ISSN: 2688-9323

Credit Card Fraud Detection based on Generative Adversarial Network

Dongyang Zhang , Panpan Quan

School of Control and Computer Engineering, North China Electric Power University , Hebei, China

Abstract

With the globalization of the economy and the international financialization, the global credit system has gradually been established, and the use of credit cards for payment has become popular. But with the rapid growth of the credit-card industry, the ways used for credit-card fraud are becoming more diverse, such as embezzlement and malicious overdraft of other people's credit cards and even counterfeiting of credit cards in private. Therefore, it is particularly important to ensure that consumers can safely and efficiently use credit cards to pay, and how to identify credit card fraudulent transaction data efficiently, quickly and accurately has become a popular concern. Credit card data are unbalanced, the number of Sample examples is very small, and the supervised classification method will be affected by the problem of non-imbalance. In this paper, a generative adversarial network model is proposed to deal with the imbalance of supervised classification in credit card fraud detection.

Keywords

Generate Adversarial network, Augmented Training Sets, Inequality of class, K-means clustering algorithm.

1. Introduction

Credit cards are now a common payment method for online payments, and effective fraud detection is crucial for all organizations that issue credit cards or manage online transactions. Fraud detection system can identify suspicious transactions from transaction data through complex analysis and data mining techniques, in which illegal transactions are mixed with legal transactions in transaction data. Machine learning is one of the methods to carry out effective fraud detection [1]. By analyzing the pre-selected sample data set, a binary classification system is formed to distinguish between fraud and non-fraud instances.

For the problem of credit card fraud detection, many transaction fraud detection is based on data mining, data warehouse can provide a lot of data available for analysis, now it is considered the most promising and most effective solution is based on the monitoring method of machine learning. Many researchers have suggested ways to solve the problem. Decision tree and boolean logic functions have already used to solve this class of problems. By analyzing the behavior of credit card fraud data, clustering the data set, using the decision tree to classify, analyzing the connection between various classes,and then according to the characteristics of relevant behavior to distinguish whether the transaction data is fraud data or non-fraud data[2]. Several data visualization methods widely used in credit card fraud detection [3]. Integrate the database with data mining and data visualization to select the most appropriate method through different visualization tools. Some scholars have proposed a credit card fraud detection method based on support vector machine (SVM) [4]. In addition, based on the traditional support vector machine, the fuzzy two-norm kernel-free quadratic surface support vector machine is proposed for credit card fraud detection [5]. Neural

ISSN: 2688-9323

networks have also been used to detect credit card fraud through large-scale data mining, such as the Falcon fraud assessment system of HNC companies, which is now used by many retail banks to detect credit card transactions [6].

In a particular application area, the fraud category has much less data than other categories due to the severe imbalance of the trained data sets, which significantly reduces the validity of the binary classifier. So as to achieve the effect of improving the credit card fraud detection problem, the generative adversarial network model captures the potential distribution of the real data sample through the generative model, generate network model can generate examples which are as possible as similiar to fraud instances , by increasing the number of fraud cases, availably solve the problem that the credit card fraud data itself has the category imbalance, realizes the synthetic minority oversampling method based on the generative countermeasure network. After putting forward the concept of generative adversarial network model. In machine learning research, the generative adversarial model is advancing in theory and application, and have already applied in a variety of imbalance problems.

2. Generate Adversarial Network

2.1. Generate Adversarial Network Model

With the rapid development of information technology, credit card has become the main media in the field of payment. At the same time, credit card fraud transactions grow at an amazing speed, and fraud techniques are constantly renovated. How to effectively prevent the risk of credit card fraud has become a research hotspot in the field of bank risk control. The most advanced fraud detection method aims to identify suspicious usage patterns from transaction logs through data analysis and mining technology. The supervised classification method in machine learning is particularly effective to solve this problem.

As the fraud data of credit card has the characteristics of category imbalance (the number of normal transactions is far greater than the number of abnormal transactions), it will cause serious damage to the performance of traditional machine learning classifiers. Sampling from the data level is the main method to solve this problem. Generative adversarial networks (GAN) is a generative model proposed by goodsell et al [7]. In 2014. Inspired by the two person zero sum game in game theory, GAN can capture the potential distribution of real data samples. At present, the generation countermeasure network has been applied to many kinds of imbalance problems, such as fraud detection [8], fault diagnosis [9], medical image diagnosis [10].

The model is inspired by the two-person zero-sum game in game theory, by analyzing the data, we can find the distribution rule of the sample [11]. The generative adversarial network model consists of a generator and a discriminator. The discriminator is a dichotomy that discriminates whether the input is a real data or a generated sample; the generator is responsible for capturing the potential distribution of the real data sample and generating new samples based on the potential distribution. The process of model training is the process of deep learning, the generation accuracy and the discriminant ability are improved by two networks fighting each other.

ISSN: 2688-9323

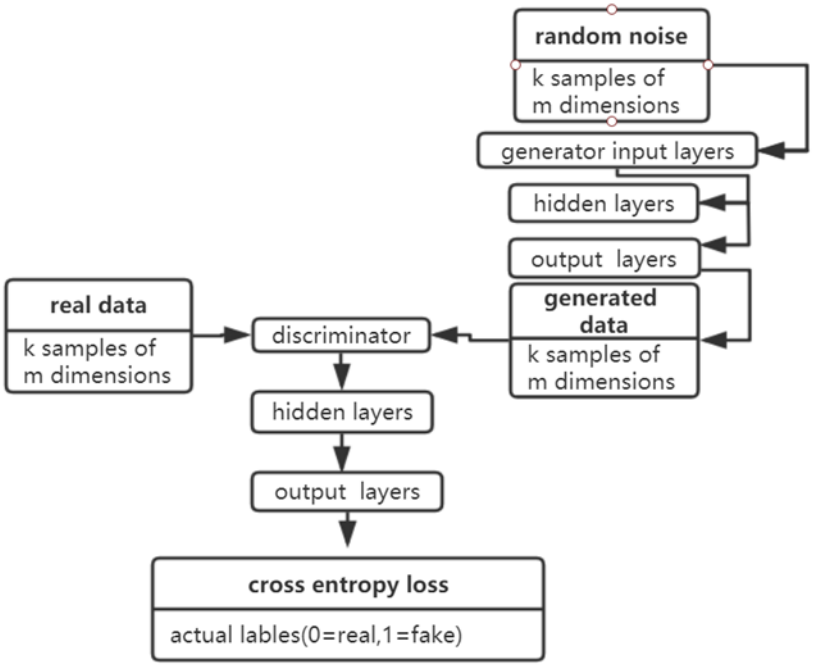

Figure 1. Generation of adversarial network models

2.1.1. Generate Network Model

The generated model can be regarded as a neural network model. The input of the generated model is a set of random numbers Z, and the output is a set of data, not a numerical value. Generating network is used to generate false samples. Its purpose is to make it impossible for discriminant network model to judge whether the generated false samples come from the original data set or the generated data set.

Each neural network consists of three layers: input layer, hidden layer and output layer. The input layer contains n nodes which are used to transfer input information to the hidden layer. The hidden layer contains m nodes, and the choice of nodes depends on the specific situation. In this experiment, the activation function of hidden layer is relu function and sigmoid function, and the function of activation function is real-time conversion of input data. The output layer contains one or more nodes whose function is to summarize the output of the hidden layer.

In this experiment, the generative network is a four-layer neural network with 128, 256, 512, 30 neurons per layer.

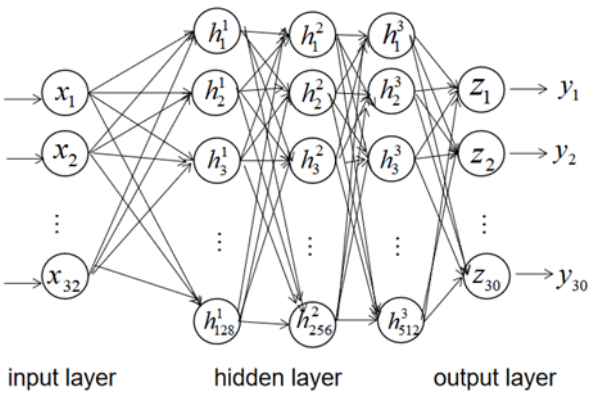

Figure 2. Generate network model

ISSN: 2688-9323

In each layer, the unit transforms its input, usually the nonlinear activation function is applied to the linear combination wu + b where u is the input vector of the unit, w is the weight matrix, and b is the additive deviation vector. In generating networks, the activation function for each layer is the relu function. The definitions are as follows:

$$\text{sigm}(v) = 1/(1 + e\text{^}(-v)) \quad (1)$$

2.1.2. Discriminate Network Model

The discriminant model can also be regarded as a neural network model.The discriminative model is in the construction process, roughly contrary to the process of generating the model.The input is a set of data, and the output is a probability value, which is used to judge the true and false use of the data. The function of discriminating network is to distinguish whether the data object belongs to the original sample set or the generated false sample set. If the input discriminates the network is the real data, the network output is close to 0, and if the input is the generated false sample, the network output will be close to 1, thus achieving a good discrimination purpose.

In this experiment, the discriminative network is also a four-layer neural network, the number of neurons in each layer is 512 ,256 ,128 ,1.

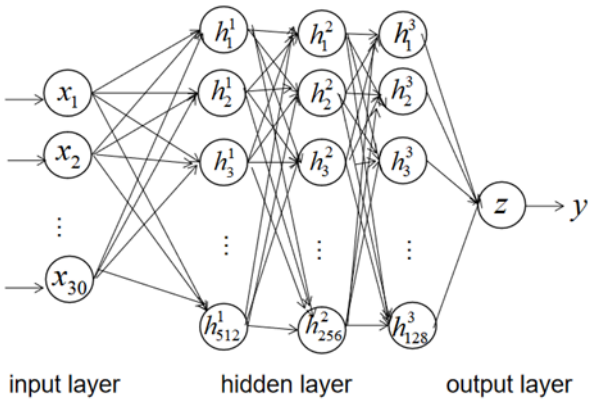

Figure 3. Discriminant network model

The activation function of the three hidden layers is relu function, and the activation function of the output layer is sigmoid function. The definitions are as follows:

$$\text{relu}(x) = \max\{0, x\} \quad (2)$$

2.1.3. Training Model

The generative and adversarial models belong to two completely independent neural network models. The method we use to train these two models is: alternate iterative training alone.

If we have built up the generative adversarial network model, as far as we know, the generation of the anti-network effect at this time is not good, now input random array, will output false sample set, at this time the discriminant network can easily distinguish the output of the generated data set is false. For the generated sample set and the original sample set, we can define the labels of these two sample sets. In this paper, we define the original sample set class label as 0 and the false sample set class label as 1.

At this point, for discriminative network, the next problem to be solved is a universal supervised dichotomy problem, which needs to be directly sent to the neural network model for training. The main function of discriminative network is to judge whether the generated samples are true. while for the generative network, its main function is to generate samples as realistic as possible. In the training model, the discriminant network should be connected in

ISSN: 2688-9323

series behind the generated network. At this point, the training of generative adversarial network model is the training of generating-discriminant network tandem. When training the generating network, the generated false samples are processed as true samples, and the labels of the generated false samples are all Set to 0.

The parameters of the discriminative network must remain unchanged when training the generation-discriminant network. Then passed the error to the generating network, and updated the parameters of the generating network.

After both the generative and discriminative networks are trained, we can generate a new false sample of the previous noise Z based on the new generative network. The process is then repeated for individual alternating training.

2.2. Model Optimization

The model of generating countermeasure network consists of a generator and a discriminator. The discriminator is a two classifier to determine whether the input is real data or generated samples; the generator is responsible for capturing the potential distribution of real data samples and generating new samples according to the potential distribution .In essence, the generation model is a kind of maximum likelihood estimation. Through the transformation of some parameters in the maximum likelihood estimation, firstly, according to the distribution of sample data, the distribution of some data characteristics is found, and then the training bias is converted to the samples with specified distribution according to the distribution of the original input data. The discriminant model is essentially a two classification model, which is responsible for judging whether the data generated by the generated model is the data in the real training data.

The discriminator is trained by minimizing its prediction error, while the generator is trained by maximizing the prediction error, which is formalized as:

$$\min\max(Ex{\sim}p_{data(x)}\left[\log\left(D(x)\right)\right] + Ez{\sim}p_z(z)\left[\log\left(1 - D\left(G(z)\right)\right)\right]) \tag{3}$$

The above formula is a process used to solve the maximum and minimum optimization problem, in fact, this process contains two optimization processes. Dismantling this formula is the following two formulas. Where first optimize the discriminator and then optimize the generator. (D for discriminator and G for generator).

Optimal discriminator :

$$maxV(D,G) = Ex{\sim}p_{data(x)}\left[\log\left(D(x)\right)\right] + Ez{\sim}p_z(z)[\log\left(1 - D\left(G(z)\right)\right)] \tag{4}$$

Optimization Generator :

$$\min V(D,G) = Ez{\sim}p_z(z)[\log\left(1 - D\left(G(z)\right)\right)] \tag{5}$$

where $G(z)$is the output result of the generated model; $D(x)$is the distribution of the original input data; $D\left(G(z)\right)$is the outp1ut result of the discriminant model and is a real value in the range of 0-1; $p_{data(x)}$and$p_{data(x)}$represent the distribution of the real data and the data distribution of the generated data.

The first term of the formula to optimize D, to make the true sample x input, the larger the result the better, because the closer the prediction result of the need for the true sample is to 1, the better. For fake samples, it is necessary to optimize the smaller the result the better, that is, the smaller the $D\left(G(z)\right)$the better. But the bigger the first item is and the smaller the

ISSN: 2688-9323

second, the contradiction arises, so change the second item to $1-\mathrm{D}(\mathrm{G}(\mathrm{z}))$so that the bigger the better.

Also in optimizing G, at this time only need to care about the fake sample, we define the label of the fake sample is 1, so it is $D(G(z))$the bigger the better, but in order to unify into the form of 1- $D(G(z))$, so change to minimize $1-D(G(z))$.

3. Experimental Process

3.1. Data Set

The experiment uses Kaggle's credit card fraud detection dataset, which contains 285000 transactions, of which only 492 are illegally traded. The 492 positive class examples represent 0.172 per cent of all transactions in the dataset. The data consist of 31 features, including:"Time","Class","Amount"as well as numeric features labeled V_1 to V_{28}. "Class" represents the category feature and is used to indicate whether the transaction is a fraud label, where "0"means normal and "1" means fraud. "Time"contains the data in the dataset, and the time of transaction delivery in seconds. The characteristic "amount" is the amount of data transaction that can be used to understand the characteristic analysis of fraudulent transaction. And features V_1, V_2...$V_{28}$is the main component obtained by PCA.

3.2. Experiment

The experiment was based on the Windows operating system, using the Anaconda software package and the Python programming language. The development tools needed for the experiment include TensorFlow, deep learning package keras, data processing package pandas, model test package scikit-learn, etc. Experiments were carried out on jupyter notebook platform to improve the effect of fraud checking on credit cards by building a model of generating anti-networks to synthesize a few types of data.

The problem of credit card fraud detection is essentially to solve an unbalanced classification problem. The main idea of this paper was to improve the related algorithms to mitigate the imbalance of fraud data sets. In the research of the previous chapter, the generative adversarial network model is established to solve the problem of unbalanced classification. The experiment of this chapter is mainly divided into two parts, the first part is the credit card fraud detection on the original training set, the second part is the credit card fraud detection on the augmented training set. Finally compare and evaluate the two classification effect.

In this paper, the experimental part of credit card fraud detection model mainly includes data analysis, model building and model training, output, testing and evaluation. The data preprocessing module first processes the original data set and analyzes the data set. In model training, the optimal parameters are selected by loss function, and the training model based on the generated countermeasure network is established. After data analysis, we learn the characteristics of the transaction data of the existing data set, generate the characteristics of fraud behavior, and then enter the training module of the generation network model against fraud, and then use the model to detect fraud. When the characteristics of data and fraud are similar or match, corresponding judgment can be made. In this experiment, we train a generation network model to output a small number of simulation examples, and then combine these generation examples and training data into an enhanced training set to improve the effectiveness of the classifier. The following diagram shows the basic flow chart of the experiment.

ISSN: 2688-9323

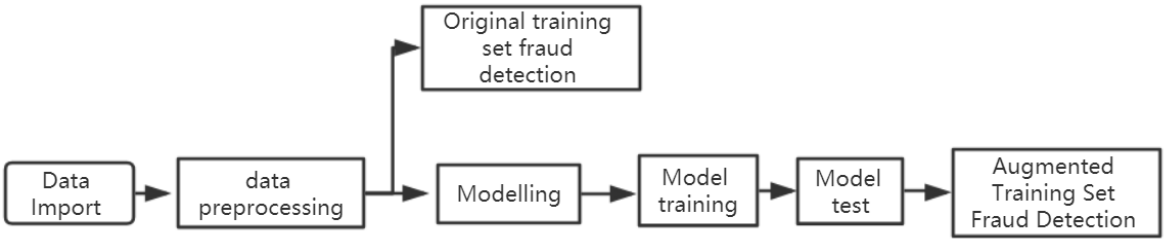

Figure 4. Flow chart of experiment

In the first experiment without using the model, I trained 70% of the data sets and tested the remaining 30%. Set the algorithm to continue until the recall rate of the test data set cannot be increased. In the original training set, xgboost algorithm is used to train binarization, and 76% of the recall rate and 94% of the accuracy rate are obtained through the test. 76% of the recall rate and 94% of the accuracy rate in the test set mean that only 6% of the predicted fraud cases are actually normal transactions.

After getting the initial accuracy, I added the generation network model. After training the model, I used half of the actual fraud data and the equivalent generation examples to train the xgboost classifier. Then I test the xgboost classifier with the other half's actual fraud data and another set of generated samples. The recall rate of the final test set is 96.7%, and the precision rate is 99.5%.

Due to the unbalanced distribution of credit card fraud data, the overall accuracy can not evaluate the classification effect well. In the experiment, the confusion matrix index is used to evaluate the experimental results. Through the performance evaluation before and after the verification set, we can see that we combine the generated examples with the original training data to get a more balanced augmented training set, so as to achieve the desired effect of the traditional classifier. The experimental results show that the classifier trained on the enhanced training set is much better than the similar classifier trained on the original data, especially in the sensitivity, thus a truly effective solution for credit card fraud detection is obtained.

4. Conclusion

Aiming at the imbalance of data, this paper uses K-means clustering algorithm to segment most of the class samples, and proposes an adversary network model to deal with the problem of class imbalance, which is used to supervise the detection and classification of credit card fraud. I set up a network model to generate antagonism, output several kinds of simulation data, and then assemble the data and the original data into an augmented training set. In the same test set, by comparing the performance of classifiers on the original training set and the augmented training set, the effectiveness of generating adversary network to deal with imbalance is proved.

In this study, the credit card fraud detection based on the generative anti network model is based on the existing data set. However, due to the rapid development of economy and finance, the fraud behavior will become more and more complex, and the amount of data will grow rapidly, so the fraud detection ability will be affected. In order to accurately predict the fraud behavior, it needs real-time adjustment Model parameters, so if we want to maintain accurate fraud detection, we must consider the real-time update of the fraud detection system.

We can train longer, larger networks and adjust the parameters of the architecture we tried in this article. Or re-examine the data cleansing we perform, perhaps design some new variables, or change whether and how we deal with the skew of features. Maybe different fraud data classification schemes will help.

ISSN: 2688-9323

Acknowledgments

This paper was financially supported by “the Fundamental Research Funds for the Central Universities(2016MS122)”.

References

[1] Gibert Daniel, Mateu Carles, Planes Jordi. The rise of machine learning for detection and classification of malware: Research developments, trends and challenges [J]. Journal of Network and Computer Applications, 2020, 153(C).

[2] Kokkinaki, A. On atypical database transactions: Identification of probable frauds using machine learning for user profiling [J]. In: Proceedings of IEEE Knowledge and Data Engineering Exchange Workshop, 1997:107-113.

[3] B.G. Becker, Using mineset for knowledge discovery [J], IEEE Comput. Graph Appl. 1997,17 (4): 75–78 .

[4] H.-C. Kim, S. Pang, H.-M. Je, D. Kim, S.Y. Bang, Constructing support vector machine ensemble, Pattern Recognit 2003, 36 (12):2757–2767.

[5] Irwin King,Jun Wang. International Conference on Neural Information Processing [J]. Neurocomputing, 2008, 71 (16).

[6] R.J. Brachman, T. Khabaza, W. Kloesgen, G. Piatetsky-Shapiro, E. Simoudis, Mining business databases, Commun. 1996, 39 (11) :42–48 .

[7] Ugo Fiore, Alfredo De Santis, Francesca Perla et al. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection[J]. Information Sciences, 2017.

[8] Fiore U , Santis A D , Perla F , et al. Using Generative Adversarial Networks for Improving Classification Effectiveness in Credit Card Fraud Detection[J]. Information Sciences, 2017: S0020025517311519.

[9] W. Mao, Y. Liu, L. Ding and Y. Li, "Imbalanced Fault Diagnosis of Rolling Bearing Based on Generative Adversarial Network: A Comparative Study," in IEEE Access, vol. 7, pp. 9515-9530, 2019.

[10] Q. Wang et al., "WGAN-Based Synthetic Minority Over-Sampling Technique: Improving Semantic Fine-Grained Classification for Lung Nodules in CT Images," in IEEE Access, vol. 7, pp. 18450-18463, 2019.

[11] Aman Gulati, Prakash Dubey, C MdFuzail et al. Credit card fraud detection using neural network and geolocation [J]. IOP Conference Series: Materials Science and Engineering, 2017, 263(4).