A Review On Credit Card Fraud Detection Using Machine Learning

Article *in* International Journal of Scientific & Technology Research · October 2019

CITATIONS
35

READS
10,903

4 authors, including:

Suresh Shirgave
DKTE Society's Textile and Engineering Institute Ichalkaranji
22 PUBLICATIONS **113** CITATIONS
SEE PROFILE

Chetan J Awati
Shivaji University, Kolhapur
54 PUBLICATIONS **137** CITATIONS
SEE PROFILE

Rashmi More
Rajendra Mane College of Engineering &Technology
2 PUBLICATIONS **47** CITATIONS
SEE PROFILE

A Review On Credit Card Fraud Detection Using Machine Learning

Suresh K Shirgave, Chetan J. Awati, Rashmi More, Sonam S. Patil

Abstract: In recent years credit card fraud has become one of the growing problem. A large financial loss has greatly affected individual person using credit card and also the merchants and banks. Machine learning is considered as one of the most successful technique to identify the fraud. This paper reviews different fraud detection techniques using machine learning and compare them using performance measure like accuracy, precision and specificity. The paper also proposes a FDS which uses supervised Random Forest algorithm. With this proposed system the accuracy of detecting fraud in credit card is increased. Further, the proposed system use learning to rank approach to rank the alert and also effectively addresses the problem concept drift in fraud detection.

Index Terms: Concept drift, credit card fraud, Machine Learning, Random Forest

1. INTRODUCTION

Credit card fraud is a major problem that involves payment card like credit card as illegal source of funds in transactions. Fraud is an illegal way to obtain goods and funds. The goal of such illegal transaction might be to get products without paying or gain an unauthorized fund from an account. Identifying such fraud is a troublesome and may risk the business and business organizations. In the real world FDS [1], investigator are not able to check all transactions. Here the Fraud Detection System monitors all the approved transactions and alerts the most suspicious one. Investigator verifies these alerts and provides FDS with feedback if the transaction was authorized or fraudulent. Verifying all the alerts everyday is a time consuming and costly process. Hence investigator is able to verify only few alerts each day. The rest of the transactions remain unchecked until customer identifies them and report them as a fraud. Also the techniques used for fraud and the cardholder spending behavior changes over time. This change in credit card transaction is called as concept drift [1] [7]. Hence most of the time it is difficult to identify the credit card fraud. Machine Learning is considered as one of the most successful technique for fraud identification. It uses classification and regression approach for recognizing fraud in credit card. The machine learning algorithms are divided into two types, supervised [14][18] and unsupervised [16] learning algorithm. Supervised learning algorithm uses labeled transactions for training the classifier whereas unsupervised learning algorithm uses peer group analysis [23] that groups customers according to their profile and identifies fraud based on customers spending behavior.

- *Dr. Suresh K. Shirgave is Associate Professor, DKTE Society's Textile and Engineering Institute, Ichalkaranji, India, E-mail: skshirgave@gmail.com*
- *Chetan J. Awati is Assistant Professor, Department of Technology, Shivaji University Kolhapur, India, E-mail: cja_tech@unishivaji.ac.in*
- *Rashmi More is M. Tech. Student, Department of Technology, Shivaji University Kolhapur, India, E-mail: rashmimore107@gmail.com*
- *Sonam S. Patil is Assistant Professor, Department of Information Technology, D Y Patil College of Engineering, Akurdi, Pune India, E-mail: skh9624@gmail.com*

Many learning algorithm have been presented for fraud detection in credit card which includes neural networks [14][19][21][22], Logistic Regression [3], decision tree [4][15], Naive Bayes [6], Support Vector Machines [5], K-Nearest Neighbors [6] and Random Forest [1][2]. This paper examines the performance of above algorithms based on their ability to classify whether the transaction was authorized or fraudulent and then compares them. The comparison is made using performance measure accuracy, specificity and precision. The result showed that Random Forest algorithm showed better accuracy and precision than other techniques.

2 RELATED WORK

There are different supervised and unsupervised learning algorithms used for fraud detection in credit card. Some important are described below. The author [1] has proposed a paper where they have first explained the proper performance measures which is used for fraud identification. The authors have structured a novel learning technique that can solve concept drift, verification latency, and class imbalance issues. The paper also showed effect of above issues in true credit card transactions. Here in paper [2] authors presented two types of classifier using random forests which are used to train the behavior features of transactions. The authors have compared the two random forests and have analyzed their performance on fraud identification in credit card. In paper [3] authors presented a FDS for credit card using Artificial Neural Network and Logistic Regression. The system used to monitor each transaction separately using classifier and then classifier would generate score for each transaction and label this transaction as legal or illegal transaction. A decision tree method was proposed in paper [4]. The method decreased overall misclassification costs and selected splitting property at each node. The author also compared the decision tree method for fraud identification with other models and proved that this approach performs well using performance measure like accuracy and genuine positive rate. The author [5] developed a FDS for credit card transaction using support vector machines and decision tree. This study built seven alternative models that were created using support vector machines and decision tree. The author also compared this classifiers performance using performance measure

accuracy. The study also showed that as size of training dataset increases the number of fraud detected by SVM are less than fraud identified by decision tree method. Here in [6] author presented fraud detection system using a Naive Bayes K-Nearest Neighbors method. The main aim of proposed system was to improve accuracy. Naive Bayes Classifier predicts probabilities of fraud in transaction while KNN classifier predicts how near the undefined sample data is to kth training dataset. The author compared both this classifier and showed that both work differently for given dataset. Most of predictive model used for detecting fraud in credit card transaction faces the issue of concept drift. The author [7] presented two FDS based on sliding window and ensemble learning and showed that classifier need to be trained separately using feedback and delayed samples. The outcome of the two was than aggregated to improve the alert precision in FDS. Thus the author showed that to solve the issue of concept drift, the feedback and delayed samples are to be handled separately.

3 COMPARISON

Performance of all learning algorithms used for fraud detection in credit card transactions are compared in table 1 The comparison is based on their accuracy, precision and specificity.

TABLE 1
COMPARISON OF MACHINE LEARNING TECHNIQUES

Classifiers	Metrics		
	Accuracy	Precision	Specificity
Random Forest	0.962	0.997	0.987
Logistic Regression	0.947	0.996	0.979
KNN	0.942	0.410	0.971
SVM	0.938	0.782	0.984
Decision Tree	0.908	0.91	0.912
Naive Bayes	0.937	0.505	0.9741

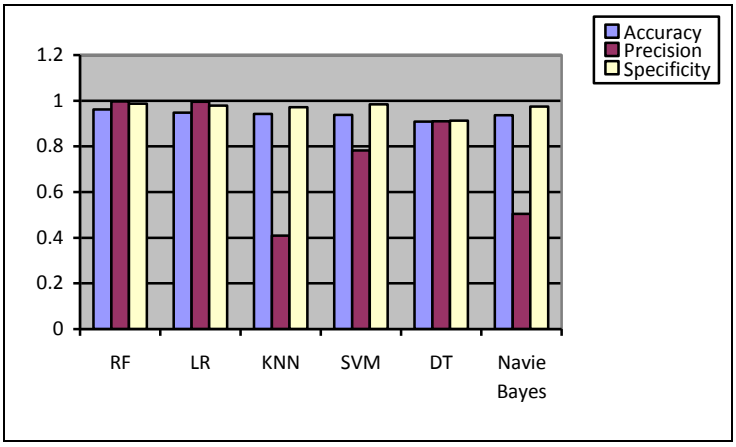

***Fig**. 1. Accuracy, precision and specificity performance of all classifier*

From table 1 we can see that accuracy of random forest is far better than the other learning algorithms. From fig. 1 we can see that precision, accuracy and specificity of Random Forest is highest followed by Logistic regression, SVM, Decision Tree, Naive Byes and KNN. Hence the proposed system using random forest will show better accuracy for larger number of training data.

4 PROPOSED SYSTEM

Today modern society is using credit cards for variety on reasons. Similarly fraud in credit card transactions has been growing in recent years. Each year, a huge amount of financial losses are caused by the illegal credit card transactions. Fraud may occur in variety of different forms and may be limited. Therefore there is need to solve the issues of fraud detection in credit card. Additionally, with the development of new technologies criminals finds new ways to commit fraud. To overcome this problem the proposed system for fraud detection in credit card transactions will be designed using ML technique that will provide investigator a small reliable fraud alerts.

4.1 Objectives

The proposed system will achieve following main objectives:

- To train the model using feedbacks and delayed samples and sum up their likelihood to identify alert.
- To implement machine learning technique to address concept drift and class imbalance issue.
- To develop a learning to rank approach to increase alert precision.
- To introduce performance measure those are considered in real-world FDS.

We propose a Fraud Detection System (FDS), which mainly focuses on data driven model and learning to rank method. It also focuses on alert feedback interaction that checks the way recent supervised samples are provided. Fig. 2 shows the block diagram of proposed system.

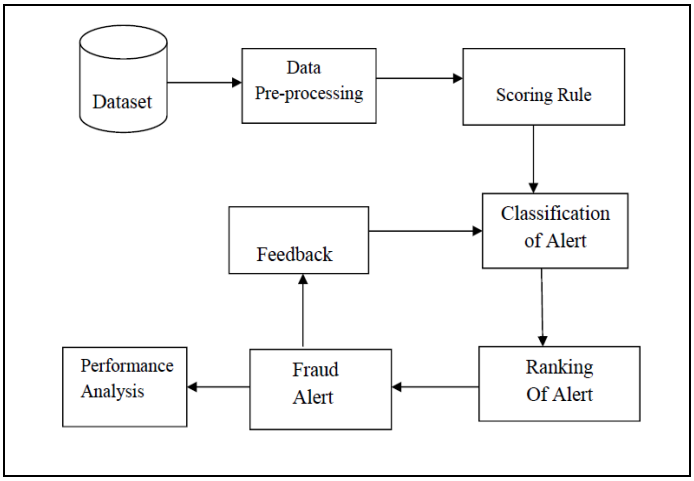

***Fig. 2**. Block Diagram of Proposed System*

4.2 Modules

The system is proposed to have the following modules along with functional requirements.

a. Data Preprocessing
b. Scoring Rule
c. Classification of Alerts
d. Ranking of Alert
e. Performance Analysis

4.2.1 Data Preprocessing

In this module selected data is formatted, cleaned and sampled. The data preprocessing steps includes following:

a. Formatting: The data which is been selected may not be in a suitable format. The data may be in a file format and we may like it in relational database or vice versa.
b. Cleaning: Removal or fixing of missing data is called as cleaning. The dataset may contain record which may be incomplete or it may have null values. Such records need to remove.
c. Sampling: As number of frauds in dataset is less than overall transaction, class distribution is unbalanced in credit card transaction. Hence sampling method is used to solve this issue.

4.2.2 Scoring Rule

Percentage of fraud in transaction is called as score. This module assigns score by matching recent transaction pattern with the past transaction pattern of cardholder. If score is greater then the transaction is considered as suspicious and further proceeding is stopped. Otherwise it is moved to next module.

4.2.3 Classification of Alert

Here machine learning model will be used that will train and update the data based on feedback and delayed samples. Classifier will be trained separately using feedback and delayed samples and their probabilities will be aggregated to identify alerts. Transaction that will be having high probability will be alerted. Hence only limited number of alerted transaction is reported to investigators.

4.2.4 Ranking of Alert

This module, rank each alert based on correctness of security question. This security questions will be created every time whenever the transaction is identified to be suspicious. The alerts are ranked using likelihood. If it is found that an alert has greater probability than other alerts then it is added to a queue and location of fraudster is tracked. This feature makes system user friendly and helps to file complaint against fraud.

5 CONCLUSION

This paper has reviewed various machine learning algorithm detect fraud in credit card transaction. The performances of all this techniques are examined based on accuracy, precision and specificity metrics. We have selected supervised learning technique Random Forest to classify the alert as fraudulent or authorized. This classifier will be trained using feedback and delayed supervised sample. Next it will aggregate each probability to detect alerts. Further we proposed learning to rank approach where alert will be ranked based on priority. The suggested method will be able to solve the class imbalance and concept drift problem. Future work will include applying semi-supervised learning methods for classification of alert in FDS.

REFERENCES

[1] Jalinus, N., Nabawi, R. A., & Mardin, A. (2017). The Seven Steps of Project-Based Learning Model to Enhance Productive Competences of Vocational Students. In 1st International Conference on Technology and Vocational Teacher (ICTVT 2017). Atlantis Press. Advances in Social Science, Education and Humanities research (Vol. 102, pp. 251-256).
[2] Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Cesare Alippi and Gianluca Botempi, "Credit card Fraud Detection : A realistic Modeling and a Novel Learning Strategy", IEEE Trans. on Neural Network and Learning system,vol.29,No.8, August 2018.
[3] Shiyang Xuan,Guanjun Liu,Zhenchuan Li,Lutao Zheng,Shuo Wang, Jiang,"Random Forest for credit card fraud detection",Int.conf.on Networking,Sensing and control,2018.
[4] Y. Sahin , and Duman,E.,(2011) "Detecting credit card fraud by ANN and logistic regression." In Innovations in Intelligent Systems and Applications(INISTA),2011 international Symposium on (pp.315-319).IEEE
[5] Y. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for fraud detection," Expert Syst. Appl., vol. 40, no. 15,pp. 5916–5923, 2013
[6] Sahin Y. and Duman E. (2011),"Detecting Credit Crad Fraud by Decision Trees and Support Vector Machines", International Multi-Conference Of Engineers and Computer Scientists(IMECS 2011),Mar 16-18,Hong Kong,Vol.1,pp.1-6
[7] Sai Kiran, Jypti Guru, Rishabh Kumar, Naveen Kumar, Deepak Katariya,"Credit card fraud detection using

Naïve Bayes model based and KNN classifier", Int. Journal of Adv. Research , Ideas and Innovations in Technology,vol.4,2018.

[8] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection and concept-drift adaptation with delayed supervised information," in Proc. Int. Joint Conf. Neural Netw., 2015,pp. 1–8.

[9] A. C. Bahnsen, D. Aouada, and B. Ottersten, "Example-dependent cost-sensitive decision trees," Expert Syst. Appl., vol. 42, no. 19,pp. 6609–6619, 2015

[10] A. Dal Pozzolo, O. Caelen, and G. Bontempi, "When is undersampling effective in unbalanced classification tasks?" in Machine Learning and Knowledge Discovery in Databases. Cambridge, U.K.: Springer, 2015

[11] N. Mahmoudi and E. Duman, "Detecting credit card fraud by modified fisher discriminant analysis," Expert Syst. Appl., vol. 42, no. 5,pp. 2510–2516, 2015

[12] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Detecting credit card fraud using periodic features," in Proc. 14th Int. Conf. Mach. Learn. Appl., Dec. 2015, pp. 208–213.

[13] Anuruddha Thennakoon, Chee Bhagyani, Sasitha Premadasa, Shalitha Mihiranga,"Realtime Credit Card Fraud Detection Using Machine Learning ,"Int. Conf. on Cloud Computing, Data Science & Engineering,2019.

[14] S. Wang, L.L. Minku, and X. Yao, "Resampling-based ensemble methods for online class imbalance learning." Trans, Knowl., Data Eng., vol 27, no. 5, pp. 1356-1368, May 2015.

[15] Jan may Kumar Behera, Suvasini Panigrahi, "Credit Card Fraud Detection: A Hybrid Approach using Fuzzy Clustering and Neural Network", 2015 IEEE Second International Conference on Advances in Computing and Communication Engineering.

[16] M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero, BankSealer: "A Decision Support System for Online Banking Fraud Analysis and Investigation", Berlin, Germany: Springer, 2014, pp. 380–394

[17] R. J. Bolton and D. J. Hand, "Unsupervised profiling methods for fraud detection," in Credit Scoring Credit Control VII. London, U.K.: Imperial College London, 2001, pp. 235–255

[18] R. Elwell and R. Polikar, "Incremental learning of concept drift in nonstationary environments," Trans. Neural Netw., vol. 22, no. 10,pp. 1517–1531, 2011.

[19] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," Decision SupportSyst., vol. 50, no. 3, pp. 602–613, 2011.

[20] Tao Guo ,Gui-Yang Li ,"Neural data mining for credit card fraud detection",Int.Conf. on Machine Learning and Cybernetics, Sept 2008

[21] J. Gao, B. Ding, W. Fan, J. Han, P.S. Yu, "Classifying data streams with skewed class distributions and concept drifts", IEEE internet comput., vol.12, no. 6, pp. 37-49, Nov 2008

[22] E. Aleskerov, B. Freisleben, and B. Rao, "CARDWATCH: A neural network based database mining system for credit card fraud detection," in Proc. IEEE/IAFE Computat. Intell. Financial Eng., Mar. 1997, pp. 220–226.

[23] J.R. Dorronsoro, F. Ginel, C. Sgnchez and C.S. Cruz, "Neural fraud detection in credit card operations", IEEE transaction neural network vol. 8, no. 4, pp. 827-834, Jul.1997.

[24] D. J. Weston, D. J. Hand, N. M. Adams, C. Whitrow, and P. Juszczak,"Plastic card fraud detection using peer group analysis," Adv. Data Anal.Classification, vol. 2, no. 1, pp. 45–62, 2008.