A Fraud Detection Approach Based on Combined Feature Weighting

Xiaoqian Liu[1,3,4](✉), Chenfei Yu[1], Bin Xia[2], Haiyan Gu[1], and Zhenli Wang[1]

[1] Department of Computer Information and Cyber Security, Jiangsu Police Institute,
Nanjing 210031, China
liuxiaoqian@jspi.edu.cn, 1578898262@qq.com, {guhaiyan,wangzhenli}@jspi.cn

[2] Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing
University of Posts and Telecommunications, Nanjing 210023, China
bxia@njupt.edu.cn

[3] Jiangsu Electronic Data Forensics and Analysis Engineering Research Center,
Nanjing, China

[4] Jiangsu Provincial Public Security Department Key Laboratory of Digital
Forensics, Nanjing, China

Abstract. Data mining technology has yielded fruitful results in the area of crime discovery and intelligent decision making. Credit card is one of the most popular payment methods, providing great convenience and efficiency. However, due to the vulnerabilities of credit card transactions, criminals are able to commit fraud to infringe on the interests of the state and citizens. How to discover potential fraudsters while guaranteeing high efficiency becomes an extremely valuable problem to solve. In this work, we talk about the advantages and disadvantages of different models to detect credit card fraud. We first introduce the data preprocessing measures for handling imbalanced fraud detection dataset. Then we compare related models to implement fraudster recognition. We also propose a feature selection approach based on combined feature weights. Some future research interests are also envisioned.

Keywords: Fraud detection · Imbalanced dataset · Fisher score · Feature weighting

1 Introduction

With the prosperity of the Internet technology, the number of netizens is rapidly increasing. According to the 45th statistical report on Internet development in China issued by China Internet Network Information Center, by March 2020, the number of Internet users in China has reached 904 million. In the meantime, the online life is significantly facilitated by credit card payment or other third party payment methods. According to the statistical data in the blue book on the development of China's bank card industry (2019), the number of credit card issuers has increased from 186 million to 970 million, and the total amount of credit card transactions has increased from 3.5 trillion yuan to 38.2 trillion yuan,

L. Qi et al. (Eds.): CloudComp 2020, LNICST 363, pp. 74–82, 2021.
https://doi.org/10.1007/978-3-030-69992-5_6

nearly 10 times more. Credit card payment has become one of the most popular payment methods.

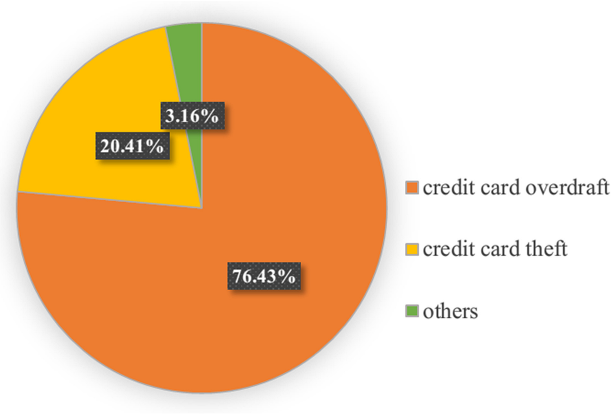

Fig. 1. The detailed proportion of credit card fraud cases

However, credit card fraud frequently happens and brings severe challenges to credit card management and seriously damages the interests of banks [1]. According to the Special report (2016 to 2018) on judicial big data of financial fraud issued by China judicial big data research institute, the number of credit card fraud is over 6 thousand. In these cases, credit card overdraft accounts for the largest proportion. Credit theft is also a major financial fraud type. More details are illustrated in Fig. 1. In comparison, according to the statistics from

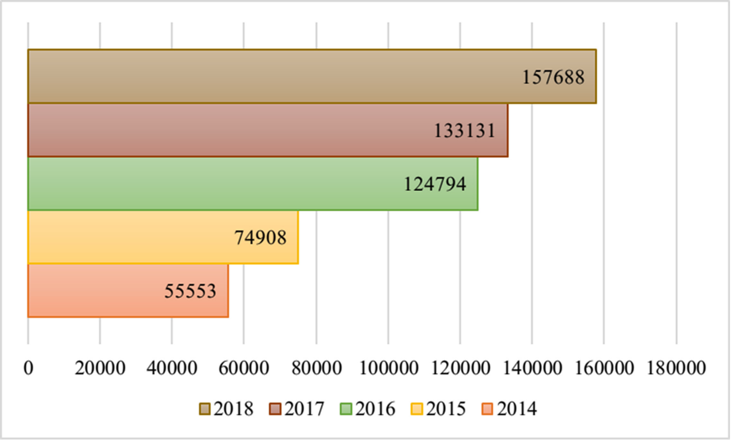

Fig. 2. The credit card fraud reports in the US from 2014 to 2018

the Shift Credit Card Processing company[1], the number of credit card fraud reported has increased from 55 thousand to more than 157 thousand as shown in Fig. 2. The US leads as the most credit fraud prone country with over 9.36 billion dollar losses in 2018. Most cases happen in the way of "card-not-present". Point-of-sale fraud and identity theft are another two main causes.

Clearly, there is a game between professional fraudsters and financial risk management party. The risk management department of Credit Card Center has summarized three main characteristics of current credit card frauds, i.e., concealment, professionalism and large-scale. Fraudsters often use professional Internet knowledge to steal card information of normal users and counterfeit individual identities. Besides, through packaging personal information, forging Internet behavior and other ways to improve personal qualifications, malicious users cheat to obtain credit cards and implement theft.

As demonstrated above, it is difficult but valuable to design accurate and efficient fraud detection methods, therefore to effectively protect the profits of card users and the banks. Intelligent credit card fraud detection is the joint area of financial risk management, information security and data mining etc., as illustrated in Fig. 3. Situations such as improper credit review and individual information breaches usually cause financial fraud crimes. To counter these conditions, data mining models are often applied to implement automatic fraud pattern discovery.

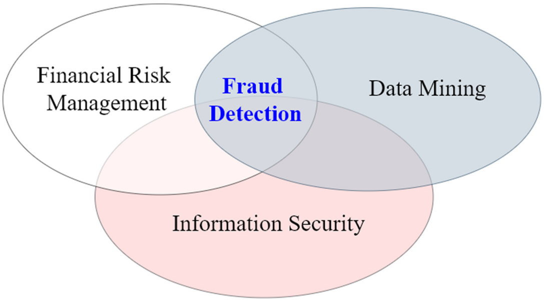

Fig. 3. The joint area of intelligent fraud detection

Fraud detection is mainly implemented based on the analysis of transaction time, amount, frequency, content and other information. Data mining models, such as decision tree, support vector machine and so on, provide automatic modeling measures to identify whether one instance should be labeled as fraud. In order to improve the capability of comprehensive fraud recognition, the algorithm should be carefully designed to handle the data preprocessing and imbalanced classification problems.

In this paper, we compare various classification models and propose a fraud detection approach through combining Fisher score [2] and feature re-weighting,

[1] https://shiftprocessing.com/credit-card-fraud-statistics/.

which improves the performance of the above mentioned classification models. Based on the experimental results, we demonstrate that the proposed feature handling approach provides satisfying accuracy and efficiency.

2 Fraud Detection Architecture and Implementing Approaches

Existing fraud discovery approaches take advantage of the advanced data mining models to solve imbalanced classification problems. In fraud detection, the target population is often very small. Misclassifying a target instance costs a lot. Therefore, the imbalanced data should be carefully preprocessed before being fed to the models. In Sect. 2.1 and Sect. 2.2, we discuss the architecture of the fraud detection models and compare the merits and demerits of each implementing approach.

2.1 Architecture

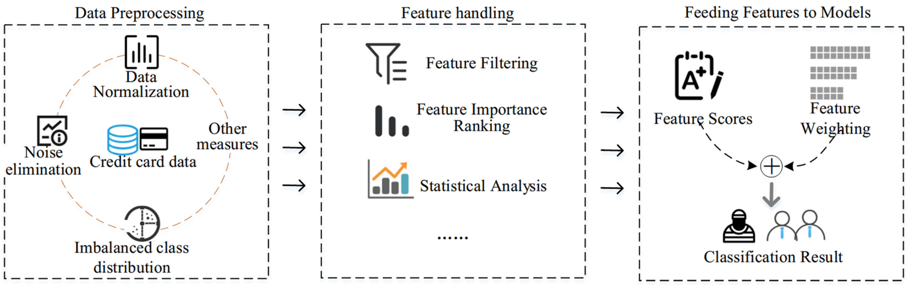

Fig. 4. The architecture of fraud detection models

Figure 4 illustrates an overview of the fraud detection architecture. In the following, we introduce each part of the architecture in detail.

Data Preprocessing. Data normalization and noise elimination are frequently applied for data preprocessing. While data imbalance is the most distinctive feature in financial fraud detection. It also widely exists in the fields of medical treatment, industry and advertising services [3]. In these areas, the true target label, such as true fraud record, severely underrepresents the other. Sampling methods, such as random oversampling and undersampling et al., attempt to balance the representative proportions of labels in the datasets [4]. In contrast, cost-sensitive learning methods consider the costs associated with misclassifying instances, therefore to improve the importance of the minority label [5,6].

Besides the sampling methods, Guo and Viktor propose the DataBoost-IM approach to adaptively generate synthetic instances to enrich the original dataset

[7]. Deep models are also talked about to learn more discriminative deep feature embeddings to maintain both inter-cluster and inter-class margins in imbalanced classification [8].

Feature Handling. In the classification task, high dimension usually infringes on both accuracy and efficiency. Dimension reduction methods are often applied to filter out the unimportant features and select the representative ones. Chandrashekar and Sahin give a comprehensive survey on feature selection focusing on Filter, Wrapper and Embedded methods [9]. Common dimension reduction methods include principal component analysis, multidimensional scaling, linear discriminant analysis, etc. [10]. In heuristic models such as decision tree, the importance of each feature is evaluated with a score, such as information gain, gini index etc. Jiang et al., propose a deep feature weighting (DFW) approach through deeply computing feature weighted frequencies from training data for the Naïve Bayes classifier [11]. In contrast, Zhang et al., propose two adaptive feature weighting approaches for Naïve Bayes text classifiers to improve model simplicity and reduce execution time [12].

In the feature handling step, crucial features are highly scored and selected. While under the premise of data privacy and security being paid more and more attention, researchers also have done a lot of work to preserve privacy in feature selection. To guarantee individual privacy, carefully generated randomness can be introduced to cover the true values without injuring classification performance [13,14].

Feeding Features to Models. Feature selection can effectively improve the training accuracy with a bit of efficiency loss. In our approach, we use a combined feature weighting strategy to prioritize features and improve the accuracy. Specifically, features are first ranked with Fisher score and then re-weighted with evaluation criteria such as information gain etc. More details about the process can be seen in Sect. 2.2.

2.2 Implementing Approaches

We compare the advantages and disadvantages of different classification models to solve the fraud detection problem in Table 1, where variables n, d, k denote the number of instances, the number of features and the number of single-trees, respectively. Both single and ensemble tree models are listed, including traditional decision tree, random forest, GBDT and XGBoost. Logistic regression and support vector machine are also compared for their simplicity and robustness, respectively.

The process of fraud detection is listed in four steps as below.

1. **Data imbalance handling**
   Perform data normalization and handle the imbalance problem through adaptively randomly sampling (cost-sensitive factor can also be introduced).
2. **Primary competitor training**
   Feed the processed data to each classification competitor derived in Step 1. and train. Calculate the AUC values and training time.

Table 1. The comparison of classification algorithms to implement fraud detection

Algorithms	Advantages	Disadvantages	Time complexity	Applicable scenarios
Decision Tree	Strong interpretability	Easy to over-fit and low accuracy	O(n*log(n)*d)	Large datasets
Random Forest	Balance errors with ensemble	Sensitive to noise	O(n*log(n)*d*k)	Large datasets
GBDT	Higher accuracy	Difficult to parallel for dependent learners	O(n*log(n)*d*k)	Large low-dimensional datasets
Logistic Regression	Small computation and low storage occupancy	Poor performance for non-linear problems	O(n*d)	Large low-dimensional datasets
Support Vector Machine	Better robustness	Difficult to handle multi-classification	O(n*n*d)	Small datasets
XGBoost	Simple model	Difficult to tune parameters	O(n*log(n)*d*k)	Large low-dimensional datasets

3. **Fisher competitor training**
Select features based on Fisher score or other criteria. Train each classification competitor with the features selected and calculate the AUC values and record training time.
4. **Re-weighted competitor training**
Weight features with a combined metric with both Fisher score and information gain etc., and train each classification competitor. Calculate the AUC values and record training time.

3 Combined Feature Weighting Approach and Evaluation Results

As mentioned above, features are evaluated with a combined metric with both Fisher score and information gain in the proposed strategy. Notice that the combined feature weighting step trades a little bit of efficiency for classification accuracy. In Sect. 3.1, we give the formal description of the combined feature weighting approach. We also give the accuracy and efficiency evaluation in Sect. 3.2.

3.1 A Feature Weighting Approach

Fisher score selects the optimal feature by calculating the inter class and intra class dispersion, which is simple and effective. The calculation of the Fisher score of feature j is shown in Eq. (1). Class labels are chosen from the set $\{0, 1, \ldots, c\}$. n_l denotes the number of instances taken label l. Specifically, let μ_l^j and σ_l^j be the mean and standard deviation of label l, corresponding to the j-th feature.

Let μ^j and σ^j denote the mean and standard deviation of the whole data set corresponding to the j-th feature [2].

$$\begin{aligned} F\left(\mathbf{x}^j\right) &= \frac{\sum_{l=0}^{c} n_l \left(\mu_l^j - \mu^j\right)^2}{\left(\sigma^j\right)^2} \quad where \\ \left(\sigma^j\right)^2 &= \sum_{l=1}^{c} n_l \left(\sigma_l^j\right)^2 \end{aligned} \tag{1}$$

Information gain quantifies the effectiveness of each feature for contributing the decrease of class distribution chaos [15]. The larger the information gain is, the more important the feature is. It is calculated with entropy and conditional entropy. Given the j-th feature with i possible values, the calculation of entropy, conditional entropy and information gain with the j-th feature are shown in Eq. (2), (3) and (4) respectively.

$$H_C(D) = -\sum_{l=0}^{c} \frac{n_l}{n} \log \frac{n_l}{n} \tag{2}$$

$$H_{C|j}(D) = \sum \frac{n_i}{n} H_C\left(D_i\right) \tag{3}$$

$$InfoGain(j, D) = H_C(D) - H_{C|j}(D) \tag{4}$$

Combining Fisher score feature selection and information gain, we have the combined ranking score of the j-th feature shown in Eq. (5).

$$score_j = F\left(\mathbf{x}^j\right) * InfoGain(j, D) \tag{5}$$

3.2 Accuracy and Efficiency Evaluation

In the imbalanced classification problem of fraud detection, AUC is more suitable than classification accuracy [4]. In this paper, we use the open credit card fraud detection dataset provided by the Kaggle platform[2]. There are 284807 instances with 29 features in the dataset and the percentage of fraudulent users and normal users was 0.17% and 99.83% respectively. Obviously, the dataset is highly imbalanced. The dataset has been collected and analyzed during a research collaboration of Worldline and the Machine Learning Group (http://mlg.ulb.ac.be) of ULB (Université Libre de Bruxelles) on big data mining and fraud detection.

In the experiments, we compare the impacts of the combined feature weighting strategy with the classification accuracy (AUC used) and training time. Each experiment has been repeated for 200 times to record the means. The experimental results with and without combined feature weighting are shown in Table 2 and Table 3. Notice that the competitor without combined feature weighting has

[2] http://www.kaggle.com/mlg-ulb/creditcardfraud.

just applied feature selection with Fisher score, as proposed by Dong et al., in [16]. In the experiments, information gain is used to further weight the features.

Table 2. AUC comparison with (without) feature weighting

Algorithms	Mean of AUC without feature weighting	Mean of AUC with feature weighting
Decision Tree	0.9038	0.9056
Random Forest	0.9717	0.9724
GBDT	0.9699	0.9701
Logistic Regression	0.9736	0.975
Support Vector Machine	0.9786	0.9796
XGBoost	0.968	0.9687

Table 3. Training time comparison with (without) feature weighting

Algorithms	Mean of training time without feature weighting(s)	Mean of training time with feature weighting(s)
Decision Tree	0.005	0.0059
Random Forest	0.1545	0.1564
GBDT	0.14	0.1474
Logistic Regression	0.0064	0.0078
Support Vector Machine	0.0038	0.004
XGBoost	0.0612	0.0663

Based on the experimental results shown in the above tables, combining Fisher score and other feature weighting metrics, such as information gain, has improved the classification performance of most compared models with a small efficiency cost.

4 Conclusions

Fraud detection is an important classification task. Fisher score can effectively shorten the training time of the classifier. To further improve the classification performance, we introduce the combined feature weighting strategy. The feature weighting approach performs especially well in logistic regression and support vector machine. In our future work, we will consider the privacy preservation of the feature selection process while balancing privacy and accuracy.

Acknowledgement. This work is supported in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant No. 19KJB510022, the Research Start-up Funds for the Introduction of High-level Talents at Jiangsu Police Institute under Grant No. JSPIGKZ, the National Natural Science Foundation of China under Grant No. 61802205, the Research Project of Higher Education Reform in Jiangsu Province under Grant No. 2019JSJG595, the 13th Five-Year Plan for Jiangsu Education Science under grant No. D/2020/01/22. We also express our thanks to the Jiangsu Electronic Data Forensics and Analysis Engineering Research Center and Jiangsu Provincial Public Security Department Key Laboratory of Digital Forensics.

References

1. Caneppele, S., Aebi, M.F.: Crime drop or police recording flop? On the relationship between the decrease of offline crime and the increase of online and hybrid crimes. Policing J. Policy Pract. **13**(1), 6–79 (2019)
2. Gu, Q., Li, Z., Han, J.: Generalized fisher score for feature selection. arXiv preprint arXiv:1202.3725 (2012)
3. Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G.: Learning from class-imbalanced data: review of methods and applications. Expert Syst. Appl. **73**, 220–239 (2017)
4. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**(9), 1263–1284 (2009)
5. Khan, S.H., Hayat, M., Bennamoun, M., Sohel, F.A., Togneri, R.: Cost-sensitive learning of deep feature representations from imbalanced data. IEEE Trans. Neural Netw. Learn. Syst. **29**(8), 3573–3587 (2017)
6. Sun, Y., Kamel, M.S., Wong, A.K., Wang, Y.: Cost-sensitive boosting for classification of imbalanced data. Pattern Recogn. **40**(12), 3358–3378 (2007)
7. Guo, H., Viktor, H.L.: Learning from imbalanced data sets with boosting and data generation: the DataBoost-IM approach. ACM SIGKDD Explor. Newslett. **6**(1), 30–39 (2004)
8. Huang, C., Li, Y., Loy, C.C., Tang, X.: Learning deep representation for imbalanced classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5375–5384 (2016)
9. Chandrashekar, G., Sahin, F.: A survey on feature selection methods. Comput. Electr. Eng. **40**(1), 16–28 (2014)
10. Zhang, T., Yang, B.: Big data dimension reduction using PCA. In: 2016 IEEE International Conference on Smart Cloud (SmartCloud), pp. 152–157. IEEE (2016)
11. Jiang, L., Li, C., Wang, S., Zhang, L.: Deep feature weighting for naive Bayes and its application to text classification. Eng. Appl. Artif. Intell. **52**, 26–39 (2016)
12. Zhang, L., Jiang, L., Li, C., Kong, G.: Two feature weighting approaches for naive Bayes text classifiers. Knowl.-Based Syst. **100**, 137–144 (2016)
13. Anandan, B., Clifton, C.: Differentially private feature selection for data mining. In: Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics, pp. 43–53 (2018)
14. Yang, J., Li, Y.: Differentially private feature selection. In: 2014 International Joint Conference on Neural Networks (IJCNN), pp. 4182–4189. IEEE (2014)
15. Quinlan, J.R.: Induction of decision trees. Mach. Learn. **1**(1), 81–106 (1986)
16. Dong, Y., Liu, X., Li, B.: Click fraud detection method based on user behavior feature selection. Comput. Sci. (10), 27 (2016)