International Journal of Pure and Applied Mathematics
Volume 118 No. 7 2018, 229-234
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

Credit Card Fraud Detection using Local Outlier Factor

Diwakar Tripathi
Department of Computer Science and Engineering,
National Institute of Technology Goa, India
Email: diwakartripathi@nitgoa.ac.in

Yograj Sharma
Department of Computer Science and Engineering,
Rustamji Institute of Technology Tekanpur, India
Email: yograjgwl@gmail.com

Tushar Lone
Department of Computer Science and Engineering,
National Institute of Technology Goa, India
Email: toshiawesome@gmail.com

Shubhra Dwivedi
Email: shubhradwivedi09@gmail.com

***Abstract*— The usage of credit card transaction for online or offline shopping has increased manifold due to fast headway in the electronic business innovation. The number of online fraudulent activities has increased proportionally because of this phenomenon. Most online financial transaction in internet is through credit or debit card transaction. Online transaction fraud detection is the biggest challenging issue for banking systems and financial institutes. We address the problem of credit card fraud over online or offline transaction. We propose the use of local outlier factor for credit and debit card fraud detection methodology and present the detailed experimental results of our proposed methodology.**

Keywords— Online transaction; credit card frauds; LOF.

I. INTRODUCTION

Credit card fraud is an illegal act of using credit card information without the knowledge of card holder. Credit card is used physically or online. In case of physical use, cardholders use their cards at merchants 'end. The fraudster has to acquire the card in its physical form through fraudulent means and use it to commit fraud. In online card transaction, information, such as, CVV code, expiry date, card number and pin code are required to commit fraud. Fraudsters acquire card information through intercepting of mails, phishing and skimming of victim's online transactions. Bhatla et al. [1] states about the types of fraud and impact of fraud on card holder, merchant and credit card institutions. The credit card fraud can be categorized as application fraud, stolen card, account takeover and counterfeit card. In application fraud fraudsters acquire card by submitting the fake personal documents or possess it from postal services or from card issuing company. In case of account takeover, the fraudsters acquire account information illegally related to card holders. The fraudster uses the information to commit the fraud. In Counterfeit card fraud, fraudsters create forged cards. Forged cards are created by erasing the magnetic strip, changing card information and creating identical fake cards. Merchant related frauds are committed by merchants or employee in two ways: the merchant in collusion with fraudster pass personal information of card holder to the latter, and in second case, merchant offers goods at highly discounted rates provoking buyer to submit personal card information to buy product online. That merchant uses the information to place the order on some other sites. Internet related frauds are site cloning, false merchant sites and fake credit card generation. In site cloning fraudsters clone the merchant site or create a false merchant site in which the customer can place the order. The merchant can use card information to purchase from other merchants or they can sell this information to other merchant. Cardholder is often unaware that his card or card information is

stolen. The best way to identify this kind of fraud is by examining the behavior of customer, on each transaction and to make sense of any abnormality regarding the "normal" behavior of that particular customer.

Detection of fraud using previous transaction history of card holder is the most convincing way for reducing fraud rate and increase the legitimate transaction rate. Every card holder has a specific purchasing behavior history which is different from another cardholder's purchasing behavior. Every cardholder's behavior history is represented by a set of patterns containing information about purchase category, time, amount and region since time of last purchase. The rapid growth of e-commerce has led to increase of usage of credit cards for online and off line shopping. The growth in fraudulent activity has also increased proportionally.

The credit card fraud and ID theft statistics [2] show a total of 13.1 million US victims suffering a loss worth $18 billion in the year 2013 and 12.7 million US victims suffering a loss of $16 billion in the year 2014. The Card fraud figures [3] in year 2013 and 2014, 1.23 and 1.28 billion fraud cases have been registered in UK. The statistics related to frauds in UK is presented in Fig.1.

Annual case volumes on UK-issued cards 2008 to 2014

It is important to note that number of cases relates to the number of accounts that have been defrauded, as opposed to the number of victims

Card fraud type	2008	2009	2010	2011	2012	2013	2014	% change 13–14
Remote purchase fraud	875,086	824,736	733,145	709,402	750,200	951,998	1,019,146	+7%
Counterfeit (skimmed/cloned) fraud	371,442	177,430	96,861	81,112	98,322	101,109	99,279	–2%
Fraud on lost or stolen cards	115,590	99,403	101,947	104,467	113,003	138,967	133,943	–4%
Card ID theft	26,488	20,736	19,555	15,420	24,078	30,718	26,542	–14%
Card non-receipt	10,839	8,302	6,622	8,536	9,018	9,125	9,302	+2%
TOTAL	1,399,445	1,113,607	958,130	918,937	994,621	1,231,917	1,288,212	+5%

Fig. 1. Annual fraud cases in UK.

Rest of the article as follows, in section 2 literature survey related to credit card fraud detection. In section 3 presents about the local outlier factor (LOF) for fraud detection. Section 4 presents experimental results followed by concussion in section 5.

II. Literature Survey

A large number of credit card fraud detection algorithms were proposed by researchers mainly based on neural network and data mining approach. Abhinav et al. [4] proposed a model which works on two phases: training and detection. In training phase the shopping behavior of credit card holder is analyzed using k-means clustering algorithm and in detection phase the sequence is constructed. If the current transaction follows the sequence then it is considered legitimate else fraudulent transaction. Alman et al. [5] have proposed two stage credit card fraud detection system. In first stage, using sequence alignment, good score is calculated by genuine card holder transaction history and behavioral change of transaction. In second stage bad score is calculated by fraudulent transaction signature generated by previous fraudulent transaction. If the difference between good score and bad score is greater than predefined threshold then this transaction is fraudulent otherwise is legitimate.

Alman et al. [6] have proposed BLAHFDS hybridization of BLASTA – SSAHA for credit card fraud detection. This is two stage model. The first stage is profile analyzer which analyzes the similarity of time and sequence of new transaction with transactional database. The second stage is deviation analyzer which analyzes the similarity of deviated time- amount sequence with fraud history database. It calculates total difference of profile score and deviation score. The total difference is used for fraud detection. Duman and Ozcelik [7] proposed hybridization of two popular algorithms, Genetic Algorithm (GA) and Scatter Search (SS) called GASS. GASS follows the steps of GA with some components of SS. The steps followed are as number of parent solution, number of children, reproduction, mutation operator, recombination and mutation probability, fitness function, selection and termination criteria. The proposed method is to improve the performance of classification cost. Li et al. [8] proposed a model for fraud detection using Bayesian classification and association rule to identify the signs and patterns of fraudulent transaction. The generated rules and patterns are used for fraud detection.

Jha et al. [9] and Whitrow et al. [10] have suggested a model using transaction aggregation strategy. In this model they aggregated the transaction and generated the customer's purchasing behavior. These behaviors are used for identifying credit card fraudulent transaction. Bahnsen et al. [11] have expanded the transaction aggregation strategy to observe the periodical spending behavior of customer. They used the feature processing and cost sensitivity to improve the fraud detection. Sanchez et al. [12] have applied association rule to generate the normal behavior patterns from fraudulent transactional database. These patterns are used for fraud detection. Jon and Sriganesh [13] used the self-organizing map (SOM) for fraud detection. SOM is used for classification and clustering the previous transactional data, deriving the hidden patterns from previous data and for filtering mechanism. Veronique et al. [14] have proposed a novel approach using network based features and intrinsic features. Intrinsic feature is how the new submitted transaction differs from earlier transaction in terms of Regency–Frequency–Monetary (RFM) of that card. The network based feature is the relationship between the merchants and cards deriving a time-dependent suspicious score for each merchant. Sahin et al. [15] had proposed a cost sensitive decision tree approach to reduce the cost of misclassification and to identify the fraudulent credit card transaction.

III. Fraud Detection using LOF

Local Outlier Factor [16, 17]: The outlier factor captures the degree by which given point can be marked as outlier. Before we jump to the definition of LOF, first we discuss some concepts that are necessary to understand LOF.

Relative density of an object o [15]: for given set of objects D, relative density of object o is denoted by distk(o), calculated as distance between o and another object, p $\in$ D such that

There are at least k objects $o' \in D - \{o\}$ such that

$$dist(o, o') \leq dist(o, p) \quad (1)$$

There are at most (k-1) objects $o'' \in$ D – {o} such that dist $(o, o'') < dist(o, p)$.

Distk(o) can be also referred as distance between o and its k-nearest neighbor. K-distance neighborhood of o contains all objects of which the distance to o is not greater than distk(o), denoted by

$$N_k = \{o' \vee o' \in D, dist(o, o') \leq dist_k(o)\} \quad (2)$$

Nk may contain more than k objects since multiple objects may be at the same distance away from o.

Local density of o can be calculated as average distance from objects in neighborhood Nk(o). If object o has very close neighbors o' that is distance between o and its neighbors is very small, the statistical fluctuations of the distance measure can be undesirable high. To overcome this problem reachability distance measure is used.

$$reachdist_k(o \leftarrow o') = max\{dist_k(o), dist(o, o')\} \quad (3)$$

The local reachability density of an object o, is calculated as

$$Ird(o) = \frac{\vee N_k(o) \vee}{\sum_{o' \in N} reachdist_k(o' \leftarrow o)} \quad (4)$$

The local outlier factor of an object o is calculated as

$$LOF_k(o) = \frac{\sum_{o' \in N_k} \frac{irdk_k(o')}{irdk_k(o)}}{\vee N_k(o) \vee} \quad (5)$$

The local outlier factor is average of the ratio of the local reachability density of o and those of its k-nearest neighbors. If $LOF_k(o) \approx 1$, that means object o has approximately same density as its neighbors, therefore we say object o is not an outlier.

Fraud detection system verifies every incoming transaction at card issuing bank. When a new transaction is initiated for execution, FDS extracts the card number and amount from incoming transaction to validate that transaction is legitimate or not. If the FDS conforms that submitted transaction is not legitimate, it raises an alert to card issuing bank. If there is an alert in that transaction, there may be the card is compromised. Fraudulent transactions are found to be deviant than that of normal transactions. Amount can be one of the examples. We capture this deviant behavior of fraudulent transaction using LOF. For the credit card fraud detection one transaction details are not sufficient. We have applied LOF because it checks for k-density reachability to detect the outlier. We have applied LOF in transaction amount because card holders are unaware of that somebody has stolen the card or card information. So we have the way to predict the fraud on transaction is spending behavior because every customer has some purchasing behavior and single transaction details of a card is not sufficient to check the behavioral matching of transaction. LOF is a method for density based outlier detection.

To map transaction processing in terms of LOF, we quantize the purchasing amount x into n ranges A1, A2, … An. these quantize values are quantized using the range clustering for example, let 1500 = [0, $1500], 3000 = [$1501, $3000] and so on. If a cardholder uses his card for purchasing with $1450, then its quantized value is $1500. In our approach if the quantized value of submitted transactional amount is k density reachable with previous quantized values of purchasing amount, then it is outlier and it is suspicious transaction else this transaction is legitimate.

IV. Experimental Results

A. Dataset

We have used two real credit card transactional datasets which have 20 fields and total 200,000 transaction details. Dataset 1 contains 100,000 transactional details of 7374 credit cards with 2659 fraudulent transaction. Dataset 2 contains 94,682 transactional details of 9810 credit cards with 2094 fraudulent transactions. That dataset is labeled as the legitimate transaction along with fraudulent transactions performed by credit card. These datasets are used in UCSD-FICO competition. The competition was organized by FICO, the leading provider of analytics and decision management technology, and the University of California, San Diego (UCSD). Dataset 1 has 100000 transactions with 2659 fraudulent transactions means in this dataset there is 100:3 ratio of legitimate and fraudulent transactions. In dataset 2 there are 94,682 transactions with 9:3 fraud cases. These dataset is of 98 days transactions and there is no difference between legitimate transaction details and fraudulent transactional details which indicates that both transaction types have similar behavior.

B. Calculation and analysis of credit card fraud detection

In case of evaluation of credit card fraud detection algorithm, we have to identify the fraudulent transaction as fraud with high probability as well as legitimate transaction as legitimate transaction based on historical transactional database.

Table 1. Results of classification as confusion matrix [18].

	Actual fraud	**Actual legitimate**
Observed as fraud	True positive (TP)	False positive (FP)
Observed as legitimate	False negative (FN)	True negative (TN)

C. Performance measures

In the use of LOF for card fraud detection we have considered some of the common measures like accuracy, true negative ratio and false positive ratio for the performance measures. Table II represents the total number of predicted as fraud or legitimate from fraudulent or legitimate transaction. In table II we have presented fraudulent transactions as F class and legitimate transactions as L class and hence the meaning of the terms fraud transactions predicted as fraud (TP), legal transactions predicted as legal (TN), legal transactions predicted as fraud (FP) and fraud transactions predicted as legal (FN). In table II FF, LL, FL and LF combinations as follows TP, TN, FP and FN. In this effort we have used the existing outlier detection technique "LOF" and we used it for fraud detection in credit card transaction. We applied this technique with different nearest neighbors (2, 3, 5, 7 and 9).

All the number of cases of TP, FP, TN and FN are listed in table II then we have calculated the false positive rate, true negative rate and accuracy using the formulas given below. These calculated values are listed below in table III. The measures of correctly identified instances are known as the accuracy of the system. The corresponding histogram for the value of accuracy, TNR and FNR of dataset 1 is shown in Fig. 2 and for dataset 2 is in Fig. 3.

$$Accuracy\,(ACC) = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{6}$$

$$True\ Negative\ Rate\ (TNR) = \frac{TN}{(FP+TN)} \tag{7}$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{(FP+TN)} \tag{8}$$

TABLE 2. TP, TN, FP AND FN WITH DIFFERENT NEAREST NEIGHBORS.

			Actual									
			Nearest neighbors									
			2		3		5		7		9	
			F	L	F	L	F	L	F	L	F	L
Predicted	DS1	F	456	28329	480	29627	528	32386	567	35137	620	37173
		L	2177	69038	2165	67636	2119	64891	2081	62145	2033	60155
	DS2	F	87	1672	83	1408	77	1191	67	927	63	829
		L	1858	85993	2011	91179	1872	86767	1882	87171	1889	87348

TABLE 3. CALCULATED ACC, TNR AND FPR.

			ACC	TNR	FPR
Nearest neighbors	DS1	2	69.49	70.90	29.09
		3	68.17	69.53	30.46
		5	65.46	66.70	33.29
		7	62.75	63.88	36.11
		9	60.78	61.80	38.19
	DS2	2	96.06	98.09	1.90
		3	96.38	98.47	1.52
		5	96.61	98.64	1.35
		7	96.88	98.94	1.05
		9	96.98	98.05	0.94

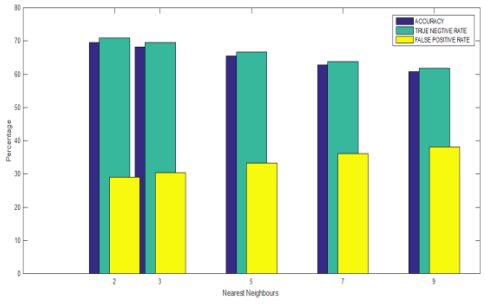

Fig. 2. Histogram for DS1.

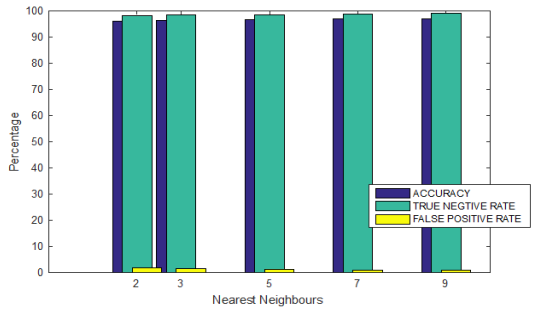

Fig. 3. Histogram for DS2.

V. Conclusion

In this paper, we have proposed use of LOF for detection of credit card fraud. We have used the purchasing amount as the examination of frauds. The proposed system is implemented in MATLAB technology, and performance of the system is evaluated over the different nearest neighbors in terms of true negative, false positive rate and accuracy of the system. Accuracy of our proposed approach is lying between to 60-69 % for dataset 1 and 96 % for dataset 2 with variation in neighbors. This system is also scalable for large volumes of transactions.

Acknowledgment

Authors would like to thank Dr. Kohei Hayashi, from Nara Institute of Science and Technology, Japan for providing the Dataset used in this experiment.

References

[1] Bhatla, Tej Paul, Vikram Prabhu, and Amit Dua. "Understanding credit card frauds." Cards business review 1.6 (2003).

[2] card fraud and ID theft statistics http://www.nasdaq.com/article/credit-card-fraud-and-id-theft-statistics-cm520388

[3] Card fraud figures "http://www.theukcardsassociation.org.uk/plastic_fraud_figures/index.asp"

[4] Srivastava, A., Kundu, A., Sural, S., & Majumdar, A. K. (2008). " Credit card fraud detection using hidden Markov model. Dependable and Secure Computing", IEEE Transactions on, 5(1), 37-48

[5] Kundu, A., Sural, S., & Majumdar, A. K. (2006). Two-stage credit card fraud detection using sequence alignment. In Information Systems Security (pp. 260-275). Springer Berlin Heidelberg.

[6] Kundu, A., Panigrahi, S., Sura, S., & Majumdar, A. K. (2009). Blast-ssaha hybridization for credit card fraud detection. Dependable and Secure Computing, IEEE Transactions on, 6(4), 309-315.

[7] Ekrem Duman, and M. Hamdi Ozcelik. "Detecting credit card fraud by genetic algorithm and scatter search." Expert Systems with Applications 38.10 (2011): 13057-13063.

[8] Li, Shing-Han, et al. "Identifying the signs of fraudulent accounts using data mining techniques." Computers in Human Behavior 28.3 (2012): 1002-1013.

[9] Jha, Sanjeev, Montserrat Guillen, and J. Christopher Westland. "Employing transaction aggregation strategy to detect credit card fraud." Expert systems with applications 39.16 (2012): 12650-12657.

[10] Whitrow, Christopher, David J. Hand, Piotr Juszczak, D. Weston, and Niall M. Adams. "Transaction aggregation as a strategy for credit card fraud detection." Data Mining and Knowledge Discovery 18, no. 1 (2009): 30-55.

[11] Bahnsen, Alejandro Correa, et al. "Feature Engineering Strategies for Credit Card Fraud Detection." Expert Systems with Applications (2016).

[12] Sánchez, Daniel, M. A. Vila, L. Cerda, and José-María Serrano. "Association rules applied to credit card fraud detection." Expert Systems with Applications 36, no. 2 (2009): 3630-3640.

[13] Quah Jon TS, and M. Sriganesh. "Real-time credit card fraud detection using computational intelligence." Expert Systems with Applications 35.4 (2008): 1721-1732.

[14] Van Vlasselaer, Véronique, Cristián Bravo, Olivier Caelen, Tina Eliassi-Rad, Leman Akoglu, Monique Snoeck, and Bart Baesens. "APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions." Decision Support Systems 75 (2015): 38-48.

[15] Sahin, Yusuf, Serol Bulkan, and Ekrem Duman. "A cost-sensitive decision tree approach for fraud detection." Expert Systems with Applications 40.15 (2013): 5916-5923.

[16] Han, Jiawei, Micheline Kamber, and Jian Pei. Data mining: concepts and techniques. Elsevier, 2011.

[17] Breunig, Markus M., Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. "LOF: identifying density-based local outliers." In ACM sigmod record, vol. 29, no. 2, pp. 93-104. ACM, 2000

[18] Bahnsen, Alejandro Correa, Djamila Aouada, Aleksandar Stojanovic, and Björn Ottersten. "Feature Engineering Strategies for Credit Card Fraud Detection." Expert Systems with Applications (2016).