

Credit Card Fraud Detection Based on ANN and XGBoost

Junfeng Shi^a

Faculty of Science and Engineering, University of Nottingham Ningbo, Ningbo, China

Keywords: Credit Card Fraud Detection, ANN, XGBoost.


Abstract: Due to the rapid expansion of electronic payment systems, the number of people using credit cards has grown significantly. However, as credit cards provide massive convenience for the public, the number of related fraud cases subsequently increased, causing significant losses for the public and the credit card issuing banks. The traditional method of fraud detection requires an extended manual analysis, which becomes almost impossible given the massive popularity of credit cards. Hence, establishing an effective credit card fraud detection (CCFD) system is imperative to reverse the situation. ANN and XGBoost are two powerful algorithms for classification problems. Their performances on balanced data sets have already been recognized, while their performances on imbalanced data sets remain unknown. To discover whether these two algorithms are suitable for CCFD, this paper applies ANN and XGBoost to the binary classification problem of CCFD and analyses their performance. The result shows that the accuracy rate of both ANN and XGBoost is as high as 99.96%. However, the f1 score of XGBoost on the minority class is higher than ANN's, indicating that XGBoost can identify the minority class more efficiently. Therefore, XGBoost is a better option for credit card detection than ANN.

1 INTRODUCTION

In this digital age, credit cards have become a popular means of payment. However, along with their convenience, fraud issues have become more serious, posing a severe challenge for consumers, financial systems, and the whole economy.

Credit card detection can be divided into inner or external card fraud. Inner card fraud refers to forging an ID card or using the loopholes in the bank system to commit a crime, and external card fraud refers to stealing cards or the card information to commit a crime (Awoyemi, Adetunmbi, and Oluwadare, 2017). So far, most fraud cases come from external card fraud, requiring a long time to analyse the cardholders' consumption patterns and the former transaction information using traditional methods (Azhan & Meraj, 2020). Therefore, data mining techniques have been increasingly utilised to optimise detecting efficiency, among which machine learning methods stand out prominently. Machine learning has been widely applied in fraud detection; methods such as ANN, Random Forest, K-Nearest Neighbour (KNN), Support Vector Machine (SVM), and PK-

XGBoost have already been commonly used (Bin Sulaiman, Schetinin, and Sant, 2022). Many researchers have also summarised and analysed the performance of different methodologies in the detection task. For example, Prajapati, Mehta, Jhaveri, and Kelkar (2021) analysed the performance of logistic regression, KNN, Random Forest, SVM, decision tree, and Naive Bayes Algorithm. In this paper, two advanced algorithms are estimated respectively: ANN and XGBoost. According to Wu, Li and Ma (2021), XGBoost performs better than ANN in the classification tasks on balanced data sets. However, the performance applied to data sets that are highly imbalanced remains unknown. Therefore, this paper compares and analyses the performance of these two algorithms applied to a highly imbalanced data set to reveal their advantages and limitations and discuss their applicability in different scenarios. Several indices will be utilised to compare the performance of the two algorithms, including recall, precision, f1-score, accuracy score and TP, FP, TN, and FN.

^a <https://orcid.org/0009-0003-6251-6432>

2 METHOD AND DATA

2.1 ANN

ANN is an algorithm utilising distributed parallel information processing inspired by simulating the structure and function of biological neural networks. It mainly consists of input layer, hidden layer, and output layer. Every layer is connected by neurons, which transform the signals. There are many definitions of neurons. This paper employs the most commonly used McCulloch-Pitts Model. In this definition, each neuron of the latter layer will give a specific weight to neurons in the former. When signals are transmitted between layers, the weighted summation of the signals in the former layer will be transformed nonlinearly through an activation function, where the result is obtained.

ANN is an algorithm applicable to credit card detection tasks. Rizki, Surjandari, and Wayasti (2017) applied ANN and SVM to detect financial fraud in Indonesian listed companies. The result shows that ANN has a 90.97% precision on data sets without feature selections, higher than that of SVM. Lin, Chiu, Huang, and Yen (2015) applied ANN, logistic regression and CART to identify fraud in financial statements. The result shows that ANN outperforms all other algorithms, approaching a 92.8% precision on the test set. Sahin and Duman (2011) discovered that ANN performs better than logistic regression. In conclusion, ANN performs well in financial fraud detection, indicating that ANN is a viable option for CCFD. The training processes for ANN are as follows in Figure 1.

The five procedures of ANN training include forward propagation, error calculation, backward propagation, gradient descent and iterative update. To begin with, a prediction result is obtained from the ANN by inputting a set of data. After that, the loss function is calculated using the prediction and actual values. The third step is backward propagation, which refers to calculating the gradient of each layer to get the partial derivative of the loss function concerning the coefficients of variables in the input layer. Finally, optimisation algorithms such as gradient descent are used to modify the weights and biases by minimising

the loss function. The process will be repeated several times until the termination condition is met.

2.2 XGBoost

Chen and Guestrin developed the XGBoost algorithm in 2016. XGBoost is an ensemble learning algorithm based on Boosting. The framework is to ensemble multiple decision trees to construct a strong learner and reduce error. In the training process, the weak learner in the latter model is used to predict the residual error in the former model to minimise the error. When the error is decreased to a specific threshold, the results in each weak model will be added to get the final prediction. Compared to traditional Gradient Boosting Decision Tree algorithms (GBDT), XGBoost improves significantly in many aspects, especially in the object function. Usually, the object function of GBDT only contains the loss function that is approximated using first-order Taylor Expansion. XGBoost improves the object function in the following two parts. Firstly, regularisation terms are added as a penalty for the complexity to prevent overfitting, enabling the model to maintain high predicting accuracy and good generalisation ability. Second-order Taylor Expansion is utilised to approximate the loss function, describing the function change more precisely, capturing more information about the learning rate and constructing a more robust model.

XGBoost performs well in many classification problems, as many researchers have confirmed. Priscilla and Prabha (2020) obtained OXGBoost by optimising the XGBoost model, discovering that the new model demonstrates comparatively high precision when dealing with imbalanced data. Liew, Hameed, and Clos (2021) combined deep-learning feature selection methods with the XGBoost classifier and applied them to the breast cancer classification task to identify cancerous cells. The result shows that the XGBoost's performance is remarkable. Hajek, Abedin, and Sivarajah (2023) presented a fraud identification framework based on XGBoost. They compared it with many other advanced machine-learning methods, discovering that XGBoost performs better than machine-learning and unsupervised techniques. In summary, XGBoost

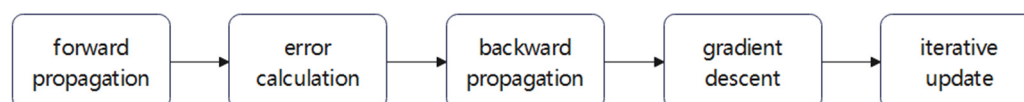


Figure 1: Flowchart of ANN Training.

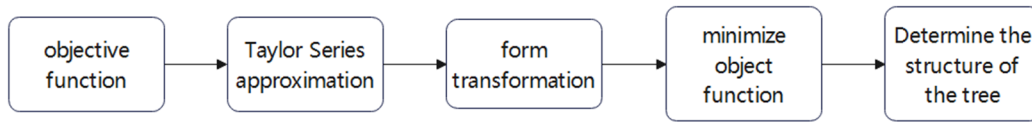


Figure 2: Flowchart of XGBoost Training.

performs well in the classification problem of the imbalanced data sets, revealing that it is very suitable for CCFD. The training processes for XGBoost are as follows in Figure 2.

In general, the principle of training XGBoost is to minimise the object function to identify the construction of the decision trees. The first step is determining the object function, which consists of regularisation terms and training loss. The training loss is calculated based on the difference between the prediction and the samples' actual values. According to the type of question, the difference can be calculated through various functions such as MSE, cross-entropy and so on. Regularisation terms usually contain the number and the weights of leaf nodes. The second step is to approximate the training loss using second-order Taylor expansion. To simplify the subsequent calculations, the approximation is usually transformed from iterating through each sample to iterating through each leaf node. The final step is to calculate the Gain of each node division to determine the decision tree's structure. For every single node, the gain refers to the difference between the values of the object functions before division and after division. A division with a difference significantly smaller than zero indicates that the division is conducive to minimising the object function and improving the model's performance. Conversely, a difference insignificantly smaller than zero or significantly larger than zero means the division is redundant or detrimental to the model's performance. Therefore, after setting the termination condition, the decision tree can be constructed according to the gain of each division.

2.3 Data Source

The data set is collected from extensive data mining and fraud detection research, which was cooperated with by the cross-border payment platform Worldline and the Free University of Brussels. It records the transaction information of European credit card holders for two days in September 2013, totalling 284,807 data entries. The data set contains the Time, Amount, Class (1: fraud, 0: legitimate) and 28 features processed by PCA. The data was split into 30% test set and 70% train set. In the train set, 20% of the data is used for validation. Data normalisation

is also used to optimise the performance of the models.

2.4 Model Construction

This paper designed a deep learning model based on ANN. The input layer, which contains 28 neurons, receives the feature data. The hidden layer utilises a series of dense layers, each containing 256 neurons, taking ReLU as the activation function. This paper adds batch normalisation and dropout layers in the hidden layers to optimise the training process, accelerate convergence, and reduce overfitting. The output layer is a single neuron layer, applying the sigmoid function to generate the probability. After clarifying the essential components, this paper chooses Adaptive Moment Estimation (Adam) to update the weights. Subsequently, binary cross entropy is decided to be the loss function.

This paper utilised its default settings for the XGBoost model. Notably, the Average Precision-Recall Curve (AUPRC) is used as the evaluation metric for its performance to prevent it from overfitting.

2.5 Result Evaluation

The result evaluation contains five tables. The first represents the accuracy score, TP, FP, TN, and FN of ANN and XGBoost during training and testing. The other four tables represent the classification report of the training and testing process.

3 RESULTS

The results demonstrate that XGBoost's performance is better than ANN's overall. In training, XGBoost achieves 100% accuracy and scores high on recall, precision, and f1-score. ANN also achieves 99.99% accuracy, but its ability to handle the minority class is still lacking compared to XGBoost. The precision, recall and f1-score of the minority using XGBoost are all 1.00, while ANN only gets 1.00, 0.95 and 0.97, respectively. Meanwhile, reducing the number of False Negatives is especially important due to the uniqueness of CCFD. The detection results of ANN

and XGBoost showed 14 and 0 false negatives. Therefore, XGBoost demonstrates a better performance than ANN in the training process.

The analysis of the testing process reaches the same conclusion. The two models achieved 99.96% accuracy in testing, indicating their effectiveness in complicated binary classification problems. However, XGBoost can detect minority classes more precisely, scoring higher than ANN on recall, precision, and f1-score in the minority class and getting fewer false negatives (See Tables 1-5).

In conclusion, this paper reveals the significant advantages of XGBoost in handling high-

dimensional, nonlinear, and class-imbalanced data. This discovery enriches the understanding of these two machine-learning models and provides valuable references and guidance for future studies in the CCFD area.

4 CONCLUSIONS

This paper applied ANN and XGBoost to CCFD problems and finally discovered that XGBoost performs better on imbalanced data. Despite XGBoost and ANN both achieving a high accuracy

Table 1: Performance Evaluation Table for ANN and XGBoost Models.

	ANN-Train	ANN-Test	XGBoost-Train	XGBoost-Test
Accuracy Score	99.99%	99.96%	100.00%	99.96%
True Positive	273	110	287	111
False Positive	0	9	0	6
True Negative	159204	85298	159240	85301
False Negative	14	26	0	25

Table 2: Classification Report of ANN-Train.

	0	1	accuracy	macro avg	weighted avg
recall	1.00	1.00	1.00	1.00	1.00
precision	1.00	0.95	1.00	0.98	1.00
f1-score	1.00	0.97	1.00	0.99	1.00
support	159204.00	287.00	1.00	159491.00	159491.00

Table 3: Classification Report of ANN-Test.

	0	1	accuracy	macro avg	weighted avg
recall	1.00	0.92	1.00	0.96	1.00
precision	1.00	0.81	1.00	0.90	1.00
f1-score	1.00	0.86	1.00	0.93	1.00
support	85307.00	136.00	1.00	85443.00	85443.00

Table 4: Classification Report of XGBoost-Train.

	0	1	accuracy	macro avg	weighted avg
recall	1.00	1.00	1.00	1.00	1.00
precision	1.00	1.00	1.00	1.00	1.00
f1-score	1.00	1.00	1.00	1.00	1.00
support	159204.00	287.00	1.00	159491.00	159491.00

Table 5: Classification Report of XGBoost-Test.

	0	1	accuracy	macro	avg
recall	1.00	0.95	1.00	0.97	1.00
precision	1.00	0.82	1.00	0.91	1.00
f1-score	1.00	0.88	1.00	0.94	1.00
support	85307.00	136.00	1.00	85443.00	85443.00

score of 99.96%, XGBoost demonstrates a better handling of the minority class, reaching a 95% precision on the minority class compared to ANN's 92% precision. Therefore, this paper concludes that XGBoost might be a more reasonable choice for CCFD.

However, there are some limitations to the collection of data sets. Due to resource constraints, this paper only utilises the transaction data of European credit card holders for two days in September 2013. The insufficiency of the samples may lead to a failure to reflect the overall situation. Moreover, although the results are apparent in the research experiments, many changes may still exist in the actual application.

In response to the limitations, the research can be improved by taking the following measures: firstly, analysing a data set that contains more samples; secondly, using different data sets to validate the universality of the result.

REFERENCES

- Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. In 2017 International Conference on Computing Networking and Informatics (ICCNI) (pp. 1-9). Lagos, Nigeria.
- Azhan, M., & Meraj, S. (2020). Credit card fraud detection using machine learning and deep learning techniques. In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS) (pp. 514-518). Thoothukudi, India.
- Bin Sulaiman, R., Schetinin, V., & Sant, P. (2022). Review of Machine Learning Approach on Credit Card Fraud Detection. *Hum-Cent Intell Syst*, 2, 55–68.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794.
- Hajek, P., Abedin, M. Z., & Sivarajah, U. (2023). Fraud Detection in Mobile Payment Systems using an XGBoost-based Framework. *Information Systems Frontiers*, 25, 1985–2003.
- Liew, X. Y., Hameed, N., & Clos, J. (2021). An investigation of XGBoost-based algorithm for breast cancer classification. *Machine Learning with Applications*, 6, 100154.
- Lin, C.-C., Chiu, A.-A., Huang, S. Y., & Yen, D. C. (2015). Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. *Knowledge-Based Systems*, 89, 459-470.
- Prajapati, D., Tripathi, A., Mehta, J., Jhaveri, K., & Kelkar, V. (2021). Credit Card Fraud Detection Using Machine Learning. In *Proceedings of the 2021 International Conference on Advances in Computing, Communication, and Control (ICAC3)*.
- Priscilla, C. V., & Prabha, D. P. (2020). Influence of Optimizing XGBoost to Handle Class Imbalance in Credit Card Fraud Detection. In *Proceedings of the 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*.
- Rizki, A. A., Surjandari, I., & Wayasti, R. A. (2017). Data mining application to detect financial fraud in Indonesia's public companies. In *Proceedings of the 2017 3rd International Conference on Science in Information Technology (ICSITech)*.
- Sahin, Y., & Duman, E. (2011). Detecting credit card fraud by ANN and logistic regression. In *Proceedings of the 2011 International Symposium on Innovations in Intelligent Systems and Applications*.
- Wu, J., Li, Y., & Ma, Y. (2021). Comparison of XGBoost and the Neural Network model on the class-balanced datasets. In *Proceedings of the 2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC)*, Greenville, SC, USA, pp. 457-461.