ACGANs-CNN: A Novel Intrusion Detection Method

Qi Zhou¹, Minsheng Tan¹*, Hewen Xi¹
¹School of Computer, University of South China, Hengyang 421001, China

*fgraw@foxmail.com

Abstract. In this paper, an intrusion detection model (ACGANs-CNN) method based on GAN and CNN fusion is proposed for the reasons that unknown attack sample data cannot be provided in training samples, the number of training samples is limited, and known attack sample types account for less such small sample data. The model converts network traffic data into grayscale images, generates the same proportion of attack samples by generating the counter network, ensures the uniform distribution of attack samples in the training set, and introduces the gradient penalty function to improve the stability of the training model. Secondly, CNN is used to better extract sample features. In order to prevent overfitting, the nonlinear activation function Relu and Dropout method are introduced. At the same time, the convergence speed of the model is accelerated, and the detection efficiency of the model is improved. Attention is introduced to highlight the key features and to classify samples based on these key features. In this paper, the KDDCUP99 data set is used for model evaluation. Experimental results show that this algorithm (ACGANs-CNN) has stronger model training stability, higher quality of generated fake samples, and better feature extraction effect in small sample data. Its detection rate and accuracy of attack types are significantly higher than that of traditional machine learning algorithms such as SVM, KNN, RF, and other CNN models.

Keywords: Intrusion detection; Conditional generation adversarial network; ACGANs-CNN.

1. Introduction

At present, traditional machine learning algorithms are mainly based on decision tree model, RF network model and support vector machine in the domain of network intrusion detection. In the literature[1], value algorithm is introduced the attribute added to calculate the best parameter value and optimize the classification accuracy, so as to reduce the false alarm rate. In the literature[2], proposed using particle swarm optimization algorithm to find the optimal feature, and using k-nearest neighbor to classify can improve the detection speed and detection rate of classification algorithm. In the literature[3], principal component analysis was used to reduce the data dimension, find the optimal attribute set, and then classify by support vector machine, which shortened the detection time and improved the detection efficiency[4][5].

Although the above methods have good detection results[6][7], in the training process, relying on manual extraction of sample features, it is easy to cause hidden correlation in the original data. In order to solve the problem of difficult feature extraction, related researchers propose to apply a convolutional neural network to network intrusion detection[8]. Literature first proposed an intrusion detection algorithm based on a convolutional neural network[9], which was applied to the domain of network intrusion detection, reducing the dependence on manual experience to extract sample features.
and innovatively introduced it into the domain of intrusion detection. Literature [10] [11] proposed that a convolution neural network can automatically propose sample features, so as to classify samples more accurately. This method does not solve the problem of overfitting and weak generalization ability of a convolutional neural network. The proposed a cross-chain aggregation method to improve the convolutional neural network and optimized the model to convergence through back-propagation, which improved the detection accuracy and false alarm rate[12]. Although it improves the ability of feature extraction to a certain extent, it cannot effectively distinguish the importance of features. Literature proposed to add attention mechanism into a convolutional neural network to calculate the importance of features, so as to obtain better classification effect[13]. However, this method does not fully consider the uneven distribution of sample data, the feature extraction of small sample data is not accurate and not obvious, which leads to the low detection rate of small sample data.

In order to solve the above problems, this paper uses the generative Adversarial network[14–21] which has good performance and robustness in the domain of image enhancement and can generate the highest degree of false samples. Using the strong feature extraction ability of the convolutional neural network in the image domain, the network traffic is converted into the image, so that the convolution neural network can better and more fully extract the image sample features. The dropout function is used to optimize the model to prevent the overfitting phenomenon, and the attention mechanism[22] [23] is introduced to give different weights to the sample features, which can effectively distinguish the importance of features and improve the characteristics. The accuracy of feature extraction. Finally, the softmax function is used to classify the features to form the probability distribution results.

2. Method

2.1. ACGANs-CNN model
In order to optimize the intrusion detection model and increase the rate of a few type test, in this paper, the original data turned gray level image processing, using the generated against Internet solve the problem of uneven distribution of the sample, CNN is utilized to extract the characteristics of the input image, and introducing the mechanism of Attention to the importance of the characteristics of the different figure to calculate, finally through the SoftMax classifier classification results are obtained. In this paper, the overall structure of the intrusion detection system combining GAN and CNN is presented in Fig. 1.

2.2. ACGANs model
The generation of the counter network can automatically learn the sample distribution and generate false sample data, which can solve the problem of the uneven sample distribution. The generation network and the discriminant network are defined in the generation adversarial network, and the generation adversarial network model is shown in Fig. 2. In the course of training, the generator network and discriminator network constantly play games and finally achieve Nash equilibrium.

Input parameters for generating network G are embedding vector V through embedding tag and Gaussian noise by the embedding layer, and then hidden vectors are obtained through the Linear layer as input parameters of internal convolutional neural network. The network is mainly composed of three layers of deconvolution neural network; each layer is subject to the deconvolution-regularization-LeakyRelu nonlinear activation function. Except that the last layer has step size 1 and the convolution kernel size 3, the others have step size 2 and the convolution kernel size 5. The last layer uses Tanh activation function to generate false samples. When training and generating a network, it is necessary to discriminate the false samples through the discriminant network, calculate Loss value through cross-entropy Loss function, and optimize the parameters of the generated network, so that the gray image input into the convolutional neural network is closer to the real value.

The purpose of network discrimination is different from that of network generation. The fundamental purpose is to distinguish between truth and falsehood and classification. It extracts the
features of the input image samples and then discriminates whether the output samples are true and the probability distribution results according to the features.

![ACGANs-CNN Intrusion Detection Model](image1)

**Fig. 1 ACGANs-CNN Intrusion Detection Model**

When training the discriminant network, the gradient penalty function is added to make the model training more stable and the image generation better. The function of the loss function is to measure the gap between performance prediction and actual data, while the purpose of network model training is to minimize the loss function. The loss function of generating a counter network mainly includes two parts, that is, the loss function of generating network G and discriminating network D. The loss functions generated against the network are shown in (1) and (2).

\[
\text{Loss}_{G} = -E_{x \sim D}(x) - r_{i}[D(x)] \quad \text{\# MERGEFORMAT (1)}
\]

\[
\text{Loss}_{D} = E_{x \sim D}(x) - E_{z \sim p_{z}(z)}[r_{i}[D(x)] + \lambda \cdot E_{x \sim p_{data}}[||D(x) - 1||^2]] \quad \text{\# MERGEFORMAT (2)}
\]

2.3. **Attention CNN model**

As a hierarchical model, the convolutional neural network is usually composed of four basic parts, namely, the input layer, the convolutional layer, the pooling layer, and the full connection layer. It can be stacked according to different requirements. In this paper, the self-attention mechanism is added to
each convolutional layer of CNN to extract features of network intrusion data and use them for classification. The computational process of its attention mechanism is shown in Fig. 3.

![Fig. 3 Attention Model](image)

3. Experiments and results

3.1. Experimental evaluation criteria

In this paper, common indicators were used to evaluate the intrusion Detection performance, including Accuracy, Detection Rate, and false alarm Rate. The performance of the intrusion detection method can be detected by these three indexes as shown in (3).

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

\[
\text{DR} = \frac{TN}{TN + FP}
\]

\[
\text{FAR} = \frac{FN}{TP + FN}
\]

Where \(TP\) means positive sample is predicted to be positive sample, \(FN\) means positive sample is predicted to be negative sample, \(TN\) means negative sample is predicted to be negative sample, \(FP\) means negative sample is predicted to be positive sample.

3.2. Dataset

In this paper, a public intrusion detection data set, KDDCUP99 is used. Each sample data in this data set contains 41 characteristic attributes and 1 tag attribute. Tag types are divided into 5 categories, including 1 type of normal data type and 4 types of abnormal data type, including DOS (denial-of-service) denial of service attack, Surveillance or Probe attack, U2R (User to Root) denial of service attack, and R2L (Remote to Local) Remote attack. The abnormal data was subdivided into 39 categories, of which 22 attack types appeared in the training set and the remaining 17 only appeared in the test set, in order to test the generalization ability of the classifier model. The data set distribution is shown in Table 1.
3.3. Results

Our proposed model and the SVM and KNN, RF theorem and other traditional machine learning algorithm are used for training and is verified on the test set, the verification results are shown in table 2 and table 3.

| Method     | Accuracy | DR   | FAR   |
|------------|----------|------|-------|
| SVM        | 72.68%   | 57.13%| 7.21% |
| KNN        | 75.11%   | 65.29%| 7.13% |
| RF         | 75.61%   | 60.19%| 2.74% |
| CGANs-DNN  | 86.98%   | 78.43%| 2.68% |
| ACGANs-CNN | 89.59%   | 86.32%| 5.16% |

**Tab.2 Test Results**

As shown in table 2. Compared to traditional models, our ACGANs-CNN model has increased by 20% in accuracy and detection rate. Although the false alarm rate has increased in a small range, it is also within the effective error interval.

| Method     | Normal  | DoS    | Probe  | U2R  | R2L  |
|------------|---------|--------|--------|------|------|
| SVM        | 72.12%  | 74.15% | 60.71% | 0.00%| 0.00%|
| KNN        | 75.33%  | 81.25% | 60.4%  | 3.52%| 3.71%|
| RF         | 80.59%  | 80.64% | 58.53% | 0.54%| 7.25%|
| CGANs-DNN  | 79.31%  | 96.23% | 76.12% | 20.97%| 59.34%|
| ACGANs-CNN | 82.15%  | 92.46% | 77.33% | 22.16%| 70.91%|

**Tab.3 Classification Results**

We can see clearly from the table 3, we proposed model not only the Accuracy above, in particular, the other three kinds of traditional machine learning algorithms, and the DR is to achieve the optimal, meanwhile, FAR is lowest than the other machine learning algorithm. Therefore, from the KDDCU99 datasets, we can draw a conclusion that our proposed model is more suitable for intrusion detection.

| Method                        | Normal | DoS    | Probe  | U2R  | R2L  |
|-------------------------------|--------|--------|--------|------|------|
| GAN-CNN                       | 75.16% | 82.58% | 65.38% | 7.45%| 12.58%|
| GAN-CNN-Attention             | 78.37% | 86.24% | 73.28% | 8.77%| 17.26%|
| ACGANs-CNN(Without Attention) | 78.89% | 88.61% | 75.46% | 15.79%| 59.63%|
| **ACGANs-CNN(Attention)**     | **82.15%** | **92.46%** | **77.33%** | **22.16%**| **70.91%**|

**Tab.4 Compare With Other Models**

In addition, in Table 4, we compared other fusion models. The Attention mechanism enhances the performance of the entire model, which helps to improve the detection rate of small samples.

4. Conclusion

This paper proposes a network intrusion detection method combining GAN and CNN. GAN is used to make up for the uneven distribution of original sample data. CNN is used to extract all sample features, and an attention mechanism is introduced to enhance the weight of important features, so as to improve detection efficiency. The experimental results show that the method has higher accuracy and a lower false alarm rate than the CNN-based method.
Acknowledgments
This work was supported by Project 18A230 of the Hunan Provincial Education Office Science Research of China, Project 20191550502 of the Hunan Provincial Office Science Research of China.

References
[1] Wang H, Chen H and Yang S 2013 An Novel Intrusion Detection System Based on Naive Bayesian Algorithm J. Computer Ence 13(10)
[2] Peng X U, Fengru J 2014 Network intrusion detection model based on particle swarm optimization and k-nearest neighbor J Computer Engineering and Applications
[3] MingYu Q I, Ming L and Yan-Ming F U 2015 Research on Network Intrusion Detection Using Support Vector Machines Based on Principal Component Analysis J Netinfo Security
[4] Zhou Y, Yu L and Liu M et al 2019 Network Intrusion Detection Based on Kernel Principal Component Analysis and Extreme Learning Machine C 2018 IEEE 18th Int. Conf. on Communication Technology (ICCT) IEEE
[5] José C , Roberto T and Josè M 2019 Garcia-Giménez,etc. Group-Wise Principal Component Analysis for Exploratory Intrusion Detection J IEEE Access
[6] Alimjan, Gulnaz and Sun et al 2018 A New Technique for Remote Sensing Image Classification Based on Combinatorial Algorithm of SVM and KNN J Int. Journal of Pattern Recognition & Artificial Intelligence
[7] Hasan M A M, Nasser M and Pal B et al 2014 Support Vector Machine and Random Forest Modeling for Intrusion Detection System (IDS) J Journal of Intelligent Learning Systems & Applications 6(1) pp44-52
[8] Vinayakumar R, Soman K P and Poornachandran P 2017 Applying convolutional neural network for network intrusion detection C 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI) IEEE pp1222-1228
[9] YU C L, ShengL L and Xing Z 2017 Intrusion Detection Algorithm Based on Convolutional Neural Network C pp1271-1275
[10] Blanco R, Cilla J J and Malagón P et al 2018 Tuning CNN input layout for IDS with genetic algorithms C Int. Conf. on Hybrid Artificial Intelligence Systems. Springer, Cham pp197-209
[11] Kim J, Kim J and Kim H et al 2020 CNN-Based Network Intrusion Detection against Denial-of-Service Attacks J Electronics 9(6) p916
[12] Hongyu Y and Feng Y W 2019 Wireless Network Intrusion Detection Based on Improved Convolutional Neural Network J IEEE Access
[13] LIU Y F, CAI S and YANG H X et al 2019 Network intrusion detection method integrating CNN and BilSTM J Computer Engineering 45(12) pp127-133
[14] Zhang Z, Pan X and Jiang S et al 2020 High-quality face image generation based on generative adversarial networks J Journal of Visual Communication and Image Representation 71 pp102719
[15] HaoChen H X Y M, Deb L D and Anil H L J L T et al 2020 Adversarial attacks and defenses in images, graphs and text: A review J Int. Journal of Automation and Computing 17(2) pp151-178
[16] Gao X, Deng F and Yue X 2020 Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty J Neurocomputing 396 pp487-494
[17] Lee M B, Kim Y H and Park K R 2019 Conditional generative adversarial network-based data augmentation for enhancement of iris recognition accuracy J IEEE Access 7 pp122134-122152
[18] Gao X, Deng F and Yue X 2020 Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty J Neurocomputing 396 pp487-494
[19] Zhu Y, Zhang Y and Zhang H et al 2020 Data Augmentation of X-Ray Images in Baggage Inspection Based on Generative Adversarial Networks J IEEE Access 8 pp86536-86544
[20] Weng Y, Zhou H 2019 Data augmentation computing model based on generative adversarial
network. *IEEE Access* 7 pp64223-64233

[21] Antoniou A, Storkey A and Edwards H 2018 Augmenting image classifiers using data augmentation generative adversarial networks *C//Int. Conf. on Artificial Neural Networks. Springer Cham* pp594-603

[22] Wang F, Jiang M and Qian C et al 2017 Residual attention network for image classification *C//Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp3156-3164

[23] Xu R, Tao Y and Lu Z et al 2018 Attention-mechanism-containing neural networks for high-resolution remote sensing image classification *J Remote Sensing* 10(10) p1602