International Journal of Industrial Engineering & Production Research (2018)

March 2018, Volume 29, Number 1
pp. 91- 101

DOI: 10.22068/ijiepr.29.1.91 http://IJIEPR.iust.ac.ir/

Detecting Frauds Using Customer Behavior Trend Analysis and Known Scenarios

Abdollah Eshghi & Mehrdad Kargari*

Abdollah Eshghi, *Industrial and systems engineering , Tarbiat Modares University*
Mehrdad Kargari, *Industrial and systems engineering , Tarbiat Modares University*

KEYWORDS

Fraud detection,
User profiles,
Fuzzy membership function,
Trend analysis,
Scenario

ABSTRACT

The present paper proposes a fraud detection method in which user behaviors are modelled through using two main components known as abnormal trend analysis component and scenario-based component. The extent of deviation of a transaction from customers' normal behavior is estimated using fuzzy membership functions. The results of applying all membership functions on a transaction will then be infused, and a final risk is determined as the basis for deciding whether to block the arrived transaction or not. An optimized threshold for the value of the final risk is estimated in order to strike a balance between fraud detection rate and alarm rate. Although the assessment of such problems is complicated, this method is shown to be useful in application according to several measures and metrics.

1. Introduction

Fraud in banking transactions has a negative impact on business, especially on electronic commerce development. There are several meanings for fraud; hence, financial fraud is also a broad term with various potential meanings. In the present study, financial fraud is defined as the intentional use of illegal methods or practices for the purpose of obtaining financial gain (Zhou and Kapoor, 2011). Banking fraud has many aspects, and the statistics about the fraud size vary widely. Billions of dollars of revenue are lost each year because of credit card fraud [1], and according to some reports, the total yearly cost imposed on the US is more than $400 billion [2].

Controlling and auditing all online transactions may aggravate the overall performance of online banking. This problem is more serious when the volume of transactions needed to be analyzed is huge, and rapid and online decision-making is required. On the other hand, as customers' behavior is ever-changing, modelling genuine behaviors is not an easy task. Thus, having an intelligent agile system for detecting fraud cases in transactions is an important necessity for any online bank.

For modelling customer behavior, a huge volume of historical data must be analyzed. Since the behavior of customers is not fixed and varies over time [3], the extracted behaviors must be re-calculated at short intervals of time [4]. The volume of data which should be analyzed is increasing over time. In addition, the data are not homogeneous and have various sources. In other words, most of the data needed for fraud detection analysis are semi-structured. For example, customers' profiles are not all the same, so handling and storing them in relational databases is labor-intensive and costly.

* **Corresponding author:** ***Mehrdad Kargari***
Email: m_kargari@modares.ac.ir
Received 18 August 2017; revised 25 February 2018; accepted 4 March 2018

Fraudsters always make their best efforts to have their behavior look legitimate making the process of fraud detection more complicated [5]. Fraudsters' behavior is based on a limitation: "A fraudster wants to gain the most benefit in the least time and with minimum risk" [6]. To this aim, they always make a transaction with a small amount initially, and if it is successful, they will make other transactions with greater amounts [4]; often, the maximum number of time a fraudster uses a card is two or three times. In addition to modelling customer behavior, modelling fraudster behavior is also needed; therefore, a genuine transaction must be as far from a fraudster pattern and as similar to his/her historical pattern as possible.

In this paper, customer behavior is modelled in different granularities. Using the principal attributes available in the dataset, the users' behavior is analyzed in three classes of "customer", "account", and "card" and three levels of "general", "business", and "individual". Next, the abnormal behavior for each class and level of each customer are modeled via fuzzy membership functions. The results of all abnormal membership functions are then infused by employing a proposed function. Fraudster' scenarios are also designed and a sequence of transactions is examined against these scenarios. At the end, an optimized threshold is gained for the final risk, in which transactions with a final risk are alarmed as fraud exceeding the threshold.

The remainder of the paper is organized as follows:

Section 2 is the literature review while section 3 describes an analysis of dissimilarity in behavior trends. Section 4 elaborates on the scenario-based component. In section 5, fuzzy functions for estimating risks are defined. Section 6 is the results and evaluation and, finally, the conclusion is presented in section 7.

2. Literature Review

While most of the real life fraud detection approaches adopt black box models [7] and practical implementations are rarely reported [5], the number of academic studies is remarkable. Most of the proposed fraud detection techniques apply a concept usually referred to as Outlier Detection. Outlier is an observation that is so different from other observations that it appears to have been generated by a different mechanism [8]; moreover, Outlier Detection refers to the process of finding outlier cases [9].

There is an increasing demand for robust and intelligent user profiling technologies [3], especially in banking business and fraud detection systems. Making profiles for user behaviors is a key technology in order to respond to the needs of real time fraud detection [3] because, according to [10], real users may gradually change their behavior over a longer period of time. In [11], the profiling method has been used for credit card fraud detection. The statistical representation of user behaviors is done through using profiles. Profile-based fraud detection systems always use thresholds for modelling. Although using such systems is easy and straightforward, to prevent the great number of false alarms prevalent in such systems due to the lack of support for different behavioral profiles between monitored accounts [12], it is necessary to calculate and assign profile thresholds for each user; hence, a proactive fraud detection system will be achievable [3]. According to [13], in order to decrease false alarm rates, the thresholds for different profiles of users must vary in time since both the legitimate and fraudulent behaviors of users change over time (e.g., interest rates, seasonal/monthly variations, new fraud attacks, etc.) [3].

Cahill [13] mentioned event-driven processing, memory, learning and self-initializing as the key elements of modern fraud detection systems. Event-driven processing facilitates the detection of fraud as it is happening rather than at fixed points in time unrelated to the account activity [3]. Memory means involving all past data in profile processing for a user (yet, not necessarily all with the same weights). Learning means the ability of the fraud detection system to adapt itself to new customer behaviors. Further, finally, self-initializing is the ability to have meaningful profile thresholds for newly opened accounts. Such issues can be solved to some extent through deploying profile-based methods.

As mentioned in [3], profile-based fraud detection methods can be categorized into two main processing models: time-oriented and action-oriented or event-driven. In time-oriented models, transactions are accumulated over a specified time period (e.g., hours, days); afterwards, batch processing will be done over them. In the event-processing model, new transactions are examined as soon as they arrive. Despite the fact that action-oriented models seem to produce better results in comparison to time-oriented models [14], the time-consuming process of reading and writing current profiles

and thresholds from and into data storage has justified the time-oriented approach, particularly for the applications in which time is not critical. However, in banking applications where time is critical, action-oriented methods may produce more false alarms.

In [14], a method has been proposed for calculating profiles. In this method, a window time of the fraud-less account activity is regarded as a base for calculating user profiles. Bolton and Hand [15] put forth a model in which break point analysis is used for detecting the trend of spending changes in customer behavior. Consequently, the aggregation of customer behavior over a series of feature variables is proposed. In [4], an experimental comparison between several algorithms has been drawn and some questions like "what algorithm to select for fraud detection?" and "what window time length is appropriate for updating the model?" have been addressed. In some other studies [16,17,18], in addition to feature variables, a sequence of actions is also considered as a base for calculating profiles. In [16], user profiles are calculated by building a contrast vector for each transaction based on its customer's historical behavior sequence, and an algorithm has been proposed for exploring contrasting patterns and distinguishing fraudulent patterns from genuine ones. In [17] and [18], a sequence of customer actions is regarded as a profile for the customer, and fraud cases are detected via Hidden Markov Model method. In [19], by combining anomaly detection and misuse detection models, a hybrid model has been propounded and the similarity of an incoming sequence of transactions to both fraudulent and non-fraudulent ones is determined through segment alignment.

With a huge growth of data over the last few years [20] and appearance of mobile and internet for doing banking transactions, fraud shapes and sizes have changed accordingly [21]. This phenomenon calls for developing new methods and tools for detecting fraud and other crimes against banks and customers [22]. In [21], a big data-based fraud detection product has been introduced.

A real-life system with the capability of action-oriented systems with a reasonable low rate of false alarm is the necessary requirement of fraud detection systems.

As it was induced from [23] and [5], there are only 10 studies to date claiming to have been implemented practically. Our research has also been implemented practically and is a semi-action-oriented system adjusting the false alarm rate to a reasonable and negligible level.

In the next section, the process of making profiles for users according to different attributes is discussed.

3. Analysis of Dissimilarity in Behavior Trends

In order to be able to analyze customer behavior trend for calculating the extent of dissimilarity, the informative profiles of customers are required to be extracted. The process is dividable into four main phases: making initial profiles, loading current profiles for an arriving transaction, calculating its deviation from the profiles, and fusing the results of all profiles for reaching a final belief (fraud or not-fraud). The process of analyzing transaction streams for fraud checking is as follows:

Making initial profiles: First, the profiles of customers are extracted through spark in an overnight processing, and the results are in another database, which is a NOSQL database. A profile consists of one or more customers' behavioral characteristics which could be extracted by analyzing their historical data. This process will estimate the normal and abnormal ranges for each characteristic and determine the related thresholds. The selected profiles for modelling the behaviors are depicted in Fig. 1. For instance, one of the profiles is "the daily trend of amounts of transactions at an individual level for card". Each profile consists of soft and hard thresholds for the abnormal behavior. This method for transaction aggregation was first used in [25] and continued to be employed in other studies such as [26] and [27].

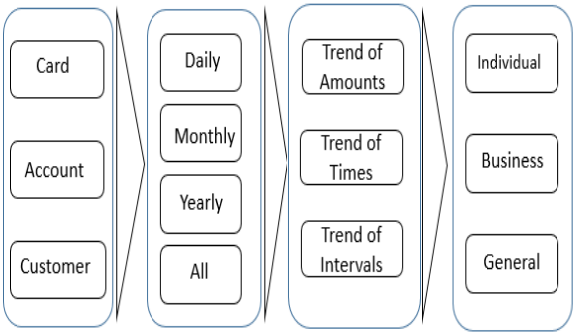


Fig. 1. Selected profiles for calculating behavior trends

Loading the current profiles: When a transaction arrives, its related profiles will be extracted and put in a buffer, which is a

temporary memory. Since the available data in the database for all arriving transactions are not the same, the number of extracted profiles will not vary: the more the available data, the more the extracted profiles. The profiles are categorized in three classes and three levels. The classes are card, account and customer, and the levels are individual, business and general.

Calculating profile deviation: As mentioned in the previous step, related profiles for the arriving transaction will be selected. The transaction will be examined against each of the selected profiles, and the extent of the deviation from normal behavior is estimated using fuzzy membership functions. The fuzzy membership functions are described in section 5.

Infusing the results: The output after calculating the profile deviation is a 3D matrix with n rows (a row for each profile), three columns (Card, Account, and Customer), and three heights (individual, business, general), as shown in Fig. 2. Card, account, and customer are considered as three main categories the profile thresholds of which may vary in level.

Customers may have several accounts. Each account may also have more than one card (because expired cards of an account have different card numbers). Each column of the matrix is a vector whose values are some numbers between 0 and 1, showing the deviation of the transaction from its historical profiles: for Card in column 1, Account in column 2, and Customer in column 3. The first page of the matrix is for general risks, the second is for business risks, and the third one is for individual risks. The number of calculated risk values for each arrived transaction is $n \times 3 \times 3$ risk values. The extracted risk values will be later infused to reach the final risk. The risks will be moderated when multiplied by their weights. Each person has some dominant behavioral characteristics. The weights of dominant characteristics are more than the others. To calculate the weights, the historical calculated risks for each behavioral characteristic will be regarded in a certain window of time (for example the customer's last 10 transactions). Since the thresholds are updated after each transaction, the previous risk values are expected to be low for genuine previous transactions, and if it is not so, the suggestion is that the customer's behavior of this profile is not certain or its profile has not been modeled correctly. In order to moderate the effect of this profile in the final estimated risk, we regard the weight as low, and vice versa. The formula for calculating the risks weight is as follows:

$$w_{P_i} = 1 - \frac{\sum_{j=1}^{j=n} R_{P_j}}{n} \quad (1)$$

where w_{P_i} is the weight for the ith characteristic of a customer profile. R_{P_j} is the risk of previous j transactions for the selected characteristic, and n is the window size. The result is always between 0 and 1. For example, if all the last n transaction risks are 0, then the weight will be 1 and, if all are 1, then the weight will be 0.

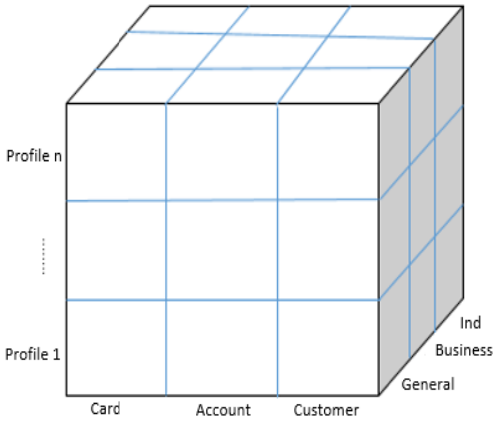


Fig. 2. Three-dimensional risk matrix based on selected profiles

The three-dimensional risk matrix was infused in 3 steps. The first step is column-by-column infusion. The process is depicted in Fig. 3. The estimated risks for profiles were divided into two main categories: strict risks and non-strict risks. Strict risks are those that unfold when some abnormal events occur (for example, if a card has two different transactions simultaneously). Non-strict risks do not necessarily occur when something bad happens; they are merely outliers and must be analyzed further to deciding whether they are bad or not.

Among strict risks, the result of the risk with the maximum value is selected. For non-strict risks, first, a list of risks which have a result more than a specific threshold (for example 0.5) is selected (the threshold is defined by experts), and their weighted average is calculated. Next, the result is multiplied by a soften factor. The soften factor formula is as follows:

$$sf = \frac{e^x - 1}{e^x} \quad (2)$$

Herein, x is the number of risks for which the risk values are more than the selected threshold. At the end, the maximum figure between strict and non-strict risks is selected as the final result.

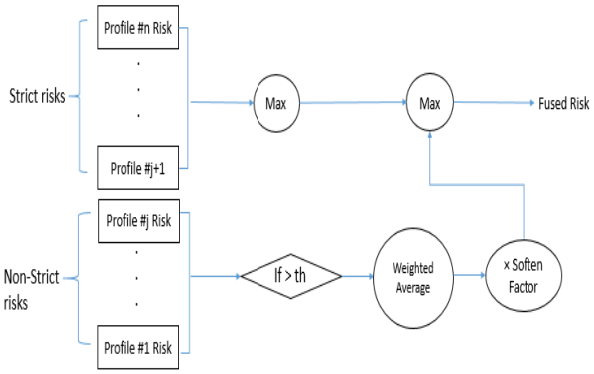


Fig. 3. The process of column-by-column risk infusion

After carrying out the infusion process for all columns, the result is a two-dimensional matrix. For each class, there is a fused risk for individual risk, one for business risk, and the other for a general fused risk. In the second step, the two-dimensional matrix is fused column by column by multiplying each level by its weight with the weights for individual, business, and general not being the same. Note that as mentioned in [9], the anomalies are categorized into three groups by point, contextual and collective anomalies. Individual, business, and general risks show point, contextual, and collective anomalies, respectively. The weights are assigned by system experts, and the final risk value for a category (e.g., Card) is obtained by calculating the average.

In the third step, fusion of 3 final risks of 3 main categories (Card, Account, and Customer) is calculated by assigning weights to each class and, then, calculating the weighted average. The final result is a number between 0 and 1.

4. Scenario-Based Component

Any fraudster seeks to make the most benefit in a short time with minimum risk [6], [28]. Based on this hypothesis, 7 scenarios are designed to be applied to arrived transactions. The list of scenarios is shown in Tab. 1.

Tab. 1. Fraud Scenarios

Scenario	Scenario Title
	Most Benefit
Scen1	Large cash withdrawals
Scen2	Relatively big sequential withdrawals
Scen3	Sequentially ascending withdrawals starting from low amounts
Scen4	Sequentially descending withdrawals starting from high amounts
Scen5	Small sequential withdrawals
	Shortest Time
Scen6	Sequential transactions at very small time intervals or simultaneous withdrawals
	Lowest Risk
Scen7	Withdrawals in non-common times

The terms used for the scenarios are defined below:

Big amount: This is an amount that is greater than the hard amount threshold calculated according to general historical data, business data or customer normal behavior data.

Relatively big amount: This refers to an amount standing between hard and soft amount thresholds and is calculated according to general historical data, business data or the customer normal behavior data.

Low amount: This is an amount less than the soft amount threshold for general, business, or customer categories.

Sequential transactions: Herein, sequential transactions represent those transactions that occur in time intervals less than or equal to the minimum time required for two transactions of a single account or card to be done sequentially, or those occurring at time intervals less than the customer normal behavior.

Non-common time: A non-common time is a time calculated as outlier time for a transaction according to general data, business data or the customer's previous normal data.

As illustrated in Fig. 4, a sequence of 4 transactions is regarded for each scenario. Each time a new transaction arrives at the system, the sequence will be updated and the 4 last transactions will be regarded as a new sequence.

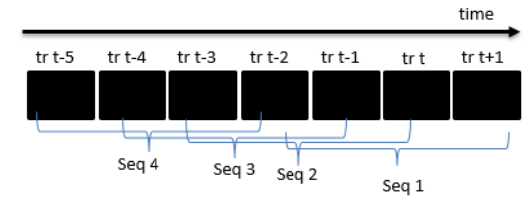


Fig. 4. Sequence of transactions for checking in scenarios

5. Fuzzy Functions for Estimating Risks

To profile the behavior of customers based on their historical transactions, the selection of relevant features for modeling behaviors is a particularly important task. Herein, as it is depicted in Fig. 1, three main obvious attributes (amount of transaction, time of transaction, and the time interval between transactions) in different periods of time and at 3 three levels for three classes are considered. According to the values of these attributes in historical transactions, a fuzzy membership risk function for each one is proposed to estimate the degree of dissimilarities.

5-1. Amount trend dissimilarity modeling

According to the customers' of spending amount habit, their outlier amounts will be specified and the thresholds will be assigned.

The risky fuzzy function for assessing the amount of risk of a transaction according to the feature amount is similar to Fig. 5. The risk of amounts smaller than the soft threshold is zero, and the risks of amounts between the soft and hard threshold are in accordance with a line and their values are between 0 and 1, while the risks of amounts are higher than hard thresholds are 1.

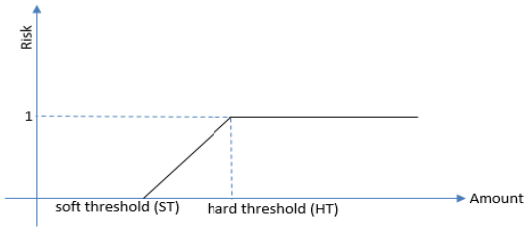


Fig. 5. Fuzzy membership function for risk of amount

To determine the thresholds (soft and hard), the box plot rule, which is a simple statistical technique, for detecting univariate and multivariate anomalies [9] is used.

Since low amounts (amounts far less than Q1) are not important for fraud detection, they were not considered in the amount of fuzzy membership function. To calculate the risk amount, the function below is applied:

$$\begin{cases} 1 & \text{amount} > \text{HT} \\ \dfrac{\text{trAmount} - \text{ST}}{\text{HT} - \text{ST}} & \text{ST} \le \text{Amount} \le \text{HT} \\ 0 & \text{o.w} \end{cases} \quad (3)$$

In the formula above, the trAmount is the amount of the transaction, and ST and HT are soft and hard thresholds, respectively. If the transaction amount is more than the hard threshold, then its risk is 1. For amounts between hard and soft thresholds, the risk is in accordance with the line connecting the two points with coordinates (0, ST) and (HT, 1). In other cases, the risks will be zero. It is natural that the steepness of the line and also the ST and HT are different for different users and even for a user in different categories and levels.

5-2. Time interval trend dissimilarity modeling

The time interval between sequenced transactions is another attribute regarded for profiling the behavior of customers. The sequenced transactions with small time intervals are more likely to be fraud; the transactions with time intervals out of customers' normal trend are also risky. The time interval risk fuzzy membership function is shown in Fig. 6.

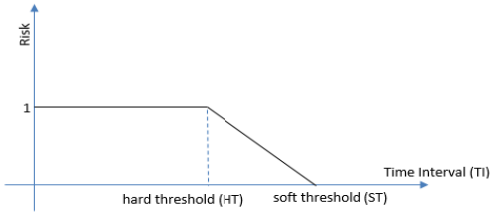


Fig. 6. Time interval risk fuzzy membership function

According to Fig. 6, the time interval risk membership function is shown below.

$$\begin{cases} 1 & TI < HT \\ \dfrac{TI - ST}{HT - ST} & \text{ST} \le \text{time} \le \text{HT} \\ 0 & o.w \end{cases} \quad (3)$$

5-3. Time of transaction trend dissimilarity modeling

The spending habit of customers at different times of the day is another important attribute which has been used as a profile in the present research. The time distribution of spending is different for different users. A customer purchase time distribution is shown in Fig. 7. In [27], the von Mises distribution is used for modelling the time distribution habit. The von Mises distribution is a distribution of a wrapped normal distributed variable across a circle [29]. Herein, a function for estimating the risks of transactions which are out of normal distribution of the times of transactions is employed. Fig. 8. shows the fuzzy membership function for time risk.

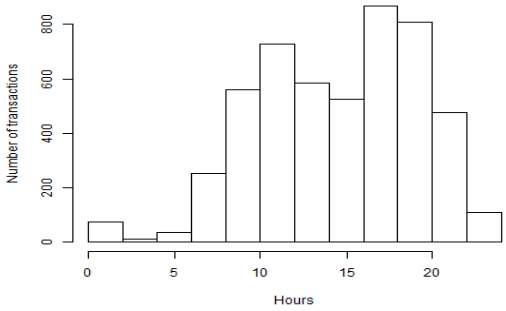


Fig. 7. Customer purchase time distribution

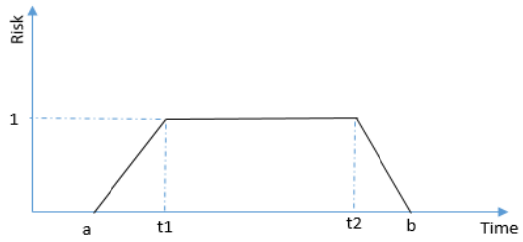


Fig. 8. Time-risk fuzzy membership function

The formula below demonstrates the fuzzy membership function for time risk. Parameters "a" and "b" are the left and right soft thresholds for time of transactions, and t1 and t2 are the left and right hard thresholds. When a transaction time is between t1 and t2, then its risk value is 1 and if the time of the transaction is between "a" and t1 or between t2 and "b", then the risk value is a number between 0 and 1.

$$\begin{cases} 1 & t1 \le t \le t2 \\ \frac{1}{t1-a}(t) - a & a \le t \le t1 \\ \frac{1}{t2-b}(t) - b & t2 \le t \le b \\ 0 & o.w \end{cases} \quad (4)$$

6. Results and Evaluation

We built the online banking risk management system. It incorporates both the trend-analysis-based and scenario-based models for online banking fraud detection.

Herein, it is intended to:

1- compare the performance and degree of accuracy of the scenario-based model with those of an existing trend analysis-based model.

2- find an optimized threshold for creating a balance between the emitted alarms and detection rate.

The evaluation of the system is difficult because, as mentioned in [7], the unsupervised analysis tool creates novel knowledge. In this paper, threshold parameters of customers for some attributes have been extracted and the risk of their behaviors has been modelled with some fuzzy functions. Then, these models have been applied to some transactions, and their high suspicious degree is certified with domain experts.

6-1. Data

The dataset used for evaluating our model in this research is the online banking transactional data from an Iranian bank. It is a sampled data of 600 customers and consists of 683100 genuine transactions and 910 fraudulent transactions during one year -from January 2015 to January 2016- and it is strongly unbalanced.

6-2. Experimental settings

As mentioned in [16], two metrics for evaluating the performance of an online banking fraud detection system are alert volume and detection rate. A fraud detection system attempts to reduce the number of alerts because every triggered alert has to be investigated manually for further investigation, and it is a labor-intensive work. On the other hand, the fraud detection system tries to increase the detection rate which is the percentage of detected fraud by the system. The modified ROC curve which is called the performance curve is another metric introduced by Hand et al. [30]. In this metric, minimizing the area under the curve is interpreted as minimizing the time needed for fraud detection [30] which is an important metric for online banking fraud detection (Fig. 10). The vertical and horizontal axes of the metric are cost and timeline, respectively, and are calculated as follows:

$$\text{cost} = \frac{\text{FP} + \text{TP}}{\text{number of frauds} + \text{number of notfrauds}} \quad (5)$$

$$\text{timeline} = \frac{\text{FN}}{\text{number of frauds}}$$

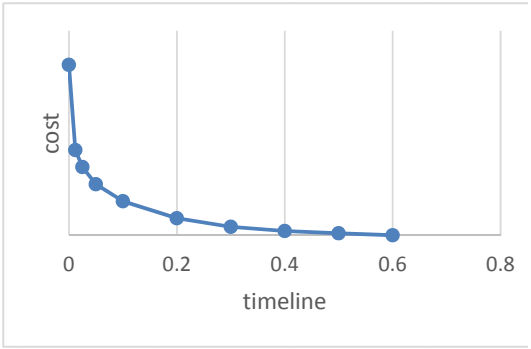


Fig. 9. An example of performance curve

In Equations 7 and 8, TP represents those cases correctly predicted as fraud, and FP and FN are those mistakenly predicted as fraud or non-fraud.

6-3. Overall performance evaluation

Fraud management system is comprised of two main components: trend analysis-based component and scenario-based component. The trend-analysis-based component investigates the input stream transaction by transaction, while the scenario-based component examines a sequence of transactions. The distribution of frauds detected by the trend-analysis-based and scenario-based components is shown in Fig. 11. By adjusting the optimized threshold to 0.8, these two components can find about 82 percent of the frauds. While the trend-analysis-based component can detect more frauds, the scenario-based component can detect frauds that are more complicated.

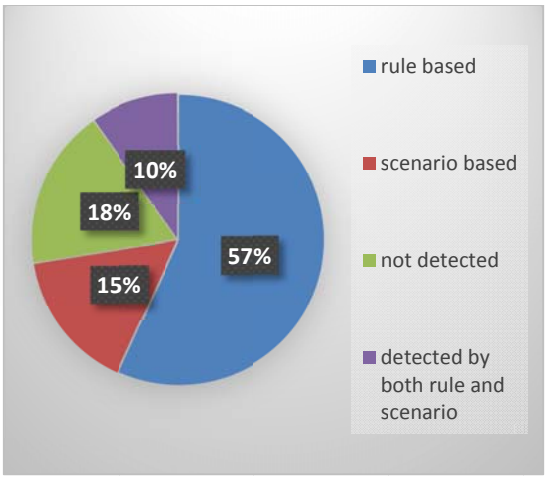


Fig. 10. Distribution of frauds detected by different fraud management system components

Fig. 11. illustrates the rate of detected frauds based on different thresholds in the infusion function. This figure shows the fraud detection rate for both the trend analysis-based system and composite trend-analysis-based and scenario-based models. While the results of these two components are almost the same in low thresholds, for higher thresholds, the composite trend-analysis-based and scenario-based have far better results. As induced from Figure 18, the best threshold for having the most detected frauds is less than 0.3. This, however, is not the only effective parameter. The number of alarms produced by any selected threshold has to be considered as well.

In Fig. 12., the fraud detection rate is shown under different alert volumes generated by the fraud detection system. As the figure demonstrates, when the alert rate is high, the detection rate is also high. However, as mentioned before, the high alert volume is both labor-intensive and also costly and causes the number of false positives to be high, which is not a desirable phenomenon.

There should exist some sort of balance between the number of produced alarms and the detection rate; a minimum number of false alarms and a maximum detection rate are, therefore, preferable.

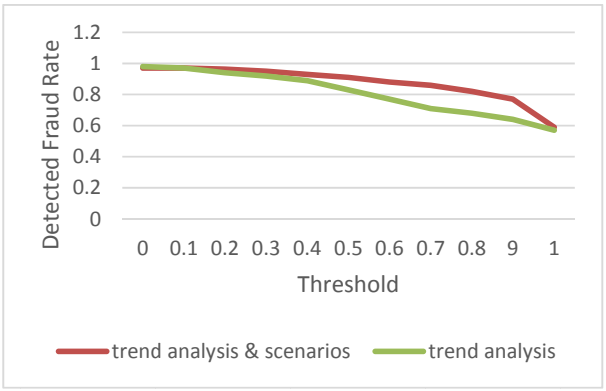
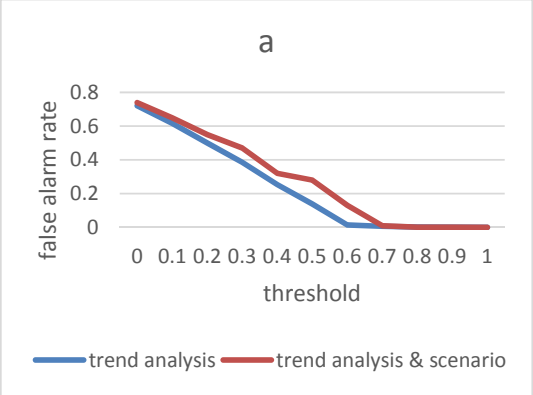


Fig. 11. Fraud detection rate according to selected threshold for infusing function

Fig. 12. Detection rate comparison among trend-analysis-based, composed trend analysis-based, and scenarios-based vs alert rate



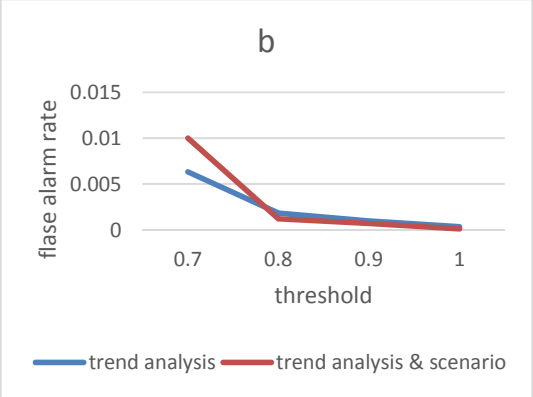


Fig. 13. False alarm rate vs selected thresholds, (b): Detailed version of (a)

In Tab. 2, a comparison among different thresholds, alarm rates and detected fraud rates is shown. According to Fig. 13., Fig. 14, and Tab. 3. , 0.8 is the desirable threshold resulting in an alarm rate equal to 0.0018 and a fraud detection rate equal to 0.82. Taking a threshold to be more than 0.8 will lead to a reduction in the fraud detection rate.

Tab. 2.Comparison among different thresholds, alarm rates and detected fraud rates in trend-analysis-based (TAB) & scenario-based (SB) components.

Threshold	Alarm Rate	Detected Fraud Rate for rules	Detected Fraud Rate for rules & scenarios
0	0.72106	0.98	0.97
0.1	0.61655	0.97	0.97
0.2	0.49997	0.94	0.963
0.3	0.38516	0.92	0.95
0.4	0.25328	0.89	0.93
0.5	0.13824	0.83	0.91
0.6	0.01440	0.77	0.88
0.7	0.00632	0.71	0.859
0.8	0.00182	0.68	0.82
0.9	0.00098	0.64	0.77
1	0.00034	0.57	0.59

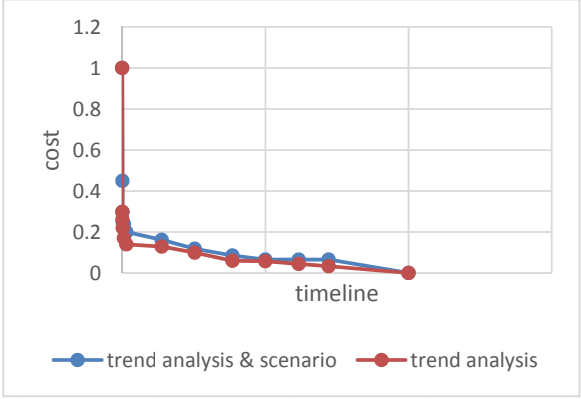


Fig. 14. Performance curve for trend-analysis-based and scenarios-based

Regarding the alarm rate and detection rate metrics, it is obvious from Fig. 12. and Fig. 13. that the composite approach has better results in comparison to only the trend analysis-based approach, but regarding the performance metric, which is shown by the performance curve in Fig. 14, it is revealed that the composite approach has lower performance compared with the trend-analysis-based approach. However, this difference is rather insignificant and can be neglected.

7. Conclusion

Several entities and events such as human wisdom, analytic tools and business systems are involved in the occurrence of a fraud. Effective and instant detection of sophisticated frauds requires more accurate analytic tools and algorithms. In this paper, both research and practice in the real world were presented, and a framework utilizing the result of behavior trend analysis and scenarios was designed for fraud detection. Herein, the customer behavior in several layers (card, account, and customer) and according to several attributes was modelled, and some associated thresholds for each customer were extracted which are later used in defined profiles for estimating the level of risk for each transaction. Additionally, based on the behavior expected from fraudsters, 7 scenarios for risky behaviors and transactions were defined. These scenarios were tested when the stream of transactions of a customer arrived at the system. The approach and the designed system were tested experimentally in a real world environment. The experiments show that our proposed system significantly improves fraud detection speed and accuracy and performs better than manually expert trend analysis-based methods.

For large-scale financial companies like international banks with multi million daily transactions, the performance is of great importance. Possibilities of using parallel processing and big data technology and designing suitable frameworks may be of great helps. Two main other important issues in fraud detection which can be regarded in next studies are determining the length and depth of required historical data for making user profiles and also finding an agile and high precision method for handling and modelling the changes in customers behaviors.

References

[1] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," *Decis. Support Syst.*, Vol. 50, No. 3, (2011), pp. 602–613.

[2] E. Kirkos, C. Spathis, and Y. Manolopoulos, "Data Mining techniques for the detection of fraudulent financial statements," *Expert Syst. Appl.*, Vol. 32, No. 4, (2007), pp. 995–1003.

[3] M. E. Edge and P. R. Falcone Sampaio, "A survey of signature based methods for financial fraud detection," *Comput. Secur.*, Vol. 28, No. 6, (2009), pp. 381–394.

[4] A. Dal Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert Syst. Appl.*, Vol. 41, No. 10, (2014), pp. 4915–4928.

[5] N. Carneiro, G. Figueira, and M. Costa, "A data mining based system for credit-card fraud detection in e-tail," *Decis. Support Syst.*, Vol. 95, (2017), pp. 91–101.

[6] V. Vatsa, S. Sural, and A. K. Majumdar, "A Game-Theoretic Approach to Credit Card Fraud Detection," in *Information Systems Security*, S. Jajodia and C. Mazumdar, Eds. Springer Berlin Heidelberg, (2005), pp. 263–276.

[7] M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero, "BankSealer: A decision support system for online banking fraud analysis and investigation," *Comput. Secur.*, Vol. 53, (2015), pp. 175–186.

[8] Hawkins, D, *Identification of Outliers*. (1980).

[9] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Comput Surv*, Vol. 41, No. 3, (2009), pp. 15:1–15:58.

[10] N. Laleh and M. A. Azgomi, "An Overview of a Hybrid Fraud Scoring and Spike Detection Technique for Fraud Detection in Streaming Data," in *Information Systems, Technology and Management*, (2009), pp. 356–357.

[11] Y. Kou, C.-T. Lu, S. Sirwongwattana, and Y.-P. Huang, "Survey of fraud detection techniques," in *2004 IEEE International Conference on Networking, Sensing and Control*, Vol. 2, (2004), pp. 749–754

[12] J. Bae, H. Bae, S.-H. Kang, and Y. Kim, "Automatic control of workflow processes using ECA rules," *IEEE Trans. Knowl. Data Eng.*, Vol. 16, No. 8, (2004), pp. 1010–1023.

[13] M. H. Cahill, D. Lambert, J. C. Pinheiro, and D. X. Sun, "Detecting Fraud in the Real World," in *Handbook of Massive Data Sets*, J. Abello, P. M. Pardalos, and M. G. C. Resende, Eds. Springer US, (2002), pp. 911–929.

[14] P. Ferreira, R. Alves, O. Belo, and L. Cortesão, "Establishing Fraud Detection Patterns Based on Signatures," in *Advances in Data Mining. Applications in Medicine, Web Mining, Marketing, Image and Signal Mining*, (2006), pp. 526–538.

[15] R. J. Bolton, D. J. Hand, and D. J. H, "Unsupervised Profiling Methods for Fraud Detection," in *Proc. Credit Scoring and Credit Control VII*, (2001), pp. 5–7.

[16] W. Wei, J. Li, L. Cao, Y. Ou, and J. Chen, "Effective detection of sophisticated online banking fraud on extremely imbalanced data," *World Wide Web*, Vol. 16, No. 4, (2012), pp. 449–475.

[17] S. S. Mhamane and L. M. R. J. Lobo, "Use of Hidden Markov Model as Internet Banking Fraud Detection," *Int. J. Comput. Appl.*, Vol. 45, No. 21, (2012), pp. 5–10.

[18] V. Bhusari and S. Patil, "Study of Hidden Markov Model in credit card fraudulent detection," in *2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave)*, (2016), pp. 1–4.

[19] A. Kundu, S. Sural, and A. K. Majumdar, "Two-Stage Credit Card Fraud Detection Using Sequence Alignment," in *Information Systems Security*, (2006), pp. 260–275.

[20] A. Siddiqa *et al.*, "A survey of big data management: Taxonomy and state-of-the-art," *J. Netw. Comput. Appl.*, Vol. 71, (2016), pp. 151–166.

[21] J. Chen, Y. Tao, H. Wang, and T. Chen, "Big data based fraud risk management at Alibaba," *J. Finance Data Sci.*, Vol. 1, No. 1, (2015), pp. 1–10.

[22] Y. Sylla and P. Morizet-Mahoudeaux, "Fraud Detection on Large Scale Social Networks," in *2013 IEEE International Congress on Big Data*, (2013), pp. 413–414.

[23] S. Wang, "A Comprehensive Survey of Data Mining-Based Accounting-Fraud Detection Research," in *2010 International Conference on Intelligent Computation Technology and Automation (ICICTA)*, Vol. 1, (2010), pp. 50–53.

[24] G. P. Gupta and M. Kulariya, "A Framework for Fast and Efficient Cyber Security Network Intrusion Detection Using Apache Spark," *Procedia Comput. Sci.*, Vol. 93, (2016), pp. 824–831.

[25] C. Whitrow, D. J. Hand, P. Juszczak, D. Weston, and N. M. Adams, "Transaction Aggregation As a Strategy for Credit Card Fraud Detection," *Data Min Knowl Discov*, Vol. 18, No. 1, (2009), pp. 30–55.

[26] S. Jha, M. Guillen, and J. Christopher Westland, "Employing transaction aggregation strategy to detect credit card fraud," *Expert Syst. Appl.*, Vol. 39, No. 16, (2012), pp. 12650–12657.

[27] A. Correa Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Syst. Appl.*, Vol. 51, (2016), pp. 134–142.

[28] A. A. I. Alnajem and N. Zhang, "A Copula-Based Fraud Detection (CFD) Method for Detecting Evasive Fraud Patterns in a Corporate Mobile Banking Context," in *2013 International Conference on IT Convergence and Security (ICITCS)*, (2013), pp. 1–4.

[29] N. Fisher, *Statistical Analysis of Circular Data*. Cambridge University Press,Cambridge, UK., (1995).

[30] P. Juszczak, N. M. Adams, D. J. Hand, C. Whitrow, and D. J. Weston, "Off-the-peg and bespoke classifiers for fraud detection," *Comput. Stat. Data Anal.*, Vol. 52, No. 9, (2008), pp. 4521–4532.

Follow This Article at The Following Site

Eshghi A, Kargari M. Detecting frauds using customer behavior trend analysis and known scenarios. IJIEPR. 2018; 29 (1) :91-101
URL: http://ijiepr.iust.ac.ir/article-1-779-en.html