
Network Traffic Recognition Technology based on Machine Learning

Dongqi Wang¹, Xuanyue Shuai²

¹ School of Control Science and Engineering, Dalian University of Technology Dalian, Liaoning 116024, China.

² School of Electrical Engineering, Dalian University of Technology Dalian, Liaoning 116024, China.

Abstract

In recent years, the rapid development of computer network technology, the arrival of the information age, the increasing use of the network, which makes the data flow of the Internet show an explosive growth; at the same time, the new application of the network continues to appear, which makes the use of network communication protocols more flexible and mixed; network communication protocol is more flexible and mixed; Viruses, eavesdropping and malicious attacks are increasing. The security of the network has become the focus of attention in the society. For these problems, network traffic identification can be well resolved. At present, there are different kinds of traffic identification methods, but from the perspective of research and application, the feasibility and effectiveness of traffic identification is the most important concern of people, that is, how to deal with massive data quickly and how to correctly recognize various applications in the network. In the face of changing network environment, this paper mainly expounds the method of network traffic identification based on machine learning (BP neural network), which aims to provide effective protection for network security.

Keywords

Machine Learning, network traffic identification, computer network, BP neural network.

1. Research Status of Network Traffic Identification

Today, we have entered the era of the rapid development of computer network technology, followed by the increasing number of network users, which makes the network data flow rapidly increasing. In this situation, people have emerged a new focus of attention: with the continuous emergence of new network applications in people's life and work, network communication protocols are also being used constantly, and the flexibility and confounding of use are constantly enhanced; meanwhile, many behaviors such as malicious attack, network virus and eavesdropping are beginning to appear. The increasing situation leads to the security of the network. This problem is a new pressure and challenge for network managers and service providers. The network traffic identification technology can improve the control ability of the telecom operators, and also play an important role in the service provider's support to the network support, so that it can provide users with reliable, high quality and safe services.

Based on the above situation, more and more people pay attention to network traffic recognition technology. At the same time, the technology of other fields is also constantly integrated with the flow recognition technology, including the following aspects: artificial neural network, machine learning, pattern recognition, data mining and so on, which further improves the feasibility and effectiveness of flow recognition.

In recent years, many academic articles on the direction of network traffic identification have begun to emerge; and international well-known communications equipment manufacturers are also committed to the launch of network traffic identification equipment, which have independent intellectual property rights. This is the highlight of the major manufacturers, including Ericsson, Qualcomm, and HUAWEI.

2. Development Status of Network Traffic Identification Technology

The network traffic identification technology is constantly updating and developing. In this process, according to the continuous changes in the network environment, the recognition algorithm and the extracted flow characteristics are constantly improving. At present, in the course of development, network traffic identification can be divided into four aspects: flow recognition based on port number mapping, flow recognition based on net load characteristics, traffic identification based on network behavior characteristics and [1] based traffic recognition method based on machine learning.

Through many analysis, the best method of extensibility is traffic recognition based on machine learning, which has the characteristics of mining the hidden characteristics of network traffic, and can identify the encrypted network flow accurately, and also has the characteristics of discovering new network applications.

3. Machine Learning Foundation

Machine learning integrates a variety of disciplines, with the advantages of various disciplines, by collecting new information from the environment, using the computer to imitate human cognitive ability, and updating the existing learning knowledge system, and constantly improving the cognitive ability of the new things [2,3]. In the field of machine learning, how to use the computer program system to work on the basis of accumulating knowledge and improving the ability of information processing, this is the difficult problem we face at present, [4]. With the advent of machine learning, artificial intelligence has been widely used in various fields in various fields.

According to different learning forms, machine learning can be divided into three categories: supervised learning, unsupervised learning and semi supervised learning.

The machine learning system is an intelligent system [5] that can continuously collect new information from the environment and update the learned knowledge system to improve the cognitive ability of new things.

In the whole machine learning system, the system is "closed loop", and develops continuously and perfected through the promotion of all elements.

Network traffic identification technology of 4. BP neural network

For the identification technology of network traffic, this paper chooses the BP neural network based on machine learning method to describe network traffic identification in detail. In the process of practice, the BP algorithm is trained to sample data, and the classification model is constructed. Finally, the recognition results are analyzed. Considering the feasibility and effectiveness of recognition, the key point of network traffic identification is that the selected algorithm must have a low computational complexity. It should be efficient and fast for the large amount of network traffic, and must have good nonlinear partitioning ability. It can be correctly recognized in a short time for the complex and multi class network data. No, no.

The advantages of machine learning BP neural network for network traffic identification are as follows:

- (1) The training mode of BP neural network is unique, and it is trained through global optimization. This makes the generalization ability of BP network higher.
- (2) The learning way of BP neural network is distributed and parallel network structure, which promotes its high fault tolerance and faster processing speed.
- (3) The unique simulation of BP neural network can simulate the nonlinear relationship between input and output very well, and improve the ability of nonlinear mapping.

(4) BP neural network has simple structure and low computational complexity and memory requirements.

The BP algorithm based on machine learning has the advantages of strong error convergence, minimum global, high security and fast speed. It is a good technique for monitoring network traffic and is highly operable in life. It is suitable to be widely used in working life.

References

- [1] Zhao Guofeng, Ji Chao Ming, Xu Chuan. Internet traffic identification technology research [J]. small microcomputer system, 2010, 31 (8): 1514-1520.
- [2] Z. Shi. Principles of Machine Learning[M]. International Academic Publishers, 1992.
- [3] H. Simon, R. S. Michalski, J. G. Carbonell, et al. et, Simon 1983.
- [4] I. Witten, E. Frank. Data Mining: Practical Machine Learning Tools and Techniques.
- [5] Wang Jue, Zhou Zhihua, Zhou Aoying. Machine learning and its application [M]. Beijing: Tsinghua University press, 2006.