Abstract

The change of payment method has promoted Internet banking, credit card, and mobile e-commerce in many countries. The mobile payment market is continuously increasing due to the popularity of smartphones in particular. However, the growth of non-face-to-face payment market has led to increased damage from abnormal payment, and many damages caused by attacks such as smishing have recently been reported in mobile transactions. Nonetheless, there are few studies on fraud detection in mobile payment services. The research and development of machine learning-based technology to detect abnormal transactions are needed to develop a payment system that can defend against intelligent attacks. This paper reviews the research trend of detection of abnormal transactions in the financial sector, analyzes the abnormal transaction patterns of mobile-based small-amount payment system. Then, this paper proposes the abnormal transaction detection technology based on decision tree machine learning, which can detect 86% of abnormal transactions in the mobile small-amount payment system.

I. Introduction

As the means of payment in many countries, many payment/settlement services are transforming from offline, face-to-face, cash-based transactions to non-face-to-face, online payment/settlement services based on various online media such as Internet, ATM, and mobile[22,23]. According to the “2015 Result of Survey on Usage Pattern of Means of Payment and Implication” published by the Payment and Settlement Systems Department of the Bank of Korea, the most widely used means of payment was the credit card, which accounted for 39.7% of the total. Small-amount payment with mobile phone also increased by as much as 9.4%, and the mobile payment market continues to grow [24].

As non-face-to-face transactions become more popular, the need to study the detection of abnormal transaction in the electronic financial transaction area has increased. FDS (Fraud Detection System) comprehensively analyzes the terminal data used in electronic transaction and the transaction details to detect suspicious transactions and frauds. FDS monitors financial transactions, establishes rules to detect suspicious transactions based on the analysis of past fraud cases, and detects and blocks frauds based on such. In that case, it is difficult to detect new types of abnormal transaction quickly. Therefore, studies applying data mining, machine learning, fuzzy logic, and others to detect suspicious transactions have recently been conducted. Such studies mainly focused on fraud detection technology in the credit card area, which has long been in use.

With the popularization of smartphone and advancement of mobile communication services, smartphones are more widely used for financial transactions such as mobile banking and online/offline payment. Korea in particular has the world’s 4th highest smartphone distribution rate, with 8 out of 10 Korean adults using smartphones [25]. The mobile small-amount payment service, which allows payment with the phone number and date of birth of the user instead of financial information, was particularly popular among various mobile-based services in Korea. As a result, considerable monetary damages including 29,761 smishing cases and loss of USD 5.18 million were recorded in 2013 [26]. Such shows the need for studies on the development of fraud detection in mobile-based payment service, but there are only a few.

This paper is consisted as follows. In Section 2, we describe the concept of abnormal transaction detection system and machine learning. In Section 3, we review the trend of studies of machine learning-based abnormal transaction technology in the financial sector. Also, we describe the mobile small-amount payment service and analyzes patterns of abnormal transactions in actual mobile-based small-amount payment system in Section 4. Then, we propose the applicable machine learning technology and test the performance of the technology in section 5. Finally, we conclude in Section 6.

II. Explanation of Fraud Detection system and Machine Learning Technology
In this section, the abnormal transaction detection system and machine learning technology will be presented.

A. Fraud Detection System

Fraud Detection System (FDS) is a system of analyzing the terminal data, IP address, and transaction details used in electronic financial transactions to detect suspicious transactions and block frauds. FDS consists of four functions: data gathering, analysis and detection, response, and monitoring and audit [27].

Each function of FDS can be described as follows:

**Data gathering function**: The user medium environment data and accident-type data are collected for accuracy of fraud detection.

**Analysis and detection function**: The data for each user type and transaction type are analyzed, and abnormal behaviors are detected through correlation analysis and rule inspection. Abnormal transactions are detected with the misuse detection model and abnormality detection model using the collected data.

- Misuse detection model: The model detects abnormal behaviors based on the known signatures of illegal behavioral patterns in the past.
- Abnormality detection model: The model detects behaviors that can cause rapid change or generate behaviors with low probability based on normal financial transactions. The data mining technique can be applied.

**Response function**: The response is carried out, such as blocking the financial transaction according to the analysis result of the requested transaction.

**Monitoring and audit function**: It executes the monitoring function, which manages all processes such as collection, analysis, and response, and the audit function that audits various factors intruding the detection system.

B. Machine Learning Technology

Machine learning technology creates the model by learning a large volume of data and predicts future behavior with the model. As an area of artificial intelligence (AI), it develops the algorithm and process to learn the data repeatedly. Thus, machine learning includes the classification and clustering of data. Machine learning is mainly divided into supervised learning, unsupervised learning, and reinforced learning.

**Supervised learning**: Supervised learning is one of the methods of mechanical learning to deduce the predictive function from training data. The training data include the attribute information of input object and indicate the value to be predicted (fraud or non-fraud) of each data. This method is mainly used to classify the data, and it includes the decision tree [14], random forest [20], support vector machine [20], neural network [4], and KNN [5], etc. Each algorithm is described as follows:

- Decision tree [14]: The tree-type detection model links the observed value and target value of an item.
- Support vector machine (SVM) [20]: The non-probability detection model judges to which category the new data belong based on the given data set.
- Random forest [20]: The model categorizes the data or outputs the average predictive value from multiple decision trees.
- Neural networks [4]: The model mimics the parallel-moving neural network to recognize patterns and find hidden facts.
- K-nearest neighbor (KNN) [5]: The model determines the characteristics of input data according to the K number of training data within a specific space.

**Unsupervised learning**: Unsupervised learning is generally used for data analysis without prediction since the attributes of input data are not provided and the function cannot be inferred using the training data. Its main purpose is clustering. In other words, it can group the input data but does not output the characteristic (fraud or non-fraud) of each group. This type includes the neural network [4], K-means clustering [18], and self-organizing map [1], etc. Each algorithm is described as follows:

- Neural networks [4]: The model mimics the parallel-moving neural network to recognize patterns and find hidden facts.
- K-means [18]: The model clusters the given data into K clusters. It measures the distance between K centroids randomly selected at the initial step and each individual data to allocate the data to each cluster and repeats the process for clustering.
- Self-organizing map [1]: The model clusters through self-learning with an artificial neural network that models the learning process of visual cortex of the cerebral cortex.

**Reinforcement learning**: Compensation is given from the external environment for each step of learning, and learning progresses in the direction of maximizing compensation by learning the optimal action in the current state. The method is applied to computer chess game and complex robot control.

III. Trend of Studies on Machine Learning-Based Fraud Detection in Finance

In this section, preceding studies on machine learning technology to detect abnormal transactions in the financial sector will be described. Studies on machine learning technology to detect fraud have mainly focused on the credit card area, but extension to other areas is possible. Therefore, in this paper, we review the trend of studies on fraud detection in the credit card area.
As mentioned in Section 2, machine learning technology can be divided into supervised learning and unsupervised learning according to the type of learned data. Cases of financial accidents are collected and analyzed to establish the rule for detecting fraud in financial transactions. Due to such environmental factors, early studies on machine learning detection of abnormal transactions mostly focused on supervised learning based on data with fraud/non-fraud identified. However, there have been recent studies on unsupervised learning since non-face-to-face mobile-based transactions are increasing and fraud or non-fraud identified data may be too small or even nonexistent realistically according to the type of financial transaction. Therefore, we seek to analyze the machine learning-based fraud detection technology according to the learning type of machine learning algorithm in this section.

In the financial area, there have been many studies on detecting fraud using the decision tree, random forest, logistic regression, KNN (K-nearest neighbor), and support vector machine among the supervised machine learning technologies. K-means and SOM (Self-Organized Map) among the unsupervised machine learning technologies were also studied. Table 1 summarizes the research trend followed by a brief description.

### Table 1. Trend of Studies on abnormal transaction Detection Technology

<table>
<thead>
<tr>
<th>Type</th>
<th>Algorithm</th>
<th>Reference No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised machine learning</td>
<td>Logistic Regression</td>
<td>[20][2][14][9][11][7]</td>
</tr>
<tr>
<td></td>
<td>Decision Tree</td>
<td>[20][14][9][12][4]</td>
</tr>
<tr>
<td></td>
<td>Random Forest</td>
<td>[20][2][9][11]</td>
</tr>
<tr>
<td></td>
<td>KNN(K-nearest Neighbor)</td>
<td>[20][9][11][5]</td>
</tr>
<tr>
<td></td>
<td>neural networks</td>
<td>[14][7][4]</td>
</tr>
<tr>
<td></td>
<td>Naive Bayes</td>
<td>[20][4]</td>
</tr>
<tr>
<td></td>
<td>Bayesian Network</td>
<td>[4]</td>
</tr>
<tr>
<td></td>
<td>Support vector Machine</td>
<td>[20][2][9-10]</td>
</tr>
<tr>
<td></td>
<td>ARIS(Artificial Immune System)</td>
<td>[4]</td>
</tr>
<tr>
<td>Unsupervised machine learning</td>
<td>K-means</td>
<td>[18,19]</td>
</tr>
<tr>
<td></td>
<td>DBSCAN</td>
<td>[19]</td>
</tr>
<tr>
<td></td>
<td>AGGOLMERATIVE</td>
<td>[19]</td>
</tr>
<tr>
<td></td>
<td>SOM(Self-Organised Map)</td>
<td>[1,13,21]</td>
</tr>
<tr>
<td></td>
<td>PCA</td>
<td>[8]</td>
</tr>
<tr>
<td></td>
<td>Simple K-means</td>
<td>[8]</td>
</tr>
<tr>
<td></td>
<td>Hidden Markov Model</td>
<td>[3],[6],[15],[16]</td>
</tr>
</tbody>
</table>
The machine learning technology can also be applied to mobile banking. Min, et al suggested detection of fraud in mobile banking using the user input pattern when using the mobile banking service and the transaction pattern [10]. Cloud funding using the social network is becoming popular as one of the fin-tech measures enabling the users to carry out financial transactions without the financial institutions. Malekipirbazari, et al suggested the method of detecting abnormal transactions by applying kNN, logistic regression, SVM, and random forest to the transaction data generated by Lending Club between 2012 and 2014 [11].

B. Studies on Unsupervised Machine Learning

In 2006, Zaslavsky, et al proposed the application of SOM (Self-organized map) machine learning algorithm to detect abnormal transactions. They showed that abnormal behaviors can be detected after analyzing the payment patterns or consumers [21]. Bansal, et al applied the SOM machine learning algorithm and tested it to detect fraud in credit card transactions. They verified the performance by comparing it with ID3, which was a type of decision tree algorithm [1]. Quah, et al applied the SOM machine learning algorithm to develop the detection mechanism to analyze fraud behaviors and real-time detection technology [13]. Vaishali studied the abnormal transaction detection using the K-means machine learning technology, conducted the performance test based on simulated credit card data, and analyzed the performance of detecting abnormal transactions based on 5 credit card data [18].

Lepoire used the PCA (Principle Component Analysis) technique and simple K-means method to detect abnormal credit card transactions. The study used the user and customer location data to increase detection accuracy. The application of the simple K-means algorithm with K=2 to the bank system data without identified fraud and no-fraud showed 3% detection rate in the no fraud environment and 100% accuracy in the fraud environment [8].

Vadoodparast, et al proposed a methodology to detect abnormal transactions based on approximately 3.6 million cases of bank transaction data. The KDA model proposed in the paper is the system of ensemble architecture using K-means, DBSCAN, and AGGOLMERATIVE, which could detect 68.75% of fraudulent online transactions and more than 81.25% of fraudulent offline transactions [19].

DHOK performed detection of fraud credit card transactions using HMM (Hidden Markov Model). HMM could detect abnormal transaction with low false positive rate [15]. The profile was generated based on the consumption pattern of card users and was applied to HMM to divide the card payment amounts into three categories (low, medium, and high). The abnormality of credit card payment was judged using it, and the system showed 80% accuracy in detecting abnormal transactions.

Similarly, Bhusari, et al applied HMM to the credit card system and showed that it was the useful algorithm for detecting frauds [3]. Other studies also applied HMM to detect abnormal transactions [6, 16].

IV. Analysis of Abnormal Transaction Patterns in Mobile Small-Amount Payment System

In this section, we briefly describe the mobile small-amount payment system and analyzes the abnormal transaction payment patterns in the actual mobile small-amount payment system.

A. Mobile Small-Amount Payment System

The mobile small-amount payment service involves the users, mobile telecommunication company, and payment agencies. It is operated as shown in (Figure 1).

![Figure 1. A Process of Mobile Small-Amount Payment Service](image)

The mobile small-amount payment service enables users make simple payment using a mobile device when purchasing merchandises. It involves the mobile telecommunications company, which authenticate the users, and the payment agencies (PG) that handle the payment. It is called small-amount payment service because only up to 300,000 won can be settled at a time.

The service process can be described as follows:
1) A user selects an item to buy and requests to make payment.
2)The payment site provides the payment window upon request for payment by the user.
3)The user enters the mobile phone number and date of birth on the payment window.
4)The payment agency checks the user’s existence using the input data and sends the success message if the user exists or the fail message as applicable.
5)The payment agency checks the user’s existence with the mobile telecommunications base station, which sends the authentication number, terminates payment, or conducts
additional authentication according to the received message.

1. The authentication number is transmitted if the success message is received from the mobile telecommunications company and the fraud detection system of the payment agency indicates normal transaction.

2. The service is terminated if the fraud detection system of the payment agency indicates abnormal transaction even when the success message was received from the mobile telecommunications company.

3. If fraud is suspected even when the success message was received from the mobile telecommunications company, and the fraud detection system of the payment agency indicates normal transaction, additional ARS authentication using voice is conducted.

6. The user inputs the received authentication number on the payment window.

7. The payment agency checks if the number issued to the user is the same as the received authentication number.

1. If they are the same, the payment agency requests settlement to the mobile telecommunications company, which in turn adds the amount to the next mobile fee bill.

2. If they are not the same, the payment is not settled. The mobile small-amount payment service enables users to make simple payment using a mobile device when purchasing merchandises. It involves the mobile telecommunications company that authenticate the users and the payment agencies (PG) that handle the payment.

B. Analysis of Abnormal Transaction Patterns

To develop the ideal abnormal transaction detection system, we analyze the consumption patterns based on actual payment data. The data consist of more than 6 million payment data of payment agency A during the period 2013~2014 when there were many smishing attacks involving small-amount payment. The analysis of normal transactions and fraudulent transactions showed different patterns of transaction time, transaction date, transaction amount, cancellation, cancellation time, cancellation date, sales type, and difference between authentication time and transaction time.

- Transaction amount: The amount of less than 10,000 won accounted for 44% in the case of normal transactions, whereas the portion of 50,000 won was highest at 36% in the case of fraudulent transactions.

- Cancellation: The cancellation rate was 4.2% in the case of normal transactions but was 3.11 times higher at 13.08% in the case of fraudulent transactions.

- Cancellation date: Most cancellations occurred on Tuesday (21%) in the case of normal transactions and on Friday (48%) in the case of fraudulent transactions.

- Cancellation time: Most cancellations occurred at 16~19 o’clock (7%) in the case of normal transactions and at 16~21 o’clock (11%) in the case of fraudulent transactions.

- Sales type: The ratio of digital to physical object is 1:1 in the case of normal transactions but is 8:2 in the case of fraudulent transactions.

- Difference between authentication time and transaction time: The 10~20 sec. section was the largest at 20% in the case of normal transactions, whereas the 20~30 sec. section was the largest at 29% in the case of fraudulent transactions.

V. Proposal and Result

In this section, we suggest an approach for applying machine learning method in financial sector. Then, we propose the machine learning algorithms that can be applied to detect abnormal transaction in the mobile small-amount payment system. Then, we show the result of test applying the decision tree algorithm among existing fraud detection technologies.

A. Approach of Machine Learning Algorithm

As described in Section 3 on the trend of fraud detection, the following technologies are widely used for the detection of abnormal transactions in the financial sector:

- Studies on supervised learning: The algorithms generally used in studies on supervised learning-based detection of abnormal transactions are logistic regression, decision tree, random forest, and support vector machine. The decision tree algorithm offers high degree of understanding from the viewpoint of application of detection technology since the detection logic can be visualized. The support vector machine (SVM) offers high accuracy in categorization because of its structural feature of selecting the optimized standard to categorize the given data accurately.

- Studies on unsupervised learning: The algorithms generally used in studies on unsupervised learning-
based detection of abnormal transactions are SOM and K-means. The K-means algorithm categorizes the input data according to the K value setting. To separate normal transactions and fraudulent transactions, K=2 will be set for grouping.

As such, algorithms generally used for each method exist, and one or more algorithms can be applied in the development of systems to detect abnormal transactions. Since the development processes involve the analysis of actual data, selection of algorithm, and selection of features, acquiring the data and analyzing them are considered the most important part.

B. The result

Considering the above mentioned in Section 5.A, we select the decision tree algorithm among machine learning algorithms. Then, we test a performance of the system with 5,000 actual mobile small-amount payment data and showed 86% sensitivity in detecting abnormal transactions.

Test overview: This test deduced two field values (payment amount and country of payment) based on 8 fields including the authentication request time, day of week, date, and tele-communication company out of 21 small-amount payment data and used them to configure the features and applied the decision tree algorithm based on the values.

Testing environment:
- Payment data: 5,000 small-amount payment transactions in 2013–2014 (4,000 normal transactions and 1,000 abnormal transactions)
- Operation System: Windows 7(64-bit) with weka 3.6(java version)

Test result: The system detected 864 cases of 1,000 abnormal transactions, indicating sensitivity value of 86%.

This study is significant since it is the first case of applying the machine learning algorithm to the mobile small-amount payment system. Further studies in this area are needed.

VI. Conclusion

In this paper, we reviewed the trend of studies of machine learning-based fraud detection technology, analyzed patterns of abnormal transactions in actual mobile small-amount payment system, and proposed the machine learning technology that can be applied to the small-amount payment system. The test result of applying the decision tree machine learning algorithm in mobile small-amount payment data showed 86% detection of abnormal payment transactions. Considering the fact that there are few studies applying machine learning technology to the mobile small-amount payment system, this study is significant since it shows the possible direction for future studies.

Acknowledgments

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.R0132-16-1004, Development of Profiling-based Techniques for Detecting and Preventing Mobile Billing Fraud Attacks)

References

Biographies

EUNYOUNG CHOI received the B.S. degree in Mathematics from Korea University, Korea in 2001, the M.S degree in Information Security from Korea University, Korea, in 2003, and the Ph.D. degree in Information Security from Korea University, Korea, 2009, respectively. Currently, She is a general researcher at Korea Internet & Security Agency in Korea. Her research areas include mobile security, cyber financial security, cryptography and information security. Dr. Eunyoung Choi may be reached at bluecey@kisa.or.kr

WOONG GO received his B.S, M.S and Ph.D degree in Information Security from Soochunhyang University, South Korea, in 2008, 2010 and 2013, respectively. Currently, he is a deputy general researcher at Korea Internet & Security Agency in Korea. His research areas include mobile security, cyber financial security, artificial intelligence, malware security and information security. Woong Go may be reached at wgo@kisa.or.kr

TAEJIN LEE received the B.E. degree in Computer Engineering from Pohang University of Science and Technology, Korea, in 2003, the M.S. degree in Computer Science from the University of YONSEI University, Korea, in 2008. Currently, he is a team manager at Korea Internet & Security Agency in South Korea. His research areas include network/system intrusion prevention, cyber threat intelligence, malware security, mobile security, web security, and cyber financial security. Taejin Lee may be reached at tjlee@kisa.or.kr

A STUDY ON FINANCIAL FRAUD DETECTION METH-OD USING MACHINE LEARNING IN MOBILE SMALL-AMOUNT PAYMENT SERVICE