ATLANTIS PRESS

Advances in Computer Science Research, volume 87

3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019)

Research on Traffic Recognition Algorithms for Industrial Control Networks based on Deep Learning

Yixiang Jiang[1, a], Wenjuan Wang[2], Chengting Zhang[3, b]

[1]Bachelor of Ningbo University of Technology, Network Engineer, China Tobacco Zhejiang Industrial CO., LTD, NingBo 315000, China

[2]Computer application and enterprise informatization, Zhejiang Tobacco Industry Co., Ltd., Ningbo 315000, China

[3]Cloud computing and smart factory, Zhejiang Tobacco Industry Co., Ltd., Ningbo 315000, China

[a]jiangyxlunwen@sina.com, [b]titanbyron@126.com

Abstract. With the development of industrial control network and the deep integration of industry and information technology, the rapid development of industrial control system has increased dramatically, which has brought huge economic and property losses to industrial control companies. Therefore, a traffic identification technology based on deep learning is proposed, which makes full use of the characteristics of industrial network traffic signs. Combined with experiments, this technology can classify network traffic and effectively identify abnormal traffic in industrial control system network. Compared with traditional classification methods, it not only improves the accuracy of traffic identification, but also reduces the time required for classification.

Keywords: Deep learning; Industrial control system; Traffic identification.

1. Introduction

With the arrival of the era of big data, how to obtain more valuable information from the growing traffic data has become one of the indispensable indicators for major operators to make development plans. As one of the key technologies of network management and network security, [1] network traffic identification can not only optimize network configuration and reduce network security risks, but also provide better quality of service according to user behavior analysis.

Shallow learning includes support vector machine, decision tree, Bayesian and k-means. At present, there are methods to realize the final traffic identification by manual selection and combination of features, but this increases the workload of network traffic identification. In addition, the WLMF is used to extract multifractal features from network traffic to describe network traffic. Then, feature selection method based on principal component analysis (PCA) is applied to these multifractal features to eliminate irrelevant and redundant features. The experimental results show that the classification accuracy of Support Vector Machine (SVM) is significantly improved compared with the existing data features of transport layer based on machine learning. The mainstream still uses PCA dimensionality reduction method to reduce the characteristics of network traffic data and realize network traffic identification. In recent years, in-depth learning has been applied to various fields, such as image recognition, speech recognition, audio processing and natural speech processing. Good results have been achieved. The deep learning classification method is static classification processing and needs to be selected manually according to the current sample, that is to say, the training can only complete the number of nodes and parameters at one time. Faced with the dynamic batch sorting process, a lot of time is wasted to update parameters and samples. [2] Aiming at the low accuracy of traffic identification in industrial networks, a deep learning structure and BP neural network traffic identification method are proposed.

To construct appropriate feature space for industrial control network traffic, establish network traffic identification model, select the number of hidden nodes and hidden layers in the model. Improve the identification accuracy of industrial network traffic model. The success of the experiment proves the feasibility of deep learning in recognition. Compared with classical machine learning methods, when the training data changes, the classification model based on deep learning can be fine-tuned based on its original optimal model. The classification model based on machine learning needs

to re-examine the characteristics and training of the changed training data in order to complete the flow recognition of new data. In view of the problems in the above research, the BP neural network proposed in this paper can implicitly learn and extract features from training data. It not only avoids the trouble of selecting artificial features, but also solves the difference of feature selection between different classification algorithms. It also improves the accuracy of flow identification.

2. Research Background

2.1 Flow Identification and Description

The overall flow of network traffic identification includes the collection of network data, the generation of data sets with accurate background information, the preprocessing of data sets, the extraction and classification of traffic characteristics. In order to verify the feasibility of the proposed algorithm, this paper uses existing data sets and traffic data sets collected through Microsoft Network Monitor to carry out experiments. According to different search strategies, the original feature set is generated into a specific feature subset, and then the feature subset is evaluated according to a certain convention. Finally, the optimal feature subset is obtained. This method requires not only extra computational overhead, but also the selection of appropriate features according to the characteristics of classification applications. Aiming at the problems of the above methods, this paper uses BP neural network to autonomously learn the features of the original data set, constructs the feature space through the deep structure of multiple hidden layers, and discovers the feature representation of a large number of data by autonomous learning to construct the appropriate feature space. This method not only solves the difficulty of feature subset selection, but also improves the efficiency of classification, which lays a foundation for real-time classification of network traffic. [3]

The standard BP algorithm is a learning algorithm based on Gradient Descent (GD). The learning process is to adjust the weights and thresholds dynamically by making the mean square error of the expected output value and the actual output value of the neural network tend to zero. However, it only uses the information of the first derivative of the mean square error function to weights and thresholds, which makes the algorithm slow in convergence and easy to fall into local minimum.

The training process of the algorithm is mainly divided into the following parts: (1) initialization of network weights, (2) forward propagation of signals in the network, (3) back propagation of errors in the network and (3) iteration until the algorithm converges.

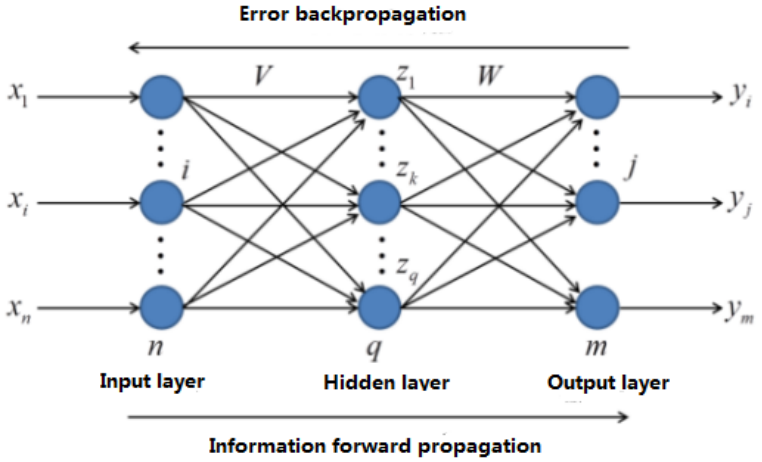


Fig. 1 Algorithm model

2.1.1 Initialization of Network Weights.

Before network training begins, the connection weights of all neurons in the network are initialized, and the ownership weights are assigned randomly or according to some statistical rules, such as assigning the connection weights to a random number in an interval (-1,1).

2.1.2 The Signal Propagates Forward in the Network.

The training samples are input in the input layer of the network, and the signals propagate backward through the input layer and the hidden layer. The output of each layer is calculated until the last output layer, and the output of the whole network is obtained. The operation of each neuron is to weigh all the inputs of the previous layer, and then obtain the output of the neuron through a non-linear activation function (usually a Sigmoid function).

2.1.3 The Error Propagates back in the Network.

Firstly, the error of the output layer is calculated according to the actual output of the network in the forward propagation stage and the ideal output (given expected value) y in the training data set (where the sum of squares of the difference between the actual output and the ideal output is an error function):

$$E = \frac{1}{2}\sum_{i=1}^{m}(y-\overline{y})^2$$

Then the original problem is transformed into the problem of minimizing the error function. The gradient descent method is used to solve the optimization problem. The weight gradient is expressed as:

$$\begin{aligned}\nabla w_i &= \frac{\partial E}{\partial w_i} = \frac{\partial(\frac{1}{2}(y-\overline{y})^2)}{\partial w_i} \\ &= (y-\overline{y})\cdot(-1)\cdot f^{'}(w_1 z_1 + w_2 z_2 + ... + w_q z_q) \\ &= -(y-\overline{y})\cdot f^{'}(w_1 z_1 + w_2 z_2 + ... + w_q z_q)\end{aligned}$$

$$\begin{aligned}\nabla w_{ij} &= \frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial h_j}\cdot\frac{\partial z_j}{\partial w_{ij}} \\ &= [(y-\overline{y})\cdot(-1)\cdot f^{'}(w_1 z_1 + w_2 z_2 + ... + w_q z_q)\cdot w_j] \\ &\quad \cdot f^{'}(w_{1j} x_1 + w_{2j} x_2 + ... + w_{ij} x_i)\end{aligned}$$

The weight updating formula is as follows:

$$w_i = w_i - \eta_1 \nabla w_i$$

$$w_{ij} = w_{ij} - \eta_2 \nabla w_{ij}$$

Among them, the calculation of the partial derivative of the output layer error to the weight of the input layer to the hidden layer is based on the chain derivative rule. First, the partial derivative of the output error to the output of the hidden layer is obtained, and then the partial derivative of the output of the hidden layer to the weight of the input layer to the hidden layer is obtained, as shown in the above formula.

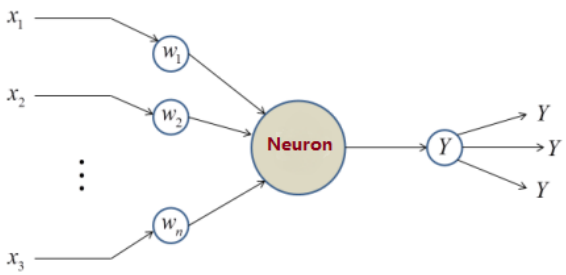


Fig. 2 Neuron model

For the first neuron in the graph, its network input can be expressed as:

$$X_i = w_{i1}x_1 + w_{i2}x_2 + \cdots + w_{in}x_n$$

In the above formula, the input signal received by the neuron is; the corresponding connection weight of the neuron is. The output of the neuron must be non-linearly activated. Generally, the activation functions include sigmoid function, hyperbolic tangent function and symbolic function. In order to ensure that the activation function can be differentiated everywhere, this paper takes sigmoid function (also known as S-type function) as an example:

$$y = f(X) = \frac{1}{1+\exp(-X)}$$

When propagating in reverse direction, the derivative operation of the upper form is needed. The derivative of the upper form is:

$$y' = \frac{1}{1+\exp(-X)} - \frac{1}{[1+\exp(-X)]^2} = y(1-y)$$

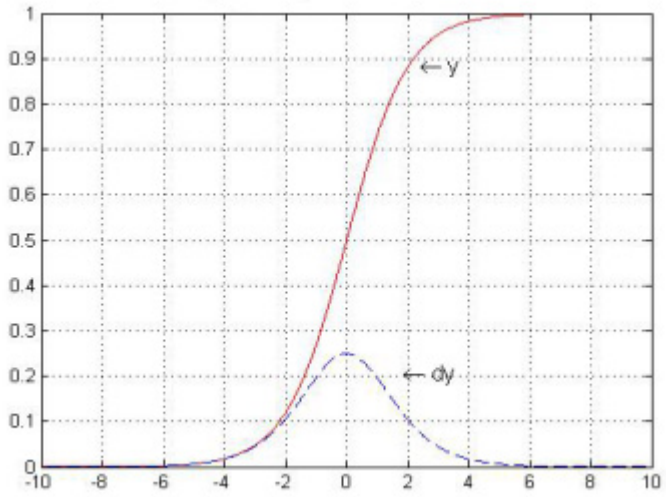


Fig. 3 Sigmoid and its derivative curve

As can be seen from the figure above, when X = 0, y is 0.5, the maximum value is 0.25, and X is sitting in the interval (-4,4), the change rate of Y is larger, but outside (-6,6), the change rate of Y is very small, tending to zero.

3. Experimental Results and Performance Analysis

3.1 Network Traffic Recognition Method based on BP Neural Network

Marking data sets is the key factor to ensure the effective use of BP neural network. We adopt and use the actual traffic data set. Its collection and labeling work are as follows.

3.1.1 Moore Data Set

Moore data sets are connected to the Internet by thousands of users over a Gigabit full-duplex Ethernet link, using network monitors to collect 24-hour network traffic. Then, using the sampling algorithm, a large number of network samples are obtained in 24-hour traffic, and classified as untyped according to the application type. Each network flow sample in the Moore data set is abstracted from a complete TCP bidirectional flow containing 249 attributes, and the last attribute is the category corresponding to each network flow. The statistical results are as follows. [4]

Table 1. Statistical results

category	Number	Proportion
WWW	327931	88.51%
FTP-DATA	3601	0.95%
ATTACK	30106	8.33%
INTERACTIVE	862	0.23%
SERVICES	1896	0.49%
MULTIMEDIA	3206	0.87%
SUM	370502	100.00%

In order to verify the validity of this method, two sets of data are obtained. The second group of data was collected from 20 September 2017 to 26 September 2017. Set them as data set 1 and data set 2 respectively. Fifty session flows are randomly selected from data set 1 as training samples to train the neural network, in which the number and weight of hidden layer neurons are determined by algorithm 1. The trained neural network is tested on dataset 2 and repeated 10 times, and different subsets of dataset 1 are used as training sets. Ten recognition rates were obtained, and the average and standard deviation of 10 recognition rates were calculated. For some traffic in Table 2, including TCP and UDP protocols, such as file sharing and audio and video in P2 P. Therefore, the method of calculating this type of traffic identification rate is as follows: the traffic ratio of TCP protocol is set to tcp_p, and that of UDP protocol is ud_p. The recognition rate of TCP session flow is tcp_r, and that of UDP session flow is udp_r. The overall recognition rate for this type of traffic is: Rate = tcp_p x tcp_r + udp_p x udp_r,tcp_p + udp_p = 1.

3.2 Determination of t Value in UDP Session Flow

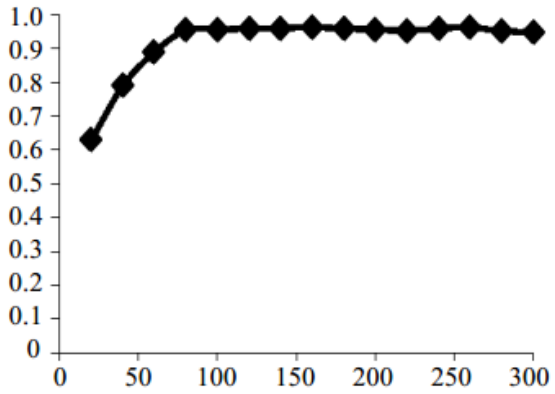
When UDP session flow is defined in Definition 2, it is the reorganization of data packets sent between the corresponding ports in t period. The determination of T value is related to protocol recognition rate and computational performance. If the value of T is too small, the information with statistical characteristics cannot be formed because of too few data packets obtained; if the value of T is too large, the calculation performance will be greatly affected. T is tested from 20 to 300 seconds, and the recognition rate of UDP session flow is shown in Figure 4.

After testing, the recognition rate tends to be stable when t is greater than 80 s, so the time t in the UDP session flow is finally chosen to be 80 s.

3.3 Contrast of Convergence Speed of Algorithms

When the number of hidden layer neurons is determined, BP algorithm is faster in training samples. The overall training accuracy is 1x10_4. The comparison of error and iteration times in training TCP and UDP protocols is shown in Fig. 5, respectively.

From the graph, BP algorithm has better convergence speed than traditional algorithm. The main reason is that the training accuracy of BP algorithm can converge faster, and then the convergence speed of traditional machine learning algorithm is improved when the initial weights have been better.



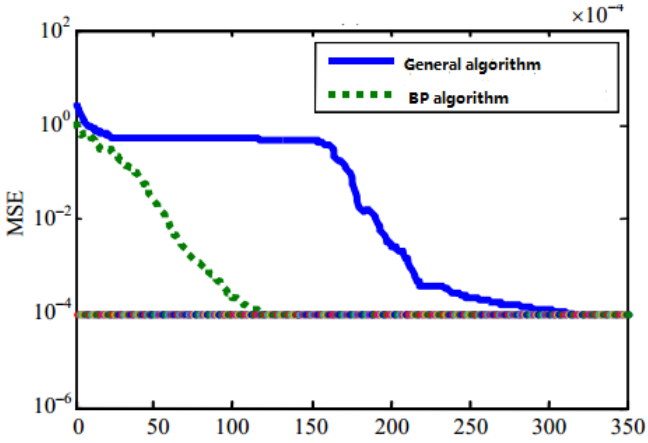


Fig. 4 Recognition rate of UDP session flow Fig.5 Contrast of Convergence Speed

3.4 Improve Recognition Rate

According to the method proposed in this paper, 50% of the samples in data set 1 are randomly selected for training, and the samples in data set 2 are tested. By measuring, when the number of hidden neurons in the neural network classifying TCP and UDP protocols is 17 and 23 respectively, the average recognition rate of each protocol is the highest. The recognition rates of eight common protocols in Table 2 are shown in Table 2.

Table 2. Recognition rates

category	Recognition rate/ (%)	Proportion
WWW	99.1±0.5	88.51%
MAIL	95.2±0.3	7.71%
FTP-DATA	96.7±0.4	0.95%
ATTACK	98.1±0.5	0.62%
INTERACTIVE	92.1±0.6	0.23%
SERVICES	92.1±0.4	0.49%
MULTIMEDIA	96.1±0.4	0.31%
SUM	42670.2	100.00%

4. Summary

The application of industrial control system has become popular, and the industrial control network has gradually begun to integrate with the Internet, making the vulnerability of industrial control system is gradually emerging. [5]

Aiming at the characteristics of industrial control network data and combining with various machine learning algorithms, this paper designs a network recognition technology based on deep learning, which can effectively detect different traffic types in industrial control network, and thus detect and defend against intrusion attacks in time. Future network attack variants will be more deceptive. More innovative research and more detailed and perfect work are needed on the issue of accuracy.

References

[1]. Gu Juntao. Research Status of Industrial Control System Security [J] Information Technology 2016, (7) 135-137.

[2]. Peng Yong, Jiang Changqing, Xie Feng, etc. Research on Information Security of Industrial Control Systems Progress [J] Journal of Tsinghua University (Natural Science Edition), 2012 (10):1396-1408.

[3]. Yang Chen, Ma Qin. Weaving Safety Net for Industrial Control System [J] Information Security and Security Communications Secrecy 2014, (6:36).

[4]. Jolliffe I. Principal component analysis[M]. John Wiley &Sons, Ltd, 2002.

[5]. Van Gestel T, De Brabanter J, De Moor B, et al. Least squares support vector machines [M]. Singapore: World Scientific, 2002.