Intrusion Detection Method Based on Deep Learning

Mingyuan Xin [1,a], Yong Wang [2,b] and lin Fan [1,c]

[1.]College of Computer and Information Engineering Heihe University Heihe City, China

[2.]President's Office Heihe University Heihe City, China

[a.]xmy5686@163.com; [b.]8150767@qq.com; [c.]1135782701@qq.com

Keywords: In-Depth learning; Intrusion detection; Artificial intelligence

Abstract. In order to solve the problem of high false alarm rate and low false alarm rate in massive network data intrusion detection, this paper proposes an intrusion detection model based on convolution neural network, which can improve the classification accuracy by selecting convolution core and data to extract local feature correlation features.

Introduction

In 2016, China's total digital economy reached 22.6 trillion yuan, accounting for more than 30% of GDP. While the Internet promotes social scientific and technological progress and simplifies users'lives, it poses unprecedented challenges to information security. (1)"China Internet Network Security Report 2017" is the National Computer Network Emergency Technology Processing and Coordination Center (hereinafter referred to as "CNCERT"). According to the situation of CNVD vulnerabilities, the number of new general software and hardware vulnerabilities has increased by more than 20% annually in the past three years. In 2017, (2)the number of CNVD "zero-day" vulnerabilities increased by 75.0% year on year.(3) Extortion software and "mining" Trojan Horse have an increasing trend. First, our country is suffering more and more network attacks. Network attacks generally refer to all attempts to bypass system security policies or penetrate the system in order to obtain, tamper with, destroy information and target networks and systems. The main reasons are network software vulnerabilities and security vulnerabilities of network protocols themselves. The goal of network intrusion detection is to identify the abnormal behavior of network and system. We can treat this problem as a classification problem to solve. The advantages of neural networks, especially convolutional neural networks in self-learning and non-linearization, have made remarkable achievements in image recognition, speech recognition and natural language processing. In view of the excellent feature classification ability of neural networks, scholars at home and abroad have applied neural networks to network attack detection and achieved some results. This paper analyses and compares the performance of current popular deep learning algorithms in NSL-KDD experiments, and proposes an intrusion detection method based on convolutional neural network.

Deep Learning Algorithms

In 2006, Hinton et al. proposed to build a multi-layer neural network model by simulating human thinking mode and building the ability to recognize things and distinguish things. This process is called in-depth learning. Deep learning has a multi-level learning framework, which makes many abstract transformations of input features. It has the characteristics of deep-seated, non-linear and layer-by-layer feature extraction. Therefore, it has a strong ability of feature expression and can solve more complex problems. Convolutional neural network is a commonly used deep learning algorithm. The difference between convolutional neural network and shallow neural network is that the feature extractor is composed of convolution layer and sub-sampling layer. The convolution layer is locally connected, so a neuron is only connected with several adjacent neurons. The main purpose is to reduce the amount of calculation. The convolution kernel values used in convolution neural networks are usually initialized in the form of decimal matrix, and the decimal matrix is generated randomly, but in the process of network training, it will gradually proceed to the direction of model

DOI: 10.25236/icmcs.2019.040

optimization, and then get a reasonable matrix value. The basic computational unit of in-depth learning is a single neuron. The composition of neurons is shown in the following figure:

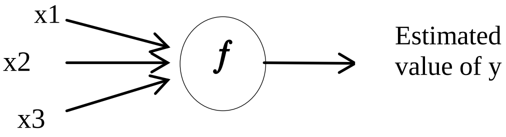

Figure. 1. Neuronal structure

The inputs of neurons are x1, x2, X3 and the outputs are g (x).

$$g(x) = \frac{1}{1+\exp(-\omega^T X)} \qquad (1)$$

We decompose formula (1-1) into formula (1-2) and formula (1-3), where formula (1-3) is the activation function Sigmoid function.

$$z = \omega^T x + b \qquad (2)$$

$$\sigma(z) = \frac{1}{1+e^{-z}} \qquad (3)$$

The commonly used activation functions are tanh (double tangent) function and ReLU (limit) function, as shown in formula (1-4) and formula (1-5).

$$\sigma(z) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \qquad (4)$$

$$\sigma(z) = \max(z, 0) \qquad (5)$$

Convolutional neural networks are mainly divided into two steps: first, bottom-up unsupervised learning, which inputs data sets into the structure of convolutional neural networks, so that labeled data can be trained layer by layer unsupervised. The output of the former layer is used as the input of the next layer, and the parameters of the former layer can be learned from the constraints of neural networks and sparsity to the data itself. Structures, so as to get more expressive features. Then, supervised learning is carried out from top to bottom, and the parameters are fine-tuned based on the previous step to achieve global optimization. The general convolution neural network is mainly composed of data layer, data normalization layer, core convolution layer, activation function, pooling layer, full connection layer, classifier and so on.The convolution layer mainly uses the characteristic matrix of the upper layer to convolute with a convolution core. After the convolution result is processed by activation function, the output of the convolution result forms the neuron of the layer. The input of each neuron is connected with the local feature of the upper layer to extract the local feature. Thus, the position relationship between the convolution and other features is determined. The convolution calculation is shown in Fig.2. The filter matrix is multiplied by the corresponding position of the input matrix.

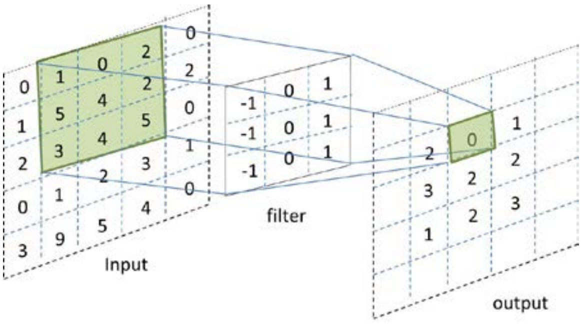

Figure. 2. Convolution

The pooling layer mainly compresses the characteristic matrix of convolution network. The calculation divides a 4 4 matrix into four groups, each with four numbers. The maximum or average value can be taken as the output, which is called maximum pooling and mean pooling respectively. Random enhances generalization, so we can pool randomly and increase diversity.

Full Connection Layer: After several convolution core pooling operations, the input signals are output as multiple sets of signals. After full connection operations, multiple sets of signals are combined into a group of signals in turn.

Model Design

KDD99 is an open source data set collected by MIT Laboratory for network security experiments, which collects a large number of network connection and system audit data. This paper uses this data set as training set and test set to design intrusion detection model. The structure of intrusion detection model proposed in this paper is divided into four parts: data preprocessing, parameter clustering, parameter selection and convolution neural network classification.

Data preprocessing.

Each record of KDD99 data set identifies a network connection, and records indicate that the network connection is normal or abnormal. It is represented by 41 eigenvalues. The input data of convolutional neural network requires floating point numbers between 0 and 1, so we need to preprocess the data.

Symbolic feature digitization.

The symbolic features of data sets mainly include protocol type, service type and flag bit Flag, which are selected by TCP, UDP and ICMP. We can use it to TCP (1,0,0), UDP (0,1,0), ICMP (0,0,1), respectively. There are 70 values for service type, corresponding to 70-dimensional vector, identifying bit has one value, corresponding to 11-dimensional vector, and finally extend 41-dimensional eigenvalue to 122-dimensional feature to complete symbols. Feature digitization.

Numerical Characteristics Normalization.

In order to eliminate the large gap of KDD99 data set characteristics, we adopt data normalization operation. Map the data between [0,1]. See formula (2-1).

$$Y_k = \frac{y - Y_{min}}{Y_{max} - Y_{min}} \tag{6}$$

Formula: Y is the processing of normalized data; Ymin is the smallest data in a dimension; Ymax is the largest data in a dimension.

parameter clustering algorithm based on dichotomous K-means algorithm.

To overcome the problem of convergence of K-means algorithm to local minimum, a bisecting K-means algorithm is proposed. Firstly, all points are divided as a cluster, and the choice of which cluster to divide depends on whether the partition can minimize the value of SSE. The partitioning process based on SSE is repeated until the number of clusters specified by users is obtaine.

- Consider all features as a cluster
- When the number of clusters is less than k
- For each cluster
- Calculating total error
- K-means clustering on a given cluster (k=2)
- Calculate the total error after dividing the cluster into two
- Select the cluster with the smallest error for partitioning operation

Parameter selection algorithm based on AdaBoost algorithm.

When making important decisions, you may consider drawing the opinions of many experts. This is the idea behind meta-algorithm. AdaBoost is the most popular meta-algorithm. AdaBoost will be applied to single-level decision tree classifier, and it will also show excellent performance in dealing with non-balanced classification problems.

The specific implementation steps are as follows:

For each iteration:

- Using buildstump () function to find the best single-level decision tree
- Add the best single-level decision tree to the single-level decision tree array
- Calculate alpha
- Calculate the new weight vector D
- Update cumulative category estimates
- If the error rate is equal to 0.0, exit the loop

Conclusion

Based on the advantages of convolution network in dealing with massive classification problems, this paper establishes a convolution neural network model in network intrusion detection data, which can overcome the challenges of huge network data and complex intrusion modes. The experimental results show that the convolutional neural network intrusion detection model proposed in this paper not only improves the detection rate, but also reduces the false alarm rate, which is basically consistent with the performance of the commonly used BP neural network and SVM. In the future, the model will be optimized, and other parameters affecting classification accuracy will be designed, implemented and evaluated.

Acknowledgment

Higher education teaching reform research project SJGY0213 in 2017; basic research business fee research project 2017-KYYWF-0353 in Heilongjiang Education Department; Backbone teachers plan in Heihe University; Heilongjiang Province youth talent plan UNPYSCT-2017104 in 2017.

References

[1] Semente: 2016 Internet Security Threat Report (ISTR), vol. 21, p.8, April 2016

[2] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision[J]. Computer Science, 2015:2818-2826.

[3] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning[J]. 2016.

[4] Taigman Y, Yang M, Ranzato M, et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014:1701-1708.

[5] Wen Y, Zhang K, Li Z, et al. A Discriminative Feature Learning Approach for Deep FaceRecognition[M]// Computer Vision – ECCV 2016. Springer International Publishing, 2016:499-515.

[6] Liu W, Wen Y, Yu Z, et al. SphereFace: Deep Hypersphere Embedding for Face Recognition[J]. 2017:6738-6746.

[7] Liu J, Deng Y, Bai T, et al. Targeting Ultimate Accuracy: Face Recognition via Deep Embedding[J]. 2015.