Journal of Risk Analysis and Crisis Response, 2022, 12(3), 110-123
https://jracr.com/
ISSN Print: 2210-8491
ISSN Online: 2210-8505

Article

CUS-RF-Based Credit Card Fraud Detection with Imbalanced Data

Wei Li [1], Cheng-shu Wu [1] and Su-mei Ruan [1,*]

[1] School of Finance, Anhui University of Finance and Economics, Bengbu (233030), Anhui, China
* Correspondence: ruansumei0116@163.com

Received: July 2, 2022; Accepted: September 4, 2022; Published: September 30, 2022

Abstract: With the continuous expansion of the banks' credit card businesses, credit card fraud has become a serious threat to banking financial institutions. So, the automatic and real-time credit card fraud detection is the meaningful research work. Because machine learning has the characteristics of non-linearity, automation, and intelligence, so that credit card fraud detection can improve the detection efficiency and accuracy. In view of this, this paper proposes a credit card fraud detection model based on heterogeneous ensemble, namely CUS-RF (cluster-based under-sampling boosting and random forest), based on clustering under-sampling and random forest algorithm. CUS-RF-based credit card fraud detection model has the following advantages. Firstly, the CUS-RF model can better overcome the issue of data imbalance. Secondly, based on the idea of heterogeneous ensemble learning, the clustering under-sampling method and random forest model are fused to achieve a better performance for credit card fraud detection. Finally, through the verification of real credit card fraud dataset, the CUS-RF model proposed in this paper has achieved better performance in credit card fraud detection compared with the benchmark model.

Keywords: Credit Card Fraud Detection; Random Forest; Imbalanced Data; Heterogeneous Ensemble; Fintech

1. Introduction

Credit card fraud has caused immense financial loss to both card-issuing banks and financial institutions. According to the statistical data from China Banking Association, by the end of 2018, credit card-based transaction in China has attained 38,200 billion Yuan RMB at a growth rate of 24.9%; 73.2% of credit cards remain activated, and unpaid balance is 6,850 billion Yuan RMB (on a year-on-year growth of 23.2%). It is especially noteworthy that credit card loss rate is 1.27%, slightly higher than 1.17% in previous year. The global credit card fraud-related loss climbs from 7.6 billion dollars in 2010 to 21.81 billion dollars in 2015, with a growth of 300% within 5 years. It is expected to reach 31.67 billion dollars by 2020.

A conventional credit card fraud detection model is usually constructed using rules provided by experts. Nevertheless, the aforesaid fraud detection models often demand manual parameter tuning and supervision from experts, which makes it impossible for financial institutions to discover fraudulent behaviors in time. Moreover, it is a heavy task to check all the transactions one by one. To overcome the shortcomings in such work, financial institutions have employed machine learning algorithm and data mining methods in setting up an artificial intelligence (AI)-based credit card fraud detection model that is different from the traditional detection models. The machine learning algorithm can help financial institutions in constructing automated detection models to significantly

improve the fraud detection efficiency and speed. Driven by machine learning, the novel credit card fraud detection model is trained and parameter-tuned in order to gain the expected effect.

However, the ability of fraud detection will be greatly impaired should severe data imbalance exist amid credit card fraud data, so that the model may fail to exert its due performance. Data and their features are the most critical factors affecting the operation of a fraud detection model. That said the category of imbalance perplexing data centralization is associated and demands analysis. Furthermore, in the real world where financial institutions apply strict censorship, fraudulent credit card swiping is far less frequent than the normal operation. Despite this, once such fraudulent transaction occurs, it will be hard for the financial institution involved to get back the money lost. Therefore, when studying credit card fraud, it is urgent to considered the data imbalance issue. Since automated credit card fraud detection system is investigated on the basis of balanced datasets when being established with machine learning, the model is hard to give a full play to its own strengths.

The problem of credit card fraud detection for imbalanced data has been studied by scholars in different countries with different ideas. After a systematic study of category imbalance processing strategies, Singh et al. (2021) compared the effectiveness and efficiency of different category imbalance processing methods and state-of-the-art classification methods, evaluating metrics in terms of Precision, Recall, K-fold Cross-validation, AUC-ROC curve and execution time, and found that oversampling and under-sampling methods performed better for integrated classification models such as AdaBoost, XGBoost and Random Forest performed better [1]. El-Naby et al. (2022) addressed the fraud data imbalanced problem by using mixed sampling and oversampling preprocessing techniques, specifically, in oversampling, SMOTE, bounded SMOTE and ADASYN were selected. In mixed sampling, SMOTEEN and SMOTETomek to eliminate the data set imbalance problem, thereby improving credit card fraud detection accuracy [2]. As for the imbalanced classification problem for credit card fraud detection, Makki et al. (2019) found LR, C5.0 decision tree algorithm, SVM and ANN to be the best methods after comparing imbalanced classification methods and based on three performance metrics: accuracy, sensitivity and average precision (AUPRC). Although the above methods improve the performance of the classifier, when the data is extremely unbalanced, these methods may create a problem of false positives and major credit fraudulent cases may remain undetected [3]. As can be seen, the data imbalance problem interferes with credit card fraud detection and even false positives, and scholars in various countries are working to improve credit card fraud prediction performance and accuracy with imbalanced data.

Therefore, in creating automated credit card fraud detection model, attention should be paid to the data imbalance issue to evade possible impaired predicting accuracy of the machine learning-based prediction model. In addition, most of sorting algorithms would be undermined in performance in such case. In view of the data imbalance issue, SMOTE (synthetic minority over-sampling technique) technology has been extensively applied to financial distress prediction, bankruptcy forecasting, and credit card fraud fields in recent years [4–8]. It exhibits better sorting performance when compared with traditional RUS (random under-sampling) and ROS (random over-sampling) technologies. In this paper, a brand-new sorting technique based on clustering under-sampling that targets imbalanced data is introduced into credit card fraud field. Related references indicate this sorting technique has demonstrated outstanding performance in dealing with financial distress prediction and so on [9, 10]. As revealed by the findings of our experiment here, the technique excels SMOTE in sorting imbalanced credit card data containing thousands of samples. It has proven application prospect in the financial field.

This paper establishes a heterogeneous ensemble model by innovatively ensemble CUS (cluster-based under-sampling) with RF (random forest) [11]. To further verify the excellency of the proposed model, CUS is blended with five classifiers respectively to form five heterogeneous ensemble models. Since each classifier has distinct theoretical backgrounds, the paper combines same imbalanced data processing technique with different base learner methods in the hope to find out which theoretical design stands out in sorting the imbalanced data concerning credit card fraud. Besides, in order to explore the performances of CUS and SMTOTE, the paper also combines SMOTE with the foresaid base learner and adopt the widely applied evaluation index system in financial field to find out which imbalanced data sorting technique is more suitable for automatic classification of credit card data. The goal of doing so is to propose the best solution for imbalanced data processing and base learner choice. Apart from that, another innovative aspect of this paper lies in its embedding of imbalance processing into base learner to form automatic sorting system. The resulting system could process imbalanced data and automatically sort base learners, which greatly cut down the time cost of learning task when compared with existing research achievements.

This paper is arranged as follows. Section 2 offers a literature review on data imbalance issue and credit card fraud detection model. Section 3 describes the CUS background model and then proposes an improved model that could deal with imbalanced data. Section 4, the proposed model is experimentally testing, and the experimental findings are analyzed. And conclusions are drawn in Section 5.

2. Literature Review

As fintech develops by leaps and bounds in recent years, a great number of fintech banks emerges [12, 13]. Consumer finance is influencing people's life in a novel way. This also gives rise to fraud issue. Detecting credit card fraud has become one of the much-concerned topics in financial industry. However, public models available for use remain rather limited. One of the major underlying reasons is that credit card transaction data are exclusively kept by the card-issuing agencies. As data owner, card-issue agencies must protect the data security and avoid leaking users' privacy, so they are not going to disclose the datasets and related models they are using [14].

There are two common types of credit card fraud: application fraud and behavioral fraud. The latter type consists of card stealing, card forgery and non-existing card [15, 16]. Fraudulent swiping is the most common type of credit card fraud [17]. In usual cases, after stealing the credit card or obtaining a temporary card, fraudster would use the card for consumption as much as possible. When committing fraud with a stolen card, the criminal usually transacts with the card at high frequency. Should the fraudster forge a fake card with the information he has collected, forgery fraud will occur. While the victim still holds his own card for legal transaction, the fraudster transacts also with the fake card. The fake card will be used only for a few times before being abandoned and realized by the victim. The third type of behavioral fraud is card not present fraud which occurs in case of remote transaction. In such case, the transaction is made on the basis of card information only such as cad number, holder name, and valid term [18, 19]. The distinction between fake card fraud and card not present fraud lies in use of solid card in the former case and use of card information only in the latter [20].

There are two primary means of data mining used in credit card fraud detection system creation, namely supervised learning and unsupervised learning [21].

Supervised learning aims to training dichotomic model, depending on the detection model trained with datasets marked as "normal" and "fraudulent" to tell fraudulent samples from normal

ones [22]. This is a most common way of fraud detection. Recently, supervise learning algorithm has been applied to the establishment of some fraud detection systems. For example, Soemers et al. proposed a dynamic model combining decision-making tree and context-based multi-arm gambler which demonstrated proven effect on identifying credit card fraud [23]. In the work by Zliobaite, an adaptive algorithm was put forward, which was able to update fraud detection model with time-dependent data flow in order to better adapt to the shifts in fraudulent transaction patterns [24]. Blending recursive feature elimination, hyper-parameter optimization and SMOTE technology, Naoufal Rtayli developed a mixed credit card fraud detection model [5]. Other supervised learning methods, including Bayes, artificial neural network [25] and support vector machine (SVM), are also frequently used in fraud detection [16, 26, 27]. Compared with semi-supervised and unsupervised fraud detection systems, supervised systems stand out with sufficient data training time that supports establishment of well-performing models [28]. The output from detection system trained with supervised learning technique has explicit meaning and can be directly applied to mode distinguishment.

In unsupervised learning, the dataset samples for constructing fraud detection model carry no tag. Instead, the unsupervised machine learning sets out to analyze data from different dimensions and resist fraud by finding out the association or difference between data [21]. For example, the GAN model could learn normal data distribution and determine whether unknown test data are normal or fraudulent samples with the proposed abnormality scoring plan. In case of insufficient label data and severely imbalanced data, unsupervised learning model will be a better choice. In addition, unsupervised learning could update model with online unlabeled data from banks or financial institutions, thus rendering it possible to detect use of fraudulent credit card. For instance, an unsupervised learning model, called Self Organizing Map (SMO), is came up with for forming an unsupervised credit card fraud detection model [29, 30]. As SMO model requires no priori information, the automated system proposed may use newly added transaction data to keep updating the model. Besides, K-means clustering algorithm sorts of transaction data according to the similarity concerning credit card fraud features and thus gets used to fraud detection model creation [31].

We have analyzed the application of ensemble to credit card fraud and related fields in recent years. A few comprehensive algorithms have become popular base learners in ensemble algorithms, among which the most noteworthy ones are Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (K-NN) and Gradient Boosting Decision Tree (GBDT). Therefore, we have built some models based on those base learners for comparison purpose. Besides, the ensemble models proposed in previous studies are limited to SMOTE and Random Oversampling when imbalanced data processing is involved. They rarely employ new techniques. This paper has applied CUS to the processing of credit card-related imbalanced data for the first time.

3. Research Design

3.1. Data

The dataset used in credit card fraud detection is provided by machine learning group (http://mlg.ulb.ac.be), which could be downloaded from Kaggle (https://www.kaggle.com/mlg-ulb/creditcardfraud). In the dataset there are data concerning credit card transactions completed by Europeans in September 2013. According to it, among 284,807 transactions within two days, 492 are fraudulent. The dataset appears quite imbalanced, as positive examples (fraudulent swiping) accounts for 0.172% of total transactions. Due to privacy protection, we cannot acquire the original

functions and more background information about related data. Features V1, V2, ..., V28 are the major constituents acquired by PCA (principal component analysis), while the features undergoing no PCA conversion are "Class" and "Amount". The feature "Class" is a response variable, which is "1" when card fraud occurs and "0" otherwise. The goal of the task is to sort out normal transaction data from abnormal ones in the dataset and predict about the test data.

Table 1. Credit Card Fraud Detection Dataset.

Instance number	Fraud samples	Normal samples	Feature number
284807	492	284315	30

As shown in Table 1, there are altogether 284,807 transaction samples in the dataset. Among them, only 492 are fraud samples, accounting for 0.17% of total dataset. the proportion of normal samples to fraud ones is as high as 578:1. In other words, this dataset features extremely imbalanced positive and negative samples. If no pre-treatment is made to improve the data imbalance here and data at primitive proportion are directly put into the classifier for training, the classifier is more like to view normal samples as white noise. This would impair the performance of whole combined fraud detection system.

Our data reveals data imbalance stays as a primary issue in fraud detection process. As a matter of fact, in an imbalanced dataset, we could find the training examples for one class variable are far less than that for the other one. Accordingly, the first one is called minority set while the second as majority set. When sorting imbalanced fraud detection transaction dataset, most models perform well in identifying the majority set but much less accurate in the minority one, suggesting they are not good at detecting the minority samples.

In order to effectively cope with the class imbalance issue in credit card fraud data, this paper introduces CUS and combines it with RF to form machine learning-based heterogeneous ensemble. It succeeds in effectively sorting the credit card fraud data. In addition to that, we use also SMOTE technology for comparison and compare data sorting performance by controlling the base learner. SMOTE is a way of oversampling that generates random examples instead of achieving oversampling by repetition or replacement alone. Furthermore, the technology can also progressively increase the learning process of fraud detection algorithm [32].

3.2. CUSBoost

CUSBoost is a combination of CUS and AdaBoost algorithm. Like RUSBoost and SMOTE-Boost, it contains key difference in sampling technique. SMOTE-Boost employs SMOTE to sample minority examples, whereas RUSBoost chooses random under-sampling over the majority ones. Based on comparison, the CUSBoost proposed by us selects the sampling from majority class based on clustering. CUSBoost separates first the majority and minority examples from the dataset first, and then applies k-means clustering algorithm to cluster majority examples to k clusters. Here, parameter k is determined through hyper-parameter optimization. Then, 50% of examples are randomized (or tuned as per field issue or dataset) with the rest being eliminated. Random under-sampling is executed to each cluster. Since clustering is applied in prior to sampling, theoretically speaking, the algorithm is expected to perform best when dataset is highly clustered. Next, those representative samples are combined with the minority ones to form a well-balanced dataset. The strength of our algorithm is displayed on the inclusion of all the subspace examples in considering the majority class, as k-means clustering contains each example in certain cluster. Other similar methods usually fail to proper represent the majority class. In Fig. 1, the CUS proposed is used to choose the majority

examples, in which red spots indicate the examples sorted out from the majority class while the black and red spots represent all the majority examples.

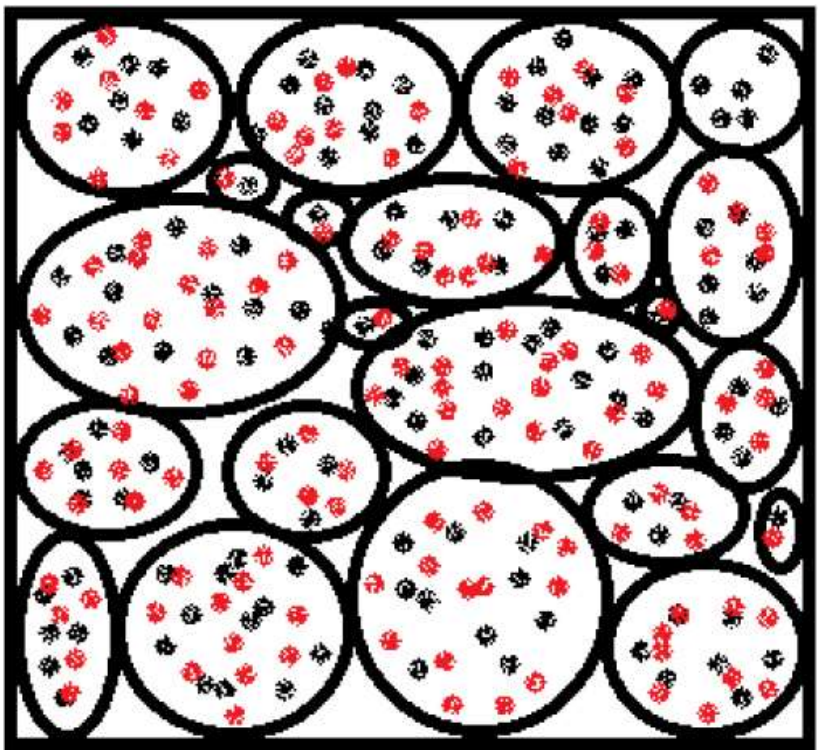

Figure 1. Cluster-based under-sampling (CUS) approach.

3.3. Random Forest Model

Random forest (RF) is an expansion variant of Bagging, which further introduces random attribution selection into decision-making tree training on the basis of base learner [33]. More specifically, traditional decision-making tree is to choose an optimal attribute from the attribute set (suppose there are d attributes) of current node when deciding to divide the attributes. In contrast, RF randomly selects one subset containing k attributes from the attribute set of each node on the base decision-making tree, and then selects one optimal attribute from the resulting subset for division purpose. Here the parameter k decides how much randomness should be introduced. If $k = d$, the base decision-making tree constructed will be nothing different from the traditional one; and if $k = 1$, one attribute will be randomly selected for division. In general cases, $k = \log_2 d$ is recommended.

Being easily and readily realizable at low computation cost, RF has exhibited strong performance in many practical tasks. Though it is adapted from Bagging by executing minor changes, its base learner is different from that in Bagging, as the diversity of its base learner comes from both sample perturbation and attribute perturbation while that in Bagging comes from sample perturbation (of initial training set sampling) alone. Therefore, it is made possible to improve the post-ensemble generalization performance by further differentiating individual learners.

RF has similar convergence as Bagging. Its initial performance could be far from being satisfactory, especially when there is only one base learner in the ensemble. However, as the number of individual learners keeps growing, RF usually will converge to lower generalization error. Notably, the training efficiency of RF is more than usual superior to that of Bagging because Bagging employs "certain" tree when constructing individual decision-making tree and has to examine all the attributes of the node in attribute classification, whereas RF adopts "random" decision-making tree

and need examine one attribute subset only [31]. In view of the strengths displayed by RF model, it is widely applied to a variety of fields including credit card fraud prediction [35–38].

3.4. Model Ensemble

In consideration of the high imbalance perplexing the dataset used here, this paper attempts to construct a model with good predicting performance on a highly imbalanced dataset. Based on CUSBoost, we replace the target of improvement for CUS-AdaBoost with RF, and thus propose a novel CUS employing Boosting (RF) and name it as CUS-RF. Different from AdaBoost, RF keeps fitting new models during the learning to generate more accurate estimation about the response variable. When constructing the decision-making tree, RF algorithm trains the tree as per the residual error of previous tree in each iteration. Finally, the output is the accumulation of all the tree classifications.

Based on Boosting thought, RF serially builds several decision-making trees to predict the data. In other words, it performs gradient boosting in the space where loss function is. In details, it views to-be-resolved decision-making tree model as parameter and fits the negative gradient of loss function in current model upon each iteration to renew the parameter to minimize the loss function.

RF could be considered as an extension of AdaBoost. The latter identifies problem on the basis of mis-classified data point and improves the model by adjusting the weight of such mis-classified data point, whereas RF finds out problem by negative gradient and improves the model through negative gradient computation. In fact, examples containing higher absolute value of negative gradient is going to gain high attention in subsequent training, because the resulting loss is likely to account for a large portion in the final loss function. Therefore, it is more depended on to diminish the loss. This is something shred by RF and AdaBoost. Compared with AdaBoost, RF could invoke more types of loss function and render more problems resolved [39–41].

Serving as the basis of CUS-RF, CUSBoost algorithm is based on AdaBoost algorithm itself. By introducing CUS, CUSBoost does some improvement towards the AdaBoost to better balance the class distribution, and AdaBoost improves the classifier performance in virtue of those balanced data. CUS-RF does achieve the same goal, but what is improved here is RF instead of AdaBoost. Results indicate this algorithm features quick model training and satisfactory performance.

Embedding CUS algorithm into RF algorithm is both implementable and effective. By following the thought of CUSBoost in algorithm improvement, we bring CUS process into RF.

3.5. Evaluation Metrics

Unfortunately, some common classifier evaluation indexes turn out to be inapplicable to imbalanced dataset regardless of their performance in dealing with balanced dataset. This happens with accuracy rate also in credit card fraud detection, a most frequently used index. It does not consider sample distribution which stays as the key issue of imbalanced dataset. Besides, accuracy rate may lead to misleading conclusion. For instance, in an imbalanced dataset, 99% of observed values are normal (meaning there is no fraud) while 1% are negative (meaning there is fraud). If taking the positive class (or majority class) prediction as the standard, this model will be deemed as having 99% of accuracy rate, because this method opts to choose the majority class and thus yield the better outcome. The ratio indicates the classifier is accurate, however it ignores the prediction of minority class (or negative class) which should be attached with highest importance in case of dataset imbalance. Thus, we need to adjust the model performance evaluation and rely on the evaluation indexes insensitive to sample distribution. We choose four evaluation indexes that are widely used

in imbalanced datasets, namely Area Under the Receiver Operating Characteristic Curve (AUC), TPR, FPR, Specificity and maximum KS value. Here we are going to sequence the samples by classifier-derived prediction outcome and predict all the samples one by one as positive examples. Each computation brings in values of two important measures which are used as horizontal and vertical coordinates for mapping. Finally, a Receiver Operating Characteristic Curve (ROC Curve) is generated, which has its vertical shaft as True Positive Rate (TPR) and horizontal one as "False Positive Rate (FPR)". Definitions of TPR and FPR are offered below:

$$TPR = \frac{TP}{TP+FN} \tag{1}$$

$$FPR = \frac{FP}{TN+FP} \tag{2}$$

When classifiers are compared in term of performance, if one classifier's ROC curve gets completely covered by that of another, it can be assumed that the latter outperforms the former. If two classifiers have their ROC curves overlap each other, it is hard to tell which one is better. In such case, we employ AUC (Area Under ROC Curve) to determine the inter-model difference in performance. AUC is the sum of all the areas under ROC. Suppose ROC is formed with points with coordinates $\{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$ that are connected in order, then

$$AUC = \frac{1}{2}\sum_{i=1}^{m-1}(x_{i+1} - x_i) \cdot (y_i + y_{i+1}) \tag{3}$$

Specificity is the ratio by which no fraud is predicted as non-fraud. It can be figured out according to following formula:

$$Specificity = TN/(TN + FP) \tag{4}$$

At last, after all the models gain the optimal hyper-parameters after parameter tuning, we re-train the models with such hyper-parameters on training and development sets before testing again the retrained models on test set. Afterwards, five indexes are invoked to evaluate the model performance respectively, as this is an effective index for imbalanced dataset.

4. Experimental Results and Discussion

4.1. Feature Selection

After finishing data preprocessing, we need to input meaningful features into the machine learning algorithm and model for training purpose. In general sense, features should be weighed from two aspects, whether the feature diverges and whether feature remains related to the goal. If a feature does not diverge, such as having a variance nearing 0, it means the samples don't differ from each other in this feature, and the feature is of little use to sample distinguishment. As for the second aspect which is more apparent, the feature highly concerned with the goal should be prioritized. Except for variance method, all other methods introduced in this paper set out from relevance.

The feature selection methods could be divided into three types by feature selection form, namely Filter, Wrapper and Embedded. Filter method scores every feature as per its divergence or relevance and sets threshold or number of threshold candidates to screen the features. Wrapper method selects a few features or eliminate some upon each time by referring to the objective function (or prediction effect score in usual cases). For Embedded method, features are trained first with certain machine learning algorithms and models to figure out their weight coefficients, and then features are selected based on the resulting coefficients in a descending order. Though having something common with Filter, the Embedded determines whether a feature is good or not through training. We select features based on XGBoost-based approach in the Feature Selection library of

Sklearn and then apply Filter method to generate a database that contains less samples but is more related to the sample type.

4.2. Parameter Tuning

To train models with better performance, we perform further optimization with parameters from such models as CUS-GBDT, SMOTE-GBDT, CUS-RF and SMOTE-RF in order to exert their best predicting performance. The post-optimization parameters are listed in Table 2.

Table 2. Optimized parameters of CUS-GBDT and XGBoost.

Classifier	Parameter	Description
CUS-GBDT SMOTE-GBDT	max_depth	Maximum depth of the individual regression estimators.
	n_estimators	The number of boosting stages to perform.
	subsample	The fraction of samples to be used for fitting the individual base learners.
	loss	Loss function to be optimized.
CUS-RF SMOTE-RF	n_estimators	The number of boosting stages to perform.
	max_depth	Maximum depth of the individual regression estimators.
	scale_pos_weight	The weight of positive samples.

4.3. Benchmark Models

Table 3. A list of the proposed method and benchmark methods.

No.	Method	Description
1	CUS-GBDT	Based on CUS-GBDT, with CUS for preprocessing class imbalance data and GBDT for sample classification.
2	CUS-KNN	Based on CUS-KNN, with CUS for preprocessing class imbalance data and KNN for sample classification.
3	CUS-LR	Based on CUS-LR, with CUS for preprocessing class imbalance data and LR for sample classification.
4	CUS-RF	Based on CUS-RF, with CUS for preprocessing class imbalance data and RF for sample classification.
5	CUS-SVM	Based on CUS-GBDT, with CUS for preprocessing class imbalance data and GBDT for sample classification.
6	SMOTE-GBDT	Based on SMOTE-GBDT, with SMOTE for preprocessing class imbalance data and GBDT for sample classification.
7	SMOTE-KNN	Based on SMOTE-KNN, with SMOTE for preprocessing class imbalance data and KNN for sample classification.
8	SMOTE-LR	Based on SMOTE-LR, with SMOTE for preprocessing class imbalance data and LR for sample classification.
9	SMOTE-RF	Based on SMOTE-RF, with SMOTE for preprocessing class imbalance data and RF for sample classification.
10	SMOTE-SVM	Based on SMOTE-SVM, with SMOTE for preprocessing class imbalance data and SVM for sample classification.

Table 3 briefly lists the proposed method and benchmark methods. In the experiment, we adopt GBDT, KNN, LR, RF and SVM as base learners to ensemble with CUS and SMOTE respectively to build heterogeneous ensemble models to determine which imbalanced data processing method or classifier performs well in dealing with imbalanced dataset containing credit card fraud data.

4.4. Results and Discussion

The data screened with feature selection method are brought into the heterogeneous ensemble models composed of imbalanced data processing and classifier. The models are appraised in terms of five indexes, namely AUC, TPR, FPR, specificity and precision. The experimental findings are illustrated in Fig. 2 and 3 and Table 4. In this paper, some common classifiers in credit card fraud field are selected as component classifiers for the heterogeneous ensemble models and then combined with two ways of imbalanced data processing CUS and SMOTE in the hope to determine which processing method outperforms and which classifier is better for classifying the credit card fraud.

Table 4. Performance between fraud detection models and benchmark models.

Models	AUC	TPR	FPR	Specificity	Precision
CUS-GBDT	0.964477	0.938776	0.107574	0.892426	0.892426
CUS-KNN	0.9999	1	0.001565	0.998435	0.998435
CUS-LR	0.860443	1	1	0	0
CUS-RF	0.999912	1	0.001565	0.998435	0.998435
CUS-SVM	0.943549	0.785714	0.001811	0.998189	0.998189
SMOTE-GBDT	0.960775	0.816327	0.030758	0.969242	0.969242
SMOTE-KNN	0.999804	1	0.002	0.998048	0.998048
SMOTE-LR	0.85719	1	1	0	0
SMOTE-RF	0.999881	1	0.001864	0.998136	0.998136
SMOTE-SVM	0.580702	0.555556	0.389474	0.610526	0.610526

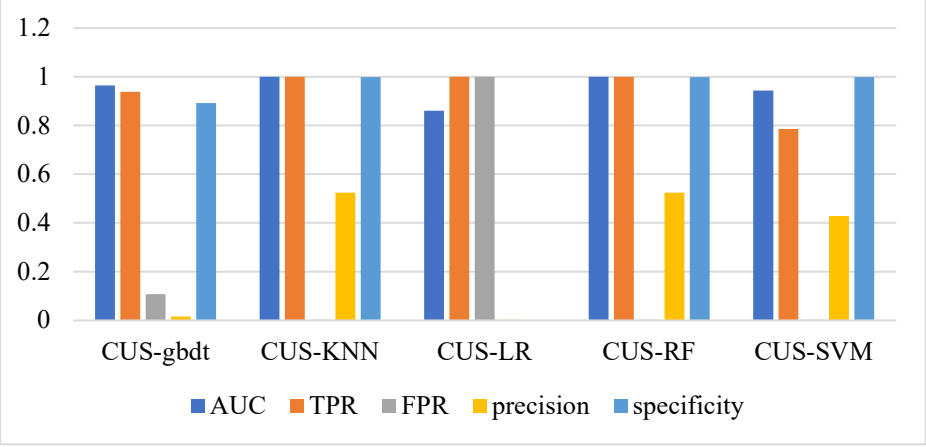

Figure 2. The column graph of the values of performance measures (CUS).

In the evaluation index system established by this paper, AUC is used to measure model's comprehensive performance in identifying credit card fraud samples, TPR could figure out the accuracy rate of model in determining sample company's financial risk, FPR acts to assess the probability for a listed company sample free from financial risk to be wrongfully identified by the model, specificity measures the accuracy rate of a model in determining whether the sample company

contains financial risk, and precision indicates the ratio of correctly classified positive samples in the positive samples identified by classifier. Among those five indexes, only FPR score is negatively related to the model's performance in identifying list company samples free from financial risk, while the scores of rests six methods are all positively related to the model's performance in the aspect involved.

Judged from the experimental findings, among five CUS-based heterogeneous ensemble models, four models (incl. CUS-KNN, CUS-LR, CUS-RF and CUS-SVM) are either better than or equal to the SMOTE-based heterogeneous ensemble models in terms of five evaluation indexes, suggesting CUS enjoys great strengths in dealing with imbalanced credit card fraud data when compared with SMOTE. In the meanwhile, the findings reveal the CUS-RF model proposed in this paper harvests the best scores in five indexes, indicating the improved model has great potential for application to credit card fraud data classification.

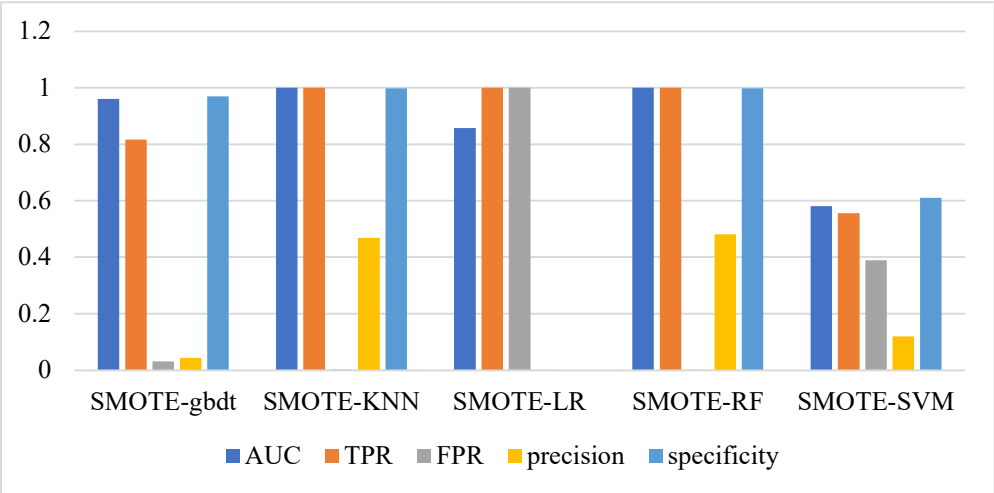

Figure 3. The column graph of the values of performance measures (SMOTE).

5. Conclusions

It is still quite tough and challenging to identify the credit card fraud samples. Through the work here, we aim to introduce novel imbalanced data processing techniques to improving the machine learning model's predicting performance during credit card fraud detection. Based on the heterogeneous ensemble principle, we have introduced CUS and RF to prediction model creation. To verify the experimental findings, we have also embedded the frequently used classifiers in this field into CUS to generate comparative models. In addition, by keeping classifier unchanged, we also create comparative models with SMOTE to prove the superiority of CUS over SMOTE in classifying imbalanced credit card fraud data. In the future, it is planned to further testify the reliability of the model with more complicated data from the real world and develop a self-adaptive credit card fraud detection system.

Acknowledgments: We would like to thank Xu-dong Du, a doctoral candidate at the School of Management, Hefei University of Technology, for providing experimental support for this paper.

Funding: This research was funded by the Philosophy and Social Science Planning Project of Anhui Province (National Social Science Fund Incubation Project), grant number AHSKF2021D07.

Conflicts of Interest: The authors declare that they have no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

[1] Singh A, Ranjan RK, Tiwari A. Credit Card Fraud Detection under Extreme Imbalanced Data: A Comparative Study of Data-level Algorithms[J]. *Journal of Experimental and Theoretical Artificial Intelligence*, 2021,34:571-598. DOI: https://doi.org/10.1080/0952813X.2021.1907795.

[2] El-Naby AA, Hemdan EED, El-Sayed A. An efficient fraud detection framework with credit card imbalanced data in financial services[J]. *Multimedia Tools and Applications*, 2022. DOI: https://doi.org/10.1007/s11042-022-13434-6.

[3] Makki S, Assaghir Z, Taher Y, Haque R, Hacid M, Zeineddine H. An Experimental Study with Imbalanced Classification Approaches for Credit Card Fraud Detection[J]. *IEEE Access*, 2019,7:93010-93022. DOI: https://doi.org/10.1109/ACCESS.2019.2927266.

[4] Sun J, Lang J, Fujita H, Li H. Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates[J]. *Information Sciences*, 2018,425:76–91. DOI: https://doi.org/10.1016/j.ins.2017.10.017.

[5] Rtayli N, Enneya N. Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization[J]. *Journal of Information Security and Applications*, 2020,55:102596. DOI: https://doi.org/10.1016/j.jisa.2020.102596.

[6] Sun J, Fujita H, Zheng Y, Ai W. Multi-class financial distress prediction based on support vector machines integrated with the decomposition and fusion methods[J]. *Information Sciences*, 2021,559:153–170. DOI: https://doi.org/10.1016/j.ins.2021.01.059.

[7] Shen F, Liu Y, Wang R, Zhou W. A dynamic financial distress forecast model with multiple forecast results under unbalanced data environment[J]. *Knowledge-Based Systems*, 2020,192:105365. DOI: https://doi.org/10.1016/j.knosys.2019.105365.

[8] Sun J, Li H, Fujita H, et al. Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting[J]. *Information Fusion*, 2020,54:128–144. DOI: https://doi.org/10.1016/j.inffus.2019.07.006.

[9] Du X, Li W, Ruan S, Li L. CUS-heterogeneous ensemble-based financial distress prediction for imbalanced dataset with ensemble feature selection[J]. *Applied Soft Computing Journal*, 2020,97. DOI: https://doi.org/10.1016/j.asoc.2020.106758.

[10] Li W, Ding S, Chen Y, et al. Transfer learning-based default prediction model for consumer credit in China[J]. *Journal of Supercomputing*, 2019,75:862–884. DOI: https://doi.org/10.1007/s11227-018-2619-8.

[11] Khan A, Rehman HU, Habib U, Ijaz U. Detecting N6-methyladenosine sites from RNA transcriptomes using random forest[J]. *Journal of Computational Science*, 2020,47:101238. DOI: https://doi.org/10.1016/j.jocs.2020.101238.

[12] Laidroo L, Koroleva E, Kliber A, et al. Business models of FinTechs – Difference in similarity[J]. *Electronic Commerce Research and Applications*, 2021,46:101034. DOI: https://doi.org/10.1016/j.elerap.2021.101034.

[13] Bollaert H, Lopez-de-Silanes F, Schwienbacher A. Fintech and access to finance[J]. *Journal of Corporate Finance*, 2021,68:101941. DOI: https://doi.org/10.1016/j.jcorpfin.2021.101941.

[14] West J, Bhattacharya M. Intelligent financial fraud detection: A comprehensive review[J]. *Computers and Security*, 2016,57:47–66. DOI: https://doi.org/10.1016/j.cose.2015.09.005.

[15] Azevedo C da S, Gonçalves RF, Gava VL, Spinola M de M. A Benford's Law based methodology for fraud detection in social welfare programs: Bolsa Familia analysis[J]. *Physica A: Statistical Mechanics and its Applications*, 2021,567:125626. DOI: https://doi.org/10.1016/j.physa.2020.125626.

[16] Forough J, Momtazi S. Ensemble of deep sequential models for credit card fraud detection[J]. *Applied Soft Computing*, 2021,99:106883. DOI: https://doi.org/10.1016/j.asoc.2020.106883.

[17] Wang D, Chen B, Chen J. Credit card fraud detection strategies with consumer incentives[J]. *Omega*, 2019,88:179–195. DOI: https://doi.org/10.1016/j.omega.2018.07.001.

[18] Soltani Halvaiee N, Akbari MK. A novel model for credit card fraud detection using Artificial Immune Systems[J]. *Applied Soft Computing Journal*, 2014,24:40–49. DOI: https://doi.org/10.1016/j.asoc.2014.06.042.

[19] Correa Bahnsen A, Aouada D, Stojanovic A, Ottersten B. Feature engineering strategies for credit card fraud detection[J]. *Expert Systems with Applications*, 2016,51:134–142. DOI: https://doi.org/10.1016/j.eswa.2015.12.030.

[20] Zhang X, Han Y, Xu W, Wang Q. HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture[J]. *Information Sciences*, 2021,557:302–316. DOI: https://doi.org/10.1016/j.ins.2019.05.023.

[21] Carcillo F, le Borgne YA, Caelen O, et al. Combining unsupervised and supervised learning in credit card fraud detection[J]. *Information Sciences*, 2021,557:317–331. DOI: https://doi.org/10.1016/j.ins.2019.05.042.

[22] Błaszczyński J, de Almeida Filho AT, Matuszyk A, et al. Auto loan fraud detection using dominance-based rough set approach versus machine learning methods[J]. *Expert Systems with Applications*, 2021,163. DOI: https://doi.org/10.1016/j.eswa.2020.113740.

[23] Soemers DJNJ, Brys T, Driessens K, et al. Adapting to concept drift in credit card transaction data streams using contextual bandits and decision trees[C]. *Proceedings of the 30th Innovative Applications of Artificial Intelligence Conference*, New Orleans Louisiana, USA, February 2 – 7,2018, IAAI 2018 7831–7836.

[24] Wang Z, Jiang C, Zhao H, Ding Y. Mining Semantic Soft Factors for Credit Risk Evaluation in Peer-to-Peer Lending[J]. *Journal of Management Information Systems*, 2020,37:282–308. DOI: https://doi.org/10.1080/07421222.2019.1705513.

[25] Pang X, Zhou Y, Wang P, et al. An innovative neural network approach for stock market prediction[J]. *Journal of Supercomputing*, 2020,76:2098–2118. DOI: https://doi.org/10.1007/s11227-017-2228-y.

[26] Craja P, Kim A, Lessmann S. Deep learning for detecting financial statement fraud[J]. *Decision Support Systems*, 2020,139:113421. DOI: https://doi.org/10.1016/j.dss.2020.113421.

[27] Chen YJ, Wu CH, Chen YM, et al. Enhancement of fraud detection for narratives in annual reports[J]. *International Journal of Accounting Information Systems*, 2017,26:32–45. DOI: https://doi.org/10.1016/j.accinf.2017.06.004.

[28] Baesens B, Höppner S, Verdonck T. Data engineering for fraud detection[J]. *Decision Support Systems*, 2021,150. DOI: https://doi.org/10.1016/j.dss.2021.113492.

[29] Olszewski D. Fraud detection using self-organizing map visualizing the user profiles[J]. *Knowledge-Based Systems*, 2014,70:324–334. DOI: https://doi.org/10.1016/j.knosys.2014.07.008.

[30] Zaslavsky V, Strizhak A. Credit Card Fraud Detection Using Self-Organizing Maps[J]. *Information & Security: An International Journal*, 2006,18:48–63. DOI: https://doi.org/10.11610/isij.1803.

[31] Srivastava A, Kundu A, Sural S, Majumdar AK. Credit card fraud detection using Hidden Markov Model[J]. *IEEE Transactions on Dependable and Secure Computing*, 2008,5:37–48. DOI: https://doi.org/10.1109/TDSC.2007.70228.

[32] Gianini G, Ghemmogne Fossi L, Mio C, et al. Managing a pool of rules for credit card fraud detection by a Game Theory based approach[J]. *Future Generation Computer Systems*, 2020,102:549–561. DOI: https://doi.org/10.1016/j.future.2019.08.028.

[33] Akila S, Srinivasulu Reddy U. Cost-sensitive Risk Induced Bayesian Inference Bagging (RIBIB) for credit card fraud detection[J]. *Journal of Computational Science*, 2018,27:247–254. DOI: https://doi.org/10.1016/j.jocs.2018.06.009.

[34] Zhou J, Li W, Wang J, et al. Default prediction in P2P lending from high-dimensional data based on machine learning[J]. *Physica A: Statistical Mechanics and its Applications*, 2019,534:122370. DOI: https://doi.org/10.1016/j.physa.2019.122370.

[35] Jurgovsky J, Granitzer M, Ziegler K, et al. Sequence classification for credit-card fraud detection[J]. *Expert Systems with Applications*, 2018,100:234–245. DOI: https://doi.org/10.1016/j.eswa.2018.01.037.

[36] Huang YP, Yen MF. A new perspective of performance comparison among machine learning algorithms for financial distress prediction[J]. *Applied Soft Computing Journal*, 2019,83:105663. DOI: https://doi.org/10.1016/j.asoc.2019.105663.

[37] Ashraf S, Félix EGS, Serrasqueiro Z. Development and testing of an augmented distress prediction model: A comparative study on a developed and an emerging market[J]. *Journal of Multinational Financial Management*, 2020,57–58,100659. DOI: https://doi.org/10.1016/j.mulfin.2020.100659.

[38] Petropoulos A, Siakoulis V, Stavroulakis E, Vlachogiannakis NE. Predicting bank insolvencies using machine learning techniques[J]. *International Journal of Forecasting*, 2020,36:1092–1113. DOI: https://doi.org/10.1016/j.ijforecast.2019.11.005.

[39] Pradeepkumar D, Ravi V. Forecasting financial time series volatility using Particle Swarm Optimization trained Quantile Regression Neural Network[J]. *Applied Soft Computing Journal*, 2017,58:35–52. DOI: https://doi.org/10.1016/j.asoc.2017.04.014.

[40] Jones S, Johnstone D, Wilson R. An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes[J]. *Journal of Banking and Finance*, 2015,56:72–85. DOI: https://doi.org/10.1016/j.jbankfin.2015.02.006.

[41] He Y, Zhang W. Probability density forecasting of wind power based on multi-core parallel quantile regression neural network[J]. *Knowledge-Based Systems*, 2020,209:106431. DOI: https://doi.org/10.1016/j.knosys.2020.10643.