVM) and Artificial Neural Networks (ANN), to deal with the fraud problem. We first collect the Questionnaire-Responded Transaction (QRT) data of users by using an online questionnaire and then divide them into different kinds of consumer types according to their consumption amounts, time and items. The data are then normalized and trained by the SVM and ANN and accordingly, consumer behavior models based on individuals are constructed. These models are then used to detect new transactions. When a new transaction is going, the model can predict whether the transaction is abnormal or not. If the prediction result is to be abnormal, the transaction is considered as a fraudulent behavior.

Support Vector Machines (SVM) and Neural Networks

Support Vector Machines: Support Vector Machines (SVM) were developed by Vapnik [7]. They can be used to do regression and classification. SVM is a method for creating a classification or a general regression function from training datasets. For classification, SVM can find a hyper-surface, which will split the negative examples from the positive examples, in the space of possible inputs. The hyper-surface will be chosen in such a way that it has the largest distance from the hyper-surface to the nearest of

Corresponding Author: Rong-Chang Chen, Department of Logistics Engineering and Management, No. 129, Sec. 3, Sanmin Road., Taichung, Taiwan 404, R.O.C.

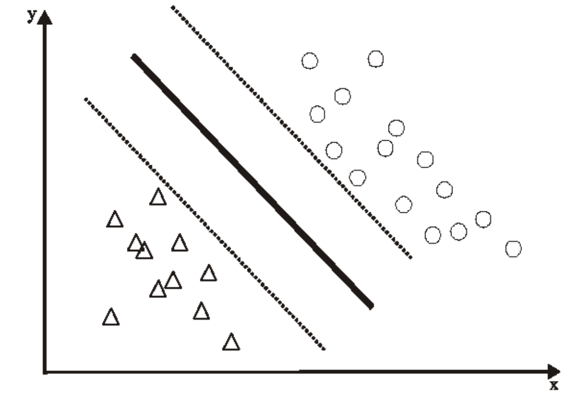

Fig. 1: Linear classifications

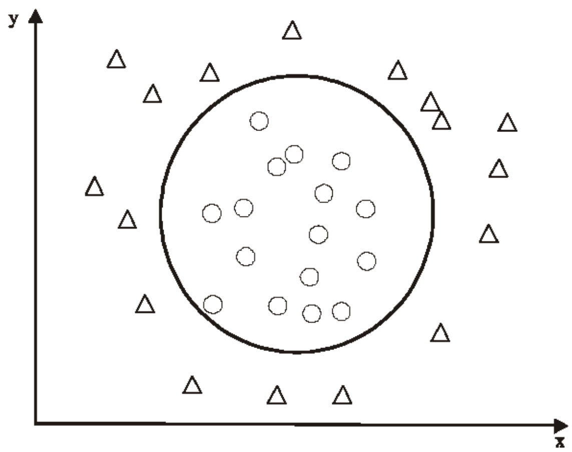

Fig. 2: Non-linear classification

the positive and negative examples [8]. This makes the classification correct for testing data that is near, but not equal to the training data. The SVM has already been successfully used for a wide variety of problems, like pattern recognition, bio-informatics, natural language learning text mining and more [9-10].

SVM can divide the data into two classes by linear or nonlinear classifiers, as shown in Fig. 1-2.

Artificial Neural Networks: Artificial Neural Networks (ANN) have become a popular tool for financial decision-making [11-12]. It is an information-processing pattern that is stimulated by the biological nervous systems like the brain[13]. The main element of this pattern is the architecture of the information-processing system. It contains a large number of highly interconnected processing elements (neurons) working in harmony to solve specific problems. ANN, like people, learns by examples. The ANN is constructed for a specific application, such as fraud detection, classification of microarrays, or pattern recognition, through a learning

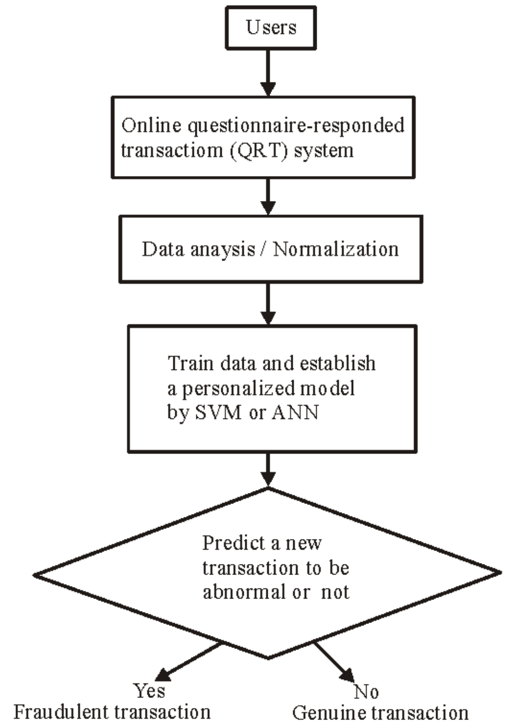

Fig.3: The procedure for predicting fraud

process. Learning in biological systems involves adaptation to the synaptic connections that exist between the neurons. This is also true for ANN.

Neural networks are an incremental mining technique that allows new data be submitted to a trained neural network to renew the previous training result. Therefore, it is suitable to use ANN to deal with the detection of credit card fraud. There are many types of ANN models. Amongst them, Back Propagation Networks (BPN) is one of the most popular models currently since it is easy to understand and can be easily implemented as a software simulation. In this paper, we will apply BPN to detection of a new transaction of a credit card.

Approach: Typically, real transaction data are utilized to establish prediction models for new transactions. This approach can differentiate fraud well in a few situations. However, for a new card user, there are no or few transaction data. Therefore, a model based on other users' transaction data is generally set up to detect fraudulent transaction for new users. This method may cause lower prediction accuracy because the consumer behavior of an individual is different from that of others. One frequent solution for issuing banks is to set a maximum transaction quantity. If the amount of a new deal is higher than this value, a warning will be produced. The disadvantage of this method is that a thief can easily commit a fraud if he knows and purchases with a price lower than this value.

To solve the problems mentioned above, we propose to establish a model for a new user before their initial use of a credit card. The proposed method is illustrated in Fig 3. First, by using an online, self-completion questionnaire system, we can obtain the transaction data of new users. We identify this kind of transaction data as questionnaire-responded transaction (QRT) data. This method is very suitable for new users who only have few transactions or without any transactions. The data are then trained by the SVM and thus a personalized model is produced. Finally, these personalized models are used to classify a new transaction as a fake or a true transaction.

Data-Collection Method: Among many ways of gathering data, two of the most popular tools are interviews and questionnaires. Both methods have pros and cons. The biggest advantages of self-completion questionnaires over structural interviews are that they can save much time and money, as well as permit much larger samples to be gained. In particular, if Internet can convey the questionnaires, the benefits of self-completion questionnaires would be much larger.

Online Questionnaire System: A successful questionnaire depends on three elements [14]:

- The questions must be comprehensible and definite.
- The appearance of questions must be standardized. • A reliable, efficient and preferably, cost-effective way of coding the data for following analysis must be used. With the three suggestions mentioned above, an efficient and cost-saving online questionnaire system is produced. Before a good final questionnaire is set up, a pilot test was performed to ensure clear questions in our study.

The design platform of online questionnaire system is Windows 2000, IIS 5.0. The program we used is ASP. The database is SQL 2000.

The questions on the questionnaires are produced according to the individual's consuming preference [15-18]. Consumer behavior changes considerably with each individual. It is thus practical to classify their behavior according to several main attributes. The gathered personal QRT data are mainly composed of several parts: gender, age, transaction time, transaction amount and transaction items. The details are described as follows.

Transaction time: Each day can be separated into some periods. Generally, 4 or 6 periods are enough to characterize the consumer behavior. Therefore, we divide each day into four periods, 00:00 – 06-00, 06:01 – 12:00, 12:01 – 18:00 and 18:01 – 24:00. Note, however, that the transaction time should be divided into more periods if the consumer behavior depends considerably on the transaction time.

Transaction item: According to the studies of consumer behavior, each individual has its preferred consuming habit. Thus, in this study, we separate transaction items into six major classes: eating, wearing, housing, transportation, education and recreation. Each major class can be further divided into more detailed subclasses.

The number of data: The number of questions that users are agreeable to answer diverges with people. Some users want to answer more to have higher prediction accuracy and they can reduce the fraud risk. On contrary, some others tolerate a loose accuracy or are unwilling to spend too much time to answer. Consequently, different users have different amount of the QRT data. To investigate the effect of the number of QRT data on the accuracy of prediction, the number of training data is varied to see their influence on the prediction accuracy.

Effect of data distribution: As mentioned by several previous studies [2-4], the real transactions data are usually skewed, i.e., positive data are much more than negative data. This results in a trivial solution for predicting majority class and low accuracy for predicting minority class. Generally, there are two basic methods for dealing with the skewed problems [19]: over-sampling, which is to duplicate the data in the minority class and under-sampling, which is to throw away part of the data in the majority class. However, both of them have known deficiency. Over-sampling increases the training size and the training time. Under-sampling, on the other hand, may remove some of potentially useful data. Contrast to the above two methods, we can collect the QRT to improve skewed distribution of transaction data and to investigate their effect on the accuracy.

EXPERIMENTAL RESULTS

In this study, we used mySVM[20] and back-propagation network (BPN) to train all personalized QRT data. The PBN tool we used is the SmartNeuron developed by Professor Chang, Department of Logistics Engineering and Management (LEM), National Taichung Institute of Technology (NTIT) in Taichung, Taiwan. SmartNeuron was developed by Visual C^{++}.

To examine the effectiveness of the proposed method based on ANN and SVM, QRT data are collected and then

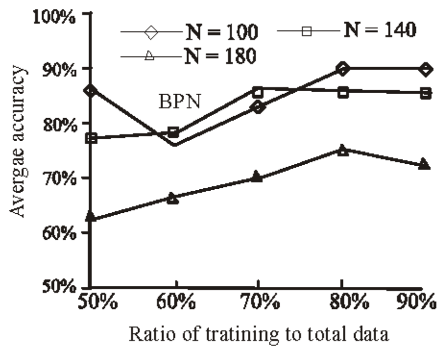

Fig 4: The influence of the ratio of training data to total data on average accuracy by BPN

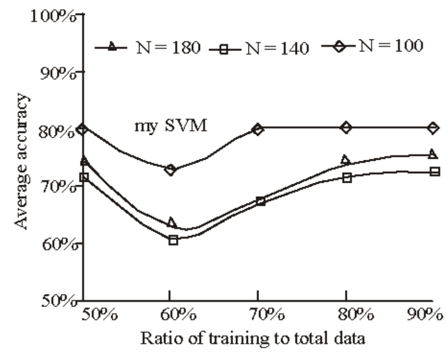

Fig 5: The influence of the ratio of tratining data to total data on average accuracy by SVM

personalized models are built up for a user. To get representative data for a good modeling of the reality, we let users choose the priority of six main classes of transaction items, which can collect different data quantity according to pre-specified ratios.

For the convenience of discussion, let us denote the number of the data as N, the number of the training data as N_{train}, the number of the test data as N_{test}, the number ratio of negative to total samples as Rn, the number ratio of training to total samples as Rt, the true negative rate as TN and the average prediction accuracy as AVG.

Back Propagation Networks: Using BPN to train data needs to set some parameters. Among the most important parameters are the numbers of hidden layer N_h, hidden nodes N_n and training epochs N_{ep}, learning rate R_l and momentum rate R_m. The setting of the parameter values remains as an art rather than a science. Complex problems can be incrementally better modeled by increasing hidden layers, but the improvement is generally accompanied by an associate cost in terms of training time and data overfitting. To improve the above problems, we selectively evaluated some parameter values in preliminary training, which decided N_h. N_n and N_{ep}, based on recommendations from previous literature and our past experience. Besides, the effect of the ratio of the number of training to total data R is investigated since the QRT data are finite.

In this study, we use the gradient descent method to minimize the total squared error of the output. According to the results from the preliminary training, we decided to use 2 hidden layers, 3 nodes in hidden layer 1, 6 or 7 nodes in hidden layer 2 in the following experiments.

Support Vector Machines: For SVM, the training and testing of QRT data were running on a Pentium III 667 PC and Windows 2000 Professional operating system. In this study, we used mySVM to train and test data.

To get optimal parameters for good results, first-round training was performed. Three types of kernel were investigated: dot, polynomial and radial. The degree of polynomial kernel was varied from 1 to 5. The preliminary results show that a radial kernel can obtain a better testing accuracy. Therefore, the base classifier we use is the radial kernel for further studies.

Influence of Training Data Ratio: When sample sizes are small, experimental results may be biased and cannot be generalized to the future. Therefore, the effect of sample sizes should be taken into consideration when sample sizes are small. Since the data number is finite in the personalized approach, the ratio of training data to total (training plus testing) data will influence the prediction accuracy. To investigate this effect, the train data ratio Rt is varied from 0.5 to 0.9, where 0.5 denotes an equal data number of training vs. testing. Three different numbers of data are studied: N = 100, N = 140 and N = 180. Rn = 0.5, indicating a balanced data distribution of positive vs. negative samples. The results in Fig. 4 are obtained by using SmartNeuron. As we can see from Fig. 4, all three cases have maximum values at the ratio of 0.8. Therefore, we will use a number ratio of training data to total data with 0.8 as a base case in the following experiments. One interesting finding is that the average accuracy increases with a decrease in data number, as we can see from Fig. 4.

The influence of the ratio of training data to total data on average accuracy by using SVM is shown in Fig. 5. Rn = 0.5. Rt ranges from 0.5 to 0.9. As Rt approaches 0.8, average accuracy AVG reaches a highest value, showing the same trend as in Fig. 5. Comparison of results between BPN and SVM shows that BPN has better average accuracy AVG for N = 100 and N = 140. Therefore,

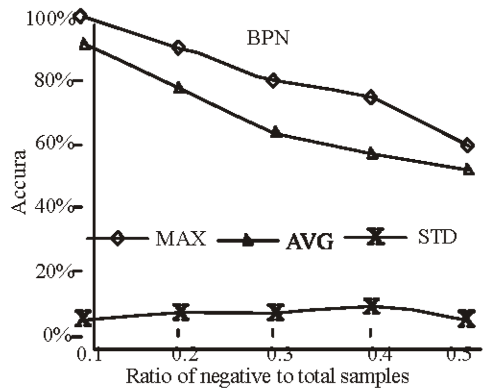

Fig 6: The effect of the ratio of abnormal sample on the accuracy

if the data number is small, saying 100 to 140, BPN can obtain better solutions.

Ratio effect of abnormal samples to normal samples: To see the effect of the ratio of abnormal samples to total samples Rn on the prediction accuracy, Rn is varied from 0.1 to 0.5. Rn = 0.5 stands for that the number of the abnormal data and the number of the normal data are equal.

With an increase in Rn, the accuracy reduces, indicating that a balanced data distribution of positive and negative samples has a worse accuracy. Both the average accuracy and maximum accuracy have the same trend, as illustrated in Fig. 6, where MAX indicates the maximum value, AVG the average accuracy and STD the standard deviation. Each result was obtained by running 10 times.

Time Effect: Consumer behavior changes with time. It is thus of great importance to investigate the time effect. Two students were asked to answer the online questionnaire weekly. The data in the first week was trained and used to predict the data in the second week. The comparison of results shows that both SVM and BPN can offer good solutions to the credit card fraud problem. However, special notes must be taken that the prediction accuracy depends on the data distribution of users. As we can find from Table 1, the performance of prediction of student A is better than that of student B.

Table 1: Comparison of predicted results between SVM and BPN

	BPN		SVM	
	AVG	TN	AVG	TN
Student_A	0.848	0.914	0.971	0.967
Student_B	0.727	0.705	0.751	0.803

CONCLUSIONS

Payment by credit cards is increasing and recently it is ousting cash as the most important payment method in some countries. Despite this advantageous trend, one of the well-known major problems for transactions using credit cards is the fraud, which causes a enormous amount of losses for issuing banks and individuals. To protect consumers and issuing banks from losses, it is, therefore, of great importance to find a solution to credit card fraud.

This research has employed a personalized approach to solve the credit card fraud problem. The approach is to prevent fraud from users' first use of their cards. Unlike the traditional way, we build up a model before use of new cards. First, we collect the Questionnaire-Responded Transaction (QRT) data of new users by using an online questionnaire system. Subsequently, the QRT data are trained by using the Support Vector Machines (SVM) and Back Propagation Networks (BPN) and the personalized models are set up.

Results from this investigation show that the proposed method can effectively detect the credit card fraud. Both SVM and BPN can have good solutions. When the data number is small, SVM can offer better performance than BPN. Moreover, the average prediction accuracy reaches a maximum when the ratio of training data to total data reaches 0.8.

REFERENCES

1. Maes, F.S., K. Tuyls, B. Vanschoenwinkel and B. Manderick, 2002. Credit card fraud detection using bayesian and neural networks, Proc. Neuro Fuzzy, Havana, Cuba.
2. Chan, P.K., W. Fan, A. L. Prodromidis and S.J. Stolfo, 1999. Distributed data mining in credit card fraud detection. IEEE Intelligence Systems, pp: 67-74.
3. Chan,P. and S. Stolfo, 1997. Toward Scalable Learning with Nonuniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection, Proc. andFourth Int'l Conf. Knowledge Discovery Data Mining. AAAI Press, Menlo Park, Calif, pp: 164-168.
4. Brause, R., T. Langsdorf and M. Hepp, 1999. Neural data mining for credit card fraud detection. IEEE Int'l Conf. Tools with Artificial Intelligence.
5. Chen, R.C., T.S. Chen, Y.E. Chien and Y.R. Yang, 2005. Novel Questionnaire-Responded Transaction Approach with SVM for Credit Card Fraud Detection, Lecture Notes in Computer Science (LNCS), pp: 916-921.

6. Chen, R.C., M.L. Chiu, Y.L. Huang and L.T. Chen, 2004. Detecting credit card fraud by using questionnaire-responded transaction model based on support vector machines and Lecture Notes in Computer Science (LNCS), pp: 800-806.
7. Vapnik, V.N. 1995. The nature of statistical learning theory, Springer.
8. http://www.kernel-machines.org/
9. Burges,C.J.C., 1998. A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery,pp: 2, 955-974.
10. Chen, R.C.J., Chen, T.S. Chen, C.H. Hsieh, T.Y. Chen and K.Y. Wu, 2005. Building an intrusion detection system based on support vector machine and genetic algorithm, Lecture Notes in Computer Science (LNCS), pp: 409-414.
11. Lam, M. 2004. Neural Network Techniques for Financial Performance Prediction Integrating Fundamental and Technical Analysis, Decision Support Systems, pp: 567-581.
12. Cheng, W.B., W.McClain and C.Kelly, 1997. Artificial neural networks make their mark as a powerful tool for investors, Review of Business, pp: 4-9.
13. http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html
14. Sapsford, R.1999. Survey Research, SAGE Publications.
15. http://www.104pool.com/
16. http://www.sino21.com/
17. http://isurvey.com.tw/
18. http://b-times.com.tw/
19. Yan, R., Y. Liu, R. Jin and A. Hauptmann,2003. On predicting rare classes with ensembles in scene classification, IEEE Int'l Conf. Acoustic, Speech and Signal Processing.
20. Rüping, S. 2000. MySVM-Manual, Computer Science Department, AI Unit University of Dortmund.