World of Computer Science and Information Technology Journal (WCSIT)
ISSN: 2221-0741
Vol. 2, No. 4, 137-141, 2012

Electronic Payment Fraud Detection Techniques

Adnan M. Al-Khatib
CIS Dept. – Faculty of Information Technology
Jerash University, Jarash - Jordan

Abstract— In this paper, we discuss the fraudulent transactions that occur in electronic payment systems. We evaluate various techniques that can be used in detecting fraudulent transactions of card-not-present payment systems. The presented evaluation based on the literature, and from our own studies for these techniques. It provides a basis for exploring the common ground between techniques and for analyzing experimental studies and scenarios in practice.

Keywords- Electronic Payment; Credit Card; Card-Not-Present; Fraud Detection; Data Mining; Machine learning; Neural Network; Expert System.

I. INTRODUCTION

Electronic Fraud is increasing with the expansion of modern technology and global communication. This increase in the fraudulent transactions, resulting in substantial losses to the businesses, and therefore, fraud detection has become an important issue to be considered. Fraud detection can be seen as a problem of classification of legitimate transactions from the fraudulent transactions.

Existing fraud detection techniques have been implemented by a number of methods such as data mining, statistics, and artificial intelligence. In general, fraud detection is a prediction problem and its objective is to maximize correct prediction and maintain incorrect predictions at an acceptable level of cost. Recent studies have shown that data mining using Artificial Intelligence (AI) techniques achieved better performance than traditional statistical methods for building prediction models [2, 12]. AI techniques, particularly rule-based expert systems, case-based reasoning systems and machine learning (ML) techniques such as neural networks have been used to support such analysis and classification problems. The major difference between traditional statistical methods and machine learning methods is that: in statistical methods usually researchers impose structures to different models, and construct the model by estimating parameters to fit the data or observation, while machine learning techniques allow learning the particular structure of the model from the data [12]. As a result, the structures of the models used in statistical methods are relatively simple, easy to interpret and tend to under-fit the data while models obtained in machine learning methods are usually very complicated, hard to explain and tend to over-fit the data. Under-fit and over-fit of the data is in fact the trade-off between the explanatory power and parsimony of a model, where explanatory power leads to high prediction accuracy and parsimony usually assures generalizability and interpretability of the model.

In this paper, section 2 presents fraudulent transaction scenario of card-not-present. Section 3 briefly discusses the various techniques that can be used in detecting fraudulent transactions of credit cards. Section 4 evaluates the presented techniques based on the literature, and from our own studies. Finally, section 5 concludes the paper with some future remarks.

II. CREDIT CARD FRAUD

Credit card fraud is divided into two types: credit-card (offline) fraud and card-not-present (online) fraud. Offline fraud is committed by using a stolen physical card at storefront or call center. Online fraud is committed via web, phone shopping or cardholder-not-present. In online fraud only the card’s details are needed, and a manual signature and card imprint are not required at the time of purchase.

Fraud detection for credit card means the process of classifying the transactions into two classes: a class of legitimate and a class of fraudulent transactions. In general, the objective of fraud detection is to maximize correct predictions and maintain incorrect predictions at an acceptable level of cost.

A. Problems with credit card fraud detection

Fraud detection system should have some properties in order to perform good results, such as:

- The system should be able to handle skewed distributions, since only a very small percentage of all credit card transactions are fraudulent [3, 7].
- The ability to handle noise (errors in the data). Noise limits the accuracy of generalization [10].
- Overlapping data: many transactions may be assumed fraudulent, while actually they are legitimate (false alarm). The opposite also may happen, when a fraudulent transaction appears to be normal (false negative). The system should

have the ability to eliminate or minimize these problems to give higher accuracy [1, 7, 10].

- The systems should be able to adapt themselves to new kinds of fraud.
- There is a need for good metrics to evaluate the classifier system. For example, overall accuracy is not suited for evaluation on a skewed distribution, since even with a very high accuracy; almost all fraudulent transactions can be misclassified [7, 11].
- The system should take into account the cost of the fraudulent behavior detected and the cost associated with stopping it. For example, no profit is made by stopping a fraudulent transaction of only a few dollars [7, 11].

III. FRAUD DETECTION TECHNIQUES

Researchers have developed two general categories of detection techniques; misuse and anomaly detections. In misuse detection, well-known fraudulent transactions are encoded into patterns, which are then used to match new transactions to identify the fraudulent ones. In anomaly detection, normal behavior of user and system activities are first summarized into normal profiles, which are then used as yardsticks, so that run-time activities that result in significant deviation from the user profiles are considered as probable fraudulent transactions. In this section we will briefly describe some current fraud detection techniques used in detecting credit cards fraud, and mention the advantages and disadvantages of these techniques:

A. Neural Networks

Neural Networks (NN) are an Artificial Intelligence (AI) techniques or methods that represent models of biological learning systems. They are networks of many simple processors or units that are connected and process numeric values [5]. It is structured as a directed graph with many nodes (processing elements) and arcs (interconnections) between them. The node computes a weighted sum of its inputs and generates an output. This output then becomes an input to other nodes in the network. The process continues until one or more outputs are generated. NN have been used as a powerful data mining technique in industry to do classification, clustering, generalization, and forecasting (predictions). For examples; it can be used to distinguish legal from fraudulent transactions, detect Internet fraud on an e-commerce site, predict which transaction may be a fraudulent transaction, etc. In addition, NN are one of the first and most successful applications in the area of detecting credit card fraud.

B. Rule Induction

Rule induction (RI) creates a decision tree (DT) or a set of decision rules from training examples with a known classification. DT is a predictive modeling technique used in classification, clustering, and prediction tasks. DT is defined as a tree where the root and each internal node are labeled with a question about an independent variable. The arcs from each node represent each possible answer to the associated question. Each leaf node represents a prediction (dependent variable) of a solution to the problem under consideration. The knowledge represented in DT can be extracted and represented in IF-THEN rules. One rule is created for each path from the root to a leaf node.

DT classification is a two step process: induction process to construct a DT using training data, and the second step, is to apply the DT to new instances or records of the data to determine its class.

C. Expert Systems

Rules can be generated from information obtained from a human expert or group of experts and stored in a rule-based system as IF-THEN rules. If information is stored in a Knowledge base (KB) then it is called a Knowledge base system (KBS), or an Expert system (ES). The rules in the ES used to perform operations on a data to inference in order to reach appropriate conclusion. ES provide powerful and flexible solutions to many application problems. In the financial area, it can be used for several applications such as financial analysis and fraud detection. Suspicious activity or transaction can be detected from deviations from "normal' spending patterns through the use of ES [6].

D. Case-based reasoning (CBR)

The basic idea of CBR is to adapt solutions that were used to solve previous problems and use them to solve new problems. In CBR, descriptions of past experience of human specialists, represented as cases, are stored in a database for later retrieval when the user encounters a new case with similar parameters. These cases can be used for classification purposes. Given a new problem, a CBR system tries to find a matching case. There are several algorithms used with this approach [9], but the nearest neighbor matching algorithm is often used. In this algorithm the training data is the model, and when a new case or instance is presented to the model, the algorithm looks at all the data to find a subset of cases that are most similar to it and uses them to predict the outcome.

E. Genetic algorithms (GAs)

GAs are search procedures based on the evolutionary computing methods. Given a population of potential problem solutions (individuals), evolutionary computing expands this population with new and potentially better solutions. In data mining, GAs may be used for clustering, prediction, and even association rules. These techniques can be used to find the fittest models from a set of models to represent the data.

F. Inductive logic programming (ILP)

ILP uses first order predicate logic to define a concept by using a set of positive and negative examples. This logic program is then used to classify new examples. In this approach of classification; complex relationship among components or attributes can be easily expressed, which improve the expressive power of the model. Domain knowledge can be easily represented in an ILP system, which improves the effectiveness of the system. The model expressed in predicate logic is also easy to understand.

G. Regression

Regression is a statistical techniques generally used to predict future values based on past values by fitting a set of points to a curve. Regression assumes that target data fit into some known type of function (e.g., linear, logistic, etc). In the banking systems, linear regression can be used to build a classification model of two classes, and then use this model to approve or reject a new loan application, or classify a transaction as fraudulent or non-fraudulent transaction.

H. Summary of the advantages and disadvantages of the techniques

Table 1 summarizes the advantages and disadvantages of the above techniques.

TABLE 1: ADVANTAGES AND DISADVANTAGES OF THE ABOVE DETECTION TECHNIQUES.

Technique	Advantages	Disadvantages
NN	Effective in dealing with noisy data, in predicting patterns, in solving complex problems, and in processing new instances. Can generate code to be used in real-time systems, highly accurate, portable, fast, and outperform other techniques	poor explanation capability, less efficient in processing large data sets, difficult to setup and operate, sensitive to data format, different data representations can produce different results, can never be exact, it is only accurate. In addition, it is only work with numeric data with values between 0 and 1; non numerical data need to be converted and normalized.
RI	Scalable, high predictive accuracy, easy to use, easy to explain results, easy to interpret the rules, easy to implement applications.	Not easy to handle continuous data, difficult to handle missing data, over-fitting problem may occur, data size and attributes to use for splitting and in which order to choose them and number of splits for each attribute impact the performance of building the tree.
ES	Easy to modify the KB, easy to develop and build the system, easy to manage complexity or missing information. Have high degree of accuracy, explanation facilities, and have good performance. Rules from other techniques such as NN and DT can be extracted, modified, and stored in the KB.	Poor in handling missing information or unexpected data values, and knowledge representation languages do not approach human flexibility.
CBR	Useful in domain that has a large number of examples, has the ability to work with incomplete or noisy data, effective, flexible, easy to update and maintain, can be used in a hybrid approach.	May suffer from the problem of incomplete or noisy data.
GA	Works well with noisy data, and easy to integrate with other systems. Usually combined into other techniques to increase the performance of those techniques.	Require extensive tool knowledge to set up and operate, and difficult to understand.
ILP	has powerful modeling language that can model complex relationships	Has low predictive accuracy, very sensitive to noise, and their performance deteriorates rapidly in the presence of spurious data.
Regression	Easy to understand, easy to build, and used with two class classification. Logistic regression can give better accuracy than some learning techniques for small data sets.	Poor with noise or outliers data, not applicable to complex applications, not work well with nonnumeric data, accuracy is good, but not high, not good for large data sets.

IV. TECHNIQUES EVALUATION

Several data mining tools for fraud detection are used widely in different organizations, but which are the most effective in terms of money, accuracy and time for a given application.

A. Factors that affect the performance of data mining techniques

There are several factors that affect the performance and accuracy of a data mining techniques. Understanding these factors is useful in evaluating and selecting an appropriate technique for an application. In this section a description of some of these factors is given:

- Noise in data: data sets often contain noise in the form of inaccuracies and inconsistencies in the data. For example, inadequate data validation procedures may allow the user to enter incorrect data.
- Missing data: attributes required for analysis may not be available. Missing data may cause problems in the training phase and in the classification process. Therefore, the ability of the technique to maintain this problem is an important factor.
- Measuring performance: The performance of classification algorithm is usually examined by evaluating the accuracy of the classification. Accuracy of a data mining technique strongly influences its effectiveness. Higher predictive accuracy with actual data is a desirable feature. Classification accuracy is usually calculated by determining the percentage of records placed in the correct class. This ignores the fact that there also may be a cost associated with an incorrect assignment to the wrong class, which also should be determined. With two classes, there are four possible outcomes. Given a class C, and a record r to be classified, the four outcomes are:

True positive (TP): r predicted to be in C and is actually in it.

False positive (FP): r predicted to be in C and is not actually in it (False alarm).

True negative (TN): r not predicted to be in C and is not actually in it.

False negative (FN): r not predicted to be in C but is actually in it.

TP and TN represent correct actions, but FP and FN represent incorrect actions. The performance of a classification could be determined by associating cost with each type of these outcomes. So maximizing TP and minimizing FP and FN are desirable characteristics of a data mining application. Some studies [1, 8] show that, overall predictive accuracy of fraud detection is inappropriate as the single measure of predictive performance. For example; If 1% of the transactions are fraudulent (usual probability of fraud), then a model that always predicts "legitimate" will be 99% accurate, and at the same time may not catch the fraudulent transactions. So of the 1% fraudulent transactions, there is a need to compute models that predict 100% of these, yet produce no false alarms (i.e. predict no legitimate transactions to be fraudulent). In addition, a model that predicts less than 100% of the fraudulent transactions may not be accurate.

For example, if a model predicts 90% of the fraudulent transactions, then it may correctly predict the lowest cost transactions, and being entirely wrong about the top 10% most expensive frauds. Therefore, there is a need for a cost model in order to best judge the success of the fraud detection technique.

- Scalability: usually data mining applications use too large data sets. These data sets loaded into RAM and may slow the processing and the running of the algorithm. In addition the network bandwidth capability of a system may affect the processing. So scalability of the data mining technique becomes an important issue.

- Different data types: Business databases or data sets contain data of various types (numeric, ordinal, and nominal etc). If a data mining technique can handle different data types, it will be more useful for business data mining.

- Explanation capability: the prediction result is more likely to be accepted by business manger, if it is explainable in business terms. Understanding the model building and involvement of the user in data loading and manipulation of the algorithm parameters will increase the accuracy and performance of the system. Therefore, the ability to explain the results is an important factor.

- Ease of integration: data mining application usually work with other information systems (IS) such as DSS or DBMS. Therefore, ease of integration with other information systems is a desirable characteristic of data mining application.

- Ease of operation: A technique that is easy to understand, easy to build and that requires fewer preprocessing activities is more useful to an end user.

- Skewed distribution: usually fraud detection data is highly skewed or imbalanced. Examples given in [10, 1, 8] show that skewed distribution of the data could be a major factor on the classifier performance. So ability of the data mining model to handle this problem is a desirable characteristic.

B. *Related work*

Several studies to evaluate the different data mining techniques for credit card fraud detection were held. Each study evaluated the different techniques according to some of the above factors.

Several machine learning algorithms (ID3, CART, BAYES, and RIPPER) and meta-learning strategies on real world credit card data (one million transactions) were tested in [10] to select the best classifier. This study shows that skewed class distribution could be a major factor on classifier performance and that True and False positive rates are the critical evaluation metrics for the accuracy of the models built. The study reported that 50%/50% distribution of fraud/non-fraud training data will generate classifiers with the highest True positive rate and low False positive rate. The best classifier in the study was a meta-classifier BAYES. The next two are CART and RIPPER. All the three trained on a 50%/50% fraud/non-fraud distribution, and each attained a True positive rate of approximately 80% and False positive rate less than 17% for base classifiers and 13% for the meta-classifier. ID3 was the last with True positive rate 76% and False positive rate 23%.

A comparative study between Bayesian Belief Network (BBN) using STAGE algorithm and Artificial Neural Network (ANN) using BP algorithm for credit card fraud detection was held in [7]. The results show that: BBNs were more accurate (in some cases 8% more of catching fraudulent transactions) and much faster to train than ANN, but ANN is much faster than BBNs when applied to new instances.

In [1]; authors employ five different inductive learning programs: Bayes (with Bayesian learning algorithm), C4.5, ID3, CART, Ripper (rule induction algorithm), and a meta-learning methods to compute accurate classification models for detecting electronic credit-card fraud. They used three metrics: the overall accuracy, the TP − FP spread and a cost model. There results show that meta-classifiers outperform all base classifiers, and in some cases by a significant margin. Also, results show that the most accurate classifiers are not necessarily the most cost effective. For example; Bayesian base classifiers are less accurate than Ripper and C4.5 base classifiers, but they are the best under the cost model. Also the results show that partitioning the large data set into smaller subsets improves the cost saving. In addition, the performance of classification was sensitive to the changes in the data sets.

An expert system model for credit-card fraud detection was presented in [6]. The model's rule base was constructed through the input of many fraud experts within a bank. The model's performance was measured based on classification accuracy and based on the cost of misclassification. It was assumed that the cost of one fraud would be approximately equal to the cost of disturbing twenty good customers. The expert model was able to classify 89.68% accuracy overall and 80.45% correct within the fraud class. This expert model was compared with other three different models of fraud detection with the same data sets and it outperforms all of them.

A comparative study to select a suitable data mining tool or product for fraud detection was carried in [4]. In this paper 40 data mining tools were chosen to be evaluated. Three stages of evaluation were done, and only the top five tools were continued to be evaluated in the third stage. In the third stage, an extensive evaluation includes the areas of client-server compliance, automation capabilities, breadth of algorithms implemented, ease of use, and overall accuracy was held. Results show that decision trees and neural networks allowed the best cross-comparison, and proved to be better than the other models. Decision Trees were better than Neural Networks at reducing false alarms and specifying misclassification costs. In addition, the pruning options for the trees were better developed than the stopping rules for the networks, so the hazard of over-fit was less.

A comparative research between three machine learning methods (ANNs, CBR, and RI) and one statistical method - least squares regression (LSR), to build prediction systems was held in [2]. The study compares the prediction systems in terms of accuracy, explanatory value and configurability. The results show that all approaches are sensitive to changes in the training data set and may not cope well with heterogeneity. ANN was the best in accuracy, but the worst in explanatory value and configurability. CBR and LSR were the best in explanatory value and configurability, and good in accuracy. RI is good in explanatory value, but the last in accuracy and configurability.

Several studies [10, 1, 8, 3] show that hybrid Techniques of classification can outperform a single classification technique with a significant margin. With hybrid architecture two or more classification techniques can be integrated or combined to get cooperative effect where the strength of one technique can compensate for the weakness of another. One study uses NN, GA, and RI to mine classification rules from a database. This approach combines the robustness and search ability of GA with high predictive accuracy of NN and interpretability of rules to create a data mining system that outperforms systems based on a single technique. Other studies [10, 1] show that combining classifiers computed by different machine learning algorithms produces a meta-classification that has the best overall performance. In [3], a study of combing BP, naïve Bayesian (NB), and C4.5 algorithms was held to build a meta-classifier to improve cost saving in credit card fraud detection. Results show that this approach performs better than the base classifiers used in the combination, and outperform the common technique (BP) used in industry. Also, results show that partitioning the data set and using multiple algorithms approach achieves higher cost savings. Combination of rule-based and case-based systems can: Offer a first check against known cases before undertaking rule-based reasoning and the associated search cost; and it can record search-based results as cases for future use, which can avoid duplicating costly search. In addition, it might be advantageous to use multiple algorithms for classification of the same problem. This will increase confidence in the results when predictions from two models are identical, and appropriately raising a flag when the two models disagree. An evaluation comparison between the above techniques is given in table 2 bellow. H(igh), M(edium), and L(ow) indicate the expected ranking for each of them.

V. CONCLUSION

In this paper, credit card fraud detection is briefly discussed. It presents the fraudulent transactions scenario of card-not-present, the various detection techniques of credit cards, and an evaluation of these techniques. Fraud detection systems for these types of fraud can be seen as a problem of classification of legitimate transactions from the fraudulent transactions.

Several data mining techniques used in different organizations for credit card fraud detection are briefly presented in this paper. Our study shows that the efficiency and performance of these techniques depend on there capability in dealing with several problems such as: noisy and missing in the data used, the performance measure used, scalability, different data types used, explanation capability of the technique, ease of integration with other systems, ease of operation, and skewed distribution of the data used.

Our study shows that the typical fraud detection techniques attempt to maximize accuracy rate and minimize false alarm rate at an acceptable level of cost. Also, it shows that hybrid or meta-learning classifier techniques can outperform a single classification technique with a significant margin. In this case the strength of one technique can compensate for the weakness of another.

Also, the study shows that partitioning the data set and using multiple algorithms approach achieves higher cost savings. In addition, it might be advantageous to use multiple algorithms for classification of the same problem. This will increase confidence in the results when predictions from two models are identical, and raising a flag when the two models disagree.

TABLE 2: A COMPARISON BETWEEN THE ABOVE DETECTION TECHNIQUES.

Characteristics	RI	NN	CBR	GA	ILP	ES	Regression
Ability to handle noisy data	M	L	M	L	M	M	H
Ability to handle missing data	M	M	L	M	H	M	H
Process large data sets (Scalability)	L	M	M	M	M	M	H
Process different data types	M	M	L	M	H	1	M
Predictive accuracy	M	L	M	M	H	M	M
Explanation capability	L	H	L	M	L	L	L
Ease of integration	M	M	L	L	L	L	M
Ease to build and operate	M	H	L	H	H	L	L

REFERENCES

[1] Andreas L. Prodromidis and Salvatore J. Stolfo; "Agent-Based Distributed Learning Applied to Fraud Detection"; Department of Computer Science- Columbia University; 2000.

[2] Carolyn Mair, Gada Kadoda, Martin Lefley, and others; "An investigation of machine learning based prediction systems"; The journal of System Software 53; 2000; pp. 23-29.

[3] Clifton Phua, Damminda Alahakoon, and Vincent Lee; "Minority Report in Fraud Detection: Classification of Skewed Data"; Sigkdd Explorations, Vol. 6, Issue 1, p.p. 50-51.

[4] Dean W. Abbott, I. philip Matkovsky, John F. Elder; "An Evaluation of High-end Data Mining Tools for Fraud Detection; Elder Research – San Diego. CA 9212; oct. 1998.

[5] Jesus Mena; "Investigative Data mining for Security and Criminal Detection"; B. H. pub. Company; 2003

[6] Kevin J. Leonard; "The development of a rule based expert system model for fraud alert consumer credit"; European journal of operational research, vol. 80, p.p. 350-356; 1995.

[7] Maes S. Tuyls K. Vanschoenwinkel B. and Manderick B.; "Credit Card Fraud Detection Using Bayesian and Neural Networks"; Vrije University Brussel – Belgium; 2002.

[8] R. Brause, T. Langsdor-f, M. Hepp; "Neural Data Mining for Credit Card Fraud Detection"; Goethe-University-Frankfurt; 1999.

[9] R. Wheeler, S. Aitken; "Multiple algorithms for fraud detection"; Knowledge-Based Systems 13; 2000; pp. 93-99.

[10] Salvatore J. Stolfo, David W. Fan, Wenke Lee and Andreas L. Prodromidis; "Credit Card Fraud Detection Using Meta-Learning: Issues and Initial Results"; Department of Computer Science- Columbia University; 1997.

[11] Salvatore J. Stolfo, Wei Fan, Wenke Lee and Andreas L. Prodromidis; "Cost-based Modeling for Fraud and Intrusion Detection: Results from the JAM Project"; Department of Computer Science- Columbia University; 0-7695-0490-6/99, 1999 IEEE.

[12] Zan Huang, Hsinchun Chen, Chia-Jung Hsu, Wun-Hwa Chen, Soushan Wu; "Credit rating analysis with support vector machines and neural networks: a market comparative study"; Decision Support Systems 37 (2004); pp. 543-558.