International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 24th Feb 2018. Link :http://www.ijiemr.org/downloads.php?vol=Volume-08&issue=ISSUE-02

Title: **ADABOOST: A NEW MECHANISM FOR IDENTIFYING CREDIT CARD FRAUD**

Volume 08, Issue 02, Pages: 127–132.

Paper Authors

MR.R.VEERABABU, P.MOUNIKA

Vignan's Lara Institute of Technology & Science

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

ADABOOST: A NEW MECHANISM FOR IDENTIFYING CREDIT CARD FRAUD

MR.R.VEERABABU[1], P.MOUNIKA[2]
Assistant Professor[1], Department of M.C.A ,Vignan's Lara Institute of Technology & Science
M.C.A Student[2],Department of M.C.A ,Vignan's Lara Institute of Technology & Science

Abstract:
Credit card fraud is a serious problem in financial services. Billions of dollars are lost due tocredit card fraud every year. There is a lack of research studies on analyzing real-world credit card dataowing to confidentiality issues. In this paper, machine learning algorithms are used to detect credit cardfraud. Standard models are firstly used. Then, hybrid methods which use AdaBoost and majority votingmethods are applied. To evaluate the model efficacy, a publicly available credit card data set is used. Then,a real-world credit card data set from a financial institution is analyzed. In addition, noise is added to thedata samples to further assess the robustness of the algorithms. The experimental results positively indicatethat the majority voting method achieves good accuracy rates in detecting fraud cases in credit cards.

Introduction

Fraud is a wrongful or criminal deception aimed to bring financial or personal gain [1]. In avoiding loss from fraud, two mechanisms can be used: fraud prevention and fraud detection. Fraud prevention is a proactive method, where it stops fraud from happening in the first place. On the other hand, fraud detection is needed when a fraudulent transaction is attempted by a fraudster. Credit card fraud is concerned with the illegal use of credit card information for purchases. Credit card transactions can be accomplished either physically or digitally [2]. In physical transactions, the credit card is involved during the transactions. In digital transactions, this can happen over the telephone or the internet. Cardholders typically provide the card number, expiry date, and card verification number through telephone or website. With the rise of e-commerce in the past decade, the use of credit cards has increased dramatically [3]. The number of credit card transactions in 2011 in Malaysia were at about 320 million, and increased in 2015 to about 360 million. Along with the rise of credit card usage, the number of fraud cases have been constantly increased. While numerous authorization techniques have been in place, credit card fraud cases have not hindered effectively. Fraudsters favour the internet as their identity and location are hidden. The rise in credit card fraud has a big impact on the financial industry. The global credit card fraud in 2015 reached to a staggering USD $21.84 billion [4]. Loss from credit card fraud affects the merchants, where they bear all costs, including card issuer fees, charges, and administrative charges [5]. Since the

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

merchants need to bear the loss, some goods are priced higher, or discounts and incentives are reduced. Therefore, it is imperative to reduce the loss, and an effective fraud detection system to reduce or eliminate fraud cases is important. There have been various studies on credit card fraud detection. Machine learning and related methods are most commonly used, which include artificial neural networks, rule-induction techniques, decision trees, logistic regression, and support vector machines [1]. These methods are used either standalone or by combining several methods together to form hybrid models. IEEE In this paper, a total of twelve machine learning algorithms are used for detecting credit card fraud. The algorithms range from standard neural networks to deep learning models. They are evaluated using both benchmark and realworld credit card data sets. In addition, the AdaBoost and majority voting methods are applied for forming hybrid models. To further evaluate the robustness and reliability of the models, noise is added to the real-world data set. The key contribution of this paper is the evaluation of a variety of machine learning models with a real-world credit card data set for fraud detection. While other researchers have used various methods on publicly available data sets, the data set used in this paper are extracted from actual credit card transaction information over three months. The organization of this paper is as follows.

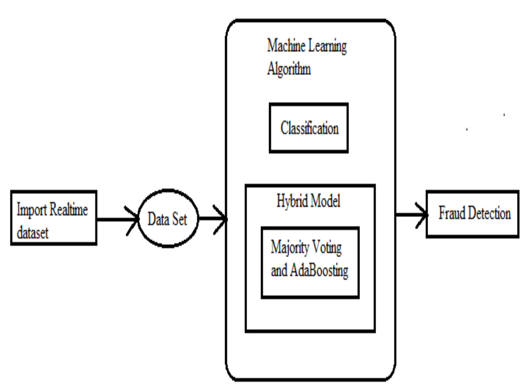
Existing system:

Three methods to detect fraud are presented. Firstly, clustering model is used to classify the legal and fraudulent transaction using data clusterization of regions of parameter value. Secondly, Gaussian mixture model is used to model the probability density of credit card user's past behavior so that the probability of current behavior can be calculated to detect any abnormalities from the past behavior. Lastly, Bayesian networks are used to describe the statistics of a particular user and the statistics of different fraud scenarios. The main task is to explore different views of the same problem and see what can be learned from the application of each different technique.

Proposed system:

Total of twelve machine learning algorithmsare used for detecting credit card fraud. The algorithmsrange from standard neural networks to deep learningmodels. They are evaluated using both benchmark and realworldcredit card data sets. In addition, the AdaBoost andmajority voting methods are applied for forming hybridmodels. To further evaluate the robustness and reliability ofthe models, noise is added to the real-world data set. Thekey contribution of this paper is the evaluation of a varietyof machine learning models with a real-world credit carddata set for fraud detection

Architecture Diagram



International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Modules:

1. **Standard Neural Networks To Deep Learning**

The Feed-Forward Neural Network (NN) uses thebackpropagation algorithm for training as well. Theconnections between the units do not form a directed cycle,and information only moves forward from the input nodes tothe output nodes, through the hidden nodes. Deep Learning(DL) is based on an MLP network trained using a stochasticgradient descent with backpropagation. It contains a largenumber of hidden layers consisting of neurons with tanh,rectifier, and maxout activation functions. Every nodecaptures a copy of the global model parameters on local data,and contributes periodically toward the global model usingmodel averaging.

2. **Forming Hybrid Models**

Adaptive Boosting or AdaBoost is used in conjunction withdifferent types of algorithms to improve their performance.The outputs are combined by using a weighted sum, whichrepresents the combined output of the boosted classifier, AdaBoost tweaks weak learners in favor of misclassifieddata samples. It is, however, sensitive to noise and outliers.As long as the classifier performance is not random,AdaBoost is able to improve the individual results fromdifferent algorithms.Majority voting is frequently used in data classification,which involves a combined model with at least twoalgorithms. Each algorithm makes its own prediction forevery test sample. The final output is for the one thatreceives the majority of the votes,

3. **Evaluate The Robustness And Reliability**

To further evaluate the robustness of the machine learningalgorithms, all real-world data samples are corrupted noise,at 10%, 20% and 30%. Noise is added to all data features.It can be seen that with the addition of noise,the fraud detection rate and MCC rates deteriorate, asexpected. The worst performance, i.e. the largest decrease inaccuracy and MCC, is from majority voting of DT+NB andNB+GBT. DS+GBT, DT+DS and DT+GBT show gradualperformance degradation, but their accuracy rates are stillabove 90% even with 30% noise in the data set.

Algorithm

1. **Machine Learning Algorithm**

A total of twelve algorithms are used in this experimentalstudy. They are used in conjunction with the AdaBoost andmajority voting methods.Naïve Bayes (NB) uses the Bayes' theorem with strong ornaïve independence assumptions for classification. Certainfeatures of a class are assumed to be not correlated to others.It requires only a small training data set for estimating themeans and variances is needed for classification.The presentation of data in form of a tree structure is usefulfor ease of interpretation by users. The Decision Tree (DT) isa collection of nodes that creates decision on featuresconnected to

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

certain classes. Every node represents a splittingrule for a feature. New nodes are established until the stoppingcriterion is met. The class label is determined based on themajority of samples that belong to a particular leaf. TheRandom Tree (RT) operates as a DT operator, with theexception that in each split, only a random subset of featuresis available. It learns from both nominal and numerical datasamples. The subset size is defined using a subset ratioparameter.

Conclusion:

A study on credit card fraud detection using machine learning algorithms has been presented in this paper. A number of standard models which include NB, SVM, and DL have been used in the empirical evaluation. A publicly available credit card data set has been used for evaluation using individual (standard) models and hybrid models using AdaBoost and majority voting combination methods. The MCC metric has been adopted as a performance measure, as it takes into account the true and false positive and negative predicted outcomes. The best MCC score is 0.823, achieved using majority voting. A real credit card data set from a financial institution has also been used for evaluation. The same individual and hybrid models have been employed. A perfect MCC score of 1 has been achieved using AdaBoost and majority voting methods. To further evaluate the hybrid models, noise from 10% to 30% has been added into the data samples. The majority voting method has yielded the best MCC score of 0.942 for 30% noise added to the data set. This shows that the majority voting method is stable in performance in the presence of noise. For future work, the methods studied in this paper will be extended to online learning models. In addition, other online learning models will be investigated. The use of online learning will enable rapid detection of fraud cases, potentially in real-time. This in turn will help detect and prevent fraudulent transactions before they take place, which will reduce the number of losses incurred every day in the financial sector.

References

[1] Y. Sahin, S. Bulkan, and E. Duman, “A cost-sensitive decision tree approach for fraud detection,” Expert Systems with Applications, vol. 40, no. 15, pp. 5916–5923, 2013.

[2] A. O. Adewumi and A. A. Akinyelu, “A survey of machine-learning and nature-inspired based credit card fraud detection techniques,” International Journal of System Assurance Engineering and Management, vol. 8, pp. 937–953, 2017.

[3] A. Srivastava, A. Kundu, S. Sural, A. Majumdar, “Credit card fraud detection using hidden Markov model,” IEEE Transactions on Dependable and Secure Computing, vol. 5, no. 1, pp. 37–48, 2008.

[4] The Nilson Report (October 2016) [Online]. Available: https://www.nilsonreport.com/upload/content_promo/The_Nilson_R eport_10-17-2016.pdf

[5] J. T. Quah, and M. Sriganesh, “Real-time credit card fraud detection using computational intelligence,” Expert Systems with Applications, vol. 35, no. 4, pp. 1721–1732, 2008.

[6] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C., “Data mining for

credit card fraud: A comparative study," Decision Support Systems, vol. 50, no. 3, pp. 602–613, 2011. [
7] N. S. Halvaiee and M. K. Akbari, "A novel model for credit card fraud detection using Artificial Immune Systems," Applied Soft Computing, vol. 24, pp. 40–49, 2014.

[8] S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, "Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning," Information Fusion, vol. 10, no. 4, pp. 354–363, 2009.

[9] N. Mahmoudi and E. Duman, "Detecting credit card fraud by modified Fisher discriminant analysis," Expert Systems with Applications, vol. 42, no. 5, pp. 2510–2516, 2015.

[10] D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, "Association rules applied to credit card fraud detection," Expert Systems with Applications, vol. 36, no. 2, pp. 3630–3640, 2009.

[11] E. Duman and M. H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search," Expert Systems with Applications, vol. 38, no. 10, pp. 13057–13063, 2011.

[12] P. Ravisankar, V. Ravi, G. R. Rao, and I. Bose, "Detection of financial statement fraud and feature selection using data mining
60techniques," Decision Support Systems, vol. 50, no. 2, pp. 491–500, 2011.

[13] E. Kirkos, C. Spathis, and Y. Manolopoulos, "Data mining techniques for the detection of fraudulent financial statements," Expert Systems with Applications, vol. 32, no. 4, pp. 995–1003, 2007.

[14] F. H. Glancy and S. B. Yadav, "A computational model for financial reporting fraud detection," Decision Support Systems, vol. 50, no. 3, pp. 595–601, 2011.

[15] D. Olszewski, "Fraud detection using self-organizing map visualizing the user profiles," Knowledge-Based Systems, vol. 70, pp. 324–334, 2014.

[16] J. T. Quah and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," Expert Systems with Applications, vol. 35, no. 4, pp. 1721–1732, 2008.

[17] E. Rahimikia, S. Mohammadi, T. Rahmani, and M. Ghazanfari, "Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran," International Journal of Accounting Information Systems, vol. 25, pp. 1–17, 2017.

[18] I. T. Christou, M. Bakopoulos, T. Dimitriou, E. Amolochitis, S. Tsekeridou, and C. Dimitriadis, "Detecting fraud in online games of chance and lotteries," Expert Systems with Applications, vol. 38, no. 10, pp. 13158–13169, 2011.

[19] C. F. Tsai, "Combining cluster analysis with classifier ensembles to predict financial distress" Information Fusion, vol. 16, pp. 46–58, 2014.

[20] F. H. Chen, D. J. Chi, and J. Y. Zhu, "Application of Random Forest, Rough Set Theory, Decision Tree and Neural Network to Detect Financial Statement Fraud–Taking Corporate Governance into Consideration," In International Conference on Intelligent Computing, pp. 221–234, Springer, 2014.

[21] Y. Li, C. Yan, W. Liu, and M. Li, "A principle component analysisbased random forest with the potential nearest neighbor method for automobile insurance fraud

identification," Applied Soft Computing, to be published. DOI: 10.1016/j.asoc.2017.07.027.

[22] S. Subudhi and S. Panigrahi, "Use of optimized Fuzzy C-Means clustering and supervised classifiers for automobile insurance fraud detection," Journal of King Saud University-Computer and Information Sciences, to be published. DOI: 10.1016/j.jksuci.2017.09.010.

[23] M. Seera, C. P. Lim, K. S. Tan, and W. S. Liew, "Classification of transcranial Doppler signals using individual and ensemble recurrent neural networks," Neurocomputing, vol. 249, pp. 337-344, 2017.

[24] E. Duman, A. Buyukkaya, and I. Elikucuk, "A novel and successful credit card fraud detection system Implemented in a Turkish Bank," In IEEE 13th International Conference on Data Mining Workshops (ICDMW), pp. 162–171, 2013.

[25] C. Phua, K. Smith-Miles, V. Lee, and R. Gayler, "Resilient identity crime detection," IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 3, pp. 533–546, 2012.

[26] M. W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" Journal of Machine Learning Technologies, vol. 2, no. 1, pp. 37–63, 2011.

[27] Credit Card Fraud Detection [Online]. Available: https://www.kaggle.com/dalpozz/creditcardfraud

[28] R. Saia and S. Carta, "Evaluating Credit Card Transactions in the Frequency Domain for a Proactive Fraud Detection Approach," In Proceedings of the 14th International Joint Conference on eBusiness and Telecommunications, vol. 4, pp. 335–342, 2017.

[29] ISO 8583-1:2003 Financial transaction card originated messages [Online]. Available: https://www.iso.org/standard/31628.html