Supervised Machine (SVM) Learning for Credit Card Fraud Detection

Sitaram patel[#1], Sunita Gond[*2],

[#1] *M.Tech Scholar BUIT, Bhopal M .P. India*
[#2]*Asst. prof., BUIT Bhopal M .P. India*

***Abstract*—: The growth of e commerce increases the money transaction via electronic network which is designed for hassle free fast & easy money transaction. The facility involves greater risk of misuse for fraud one of them is credit card fraud which can happen by many types as by stolen card, by INTERNET hackers who can hack your system & get important information about your card or by information leakage during the transaction. Several researchers have proposed their work for credit card fraud detection by characterizing the user spending profile. In this thesis we are proposing the SVM (Support Vector Machine) based method with multiple kernel involvement which also includes several fields of user profile instead of only spending profile. The simulation result shows improvement in TP (true positive), TN (true negative) rate, & also decreases the FP (false positive) & FN (false negative) rate.**

***Keywords*— MATLAB, e-commerce, online banking, classification support vector machine**

I. INTRODUCTION

The prediction of user behaviour 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. 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 authorised. 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. 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.

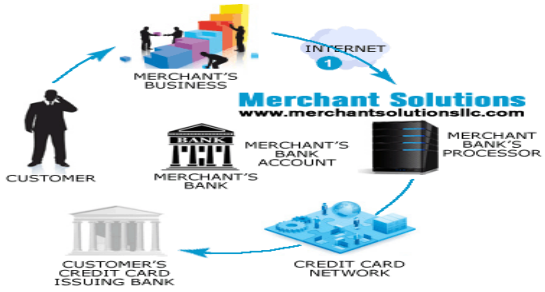


Figure 1: Credit Card Blocks

II. CREDIT CARD FRAUD DETECTION

Growth in communication network, increased internet speed, easy wireless connectivity & lack of time causes the people to buy through electronic network. Here are some statistics and projections of the Indian credit card industry (http://hubpages.com/hub/Indian-Credit-card-Industry) to show importance of the topic.

- India is currently the fastest growing Mobile Market in the world and is also among the fastest growing credit card markets in the world.
- India has a total approx.75 million cards under circulation (25 million credit and 50 million debit) and a 30% year-on-year growth.
- With 87% of all transactions in plastic money happening through credit cards, debit cards in India continue to be used largely for cash withdrawals.
- Though Visa, which accounts for 70% of the total card industry is the market leader in India; MasterCard is fast catching up.
- Every transaction involves payment of an interchange charge to MasterCard or Visa for settlement, which amounted to about $50 million during the year.
- Internal estimates of Barclaycard have pegged the Indian market with potential to grow to at least 55million credit cards by 2010-11.

III. CARD CLASSIFICATIONS

Qualified: The lowest retail processing rate category, In order for a transaction to qualify for this category, the credit card must be swiped though the terminal, the contents of the

magnetic stripe transmitted, and an authorization must be received. The transaction must also be settled or batched out of the terminal within 24 hours of the authorization.

Mid-Qualified: The lowest rate category for which key-entered transactions can qualify. This category requires that the billing address of the cardholder be verified with a match of their zip code. Once again, an authorization number must be received, and the transaction must be settled or batched out of the terminal within 24 hours of the authorization.

Non-Qualified: The highest rate category in the processing environment. Generally speaking, this rate applies to key-entered transactions where the address is not verified with a zip code match, or transactions that are not settled or batched out of the terminal within 24 hours of their authorization. Most business cards also fall into the category of pricing.

IV. PROPOSED ALGORITHM AND IMPLEMENTATION

Firstly input the transaction probability for which or for what number of synthetic data you want to generate. It will be taken as the number of day you have done. Some transaction on your account according to the input the number of days will be selected randomly. Once we get the synthetic data generate. Here we detail the proposed algorithm for classification of Fraud Transactions.

- Read the given data.
- Re-categorize the data into five groups as transaction month, date, day, amount of transaction & difference between successive transaction amounts.
- Make each transaction data as vector of five fields.
- 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).
- Select one of three kernels (Linear, Quadratic, and RBF).
- Train SVM.
- Save the classifier.
- Read the current Transaction.
- Repeat the process from step1 to step3 for current transaction data only.
- Place the saved classifier & currently generated vector in classifier.
- Take the generated decision from the classifier.

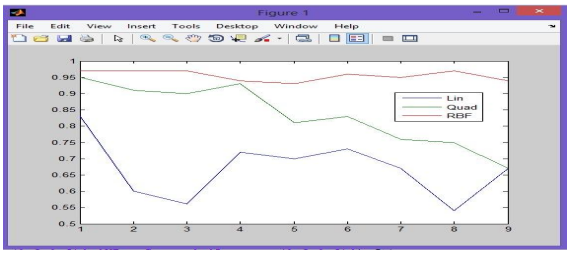
We describe the techniques which are used in implementing our model. The forms as well as the output screens are also shown to explain the working of our implemented algorithm.

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

The complete simulation & comparison of all three kernels are performed in MATLAB 7.5 environment. The MATLAB 7.5 is preferred because of its authenticity proven by time & simple programming. It also contains specific tool box for machine learning functions called bioinformatics which contains complete set of function required by our program this reduces the programming effort & increase the precision. The main function from SVM tool box is used to train a machine by passing machine parameters. Stretcher of train machine which can be used for data classification in future through SVM classify.

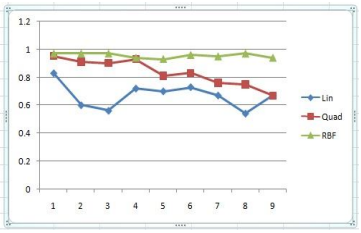
V. RESULTS ANALYSIS

The results are simulated for three different Fraud probabilities Kernel Type: Linear, Quadratic and RBF from 0.30 to 0.50 & changing the training data size from 30 to 100.



.

Lin	Quad	RBF
0.83	0.95	0.97
0.6	0.91	0.97
0.56	0.9	0.97
0.72	0.93	0.94
0.7	0.81	0.93
0.73	0.83	0.96
0.67	0.76	0.95
0.54	0.75	0.97
0.67	0.67	0.94



This shows that the RBF kernel outperform to Linear & quadratic kernel in all fields of comparison it has maximum accuracy up to 97%, maximum TPR(99%),maximum TNR(98%) & maximum FPR(7%),maximum FNR(6%), it also behaves almost same for all type of data set generated(having very low fraud data & high fraud data).

VI. CONCLUSION & FUTURE WORK

Referring to results we can say that proposed algorithm with RBF kernel gives the better accuracy in comparison with the linear, Quadratic. In previous papers HMM model was used. The SVM classifier scheme is a novel scheme used by us. In this we have compared performance in three different kernels namely linear, Quadratic, RBF. As shown in the result our approach is better than the previous approaches, hence it can be used for automatic Credit card Fraud detection with excellent accuracy & minimum false alarm.

A. Limitations

- None of the machine learning classifier algorithms evaluated was able to perform detection of user-to-root attack categories significantly.

- ❖ No procedure was devised for kernel width delimitation that is to obtain the best kernel function.

B. Future Work

- ❖ Enhance this model for dynamic improvements in training of classifiers using different SVM models like incremental SVM, detrimental SVM and evolutionary SVM etc.
- ❖ Explore continuous updating and adaption of subsystems and combiner.
- ❖ Extend adaptive classifier to finite mixture model (more flexible), approximate logistic regression and RBF networks.
- ❖ More realistically handle the delayed fraud label.

REFERENCES

[1] Abhinav Srivastava, Amlan Kundu, Shamik Sural, Arun Majumdar,(Jan.-Mar. 2008) "Credit Card Fraud Detection Using Hidden Markov Model," IEEE Transactions on Dependable and Secure Computing, vol. 5, no. 1, pp. 37-48,.

[2] Wen-Fang Yu, Na Wang,(2009.)"Research on Credit Card Fraud Detection Model Based on Distance Sum International Joint Conference on Artificial Intelligence ," JCAI, pp.353-356,

[3] Sushmito Ghosh and Douglas L. Reilly Nestor,(1994.) "Credit Card Fraud Detection with a Neural-Network," Inc. Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System Sciences,pp.621-630 ,

[4] Aihua Shen, Rencheng Tong (2007), Yaochen Deng School of Management, Graduate University of the Chinese Academy of Sciences, Beijing, 100084, China, "Application of Classification Models on Credit Card Fraud Detection",pp.1-4,IEEE.

[5] Chun-Hua JU, Na Wang,(2009) " Credit Card Fraud Detection Model Based on Similar Coefficient Sum," First International Workshop on Database Technology and Applications ,pp.295 - 298 , Database Technology and Applications, First International Workshop on Wuhan, Hubei .

[6] V. N. Vapnik,(1995.)" The Nature of Statistical Learning Theory," New York: Springer-Verlag,Second Edition, Technical report,

[7] V. Kecman,(2001)"Learning and Soft Computing: Support Vector Machines, Neutral Networks and Fuzzy Logic Models,"Cambridge, MA: MIT Press,Technical report,

[8] Research Article Bird Species Recognition Using Support Vector Machines Hindawi(2007) Publishing Corporation EURASIP Journal on Advances in Signal Processing Volume, Article ID 8637,8pages doi : 10.1155 / 2007 / 38637.

[9] "Global Consumer Attitude Towards On-Line Shopping,"(Mar. 2007).http://www2.acnielsen.com/reports/documents/2005_cc_online shopping.pdf,

[10] D.J. Hand, G. Blunt, M.G. Kelly, and N.M. Adams,(2000.) "Data Miningfor Fun and Profit," Statistical Science, vol. 15, no. 2, pp. 111-131,

[11] "Statistics for General and On-Line Card Fraud,(Mar. 2007) "http://www. epaynews.com/statistics/fraud.html,.

[12] S. Ghosh and D.L. Reilly(1994.), "Credit Card Fraud Detection with a Neural-Network," Proc. 27th Hawaii Int'l Conf. System Sciences:Information Systems: Decision Support and Knowledge-Based Systems,vol. 3, pp. 621-630,

[13] M. Syeda, Y.Q. Zhang, and Y. Pan,(2002). "Parallel Granular Networks for Fast Credit Card Fraud Detection," Proc. IEEE Int'l Conf. Fuzzy Systems, pp. 572-577,

[14] S.J. Stolfo, D.W. Fan, W. Lee, A.L. Prodromidis, and P.K. Chan(1997),"Credit Card Fraud Detection Using Meta-Learning: Issues and Initial Results," Proc. AAAI Workshop AI Methods in Fraud and Risk Management, pp. 83-90,.

[15] S.J. Stolfo, D.W. Fan, W. Lee, A. Prodromidis, and P.K. Chan, (2000) "Cost-Based Modeling for Fraud and Intrusion Detection: Results from the JAM Project," Proc. DARPA Information Survivability Conf. and Exposition, vol. 2, pp. 130-144,.