IMPACT: International Journal of Research in Engineering & Technology
ISSN(P): 2347-4599; ISSN(E): 2321-8843
Vol. 6, Issue 1, Jan 2018, 41-46
© Impact Journals

FRAUD DETECTION OF CREDIT CARD BY USING HMM MODEL

***Sadhana Yadav*[1] *& Siddartha*[2]**

[1]*Research Scholar, Department of Computer Science, Kakatiya University, Warangal, Telangana, India*

[2]*Research Scholar, Department of Mechanical Engineering, Kakatiya University, Warangal, Telangana, India*

Received: 22 Dec 2017 **Accepted: 27 Dec 2017** **Published: 30 Jan 2018**

ABSTRACT

Credit card fraud is a serious and growing problem. 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. In this paper, we model the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and show how it can be used for the detection of frauds. An HMM is initially trained with the normal behavior of a cardholder. If an incoming credit card transaction is not accepted by the trained HMM with a sufficiently high probability, it is considered to be fraudulent. At the same time, we try to ensure that genuine transactions are not rejected. We present detailed experimental results to show the effectiveness of our approach and compare it with other techniques available in the literature.

KEYWORDS: *Credit Card Fraud, Regular Purchase, Credit Card Transaction*

INTRODUCTION

So much money is lost due to the credit card fraud. Over the years, along with the evolution of fraud detection methods, perpetrators of fraud have also been evolving their fraud practices to avoid detection. In this study, we evaluate two advanced data mining approaches: Super vector approach and Random forests approach. This study is based on real-life of data transactions from international credit card operation.

Credit card purchase frauds are essentially two types: I) Application fraud and ii) Behavioral fraud. In an Application Fraud based on 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. In Behavioral fraud, 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.

Purchase data of the cardholder are 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.

Predictive models for credit card fraud detection are in active use in practice. Considering the profusion of data mining techniques and applications in recent years, however, there have been relatively few reported studies of data mining for credit card fraud detection.

Among these, most papers have examined neural networks not surprising, given their popularity in the 1990s. A summary of these is given in, which reviews analytic techniques for general fraud detection, including credit card fraud.

Other techniques reported for credit card fraud detection include case based reasoning and more recently, Hidden Markov Models are sent. There are several techniques available, including Support Vector machines and Random Forests for predicting credit card fraud. Their study focuses on the impact of aggregating transaction level data on fraud prediction performance. It examines aggregation over different time periods on two real-life datasets and that aggregation can be advantageous, with the aggregation period length being an important factor. Aggregation was found to be especially effective with random forests. Random forests were noted to show better performance in relation to the other techniques, though logistic regression and support vector machines also performed well. Support Vector machines and Random Forests are sophisticated data mining techniques which have been noted in recent years to show superior performance across different applications. The choice of these two techniques, together with logistic regression, for this study is based on their accessibility for practitioners, ease of use, and noted performance advantages in the literature.

DATA-MINING TECHNIQUES

As stated above, we investigated the performance of the three techniques in predicting fraud: Logistic Regression (LR), Support Vector Machines (SVM), and Random Forest (RF). In the paragraphs below, we briefly describe the three techniques employed in this study.

Logistic Regression

Qualitative response models are appropriate when the dependent variable is categorical. In this study, our dependent variable fraud is binary, and logistic regression is a widely used technique in such problems. Binary choice models have been used in studying fraud. For example, used binary choice models in the case of insurance frauds to predict the likelihood of a claim being fraudulent. In case of insurance fraud, investigators use the estimated probabilities to flag individuals that are more likely to submit a fraudulent claim.

Support Vector Machines

Support vector machines (SVMs) are statistical learning techniques that have been found to be very successful in a variety of classification tasks. Several unique features of these algorithms make them especially suitable for binary classification problems like fraud detection. SVMs are linear classifiers that work in a high-dimensional feature space that is a non-linear mapping of the input space of the problem at hand. An advantage of working in a high-dimensional feature space is that, in many problems the non-linear classification task in the original input space becomes a linear classification task in the high-dimensional feature space. SVMs work in the high-dimensional feature space without incorporating any additional computational complexity. The simplicity of a linear classifier and the capability to work in a feature rich space make SVMs attractive for fraud detection tasks where highly unbalanced nature of the data (fraud and non-fraud cases) make extraction of meaningful features critical to the detection of fraudulent transactions is difficult to achieve. Applications of SVMs included bio informatics, machine vision, text categorization, and time series analysis.

‘m’.

Random Forests

Random forests have been popular in application in recent years. They are easy to use, with only two adjustable parameters, the number of trees (T) in the ensemble and the attribute subset size (b), with robust performance noted for typical parameter values. They have been found to perform favorably in comparison with Support Vector Machine and other current techniques. Other studies comparing the performance of different learning algorithms over multiple datasets have found Random Forest to show good overall performance. Random forests have been applied in recent years across varied domains from predicting customer churn, image classification, to various bio-medical problems. While many papers not either excellent classification performance in comparison with other techniques including SVM, a recent study finds SVM to outperform random forests for gene expression micro array data classification. The application of random forests to fraud detection is relatively new, with few reported studies.

LITERATURE SURVEY

Abhinav Srivastava [1] et al describes the “Credit card fraud detection method by using Hidden Markov Model (HMM)”. In this paper, they model the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and show how it can be used for the detection of frauds. An HMM is initially trained with the normal behavior of a cardholder.

S. Ghosh and Douglas L. Reilly[2] et al describes the “Credit card fraud detection With Neural Network”. In this paper they use data from a credit card issuer, a neural network based fraud detection system was trained on a large sample of labeled credit card account transactions and tested on a holdout data set that consisted of all account activity over a subsequent two-month period of time. The neural network was trained on examples of fraud due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud and NRI (non-received issue) fraud. The network detected significantly more fraud accounts (an order of magnitude more) with significantly fewer false positives (reduced by a factor of 20) over rule based fraud detection procedures.

ALGORITHM USED

HMM Model

To map the credit card transaction processing operation in terms of an HMM, we start by first deciding the observation symbols in our model. We quantize the purchase values x into M price ranges V1; V2... VM, forming the observation symbols at the issuing bank. The actual price range for each symbol is configurable based on the spending habit of individual cardholders. These price ranges can be determined dynamically by applying a clustering algorithm on the values of each cardholder’s transactions. We use Vk, k ¼ 1; 2;...M, to represent both the observation symbol, as well as the corresponding price range.

In this work, we consider only three price ranges, namely, low (l), medium (m), and high (h). Our set of observation symbols is, therefore, V ¼ fl; m; hg making M ¼ 3. For example, let l= (0, $100], m =($100,$500], and h=($500, credit card limit]. If a card holder performs a transaction of $190, then the corresponding observation symbol is ‘m’.

A credit card holder makes different kinds of purchases of different amounts over a period of time. One possibility is to consider the sequence of transaction amounts and look for deviations in them. However, the sequence of types of purchase is more stable compared to the sequence of transaction amounts. The reason is that, a cardholder makes purchases depending on his need for procuring different types of items over a period of time. This, in turn, generates a sequence of transaction amounts. Each individual transaction amount usually depends on the corresponding type of purchase. Hence, we consider the transition in the type of purchase as state transition in our model. The type of each purchase is linked to the line of business of the corresponding merchant. This information about the merchant's line of business is not known to the issuing bank running the FDS. Thus, the type of purchase of the cardholder is hidden from the FDS. The set of all possible types of purchase and, equivalently, the set of all possible lines of business of merchants forms the set of hidden states of the HMM. It should be noted at this stage that the line of business of the merchant is known to the acquiring bank, since this information is furnished at the time of registration of a merchant. Also, some merchants may be dealing in various types of commodities (For example, Wal-Mart, K-Mart, or Target sells tens of thousands of different items). Such types of line of business are considered as Miscellaneous, and we do not attempt to determine the actual types of items purchased in these transactions. Any assumption about the availability of this information with the issuing bank and, hence, with the FDS, is not practical and, therefore, would not have been valid.

Advantages

- Highly Security from unauthorized use of credit card.
- Avoids fraud usage of card through online transactions.
- Detect if card used by others if card lost.

PROBLEM DEFINITION

The proposed system we introduce a new technology to protect the network. This is achieved by the following way. Realizing wide spread adoption of such applications mandates sufficiently trust worthy computers that can be realized at low cost. Apart from facilitating deployment of futuristic applications, the ability to realize trustworthy computers at low cost can also addresses many of the security issues that plague our existing network infrastructure. Although, at first sight, “inexpensive” and “trust worthy”. May seem mutually exclusive, a possible strategy is to reduce the complexity of the components inside the trusted boundary. The often heard statement that “complexity is the enemy of security” is far from dogmatic. For one, lower complexity implies better verifiability of compliance. Furthermore, keeping the complexity inside the trust boundary at low levels can obviate the need for proactive measures for heat dissipation. Strategies constrained to simultaneously facilitate shielding and heat dissipation tend to be expensive. On the other hand, unconstrained shielding strategies can be reliable and inexpensive to facilitate.

SYSTEM ANALYSIS

System Feasibility Study

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are:

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

Economical Feasibility

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system is well within the budget and this was achieved, because most of the technologies used are freely available. Only the customized products had to be purchased.

Technical Feasibility

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

System Implementation

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus, it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

The implementation stage involves careful planning, investigation of the existing system and its constraints on the implementation, designing of methods to achieve changeover and evaluation of changeover methods.

Implementation is the process of converting a new system design into operation. It is the phase that focuses on user training, site preparation and file conversion for installing a candidate system. The important factor that should be considered here is that the conversion should not disrupt the functioning of the organization.

CONCLUSIONS

We have proposed an application of HMM in credit card fraud detection. The different steps in credit card transaction processing are represented as the underlying stochastic process of an HMM. We have used the ranges of the transaction amount as the observation symbols, whereas the types of item have been considered to be state of the HMM. We have suggested a method for finding the spending profile of cardholders, as well as application of this knowledge in deciding the value of observation symbols and an initial estimate of the model parameters. It has also been explained how the HMM can detect whether an incoming transaction is fraudulent or not. Experimental results show the performance and effectiveness of our system and demonstrate the usefulness of learning the spending profile of the cardholders. Comparative studies reveal that the Accuracy of the system is close to 80 percent over a wide variation in the input data. The system is also scalable for handling large volumes of transactions.

REFERENCES

1. A. Srivastava, A. Kundu, S. Sural, A. Majumdar, Credit card fraud detection using hidden Markov model, IEEETransactions on Dependable andSecure Computing 5(1) (2008) 3748.
2. Smriti Srivastava & Anchal Garg, Data Mining for Credit Card Risk Analysis: A Review, International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR), Volume 3, Issue 2, May-June 2013, pp. 193-200
3. S. Ghosh, D.L. Reilly, Credit card fraud detection with a neural-network, in: J.F. Nunamaker, R.H. Sprague (Eds.), Proceedings of the 27th Annual Hawaii International Conference on System Science, Vol 3, Information Systems: DSS/ Knowledge-based Systems, Los Alamitos, CA, USA, 1994.
4. A. Ashok Kumar, M. Poornashri, A. Karunya Priyanka, G.Arun & D. Gowrishankar, Opinion of Customers Towards the Promotion Schemes Offered for Credit Cards With Special Reference to Icici Bank, Dharmapuri, Tamil Nadu, International Journal of Business and General Management (IJBGM), Volume 5, Issue 3, April-May 2016, pp. 21-28