Research Article *ISSN: 2319-507X*
Ranjit Kumar, IJPRET, 2012; Volume 1(4): 20-25 *IJPRET*

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

CREDIT CARD TRANSACTION FRAUD DETECTION METHOD BASED ON SUPPORT VECTOR MACHINE

IJPRET-QR CODE

RANJIT KUMAR, SANDEEP RAJ

PAPER-QR CODE

1. **Assistant Professor, Galgotias College of Engineering and Technology, Greater Noida.**

Accepted Date:

30/10/2012

Publish Date:

01/12/2012

Keywords

Credit Card

Fraud

Transaction

Corresponding Author

Mr. Ranjit Kumar

Assistant Professor, Galgotias College of Engineering and Technology, Greater Noida.

Abstract

Due to a rapid advancement in the electronic commerce technology, the use of credit cards has dramatically increased. As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of fraud associated with it are also rising. Credit-card-based purchases can be categorized into two types: 1) physical card and 2) virtual card. In a physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. To carry out fraudulent transactions in this kind of purchase, an attacker has to steal the credit card. If the cardholder does not realize the loss of card, it can lead to a substantial financial loss to the credit card company. In the second kind of purchase, only some important information about a card (card number, expiration date, secure code) is required to make the payment. Such purchases are normally done on the Internet or over the telephone. To commit fraud in these types of purchases, a fraudster simply needs to know the card details. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information. The only way to detect this kind of fraud is to analyze the spending patterns on every card and to figure out any inconsistency with respect to the "usual" spending patterns. Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce the rate of successful credit card frauds. Since humans tend to exhibit specific behaviorist profiles, every cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc. Deviation from such patterns is a potential threat to the system.

INTRODUCTION

The prediction of user behavior in financial systems can be used in many situations. Predicting client migration, marketing or public relations can save a lot of money and other resources[2]. One of the most interesting fields of prediction is the fraud of credit lines, especially credit card payments. For the high data traffic of 400,000 transactions per day, a reduction of 2.5% of fraud triggers a saving of one million dollars per year. Certainly, all transactions which deal with accounts of known misuse are not authorized. Nevertheless, there are transactions which are formally valid, but experienced people can tell that these transactions are probably misused, caused by stolen cards or fake merchants[6]. So, the task is to avoid a fraud by a credit card transaction *before* it is known as "illegal". With an increasing number of transactions people can no longer control all of them. As remedy, one may catch the experience of the experts and put it into an expert system. This traditional approach has the disadvantage that the expert's knowledge, even when it can be extracted explicitly, changes rapidly with new kinds of organized attacks and patterns of credit card fraud. In order to keep track with this, no predefined fraud models as in but automatic learning algorithms are needed[9].This paper deals with the problems specific to this special data mining application and tries to solve them by a combined probabilistic and SVM approach for a given data base of credit card transactions of the GZS.

SVM Background

Support Vector Machines (SVMs) have been developed from Statistical Learning Theory. The SVM are widely applied to fields such as character, handwriting digit and text recognition, and more recently to satellite image classification[5]. SVMs, like ANN and other nonparametric classifiers have a reputation for being robust. SVMs function by nonlinearly projecting the training data in the input space to a feature space of higher dimension by use of a kernel function. This results in a linearly separable dataset that can be separated by a linear classifier. This process enables the classification of datasets which are usually nonlinearly separable in the input space. The functions used to project the data from

input space to feature space are called kernels (or kernel machines), examples of which include polynomial, Gaussian (more commonly referred to as radial basis functions) and quadratic functions. Each function has unique parameters which have to be determined prior to classification and they are also usually determined through a cross validation process. A deeper mathematical treatise of SVMs can be found in this paper.

By their nature SVMs are intrinsically binary classifiers however exist strategies by which they can be adapted to multi-class tasks. But in our case we not need multi-class classification.

Proposed Algorithm

Here we detail the proposed algorithm for classification of Fraud Transactions.

Step 1: Generate the synthetic data according to given Probability. Use to separate distribution for Genuine and Fraud transactions.

Step 2: Read the generated data.

Step 3: Re-categorize the data into five groups as transaction month, date, day, amount of transaction & difference between successive transaction amounts.

Step 4: Make each transaction data as vector of five fields.

Step 5: Make two separate groups of data named True & False transaction group (if false transaction data is not available add randomly generate data in this group).

Step 6: Select one of three kernels (Linear, Quadratic, and RBF).

Step 7: Save the trained matrix.

Step 8: Read the current Transaction.

Step 9: Repeat the process from step 1 to step3 for current transaction data only.

Step 10: Place the saved Matrix & currently generated vector in classifier.

Step 11: Take the generated decision from the classifier.

Kernal Type: Linear

Total DATA	Fraud Prob.	TPR	TNR	FPR	FNR	Accuracy
30	0.30	0.90	0.72	0.28	0.10	0.83
30	0.40	0.61	0.59	0.41	0.39	0.60
30	0.50	0.26	0.77	0.23	0.74	0.56
60	0.30	0.98	0.22	0.38	0.03	0.72
60	0.40	0.77	0.61	0.32	0.26	0.70
60	0.50	0.70	0.75	0.29	0.24	0.73
100	0.30	0.89	0.27	0.39	0.20	0.67
100	0.40	0.65	0.43	0.51	0.38	0.54
100	0.50	0.81	0.48	0.38	0.24	0.67

Kernal Type: Quadratic

Total DATA	Fraud Prob.	TPR	TNR	FPR	FNR	Accuracy
30	0.30	0.96	0.93	0.03	0.06	0.95
30	0.40	0.95	0.81	0.08	0.09	0.91
30	0.50	0.91	0.88	0.08	0.11	0.90
60	0.30	0.98	0.75	0.06	0.08	0.93
60	0.40	0.91	0.67	0.25	0.10	0.81
60	0.50	0.89	0.75	0.18	0.14	0.83
100	0.30	0.92	0.40	0.27	0.16	0.76
100	0.40	0.87	0.54	0.26	0.21	0.75
100	0.5	0.64	0.69	0.35	0.30	0.67

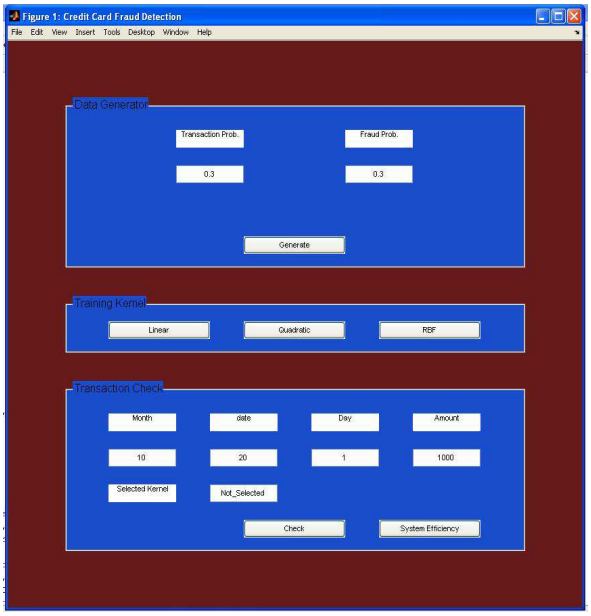
Kernal Type: RBF

Total DATA	Fraud Prob.	TPR	TNR	FPR	FNR	Accuracy
30	0.30	0.98	0.92	0.03	0.01	0.97
30	0.40	0.98	0.94	0.03	0.01	0.97
30	0.50	0.99	0.94	0.05	0.01	0.97
60	0.30	0.98	0.89	0.06	0.03	0.94
60	0.40	0.94	0.93	0.07	0.04	0.93
60	0.50	0.97	0.94	0.05	0.03	0.96
100	0.30	0.98	0.90	0.05	0.02	0.95
100	0.40	0.97	0.98	0.02	0.04	0.97
100	0.50	0.95	0.93	0.05	0.06	0.94

Implementation

Since there is no real data is available because of privacy maintained by banks. Hence for testing of implementation of our algorithm we generated the data of true & false Transaction using different mean & variance & then mixed them with different probability. We used the MATLAB for the implementation of the algorithm because of its rich sets of mathematical functions and also supporting the inbuilt functions for SVM.

Output Screen



Simulation Results

The results are simulated for five different Fraud probabilities from 0.3 to 0.5 & changing the training data size from 30 to 100, and then according to output of the program, following tables is drawn.

This shows that the RBF kernel outperform to Linear & quadratic kernel in all fields of comparison it has maximum accuracy up to 97%, & maximum training time 1.2 seconds & maximum matching time of 0.17 seconds in P4 system with 2GB of RAM.

CONCLUSION

Referring to results we can say that proposed algorithm can be used for automatic Fraud transaction classification with excellent accuracy & negligible delay. We can enhance this model for dynamic improvements in training of SVM.

REFERENCES

1. Wang Xi. Some Ideas about Credit Card Fraud Prediction China Trial. 2008: 74-75.

2. Chen Lei, Fraud and Prevention of International Credit Card. China Credit Card. Jun. 2004; 294: 43-47.

3. Liu Ren, Zhang Liping, Zhan Yinqiang. A Study on Construction of Analysis Based CRM System. Computer Applications and Software. 2004; 21: 46-47.

4. Han JW, Kamber M. Data Mining: Concepts and Techniques. Beijing: Higher Education Pr. and Morgan Kaufmann Publishers, 2007.

5. Barnett V, Lewis T. Outliers in Statistical Data New York: John Wiley & Sons, 1994.

6. Knorr E, Ng R. A Unified Notion of Outliers: Properties and Computation in proc. 1997 Int. Conf. Knowledge Discovery and Data Mining (KDD 97), Newport Beach, CA, 1997: 219-222.

7. Arning A, Agrawal R, Raghavwn P. A Linear Method for Deviation Detection in Large Database. In Proc. 1996 Int. Conf. Data Mining and Knowledge Discovery (KDD07), Portland, 1996: 164-169.

8. E. Aleskerov, B. Freisleben and B. Rao, CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection, Proc. IEEE/IAFE: Computational Intelligence for Financial Eng., 1997: 220-226.

9. MJ Kim and TS Kim, A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection, Proc. Int'l Conf. Intelligent Data Eng. and Automated Learning, 2002: 378-383.

10. W. Fan, AL Prodromidis and S.J. Stolfo, Distributed Data Mining in Credit Card Fraud Detection, IEEE Intelligent Systems, 1999; 14(6): 67-74.

11. R. Brause, T. Langsdorf and M. Hepp, Neural Data Mining for Credit Card Fraud Detection, Proc. IEEE Int'l Conf. Tools with Artificial Intelligence, 1999: 103-106.

12. C. Chiu and C. Tsai, A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection, Proc. IEEE Int'l Conf. e-Technology, e-Commerce and e-Service, 2004; 177-181.

13. C. Phua, V. Lee, K. Smith and R. Gayler, A Comprehensive Survey of Data Mining-Based Fraud Detection Research, http://w.bsys.monash.edu.au/people/cphua/, Mar. 2007.

14. S. Stolfo and AL Prodromidis, Agent-Based Distributed Learn-ing Applied to Fraud Detection, Technical Report CUCS-Columbia Univ., 1999.

15. C. Phua, D. Alahakoon and V. Lee, Minority Report in Fraud Detection: Classification of Skewed Data, ACM SIGKDD Explora-tions Newsletter, 2004; 6(1): 50-59.

16. V. Vatsa, S. Sural and AK Majumdar, A Game-theoretic Approach to Credit Card Fraud Detection, Proc. First Int'l Conf. Information Systems Security, 2005; 263-276.

17. S. Axelsson, the Base-Rate Fallacy and the Difficulty of Intrusion Detection, ACM Trans. Information and System Security, 2000; 3(3): 186-205.

18. LR Rabiner, a Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proc. IEEE, 1989; 77(2): 257-286.