Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 48 (2015) 679 – 685

Procedia Computer Science

International Conference on Intelligent Computing, Communication & Convergence (ICCC-2015)

Conference Organized by Interscience Institute of Management and Technology, Bhubaneswar, Odisha, India

Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier

Masoumeh Zareapoor[a], Pourya Shamsolmoali[a,b]

[a] *Department of Computer science, Jamia Hamdard University, New Delhi, India*
[b] *Department of Computer Science, *Baghin University, Kerman, Iran*

Abstract

Credit card fraud is increasing considerably with the development of modern technology and the global superhighways of communication. Credit card fraud costs consumers and the financial company billions of dollars annually, and fraudsters continuously try to find new rules and tactics to commit illegal actions. Thus, fraud detection systems have become essential for banks and financial institution, to minimize their losses. However, there is a lack of published literature on credit card fraud detection techniques, due to the unavailable credit card transactions dataset for researchers. The most commonly techniques used construct the fraud detection model. The performance evaluation is performed on real life credit card transactions dataset todemonstrate the benefit of the bagging ensemble algorithm.

Peer-review under responsibility of scientific committee of International Conference on Computer, Communication and Convergence (ICCC 2015)

Peer-review under responsibility of scientific committee of International Conference on Computer, Communication and Convergence (ICCC 2015)
doi:10.1016/j.procs.2015.04.201

fraud detection methods are Naïve Bayes (NB), Support Vector Machines (SVM), K-Nearest Neighbor algorithms (KNN). These techniques can be used alone or in collaboration using ensemble or meta-learning techniques to build classifiers. But amongst all existing method, ensemble learning methods are identified as popular and common method, not because of its quite straightforward implementation, but also due to its exceptional predictive performance on practical problems. In this paper we trained various data mining techniques used in credit card fraud detection and evaluate each methodology based on certain design criteria. After several trial and comparisons; we introduced the bagging classifier based on decision three, as the best classifier to construct the fraud detection model. The performance evaluation is performed on real life credit card transactions dataset to demonstrate the benefit of the bagging ensemble algorithm.

*Keywords:*Credit card fraud; Fraud detection technique; Data mining; Bagging ensemble classifer.

1.Introduction

Credit card is considered as a "nice target of fraud" since in a very short time attackers can get lots of money without much risk and most of the time the fraud is discovered after few days. To commit the credit card fraud either offline or online, fraudsters are looking for sensitive information such as credit card number, bank account and social security numbers. In case of offline payment (which using credit card physically) to perform the fraudulent transactions, an attacker has to steal the credit card itself, while in the case of online payment (which occurs over internet or phone), the fraudsters should be steal costumer's identity. Credit card fraud is a significant issue and has considerable cost for banks and card issuer companies. Thus, with this massive problem in transaction system, banks take credit card fraud very seriously, and have highly sophisticated security systems to monitor transactions and detect the frauds as quickly as possible once it is committed. A secured and trusted banking payment system requires high speed verification and authentication mechanisms that let legitimate users easy conduct their business, while flagging and detecting suspicious transaction attempts by others. Fraud detection has become a vital activity in order to decrease the impact of fraudulent transactions on service delivery, costs, and reputation of the company. According to Kount, one of the top five fraud detection consultants revealed by *topcreditcardprocessors.com* in the month of August 2013[1], 40% of the total financial fraud is related to credit card and the loss of amount due to credit card fraud worldwide is $5.55 billion. There are various methods used for fraud detection each of them tries to increase the detection rate while keeping false alarm rate at minimum. Different methods have been used for fraud detection including such as, Bayesian algorithm [15][9], K-Nearest Neighbor[17][7], Support Vector Machine [17][6] etc. Statistical fraud detection methods have been divided into two broad categories [3][11]: supervised and unsupervised. In supervised fraud detection methods, models are estimated based on the samples of fraudulent and legitimate transactions, to classify new transactions as fraudulent or legitimate. While in unsupervised fraud detection, outliers or unusual transactions are identified as potential cases of fraudulent transactions. Both these fraud detection methods predict the probability of fraud in any given transaction. The objective of this paper is to perform a comprehensive review of various fraud detection methods and selects some innovative method and technique for discussion.

2. Challenges in credit card fraud detection

Credit card fraud detection is one of the most explored domains of fraud detection and relies on the automatic analysis of recorded transactions to detect fraudulent behaviour. Furthermore the problem of credit card fraud detection has many constraints. Here we discussed various challenges in credit card fraud detection are:

2.1 Non-availability of real data set

One of the biggest issues associated with credit card fraud detection is the unavailability of dataset that researchers can perform the research on as it mentioned by many authors [11][14][16]. The reason of unavailability of real world data is, banks and financial institution are not ready to reveal their sensitive customer transaction data due to privacy reasons.

2.2 Unbalanced Data Set

Credit card fraud datasets are highly skewed data (where many more of data is legitimate, and a few of

[1] www.prweb.com/releases/2013/8

them is fraudulent), and the legal and fraud transactions vary at least hundred times [14]. Generally, in real case 98% of the transactions are legal while only 2% of them are fraud [13].

2.3 Size of the Data Set

Chan et al point out in [5] that, millions of credit card transactions are processed every day. Analyzing such enormous amounts of transactions requires highly competent techniques that scale well, as well as requiring considerable computing power. It creates certain restrictions for the researchers.

2.4 Determining the appropriate evaluation parameters

There are two very common measures for the fraud detection techniques: false-positive and false-negative rates. These two measures have an opposite relationship, one decrease and other one increase. Accuracy is not a suitable metrics for credit card fraud detection technique since; the dataset is highly imbalanced [3]. Therefore with very high accuracy all fraudulent transactions can be misclassified. The error cost of misclassifying fraudulent instances is higher than the error cost of misclassifying legitimate instances, it is important to study not only the precision (correct classified instances) but also the sensibility (correct classified fraudulent instances) of each case.

2.5 Dynamic behaviour of fraudster

Fraudsters having dynamic behaviour mean that the fraudsters change their behaviour over the time to get through any new detection system and modify fraud styles. So, fraud is becoming increasingly more complex and sophisticated which which is not even predictable by human experts.

But with these challenges, credit card fraud detection is still a fashion and hard research topic.

3. Data mining techniques

We investigated the performance of five states of art techniques in predicting credit card fraud: Support Vector Machines (SVM), Naïve Bayes (NB), KNN and Bagging ensemble classifier. In the paragraphs below, we briefly describe the techniques employed in this study.

3.1 Naive Bayes classifiers

Naïve Bayes is a supervised machine learning method that uses a training dataset with known target classes to predict the class of future instances. This algorithm was first introduced by John and Langley (1995) [2]. In simplest terms, a Naïve Bayes method assumes that the "presence or absence" of a particular attribute of a set is not based on the presence or absence of any other attributes in the same set. Experiments on the real world dataset have shown that the NB algorithm performs comparably well. However, this technique is named by the name "Naïve" because it naively assumes independence of the attributes given the class [17][2]. Then the classification is done by applying "Bayes" rule to calculate the probability of the correct class which is the particular attributes of the credit card transaction. Bayes theory is calculated as:

$$\textit{Prior Probability of Z: } \frac{\text{Number of Z instances}}{\text{total number of instances}} \qquad \textit{Likelihood of Y given Z: } \frac{\text{Number of Z in vicity of Y}}{\text{total number of Z}}$$

In the Bayesian theory, the final classification is produced by combining both information (likelihood, priori), to form a *posterior probability* which is called Bayes rule.

$$\textit{Posterior} = (\text{Prior} * \text{Likelihood}) / (\text{Evidence})$$

It has a good performance with small amount of training data. It is used to solve both binary classification problem and multiclass classification problem.

3.2 K-Nearest Neighbor algorithm

The k-Nearest Neighbor (KNN) technique is a simplest technique that stores all available instances and then

classifies any new instances based on a similarity measure. KNN has been used in statistical estimation and pattern recognition in the beginning of 1970's. The KNN algorithm is an example of an instance based learner. In other word, all of the learning models are "instance based," as well, because they start with a set of instances as the initial training information. In the nearest-neighbour classification method, each new instance is compared with existing ones by using a distance metric, and the closest existing instance is used to assign the class to the new one. Sometimes more than one nearest neighbour is used, and the majority class of the closest *k* neighbours is assigned to the new instance. The concept of the instance-based nearest-neighbour algorithm was first introduced by Aha, Kibler, and Albert (1991) [4]. K- Nearest neighbour based credit card fraud detection techniques require a distance or similarity measure defined between two data instances [17][1]. In process of KNN, we classify any incoming transaction by calculating a nearest point to the new incoming transaction. Then if the nearest neighbour be fraudulent, then the transaction indicates as a fraud. The value of K is generally small and odd to break the ties (typically 1or 3). Larger K values can help to reduce the effect of noisy data set. In this algorithm, distance between two data instances can be calculated in different ways. For continuous attributes, Euclidean distance is a good choice. For categorical attributes, a simple matching coefficient is often used. The most important pitfall of KNN algorithm is that unrelated attributes have a large negative impact on the training process of the K-Nearest neighbour and because of these irrelevant attributes the training of classifiers based on these algorithm can often be inefficient and impractical [17].

3.3 Support Vector Machines (SVMs)

The Support Vector Machine (SVM) is statistical learning technique which is especially suitable for binary classification technique [14][6][17] such as credit card fraud detection techniques which only two classes are needed, namely the legitimate and fraudulent class. The goal of the SVM method is to construct a "hyperplane" which do separate the data instances into two classes:- positive and negative [8][30]. The strength of SVMs comes from two main properties:- kernel representation and margin optimization. Kernels, such as radial basis function (RBF) kernel, can be used to learn complex regions. A kernel function represents the dot product of projections of two data instance in a high dimensional feature space. The basic technique finds the smallest "hypersphere" in the kernel space that contains all training instances, and then determines on which side of "hypersphere" a test instance lies. This classifier finds the maximum margin hyper plane, and it classifies all training instances correctly by separating them into correct classes through a hyper plane. The maximum margin hyper plane is the one that gives the greatest separation between the classes. The instances that are nearest to the maximum margin hyper plane are called support vectors. In credit card fraud detection if a test instance lies within the learned region, it is stated as normal; else it is declared as anomalous. SVM methods require large training dataset sizes in order to achieve maximum prediction accuracy. However, regular SVM method is invalid to the imbalanced data sets. Because in imbalanced data sets, the learned boundary is close to the minority instances, so SVM should be biased in a way that will push the boundary away from the positive samples [17].

3.4 Bagging ensemble classifier based on decision tree

Bagging classifier is an ensemble technique which was proposed by Leo Breiman in 1994 [12]. It can be handle classification and regression methods. It is designed to improve the stability and accuracy of machine learning algorithms used in classification and regression. It works by combining classifications of randomly generated training sets to form a final prediction. Such techniques can typically be used as a variance reduction technique by randomization into its construction procedure and then creating an ensemble out of it. Bagging classifier has attracted much attention, due to its simple implementation and the improving accuracy. Thus, we can call bagging as a "smoothing operation" that has advantage when intending to improve the predictive performance of regression or classification trees. The basic principle behind of this ensemble method is that a group of "weak learners" can come together to form a "strong learner". Bagging grows many decision trees. Here each individual decision tree is a "weak learner", while all the decision trees taken together are a "strong learner". When a new instance has to be classified, it is done repeatedly to each of the trees in the ensemble. Each tree gives a "vote" for a class. The final prediction for the new instance's class is gained by the class having maximum votes. In this paper we used bagging classifier [12], with the decision tree algorithm J48 based on the C4.5 model as the single classifier to construct the ensemble. The reason for selecting decision three as a single classifier for our ensemble is that, our dataset is highly imbalanced, so decision three algorithm presents a very good behavior by weighting the results of the trees and reducing the variance of the dataset and the overfitting. Bagging ensemble classifier is fast and they can efficiently

handle unbalanced and large databases with thousands of features.

4. Experimental setup

The objective of this paper is to examine the evaluation performance of three advanced data mining techniques, with the well known and proposed bagging ensemble classifier, for credit card fraud detection technique. In this work we used 10 fold cross validation techniques.

4.1 Dataset

In the field of credit card fraud detection technique there are different types of datasets with different fraud properties, e.g. type of fraud, number of fraudulent records, variety of fraud, the distribution of fraudulent transactions among legal transactions. In this paper for evaluating the state of art techniques we used the real world credit card dataset which is obtained from 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). The dataset is a real dataset of e-commerce transactions and the objective was to detect anomalous e-commerce transactions. The obtained dataset include of 100,000 records of credit card transactions. Each record has 20 fields. The data given to us was already labeled by bank, as legitimate and fraudulent. The ratio of legitimate transactions to fraudulent transactions approximately is 100:3. This means that among the 100,000 records, 2.8% (2293 records) are fraudulent transactions, and 97.2% (records) are legitimate transactions. The data are sampled from 98 days period. In our dataset we couldn't find any difference between the attributes in legitimate and fraudulent transactions, due to appearance of fraudulent behavior more and more like legitimate ones. Figure 1 present the credit card transactions dataset available in this paper as follow:

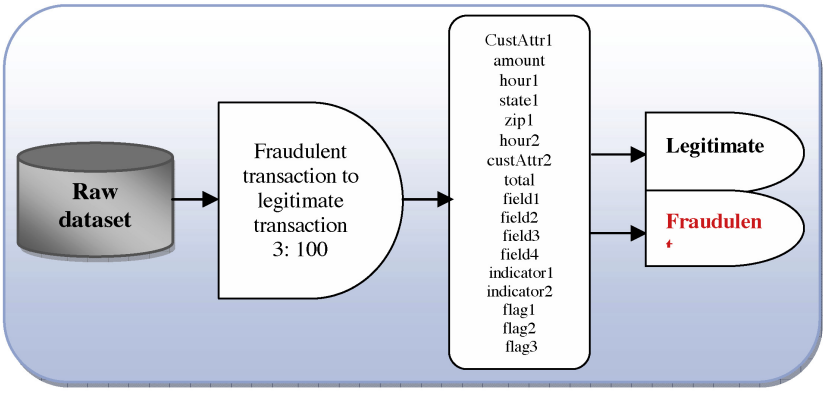

Fig. 1. Dataset description

4.2 Training and test data

As mentioned in introduction section, the given dataset in this paper is highly imbalanced data which is the nature of credit card transaction dataset. We divided our dataset into 4 groups. As shown in Table 1, the fraud rates in these modeling datasets Df1, Df2, Df3, Df4 is approximately 20%, 15%, 10%, 3% respectively.

5.1 Performance measures

In this paper we didn't consider some of common metrics like accuracy and error rate, since they are known to be bias metrics in the case of imbalanced dataset. For fraud detection domain, the "fraud catching rate" and "false alarm rate" are the criteria metrics. We are evaluated the performance of the various techniques in terms of 4 classification metrics relevant to credit card fraud detection [18] – Fraud Catching Rate, False Alarm Rate, Balanced Classification Rate and Matthews Correlation Coefficient. Here, fraud is considered as positive class and legal as negative class and hence the meaning of the terms TP, TN, FP and FN are defined as follows:

True Positive (TP) = Number of fraud transactions predicted as fraud

True Negative (TN) = Number of legal transactions predicted as legal

False Positive (FP) = Number of legal transactions predicted as fraud

False Negatives (FN) = Number of fraud transactions predicted as legal

6. Results

This section presents result of our evaluation performance model developed from the dataset. In this paper, we compare several standard classifiers with the bagging ensemble classifier which is novel technique in credit card fraud detection technique.

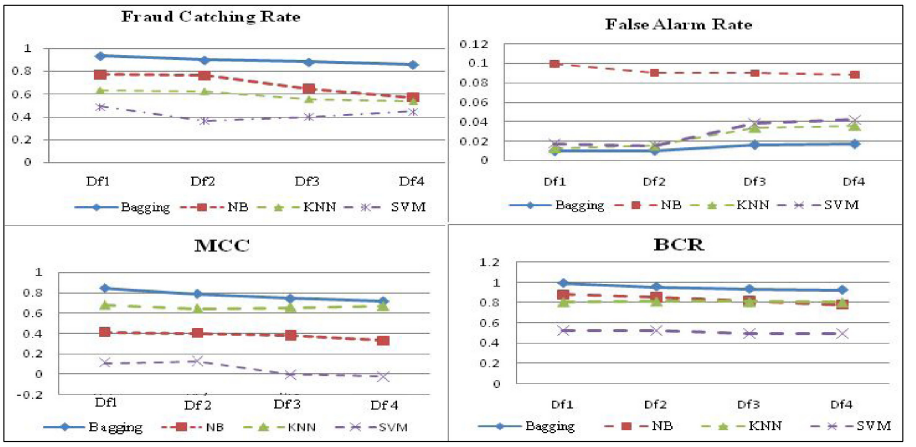

Fig. 2. Performance of all classifies in terms of MCC, BCR, FAR, FCR LS using real world credit card dataset

As shown in figure 2, bagging classifier based on decision three, very well detect the fraudulent transactions, the fraud catching rate is high keeping and the false alarm rate very low. While other methods have problem to detect the fraudulent transactions by increasing the number of false alarm rate. Also, the results show the superiority of the bagging classifier based decision three, irrespective to the rate of fraud. It means that, the bagging ensemble classifier has stable performance in during the training and testing evaluation, and is independent to rate of frauds. Even it shows the better result with highly imbalanced dataset (Df4). When we compare the methods in terms of false alarm rate, we observe that the behaviour of the bagging technique is similar in duration of evaluation (for Df1, Df2, Df3, and Df4). While the performance of other classifiers is lower than bagging classifier. Two balanced metrics-BCR & MCC are used to evaluate the capability of various techniques for handling class imbalance and figure 2, shown bagging classifier has a very good performance according to these measures compared with other classifiers. This is explained by the fact that bagging ensemble classifier is more suitable for credit card fraud detection, since the nature of dataset is highly imbalanced, and it has capability to handle the imbalanced dataset. But other standard classifiers are known to be bias classifiers.

7. Conclusion

This paper examined the performance of three states of art data mining techniques, with bagging ensemble classifier based on decision three algorithms which is a novel technique in area of credit card fraud detection system. A real life dataset on credit card transactions is used for our evaluation. And we found that, the bagging classifier based on decision three works well with this kind of data since it is independent of attribute values. The second feature of this novel technique in credit card fraud detection is its ability to handle class imbalance.

This is incorporated in the model by creating four sets of dataset (Df1, Df2, Df3, DF4) which the fraud rate in each of them were 20%, 15%, 10%, 3% respectively. Bagging classifier based decision three algorithm performance is found to be stable gradually during the evaluation. More over the bagging ensemble method takes very less time, which is also an important parameter of this real time application, because in fraud detection domain time is known one of the important parameter.

ACKNOWLEDGEMENTS

The authors wish to thank Dr Kohei Hayashi, Nara Institute of Science and Technology, Japan, and Dr.Haiqin Yang, Chinese University of Hong Kong for providing the Dataset.

Table1. Training and testing dataset

Group of dataset	Legal	Fraud	Total	Fraud rate
Df1	14170	2834	17004	**20%**
Df2	18895	2834	21729	**15%**
Df3	28340	2834	31174	**10%**
Df4	97166	2834	10000	**3%**

References

[1] S. Kotsiantis, D. Kanellopoulos, P. Pintelas (2006). Handling imbalanced datasets: A review. International Transactions on Computer Science and Engineering.
[2] G.H. John, P. Langley (1995). Estimating continuous distributions in Bayesian classifiers. in: Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, (1995); 338 – 345.
[3] R.J. Bolton, D.J. Hand (2001). Unsupervised profiling methods for fraud detection. *In Conference on credit scoring and credit control*, Edinburgh.
[4] D. Kibler, D.W. Aha, M. Albert (1989). Instance-based prediction of real-valued attributes. Computational Intelligence, Vol(5); 51-57.
[5] P.K. Chan, W. Fan, A.L. Prodromidis, S.J. Stolfo (1999). Distributed Data Mining in Credit Card Fraud Detection. *IEEE Intelligent Systems*, pp 67–74.
[6] C. Cortes, V. Vapnik (1995). Support vector networks. Machine Learning. , 20:273–297.
[7] T.M. Cover, P.E. Hart (1967). Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory*, 13(1):21–27.
[8] G. Potamitis (2013). Design and Implementation of a Fraud Detection Expert System using Ontology-Based Techniques. A dissertation submitted to the University of Manchester for the degree of Master of Science in the Faculty of Engineering and Physical Sciences.
[9] E. David (2012). Bayesian inference-the future of online fraud protection. *Computer Fraud & Security*, 8-11.
[10] S. Ghosh, D.L. Reilly. (1994). Credit Card Fraud Detection with a Neural- Network. *In Proceedings of the International Conference on System Science*, pages 621-630.
[11] Jha.Sanjeev, G. Montserrat, J.C.Westland (2012). Employing transaction aggregation strategy to detect credit card fraud. *Expert system with application*, 39: 12650-12657.
[12] L. Breiman. Random forests. Machine Learning, (2001). Vol(45); 5–32.
[13] J. Piotr., A.M. Niall, J.D. Hand, C. Whitrow, J. David (2008). Off the peg and bespoke classifiers for fraud detection. *Computational Statistics and Data Analysis,* 52, pp:4521-4532.
[14] L. Qibei. & J. Chunhua. (2011). Research on Credit Card Fraud Detection Model Based on Class Weighted Support Vector Machine. *Journal of Convergence Information Technology*, 6(1), 62-68.
[15] S. Maes, K.Tuyls, B.Vanschoenwinkel, B.Manderick (1993). Credit card fraud detection using Bayesian and neural networks. *In Proceedings of the First International NAISO Congress on Neuro Fuzzy Technologies*, pages 261-270.
[16] E.W.T.Ngai, H.Yong., Y.H.Wong, Y.Chen, X. Sun (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50:559–569
[17] M. Zareapoor, Seeja, K.R. & M. Alam ,Afshar (2012).Analyzing Credit Card:Fraud Detection Techniques Based On Certain Design Criteria. *International Journal of Computer Application*, 52(3):35-42.
[18] http://www.damienfrancois.be/blog/files/modelperfcheatsheet.pdf