Network Security State Identification Based on Neural Network Optimized by Ant Colony Algorithm

Chenxiang Zhang[1*]
[1]SuZhou Industrial Park Institute of Services Outsourcing, Suzhou 215123 China
**Corresponding Author.*

Abstract

Intrusion detection is the key technology to ensure network security. In order to solve the problem of parameter optimization in the application of neural network in intrusion detection, a network intrusion detection model based on ant colony algorithm is proposed in this paper. Firstly, this paper describes the relationship between ant colony algorithm and neural network parameters, and establishes the objective function of neural network parameter selection. Then the ant colony algorithm is used to search the optimal solution of the objective function and determine the optimal parameters of the neural network. Finally, this paper realizes the construction of intrusion detection classifier through neural network self-organizing learning. The results show that the model solves the problem of parameter optimization of neural network in intrusion detection. The classification results and classification speed have significant advantages over the typical model.

Keywords*: deep learning, massive data, behavior prediction, normalized coding algorithm*

I. Introduction

Ant system (ant system or ant colony system) was first proposed by Italian scholars Dorigo, ,Maniezzo and others in the 1990s [1-2]. In the process of studying ant foraging, they found that the behavior of a single ant is relatively simple, but the ant colony as a whole can reflect some intelligent behavior. For example, ant colony can find the shortest path to the food source in different environments. This is because the ants in the ant colony can transmit information through some information mechanism. After further research, it is found that ants will release a substance called "pheromone" on their path. Ants in the ant colony have the ability to perceive the "pheromone" [3-5]. They will walk along the path with high concentration of "pheromone", and each passing ant will leave "pheromone" on the road, which forms a mechanism similar to positive feedback, In this way, after a period of time, the whole ant colony will reach the food source along the shortest path.

The basic idea of applying ant colony algorithm to solve the optimization problem is: the walking path of ants is used to represent the feasible solution of the problem to be optimized, and all paths of the whole ant colony constitute the solution space of the problem to be optimized. The amount of pheromone released by ants with short path is more [6]. With the advance of time, the pheromone concentration accumulated on the short path gradually increases, and the number of ants choosing the path is also more and more. Finally, the whole ant will focus on the best path under the action of positive feedback. At this time, the corresponding is the optimal solution of the problem to be optimized [7].

Digitization is an important trend in today's world, and its advantages are beyond doubt: intelligent communication, intelligent measurement and other digital technologies have greatly improved the efficiency and reliability of the energy system [8-10]. But digitization also opens the door for the emergence of network security threats. Nowadays, digital equipment has penetrated into every field and link of the energy industry, especially the wide application of industrial automatic control system, and has gradually become the control center and core of the energy system, which may bring huge potential security risks. Once hackers invade and control these devices, in theory, they can control the energy system and "do whatever they want", such as opening and closing all kinds of switches and valves at will, changing the operation state of the equipment, adjusting the setting of the early warning system, etc., thus leading to the interruption of energy supply or physical damage such as explosion and fire.

II. The Model of Algorithm

Ants find the shortest path thanks to pheromones and the environment. Suppose there are two roads from the ant nest to food [11]. At the beginning, the number of ants on the two roads is almost the same: when the ants reach the end point, they will return immediately. Ants on the short distance road have a short round-trip time and a fast repetition rate. There are more ants and more pheromones in unit time, It will attract more ants and leave more pheromones [12-13]. The long distance is just the opposite, so more and more ants gather on the shortest path.

The intelligent behavior of ants benefits from its simple behavior rules, which make them have diversity and positive feedback. When foraging, diversity makes ants not go into a dead end and cycle indefinitely, which is an innovative ability; Positive feedback keeps good information, which is a kind of learning ability. The clever combination of the two makes intelligent behaviors emerge. If the diversity is excessive and the system is too active, it will lead to too much random motion and fall into chaos; If the diversity is not enough and the positive feedback is too strong, it will lead to rigidity. When the environment changes, the ant colony cannot adjust accordingly.

2.1 Deep Learning Method

Figure 1 shows the popularity trend of the word "deep learning" in Google search in the last decade.

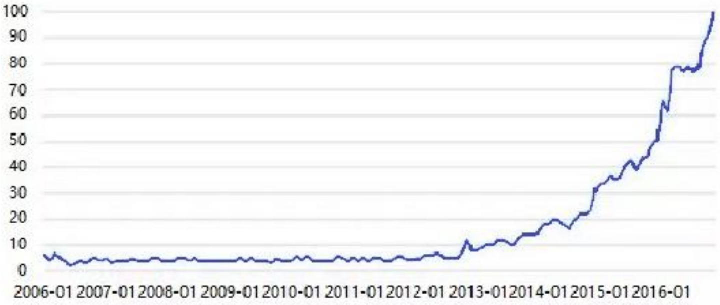


Fig. 1 The popularity trend of "deep learning" in Google search in recent ten years.

As can be seen from the figure, the popularity of deep learning has increased exponentially since 2012. By 2016, deep learning has become the most popular search term on Google. Deep learning is not a new word, it is basically synonymous with deep neural network [14-15]. Inspired by the structure of human brain, the computational model of neural network was first proposed in 1943. After the invention of perceptron, neural network becomes a model that can "learn" from data. But because the structure of the sensor network is too simple, it can not solve the problem of linear indivisibility. In addition, the amount of computation required by the neural network is too large, the computer at that time can not meet the needs of computation, making the research of neural network into the first winter. By the 1980s, the deep neural network and back propagation algorithm have solved these problems well, and let neural network enter the second period of rapid development.

In 2006, Hinton et al. Put forward unsupervised greedy layer by layer training algorithm to solve the optimization problems related to deep structure, and the concept of deep learning was proposed [9]. The CIFAR affiliate team said the same strategy could be used to train other types of deep networks and help systematically improve generalization on test samples. In addition, Lecun established the first "real" convolutional neural network model with multi-layer structure in Bell laboratory [10]. With the rapid development of Internet industry, the computer software and hardware infrastructure for deep learning has been improved, the amount of available training data is increasing, and the scale of deep learning model also increases.

2.2 TensorFlow Deep Learning Framework

TensorFlow is a distributed machine learning platform, and its main architecture is shown in Figure 2. RPC and RDMA are network layers, which are mainly responsible for transmitting neural network algorithm parameters. CPU and GPU are the device layers, which are mainly responsible for the specific operation in neural network algorithm. Kernel is the concrete implementation of algorithm operation in TensorFlow, such as convolution operation and activation operation. Distributed Master is used to build subgraph and cut subgraph into multiple slices. Different subgraph slices run on different devices. Master is responsible for distributing subgraph slices to the Executor/Work. Executor/Work is on the equipment (CPUs, GPUs, etc.), and is responsible for sending and receiving the running results of graph operation to other Worker. C API divides TensorFlow into front end and back end, and the front end (Python/C++/Java Client) triggers TensorFlow back end program to run based on C API. Training libraries and Inference libs are library functions for model training and derivation, which are used by users to develop application models.

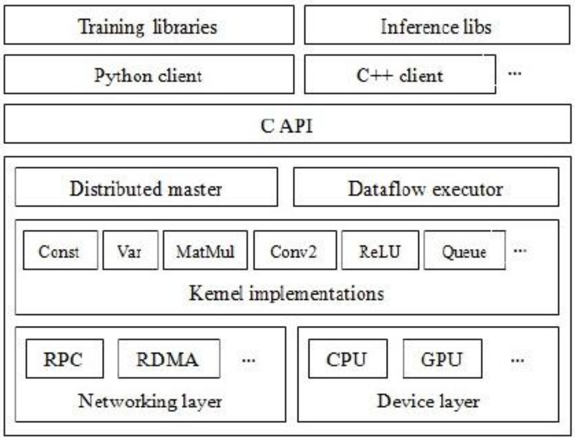


Fig. 2 TensorFlow system architecture.

III. Deep learning algorithm and experimental analysis of intrusion detection

3.1 Learning Algorithm

1) Several common classification algorithms

Classification is a kind of data mining. Classification is to predict the value of a specific attribute according to the value of other attributes. The value of a specific attribute determines that it belongs to one of several categories. In other words, classification belongs to the prediction task, which is to get an objective function f through the learning of existing data sets, and map each attribute set x to the objective attribute y, and y must be discrete. It is difficult to give the rules of classification algorithm directly by programming, but it is easy to get them by learning algorithm. The classification process first needs to process the actual data into data that can be understood by computer (data preprocessing), generally in the form of table. If there are too many features in the learning data, we may need to select the most representative features from the feature set. Feature selection can reduce the training time, improve the performance of the learning algorithm, and avoid the dimension disaster problem.Classification algorithms are based on mathematics, through a variety of methods to analyze the data and make predictions. Here are several common classification algorithms:

(1) Naive Bayesian model (NB)

Bayesian classification is a classification method based on Bayesian theorem and independent assumption of feature conditions. It classifies by calculating the probability that a given tuple belongs to a specific class. The advantage of this method is that it is not sensitive to missing data and needs less parameters to be estimated. The disadvantage is that all attributes are required to be independent of each other (this requirement is difficult to meet in practical problems), and need to know the prior probability.

(2) Decision tree model (DT)

Decision tree is a widely used classifier. It uses training data to construct decision tree, and then realizes data classification through decision tree. The decision tree has a tree structure. Each internal node is used to test an attribute. The branch of each node represents the test result. Each leaf node of the tree represents a classification category. The decision tree starts from the root node, tests step by step, selects step by step, and gets the classification results when it reaches the leaf node. The advantage of this method is that it doesn't need any domain knowledge or parameter assumption, and is suitable for high-dimensional data, processing a large number of data in a short time, and can get better results. The disadvantage is that it is easy to over fit and does not support online learning. Some data have different number of samples for each category, and the information gain tends to those features with more values.

(3) K-nearest neighbor algorithm (KNN)

KNN is one of the simplest classification algorithms. The idea of the algorithm is to find the nearest K samples in the sample space. If most of the K samples belong to a certain class, then the samples also belong to this class. The algorithm is simple, easy to understand, easy to implement, without training. The disadvantage is that when the samples are unbalanced (the number of samples belonging to some classes is much more than other classes), it is easy to cause judgment bias.

2) Softmax regression learning method

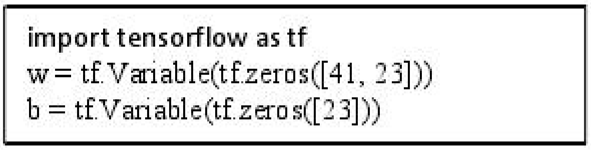
Softmax regression, namely multiple Logistic regression, is a commonly used multi-class classifier. In this summary, we will learn and classify intrusion detection data based on Tensorflow using Softmax regression method, and analyze its classification effect. Softmax function is a normalized exponential function and is defined as follows:

$$y_c = \varphi\left(Z\right)_c = \frac{e^x c}{\sum_{d-1}^{c} e^z d} \quad (1)$$

Among them, φ Represents the softmax function. The input z is a C-dimensional vector and the output y is also a c-dimensional vector. The denominator in the formula acts as a regular term, which makes:

$$\sum_{j=1}^{C} y_j = 1 \quad (2)$$

Tensorflow provides an embedded softmax implementation function. In order to construct the softmax classification model, it is necessary to establish the full connection between the input vector and the output category, and train the weight of each connection and the bias vector of the classification. For this reason, we need to define the weight matrix and bias term vector and give the initial value. Figure 3 is a python program that defines the weight matrix and the offset term vector.



```
import tensorflow as tf
w = tf.Variable(tf.zeros([41, 23]))
b = tf.Variable(tf.zeros([23]))
```

Fig. 3 Definition of weights and offsets.

3.2 Feature Learning Algorithm Based on Depth Structure

Convolutional neural network is a deep learning structure inspired by visual perception mechanism. It uses multi-layer network model to extract the features of things, and then classifies, recognizes, predicts or makes decisions according to the features. It has a wide range of applications, including image classification, target detection, target recognition, target tracking, text detection and recognition, position estimation and so on. The basic structure of CNN generally includes input layer, convolution layer, pooling layer and full connection layer. Each node of convolution layer is connected with a region of the upper layer by convolution kernel. In the same convolution layer, the weights of all neurons are the same. The pooling layer is sandwiched in the middle of the convolution layer, and its main function is to gradually compress, reduce the number of data and parameters, and also reduce the over fitting phenomenon to a certain extent, and compress a certain area of the input data of the upper layer into a value. The full join layer is mainly used for learning, mapping the learned feature representation to the sample label space. Based on the analysis of CNN and test data set, we designed CNN training model, as shown in Figure 4.

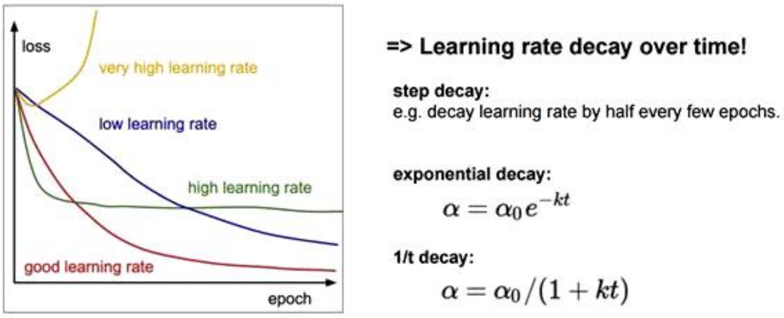


Fig. 4 CNN training model.

In many deep learning network models, sparse self encoder is one of the effective algorithms for feature extraction. The sparse self coding model considers that the input data can be transformed into a weighted representation of a group of bases. For example, the integer can be expressed as a weighted representation of the group of bases t = [number, ten, hundred, thousand, ten thousand, one hundred thousand...], that is, any integer k can be expressed as:

$$k = \sum x_i \times T_i \quad (3)$$

Where x is the weighted vector.

SAE, a neural network with multiple hidden layers, is an unsupervised learning method, which uses back propagation algorithm. The idea is to make the output equal to the input, and let the encoder find the hidden features (that is, the set of bases) in the input data. SAE is generally divided into coding process and decoding process. The decoding process is the reverse process of coding process, but it does not require that the decoding weight is the same as the coding weight, but is approaching through learning. The encoding process is to find t and express the input k as ∑xi × Ti, while the decoding process is to express ∑xi × Ti as k again. Learn from the errors of output and input in every encoding and decoding process.

SAE requires that the output is equal to the input, which is different from the requirement of intrusion detection (the input of intrusion detection is 41-dimensional network data, and the output is 23 categories represented by 23-dimensional vector). Therefore, we changed the last decoding of SAE model into mapping to 23-dimensional vector, compared the mapping result with the actual classification result, and used its error to learn. In this way, unsupervised learning methods have become supervised learning methods.

Compared with other optimization algorithms, ant colony algorithm has the following characteristics:

(1) The positive feedback mechanism is adopted to make the search process converge and finally approach the optimal solution.

(2) Each individual can change the surrounding environment by releasing pheromones, and each individual can perceive the real-time changes of the surrounding environment, and individuals communicate indirectly through the environment.

(3) The search process adopts distributed computing, and multiple individuals perform parallel computing at the same time, which greatly improves the computing power and operation efficiency of the algorithm.

(4) The heuristic probability search method is not easy to fall into local optimization, and it is easy to find the global optimal solution.

IV. Conclusion

In recent years, with the rapid development of Internet technology, network security has become an important issue that must be paid attention to in all fields, especially in the fields of industrial control, intelligent technology, mobile payment and cloud computing. At the same time, hackers and network terrorist organizations and other groups launched a variety of network attacks are more and more influential and destructive, China's network security situation is more and more severe. With the maturity and popularization of deep learning technology, especially the excellent performance in the field of feature learning, we have found a breakthrough. In this paper, the research of deep learning oriented to network security detection is taken as the exploration direction, focusing on the intrusion detection algorithm based on deep learning, the traditional classification detection algorithm is analyzed and improved, and its implementation in tensorflow machine learning platform is explored.

Acknowledgment

The paper is supported by research results on the objects of “Cyan Engineering” supported by Jiangsu Universities in 2019 (No. JSQL201901).

References

[1] Jiang Wei, Fang Binxing, Tian Zhihong. Network Security Evaluation and Optimal Active Defense Based on Attack Defense Game Model. Acta Computer Sinica, 2009, 32 (004): 817-827
[2] Miao Yongqing. Stochastic Model Method and Evaluation Technology of Network Security. China Science and Technology Investment, 2017, 4: 314
[3] Yi Hua Zhou, Wei Min Shi, Wei Ma. Research on Computer Network Security Teaching Mode for Postgraduates Under the Background of New Engineering. Innovation and Practice of Teaching

Methods, 2020, 3 (14): 169

[4] Bao Xiuguo, Hu Mingzeng, Zhang Hongli. Two Quantitative Analysis Methods for Survivability of Network Security Management Systems. Acta Communication Sinica, 2004, 25 (9): 34-41

[5] Yang Yi, Bian Yuan, Zhang Tianqiao. Network Security Situation Awareness Based on Machine Learning. Computer Science and Application, 2020, 10 (12): 8

[6] Li Zhiyong. Hierarchical Network Security Threat Situation Quantitative Assessment Method. Communication World, 2016, 23: 70-70

[7] Hu Wenji, Xu Mingwei. Analysis of Secure Routing Protocols for Wireless Sensor Networks. Journal of Beijing University of Posts and Telecommunications, 2006, 29 (s1): 107-111

[8] Wei Yonglian, Yi Feng, Feng Dengguo, Yong W, Yifeng L. Network Security Situation Assessment Model Based on Information Fusion. Computer Research and Development, 2009, 46 (3): 353-362

[9] Gong Xiayi, Li Bohu, Chai Xudong, et al. Overview of big data platform technology, Journal of System Simulation, 2014,26 (3): 489-496

[10] Zhou Lei, Huang Haitao, Huang Lin, et al. Research on Guizhou transportation assistant decision analysis system based on big data technology. China Transportation Informatization, 2018 (S1): 31-35.

[11] Sun Xuan, Sun Tao. Urban visual governance decision support model and application based on big data. Journal of Public Management, 2018 (2): 120-129158-159

[12] Guan Xin, Shao Chang'an. Process model construction and data problem analysis of network big data application. Library and Information Work, 2017 (5): 50-56

[13] Hong Zhixu, Chen Hao, Cheng Liang. Data integration and decision analysis method of social governance based on big data. Journal of Tsinghua University (NATURAL SCIENCE EDITION), 2017 (3): 1264-269.

[14] Cheng Lin, Zhu Xiaofeng, Lu Jingyun. Research on sharing logistics information platform model based on big data. Science and Technology Management Research, 2018 (15): 234-238

[15] Xu man, Shen Jiang, Yu Haiyan. Review of data driven medical and health decision support. Industrial Engineering and Management, 2017 (1): 1-13