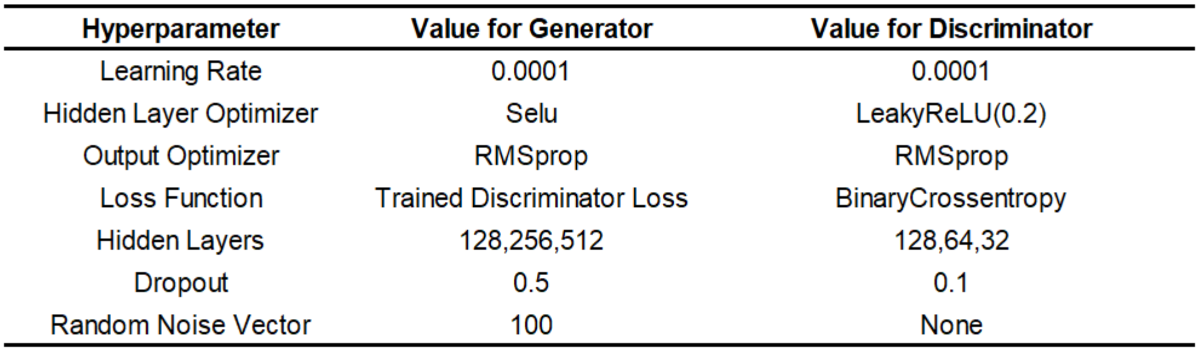
GANs-Based Data Augmentation in Credit Card Fraud Detection

Abstract

The landscape of credit card fraud is evolving rapidly, with the emergence of increasingly sophisticated fraudulent methods. This trend has resulted in a significant uptick in financial losses incurred by both businesses and consumers. To address the challenge of credit card fraud detection, the industry has widely adopted machine learning models. However, building effective models is hindered by limited real-world data and the severe imbalance between fraudulent and genuine transactions. In this work, we explore the application of Generative Adversarial Networks (GANs) to synthesize fraudulent samples and their superiority compared with traditional data augmentation techniques, extending this comparative analysis to 5 different augmentation ratios. To mitigate potential biases introduced by a single classification algorithm, Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGB) are employed, representing different modeling paradigms. our experiments show that models trained on the GANs-based synthetic data exhibit superior generalization capabilities and a stronger ability to discriminate between different classes.

1. GANs Architecture and Hyperparameters in this research

To build the GANs model for this research, we choose the vanilla GANs architecture due to its simplicity and representation. Both the generator and the discriminator comprise of 4 dense layers. For the first third dense layers, each is followed by a dropout layer and a batch normalization layer. Adding dropout layers is to prevent overfitting while batch normalization layer to speed up the training process. The last dense layer works as an output layer with the sigmoid activation function. Scaled Exponential Linear Unit(SeLu) is selected as the activation function for the Generator because of its good properties in maintaining gradient stability and enhancing generalization so that it can help the generator produce more diverse and high-quality samples. As for the Discriminator, we use the Leaky Rectified Linear Unit (LeakyReLU) as the activation function because it allows a small gradient for negative values which helps to reduce the risk of overfitting. After 100 epochs of training (batch_size=32), the loss of the generator and discriminator tend to be stable and the generator is used to augment the training dataset. Table 1 and Figure 2 present the hyperparameters of the generator and the discriminator and their losses during the training process.



Hyperparameter	Value for Generator	Value for Discriminator
Learning Rate	0.0001	0.0001
Hidden Layer Optimizer	Selu	LeakyReLU(0.2)
Output Optimizer	RMSprop	RMSprop
Loss Function	Trained Discriminator Loss	BinaryCrossentropy
Hidden Layers	128,256,512	128,64,32
Dropout	0.5	0.1
Random Noise Vector	100	None

Table 1: hyperparameters of the generator and the discriminator

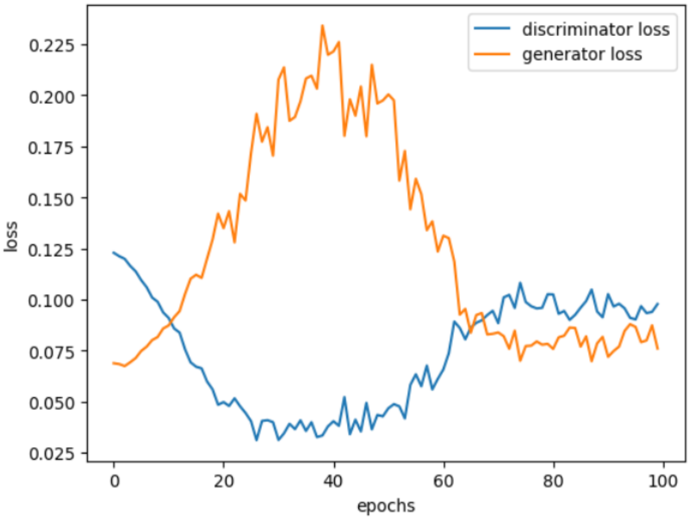


Figure 2: the losses of the generator and the discriminator

2. Experimental Protocol

This study utilizes the publicly available dataset of European cardholders in September 2013 downloaded from Kaggle [21], which is commonly used in credit card fraud research. The dataset covers transactions over two days, where 492 frauds are out of a total of 284,870 transactions. The dataset is highly unbalanced, the positive class only account for 0.172% of all transactions [21]. All 30 input variables are numerical with no missing or erroneous values. Feature “Time” contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature “Amount” is the transaction Amount [21]. Apart from these two variables, the remaining 28 variables are derived from the PCA transformation due to confidentiality concerns on the original information.

The original dataset is divided into training and testing dataset at a ratio of 8:2. All data augmentation techniques are exclusively applied to the training dataset to ensure that the testing dataset remained completely unseen, serving as a benchmark for evaluating the performance of all models. In this experiment, we employ GANs, SMOTE, and ADASYN to augment the training dataset by increasing the proportion of fraudulent samples to 10%, 30%, 50%, 80%, and 100% of the genuine samples, in order to extend the research under varying sample ratios. Logistic Regression, Random Forest, and XGB are performed to fit models separately on different ratios of dataset, 5 times repeatedly, totaling 75 fits. Then the mean and standard deviation of Precision, Recall, and F1-Score and ROC-AUC are extracted. Figure 3 presents the correlation matrix of the original training dataset and the three balanced training datasets.

3. Results Analysis（Not on display）