IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

Survey on Credit Card Fraud Detection using Recurrent Attributes

Navneet Jain[1], Vasima Khan[2]

UIT BU Computer Science Department, Barkatullah University, Bhopal, India [1, 2]

Abstract: It is very important to extract the right features from transactional data in implementing a credit card fraud detection model. It is normally done by combining the transactions in order to observe the spending patterns of the customers. We propose to create a new set of features based on analyzing the periodic behavior of the time of a transaction using the von Mises distribution in this paper. We compare credit card fraud detection models, and evaluate how the different sets of features have an impact on the results with the help of a real credit card fraud dataset provided by a large European card processing company. The results show an average increase in savings of 13% by including the proposed periodic features into the methods. The methodology proposed in this paper is currently being incorporated into the fraud detection system of aforementioned card processing company.

Keywords: Fraud detection; von Mises distribution; Cost sensitive learning.

I. INTRODUCTION

Credit card fraud has been an increasing problem worldwide. Representing an increase of 14.8% compared with 2011 [1], the total level of fraud reached 1.33 billion Euros in the Single Euro Payments Area during 2012. Moreover, payments across non-traditional channels (mobile, internet,) accounted for 60% of the fraud, whereas it was 46% in 2008. This opens new challenges as new fraud patterns arise, and current fraud detection systems are less capable in preventing these frauds. Furthermore, to avoid being detected, fraudsters regularly change their strategies, something that makes traditional deceit recognition tools, such as expert rules, insufficient [2].

In fraud detection, the use of machine learning has been an interesting topic in recent years. Based on machine learning techniques, different detection systems have been successfully used for this problem, in particular: neural networks [3], Bayesian learning [4], artificial immune systems, hybrid models, support vector machines, peer group analysis, online learning and social network analysis.

Nowadays, enterprises and public institutions need automatic systems to implement fraud detection and have to face a growing presence of fraud initiatives. Since it is not always possible or easy for a human analyst to detect fraudulent patterns in transaction datasets, constantly characterized by a large number of samples, many dimensions and online updates, automatic systems are imperative. Also, the cardholder is not reliable in reporting the theft, loss or fraudulent use of a card [5]. Since the number of fraudulent transactions is much smaller than the legitimate ones, the data distribution is unbalanced, i.e. skewed towards non-fraudulent observations. Methods have been proposed to improve the performances of many learning algorithms which underperform when used for unbalanced dataset. Many other factors other than unbalancedness determine the difficulty of a classification/detection task. Another in influential factor is the amount of overlapping of the classes of interest due to limited information that transaction records provide about the nature of the process [6].

It is very important to use those features that allow precise classification when constructing a credit card fraud detection model. Raw transactional features, such as time, amount, and place of the transaction are only used by typical models. However, the spending behavior of the customer, which is expected to help discover fraud patterns [7], is not taken into account by these approaches. In [8], where Whitrow et al. proposed a transaction aggregation strategy in order to take into account a customer spending behavior is a standard way to include these behavioral spending patterns? In classification of the transactions made during the last given number of hours, first by card or account number, then by transaction type, merchant group, country or other, followed by calculating the number of transactions or the total amount spent on those transactions, the computation of the aggregated features consists.

In many situations, the prediction of user behaviour in financial systems can be used. A lot of money and other resources can be saved by predicting client migration, marketing or public relations. The fraud of credit lines, especially credit card payments is one of the most interesting fields of prediction. A reduction of 2.5% of fraud triggers a saving of one million dollars per year [9] for the high data traffic of 400,000 transactions per day. Certainly, all transactions which deal with accounts of known misuse are not authorized. Experienced people can tell that the transactions are probably misused, caused by stolen cards or fake merchants even if they are formally valid. So, a credit card transaction before it is known as “illegal”, fraud has to be avoided.

Copyright to IJARCCE

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

People can no longer control all increasing number of transactions. As solution one can put the experience of expert into an expert system. The expert's knowledge, even when it can be obtained clearly, changes regularly with new kinds of organized attacks and patterns of credit card fraud this is the disadvantage of this traditional approach. No predefined fraud models as in [10] but automatic learning algorithms are needed in order to keep track with this.

Along with an increasing volume of payment traffic, advancement and expansion of modern technology and sophistication of fraudulent tactics, credit card fraud is growing. Significant losses and great inconvenience to issuing companies, merchants and customers world-wide is caused by it. Total card fraud losses on UK issued cards increased by 25% from the previous year and amounted to £535 million (APACS, 2008) in year 2007. Within the following categories the range of fraud tactics observed in the industry can be broadly described: In response to practices of issuing companies and merchants to protect against identified tactics in the future this list evolves over time as fraudsters adapt new strategies like lost and stolen card fraud, counterfeit card fraud, card not present fraud, mail non-receipt card fraud, account takeover fraud and application fraud. Currently in the UK, Card-not-Present fraud, where the physical card is not present at the point-of-sale is the largest type of credit card fraud. This includes fraud conducted over the Internet, by telephone, fax and mail order and amounts to 54% of all fraud on UK cards. As face-to-face fraudulent transactions become increasingly difficult, it is expected that the volume of CNP fraud will continue to grow.

A number of challenges for designing a fraud detection system are presented by the nature of transaction data and some particular operational issues:

- Each transaction contains more than 70 fields of coded information furthermore the number of transactions processed by credit card issuers daily is high. Transaction data is heterogeneous and changing time to time within and between accounts. For different groups of merchants, holiday seasons and geographical regions patterns and trends vary expressively.
- Within the credit card industry the generally accepted fraud rate is 0.1–0.2%, i.e. the occurrence of fraud is relatively rare. Repeatedly this leads to the problem that the majority of cases detected by the fraud detection system as being potentially fraudulent are in fact legal. This type of error is referred to as false positive (FP). The associated costs and customer inconvenience increased as the number of FPs increase.
- For further investigation, alerts emerging from the fraud detection system are usually passed on to the fraud department. For verification of the transactions, where it is required by the bank policy, the suspected cases are followed up with a call to a cardholder. As a result of this, the number of alerts should be kept at a level such that it can be handled by the available number of investigators and fraud analysts.
- When the cardholder identifies that their account has been compromised fraudulent cases missed by the fraud detection system are reported to the issuing company. Resulting in a delay in correctly labelling each case, this can take up to several months. Some fraudulent cases are mislabelled because they remain unidentified. Thus, a fraud detection model is almost certainly trained on noisy data.

Based on analyzing the time of a transaction, we propose a new set of features in this paper. At similar hours, it is expected from customer to make transactions. This is the logic behind it. Hence, based on the periodic behavior of a transaction time, using the von Mises distribution [11] a new method for creating features is proposed. In particular, if the time of a new transaction is within the confidence interval of the previous transaction time new time features should estimate.

Furthermore, using two kinds of classification algorithms; cost-insensitive [12] and example-dependent cost-sensitive using a real credit card fraud dataset provided by a large European card company we compare various sets of features (raw, aggregated and periodic). By using the proposed periodic features the results show an average increase in the savings of 13%. Additionally, to implement a state-of-the-art fraud detection system that will help to combat fraud the outcome of this paper is being currently used once the implementation stage is finished.

The remainder of the paper is organized as follows. In Section 2, we discuss current approaches to create the features used in fraud detection models. We present our proposed methodology to create periodic features in section 3. Afterwards, the experimental setup and the results are given in Sections 4 and 5. Finally, conclusions and discussions of the paper are presented in Section 6.

II. STUDY ON METHODS

In this section, various techniques of credit card fraud detection are discussed and analyzed.

V. Van Vlasselaer, C. Bravo, O. Caelen, T. Eliassi-Rad, L. Akoglu, M. Snoeck, and B. Baesens (2015)

In [2], authors have introduced a novel approach to detect fraudulent credit card transactions conducted in online stores. Their approach combines (1) intrinsic attributes derived from the properties of incoming transactions and the customer spending patterns with the help of the fundamentals of RFM (Recency–Frequency–Monetary); and (2) network related attributes by studying the network of credit card holders and merchants and developing a time-dependent suspiciousness score for each network object. Their results show that both intrinsic and network-based features are two strongly intertwined sides of the same picture. The best performing models which reach AUC-scores higher than 0.98 are generated by the combination of these two types of features.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

A. D. Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi (2014)

In this paper author provide some answers from the practitioner's perspective by focusing on three crucial issues: unbalancedness, non-stationary and assessment. The analysis is made possible by a real credit card dataset provided by our industrial partner. Billions of dollars of loss are caused every year due to fraudulent credit card transactions. For reducing these losses, more algorithms depend on advanced machine learning methods to help fraud investigators the design of better fraud detection algorithms is the key. Due to non-stationary distribution of the data, highly imbalanced classes distributions and continuous streams of transactions the design of fraud detection algorithms is however particularly challenging.
For confidentiality issues, leaving unanswered many questions about which is the best strategy to deal with them time public data are scarcely available at the same time.

A.Correa Bahnsen, A. Stojanovic, D. Aouada, and B. Ottersten (2013)

Author proposes an evaluation measure that realistically represents the monetary gains and losses due to fraud and its detection. Moreover, he present a Bays minimum risk classifier including the real financial costs of credit card fraud detection in order to have a cost sensitive detection system. Affecting card holders around the world credit card fraud is a growing problem. Fraud detection has been an interesting topic in machine learning. Nevertheless, current state of the art credit card fraud detection algorithms miss to include the real costs of credit card fraud as a measure to evaluate algorithms.

Alejandro Correa Bahnsen (2014)

With the objective of finding the model that minimizes the real losses due to fraud two different methods for calibrating probabilities are evaluated and analyzed in the context of credit card fraud detection in this paper. It is shown that when probabilistic models are used to make decisions based on minimizing risk, using the full dataset provides expressively better results even though under-sampling is often used in the context of classification with unbalanced datasets. In order to test the algorithms, a real dataset provided by a large European card processing company is used. It is shown that the losses due to fraud are reduced by calibrating the probabilities and then using Bays minimum Risk. Furthermore the aforementioned card processing company is currently incorporating the methodology proposed in this paper into their fraud detection system due to good overall results. Lastly, the methodology has been tested on a different application, namely, direct marketing.

A. Correa Bahnsen, D. Aouada, and B. Ottersten (2015)

By incorporating the different example-dependent costs into a new cost-based impurity measure and a new cost-based pruning criteria authors propose an example-dependent cost-sensitive decision tree algorithm. Then, using three different databases, from three real-world applications: credit card fraud detection, credit scoring and direct marketing, they evaluate the proposed method. The results show that for all databases, the proposed algorithm is the best performing method. Furthermore, while having a superior performance measured by cost savings, leading to a method that not only has more business-related results, but also a method that develops easier models that are easier to analyze when compared against a standard decision tree our method builds significantly smaller trees in only a fifth of the time,.

A.Correa Bahnsen, D. Aouada, and B. Ottersten (2014)

Author proposed a new example-dependent cost matrix for credit scoring in this paper. Furthermore, they propose an algorithm that introduces the example-dependent costs into a logistic regression. Authors compare our proposed method against state-of-the-art example-dependent cost-sensitive algorithms using two publicly available datasets. The results highlight the importance of using real financial costs. Moreover, significant improvements are made in the sense of higher savings by using the proposed cost-sensitive logistic regression.

III. COMPARATIVE ANALYSIS

The methods studied above are compared in terms of advantages, disadvantages, techniques and accuracy performance. Table 1 is showing the comparative study among these methods.

TABLE 1 COMPARATIVE STUDY OF CREDIT CARD FRAUD DETECTION METHODS

Paper Title	Key Techniques and Methods	Advantages	Disadvantages
APATE: A Novel Approach for Automated Credit Card Transaction Fraud Detection using Network Based Extensions	APATE, RFM (Recency Frequency Monetary), supervised learning	Automatically Detect online fraudulent transactions.	Processing Time is not evaluated.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

Learned lessons in credit card fraud detection from a practitioner perspective	Incremental learning; Unbalanced data	Has formalised the fraud detection problem and proposed total detection cost as the correct metrics for measuring the Detection performance.	Less accuracy and processing time is not evaluate
Cost Sensitive Credit Card Fraud Detection Using Bays Minimum Risk	Bayesian decision Theory, Cost sensitive classification	Robust and simple method	Processing time is not evaluated and very complex method
Improving Credit Card Fraud Detection with Calibrated Probabilities	Calibrated probabilities, probabilistic models	Using Bays minimum Risk the losses due to fraud are Reduced.	Processing time is not evaluated.
Example-Dependent Cost-Sensitive Decision Trees	Cost-sensitive learning, Cost-Sensitive Classifier, Credit scoring	Having good Accuracy	Processing time is not evaluated and conflicts resulted sometimes while classification process.
Example-Dependent Cost-Sensitive Logistic Regression for Credit Scoring	Cost-sensitive learning, Logistic Regression	Evaluated and compared in terms of precision, recall and accuracy rates.	This is complex method and processing time is not evaluated.

IV. RESEARCH PROBLEM

After studying the recent methods and comparing their performances, in this section the current limitations and research challenges are highlighted for future work. Working on credit card fraud detection is very essential now days, hence its must that method should be efficient and robust in all aspects. Recently many researchers did work to deliver the best solution to detect forgery in images, but we had below observations through our study:

- Most of existing methods are not consider and evaluated the processing time and complexity parameters.
- The methods with best accuracy are having very complex procedure for forgery detection.

Some methods are designed and evaluated by considering on accuracy metrics while precision, recall and complexity are equally important for evaluation purpose.

V. CONCLUSION AND FUTURE WORK

In this paper, introduction to credit card fraud detection using recurrent features is presented and explained at first, and then importance of detecting frauds on digital transactions is given. The different types of credit card frauds and different types of methods explained. Basically this paper is aimed to present the study on all recent 2013 to 2016 credit card fraud detection methods with comparative analysis. Section II and III, presented the detailed study on all recent techniques and compare them accuracy wise. Finally, the research limitations and problems have been pointed out in section IV. For future work, we suggest to work on addressing the current research problems.

REFERENCES

[1] S. Panigrahi, A. Kundu, S. Sural, and A. Majumdar, "Credit card fraud detection: A fusion approach using Dempster Shafer theory and Bayesian learning," Information Fusion, vol. 10, no. 4, pp. 354–363, Oct. 2009.
[2] M. Krivko, "A hybrid model for plastic card fraud detection systems," Expert Systems with Applications, vol. 37, no. 8, pp. 6070–6076, Aug. 2010.
[3] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," Decision Support Systems, vol. 50, no. 3, pp. 602–613, Feb. 2011.
[4] D. J. Weston, D. J. Hand, N. M. Adams, C. Whitrow, and P. Juszczak, "Plastic card fraud detection using peer group analysis," Advances in Data Analysis and Classification, vol. 2, no. 1, pp. 45–62, Mar. 2008.
[5] C. Whitrow, D. J. Hand, P. Juszczak, D. J. Weston, and N. M. Adams, "Transaction aggregation as a strategy for credit card fraud detection," Data Mining and Knowledge Discovery, vol. 18, no. 1, pp. 30–55, Jul. 2008.
[6] European Central Bank, "Third report on card fraud," European Central Bank, Tech. Rep., 2014.
[7] V. Van Vlasselaer, C. Bravo, O. Caelen, T. Eliassi-Rad, L. Akoglu, M. Snoeck, and B. Baesens, "APATE: A Novel Approach for Automated Credit Card Transaction Fraud Detection using Network Based Extensions," Decision Support Systems, vol. 75, pp. 38–48, 2015.
[8] A. D. Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," Expert Systems with Applications, vol. 41, no. 10, pp. 4915–4928, Aug. 2014.
[9] A. Correa Bahnsen, A. Stojanovic, D. Aouada, and B. Ottersten, "Cost Sensitive Credit Card Fraud Detection Using Bays Minimum Risk," in 2013 12th International Conference on Machine Learning and Applications. Miami, USA: IEEE, Dec. 2013, pp. 333–338.
[10] Improving Credit Card Fraud Detection with Calibrated Probabilities," in Proceedings of the fourteenth SIAM International Conference on Data Mining, Philadelphia, USA, 2014, pp. 677 – 685.
[11] A. Correa Bahnsen, D. Aouada, and B. Ottersten, "Example Dependent Cost-Sensitive Decision Trees," Expert Systems with Applications, vol. 42, no. 19, pp. 6609–6619, 2015.
[12] A. Correa Bahnsen, D. Aouada, and B. Ottersten, "Example-Dependent Cost-Sensitive Logistic Regression for Credit Scoring," in 2014 13th International Conference on Machine Learning and Applications. Detroit, USA: IEEE, 2014, pp. 263–269.