International Journal of Security and Its Applications
Vol. 10, No. 3 (2016), pp.429-438
http://dx.doi.org/10.14257/ijsia.2016.10.3.37

SVDD-Based Financial Fraud Detection MethodThrough Respective Learnings of Normal/Abnormal Behaviors

Mun-Kweon Jeong, Seong-Ho An and Kihyo Nam

UMLogics Co., Ltd., 17, Techno 2-ro, Yuseong-gu, Daejeon, Republic of Korea
{jmk, ash, nkh}@umlogics.com

Abstract

This thesis proposes a method to detect financial fraudby dividing users' financial transactions into a normal area and an abnormal area,using SVDD and train the areas as such fraud evolves in terms of complexity. The existing financial industry detects electronic financial frauds using FDS, but its false positive rate is high enough to require additional authentications of user information. It causes customers inconveniences and does not always detect those sophisticated financial frauds. In order to resolve the aforementioned issues, this study proposes a method to detect such potential frauds by profiling user financial transaction data including user activities, device information, andtransaction patterns and vectorizing them into a normal area and an abnormal area using SVDD.

Keywords*: FDS, SVDD*

1. Introduction

The financial industry can stay as a going concern by maintaining steady customer relationships based on trust.It is the very first step to establish trust with customers that the industry should create safe financial environments to protect customers. But, as electronic financial services like Internet banking and phone banking have expanded and non face-to-face transactions have grown faster due to the popularity of FinTech, the number of financial frauds has risen as well.So far,regulatory agencies like Korean Financial Supervisory Service and financial institutes have made a lot of effort to promote electronic finance and the feasibility of its system.However, electronic financial frauds like voice phishing, pharming, smishing, and memory hacking have also matured, and recent criminal activities have become sophisticated enough to generate new victims. [1]

In order to handle increasing electronic financial frauds, the Korean Financial Supervisory Service encourages the financial industry to set up FDS (Fraud Detection System)[2] to protect customers and prevent any possible incidents of electronic financial transaction. Credit card companies have already set up FDS, and it has turned out to be effective. Because of this, the Korean Financial Supervisory Service required the financial industry to set up FDS in 2014.[2]

Analysis ability to use inside knowledge is absolutely needed to effectively operate FDS, but Korean banks still have few operation experiences and have not accumulated enough history to detect electronic financial frauds. It leads to a high rate of false positive, more customer inconveniences, and more customer complaints. Additional costs are incurred on more rigorous customer verification including additional authentications or outbound calls due to those false positives.[1]

Higher sensitivity of detection rules of financial frauds results in a wider area of detection, but more false positives lead to more customer complaints. It is very important to create an algorithm that minimizes false negatives with no false positives. This thesis proposes an SVDD-based financial detection method for ever-evolving financial frauds.

ISSN: 1738-9976 IJSIA

It profiles users' normal activities and abnormal activities and vectorizes them using SVDD to minimize customer inconvenience.

2. Data Profiling of Financial Transaction

2.1 Research Status of Detection of Financial Frauds

Much research on detection techniques of financial frauds has been done in fields related to credit card and insurance fraud.In reference to Nagi's work[3], many of the previous research about bank fraud are algorithm research about effective classification.According to Jha's work [4], research of detection of financial frauds can be categorized into statistic fraud detection, rule-based detection, and data mining-based detection.

Artificial intelligence detection method can be classified into data mining, pattern recognition, and machine learning. The detection method using data mining sorts or assembles data, and then automatically breaks it down into invalid transaction and valid transaction. [4] The method using pattern recognition and machine learning identifies frauds (i.e. invalid transactions) by figuring out characteristics of valid transactions and invalid ones based on transaction information. Leading methods are ones using artificial neural network and Bayesian modeling. [5]

SVDD, a method using data mining, is a derived form of nonlinear SVM (Support Vector Machine) and is the most common method for one-class classification. It is a one-class classification to identify invalid vectors in a specified vector group, first proposed by David [6]. This method, as shown below in Figure 1, maps data into higher dimension vector space through nonlinear transformation and produces the one-class classification of the same characteristics by producing boundaries of a hypersphere with the minimum radius that contains all of the mapped data. SVDD searches for a hypersphere and maps data in the hypersphere, so it is important to create an optimized hypersphere that contains most mapped data with the minimum radius.It is important to find a hypersphere with its center a and radius $\mathcal{R}$ in SVDD. Suppose the researchers have a hypersphere with the set D, the hypersphere needs to be as small as possible and also needs to contain as many mapped data as possible. Once the hypersphere is formed with the mapping, SVDD determines validity by spotting incoming data inside the hypersphere.

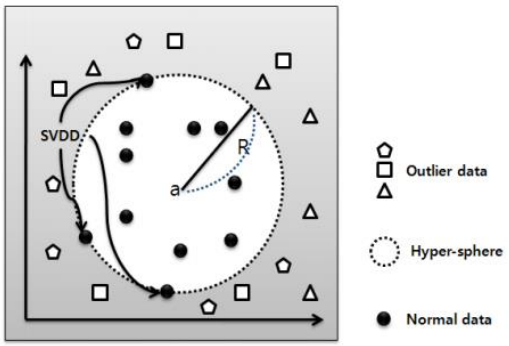


Figure 1. Basic Concept of SVDD

2.2 Financial Transaction user Action-based SVDD Profiling

The researchers produced a hypersphere of its center a and radius $\mathcal{R}$ with profiled data sets of valid financial transactions and invalid ones through SVDD learning (or training). They profiled subsequent transactions and determined validity of those subsequent ones by comparing a normal area and an abnormal area generated through SVDD learning. For

this, they extracted feature values by profiling information from financial transactions by each user and create vectors for learning.

In previous research, the researchers came up with customers' pattern information and profile information from electronic financial incidents in a Korean bank and set a detection rule based on the output. [7] They measured a starting coordinate, an ending coordinate, an inclination of fingers, and a scroll speed on the touchscreen of a smartphone and created a detection rule by applying them to a data-mining algorithm. [8]

In this research, the researchers have Table I, which contains information including financial transactions and setting information like device information based on the data used in the previous researches. Table I also includes profiling information to handle potential invalid transactions. Data from the transaction information, device information, and activity information is used as feature values for a normal area or an abnormal area of SVDD.

Table 1. User Financial Transaction Information Profile

Type	Variable	Remarks
Transaction information	Transfer_location	Location of transfer
	Start_time	Start time of transfer
	Duration	Duration of transfer
	Bank_code	Bank code of transfer
	Bank_account	Whether the account has been used before
	Telephone_authentication	Whether the telephone authentication is successful per security rules
	Transfer_amount	Whether the amount is in the range of previous transfer amounts
Device information	Device_information	Unique information of the device(PC/smartphone)
	OS	For PC or smartphone
	IP_address	Location information of the device
	MAC_address	Unique information of the device
	Country_code	Area code where the transaction took place
	Proxy_IP_address	Whether a proxy has been used
Activity information	Movement_speed	Usual user behavior
	Keyboard_typing_speed	Usual user behavior
	Failed_password_attempts	Determine invalidity regarding consecutive failed attempts
	Secret_card	Determine invalidity regarding consecutive failed attempts
	Certificate_used	Usual user behavior
	ID_PW	Usual user behavior

3. Detection of user Financial Frauds using SVDD

3.1Procedures to Detect Financial Frauds

(Step 1)Creating a user profile vector of legitimate financial transactions

Compile n units of profile information of legitimate financial transactions based on Table 1. Produce one vector $\mathbb{X}a$ that includes n numbers converted from the n units of profile information.

$$\mathbb{X}a = \{x_1, x_2, x_3, \cdots, x_n\}$$

Repeat the step κ times and produce a vector group $\mathcal{D}a$ that contains κ$\mathbb{X}$ vectors of legitimate financial transactions.

$$\mathcal{D}a = \{\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_{3,} \cdots, \mathbb{X}_\kappa\}$$

(Step 2) Creating a user profile vector of illegitimate financial transactions

Compile n units of profile information of illegitimate financial transactions based on Table I. Produce one vector $\mathbb{X}b$ that includes n numbers converted from the n units of profile information.

$$\mathbb{X}b = \{x_1, x_2, x_3, \cdots, x_n\}$$

Repeat the step κ times and produce a vector group $\mathcal{D}b$ that contains κ$\mathbb{X}$ vectors of illegitimate financial transactions.

$$\mathcal{D}b = \{\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_{3,} \cdots, \mathbb{X}_\kappa\}$$

(Step 3) Defining an area of legitimate financial transactions and another area of illegitimate financial transactions through SVDD learning.

Produce an optimized hypersphere by running the SVDD algorithm with the vector group $\mathcal{D}a$ from the step 1. The results, center a and radius $\mathcal{R}a$, are bases to detect legitimate financial transactions. Produce another optimized hypersphere by running the SVDD algorithm with the vector group $\mathcal{D}b$from the step 2.The results, center b and radius $\mathcal{R}b$, are bases to detect illegitimate financial transactions.

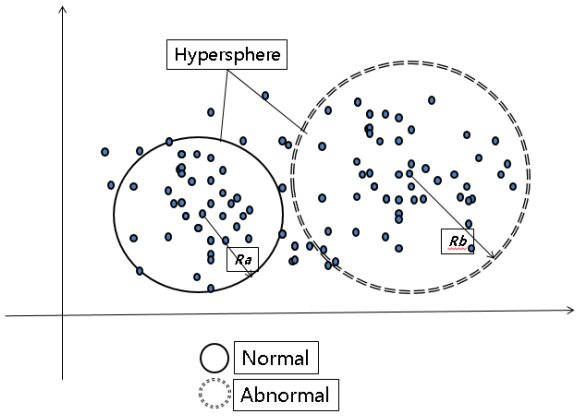


Figure 2. Hypersphere Creation through SVDD Learning

(Step 4) Detecting financial frauds

The vector z of users' financial transactions is applied to SVDD based on the vector groups $\mathcal{D}a$ and $\mathcal{D}b$ to determine its validity.The researchers need to calculate an area value of the vector z and find distances from the center of the vector groups $\mathcal{D}a$ and $\mathcal{D}b$, respectively.They then apply the distances to the following formulae.

The researchers apply the following formula of the normal area to determine whether the vector z is valid (i.e. legitimate).

$$\| z - a \|^2 = K(z,z) - 2\sum_i a_i K(z, d_i) + \sum_{i,j} a_i a_j K(d_i, d_j) \leq \mathcal{R}a^2$$

The researchersalso apply the following formula of the abnormal area to determine whether the vector z is invalid (i.e. illegitimate).

$$\| z - b \|^2 = K(z,z) - 2\sum_i b_i K(z, d_i) + \sum_{i,j} b_i b_j K(d_i, d_j) \leq \mathcal{R}b^2$$

If the vector z is out of range from both the normal area and the abnormal one, the researchers apply the following formula to determine its validity.

$$\| z - a - Ra \|^2 - \| z - b - Rb \|^2 \leq 0$$

If it satisfies the condition, it is valid. Otherwise, it is invalid.

4. Performance Evaluation of SVDD-based FDS

4.1. Evaluation Criteria

4.1.1 Evaluation Method

To evaluate the financial fraud detection method proposed in this thesis, the researchers found a hypersphere with its center a and radius $\mathcal{R}a$. The hypersphere is a normal area created based on users' profile data by SVDD. They also found another hypersphere with its center b and radius $\mathcal{R}b$. The hypersphere is an abnormal area created based on users' profile data by SVDD. Applying n vectors of legitimate transactions and m vectors of illegitimate transactions, they then measure such a ratio that accurately determines validity.

4.1.2 Evaluation Data

For evaluation purposes, the researchers used500 legitimate financial transactionsto be "trained" through SVDD and 100 other normal data to determine validity of users' financial transactions. They used500 illegitimate financial transactions to be "trained" through SVDD and 100 other abnormal data to determine validity of users' financial transactions.

The researchers created 500 vectors of legitimate financial transactions and 500 other vectors of illegitimate financial transactions based on Table I. They created a hypersphere of normal area with center a and radius $\mathcal{R}a$ by running the SVDD algorithm on a vector group of those 500 legitimate vectors.They also created another hypersphere of abnormal area with center b and radius $\mathcal{R}b$ by running the SVDD algorithm on a vector group of those 500 illegitimate vectors.

Jae Hoon Park's thesis, "Effective Normalization Method for Fraud Detection Using a Decision Tree", mentions signs of possible illegitimate transactions regarding financial frauds as shown in Table II. The table is a result of analysis of patterns of actual 500 financial frauds in bank A since 2013.

Table 2. Pattern of FDS Construction

Classification	Item	Analysis
Transaction period of attacked users	Transaction Time	Midnight transaction (0 am ~ 4am)
	Initial / Final Transaction	Deviation from normal transaction period
Mediums	New medium	Access with new medium
	Number of mediums	Using 2 or more mediums for attack
	Local	Access from outside of usual local
Daily Transaction to other banks	Daily Transaction frequency	Exceeding daily transaction frequency limit
	Daily Transaction Amount	Exceeding daily transaction Amount limit
Remittance bank	Initial remittance bank	Transfer to unprecedented bank(more than 300,000 KRW)
Attacked Saving Account	Withdrawal account balance	Withdrawal minimum balance of Savings Account

The researchers created a vector $\mathbb{X}$ of potential financial frauds using and Table I, Table II, and normal data $\mathcal{D}$.

The researchers repeated the procedure 5 times and produced profile data for 5 users.

4.2. Evaluation Result

In this thesis, the researchers calculated feature values from profiles of users' activities and device information through financial transactions and vectorized the data into normal area and abnormal area. They had the data for each area trained by SVDD and used LIBSVM (Library for Support Vector Machines) [9] to empirically evaluate validity of financial transactions.

To evaluate detection accuracy of legitimate transactions, the researchers readied 150 vectors of legitimate transactions and 150 other vectors of illegitimate transactions. They then createdthree vector groups of those 150 vectors of illegitimate transactions in varying ranges from the Feature set. They used the data to find the detection accuracies and repeated it 5 times overall.

The researchers derived the results in Table III from the trained center a and radius $\mathcal{R}$ by SVDD.

Table 3. Test Results of Detecting Financial Frauds

	Legitimate transaction Detection accuracy (%)	Illegitimate transaction Detection accuracy (%) [Group A]	Illegitimate transaction Detection accuracy (%) [Group B]	Illegitimate transaction Detection accuracy (%) [Group C]
1st run	99	90	93	95
2nd run	98	86	92	98
3rd run	97	88	94	97
4th run	98	87	92	98
5th run	99	90	95	98
Average	98.2	88.2	93.2	97.2

The false positiverate for legitimate transactions is about 98.2%. The false positive rate for illegitimate transactions in each group is 88.2%, 93.2%, and 97.2%, respectively.

5. Conclusion

This thesis examinesa financial fraud detection method that performs with a normal area and an abnormal area trained by SVDD. The areas are created based on users' activities of financial activities and device information. Its main objective is to detect financial frauds that have kept growing more sophisticated and to minimize customers' inconvenience from false positives.The method runs SVDD for legitimate transactions and illegitimate ones separately to measure users' financial activities for validity in multiple areas. Its purpose is to minimize false positives in detecting financial frauds.To detect users' financial frauds, the method performs profiling of users' devices and activities and vectorizes them. It then trains them using SVDD to promptly detect financial frauds.

The process of research results is to create an optimized hypersphere based on the vector group $\mathcal{D}$ of profile of legitimate transactions and SVDD. It also creates another hypersphere based on illegitimate transactions. The researchers then measured a ratio

using n vectors of legitimate ones and m vectors of frauds. For evaluation purposes, they used 500 legitimate financial transactions to be "trained" and 150other vectors to evaluate and ran LIBSVM to execute the evaluation.

In this thesis, the researchers used a limited number of features of users' activities. If the researchers use all available feature data, including but not limited to user information, device information, and user activity as in real finance transaction scenarios, the method would detect financial frauds more accurately with an optimized hypersphere trained by SVDD. Future related researches could expand to coverdifferent businesses like electronic commerce and online gaming with corresponding feature data.

Acknowledgements

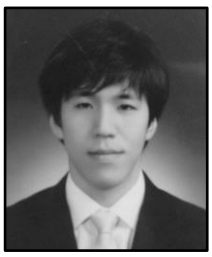
This work was supported by the Institute for Information &Communications Technology Promotion (IITP), grant funded by the government of Korea (MSIP) (No.R-20150521-001431, Fraud Detection System development by analysis of user action pattern in various web environment)

References

[1] E.-S. Choi, "A Study on Improvement of Effectiveness Using Anomaly Analysis rule modification in Electonic Finance Trading", Jounal of The Korea Institute of Information Security & Cryptology, vol. 25, no. 3, **(2015)**.
[2] S. H. Jeong, "A Survey of Fraud Detection Research based on Transaction Analysis and Data Mining Technique", Jounal of The Korea Institute of Information Security & Cryptology, vol. 25, no. 6, **(2015)**.
[3] E. Ngai, "The application of data mining techiques in financial fraud detection: A classification framework and an academic review of literature", Decision Support Systems, vol. 50, no. 3, **(2011)**.
[4] S. Jha, "Employing transaction aggregation strategy to detect credit card fraud", Expert Systems with applications, vol. 39, no. 16, **(2012)**.
[5] C. Phua, "A Comprehensive Survey of Data Mining-based Fraud Detection Research", Intelligent Computation Technology and Automation (ICICTA), **(2010)**, pp. 50 – 53.
[6] T. David and D. Robert, "Support vector data description", Machine Learning, vol. 54, no. 1, **(2004)**, pp. 45-66.
[7] J. H. Park, "Effective Normalization Method for Fraud Detection Using a Decision Tree", Jounal of the Korea Institute of Information Security & Cryptology, vol. 25, no. 1, **(2015)**.
[8] H. Y. Min, "Outlier Detection Method for Mobile Banking with User Input Pattern and E-finance Transaction Patter", Journal of Korean Society for Internet Information, vol. 15, no. 1, **(2014)**, pp. 157-170.
[9] http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Authors

Mun-Kweon Jeong
Master of Information Industrial Engineering(Chungbuk University)
Director-General of UMLogics Co., Ltd.
Republic of Korea



Seong-Ho An
Master of Science in Computer Engineering
Daejeon University
Republic of Korea

Kihyo Nam
Ph.D of Industrial Engineering(Korea University)
CISA(Certified Information Systems Auditor)
CISSP(Certified Information Systems Security Professional)
Adjunct Professor(Konkuk University)
Republic of Korea