Research on the Application of Deep Learning in Computer Network Information Security

Han Zeng¹, zixi Liu² Honghao Cai³

¹College of Information Engineering & Art Design, Zhejiang University of Water Resources and Electric Power, Hangzhou, Zhejiang, 310018
²School of Physical Science and Technology, Southwest Jiaotong University, Chengdu, Sichuan, 610000
³Southwest University of Science And Technology, School of Computer Science and Technology, Mianyang, Sichuan, 621010

Abstract. With the development of current information technology, computer networks are playing an increasingly important role in people's daily work and life. While the network brings great convenience to people, it also puts people at risk. A large number of information is stored on the network, which may be illegally invaded at any time, and there are serious security risks. Therefore, research on computer network information security is of great significance. Based on this, this article analyzes the application of deep learning in computer network information security. First, it gives an overview of network information security, introduces its concepts and current development status in China, then explains the concepts and principles of deep learning, and then analyzes the application of deep learning in network information security, including the current application status and its application in 7 aspects in details. This article has certain significance for deepening the research and application of deep learning in the field of information security.

1. Introduction

With the continuous development of computer technology, computer networks have become an important feature of the information age. The network is a product of computer technology and communication technology and has been developed in response to the requirements of society for information sharing and information transfer. Countries are building their own information expressways. China's computer network has developed rapidly in recent years, and has been widely used in national defense, telecommunications, banking, broadcasting, etc. It is because of the wide application of network and the important role it plays in life. Its security cannot be ignored, and it is the key to whether the network can withstand the test. As an essential link in life, network information security must be paid corresponding attention.

2. Network information security overview

2.1. Network Information Security Concept

Information security refers to the technical and management security protection adopted to establish an information processing system to achieve the confidentiality, integrity, availability, controllability and
non-repudiation of electronic information [1]. In the current information age, computers network has become an indispensable tool for information exchange. However, due to the openness, interconnectivity, diversity of connection methods, and heterogeneity of terminal distribution of computer networks, coupled with existing technical weaknesses and human negligence, make the network susceptible to computer viruses, hackers or malware. In the face of various threats to network security, information security must be considered as a crucial issue. Network information security is divided into network security and information security. Network security including system security, i.e. hardware platform, operating system, application software; running service security, i.e. ensuring service continuity and high efficiency. Information security mainly refers to data security, including data encryption, backup, programs, etc.

2.2. Development Status of China's Network Information Security

Compared with foreign countries, the development of China's network information technology has a late start, backward technology, and lack of talents, which has resulted in the relatively fragile stage of China's network information security [2]. However, the problem of network information security cannot be ignored.

(1) In recent years, cyber threats have been diversified, and the biggest temptation to carry out cyber attacks is economic benefits.

Tighter organization of cyber attacks, specific targets, and behavioralization have become a major part of current cyber attacks trends. Internet deception methods are constantly upgraded. Ransomware, online game hacking and online banking hacking Trojans are everywhere in the current network spoofing, which fully demonstrates that certain criminals carry out network attacks driven by economic benefits. In addition, some hackers are now joining forces form a group or organization, when carrying out a cyber attack, its internal staff has a clear division of labor, making malicious code or even a destructive virus, and disseminating the code to obtain the required information to achieve the purpose of the attack. The hobby of technology has shifted to illegally seeking economic benefits.

(2) At present, network security vulnerabilities are still high.

Often after the old vulnerabilities are patched by management personnel or network technicians, new and more serious security vulnerabilities will appear [3]. Moreover, when some hacking organizations or network technicians discover new security vulnerabilities, they do not announce them immediately, but use them by themselves and wait until the time is right to release them. On the other hand, managers do not always upgrade and maintain network systems, which causes security breach.

(3) Diversity of the form of spreading viruses.

Now the functions of viruses are becoming more and more powerful, and their ability to hide and protect themselves is also becoming more and more advanced, so that they can continue to spread through network systems and removable devices. The method of traditionally rely on email to spread viruses is not often used. Hackers usually use network security holes to hang viruses on websites, and then make users attack or spread the virus.

(4) Botnets have a growing trend.

Botnets are new terms that have appeared in recent years, mainly referring to hackers controlling a large number of computers on the network and using these computers to form large-scale network pairs. The target to be attacked is a cyber attack. A botnet is extremely harmful, on the one hand, it is difficult to find the back of the botnet; on the other hand, using a botnet to attack is fast and effective, and can be performed in minutes or even seconds. The other party's network was paralyzed within the time.

(5) Rogue software is now increasing, disrupting the order of the network.

Rogue software refers to software that provides normal functions, but when installing or downloading, it starts to install or upload certain files without the user's consent. This kind of software will do something unknown to the user without knowing it, causing substantial harm to the user. The typical ones are Thunder, QQ real-time chat tools and some players.

(6) The society's awareness of network information security is weak.
At present, although society often emphasizes the need to improve information security awareness, it has not specifically applied it to practical problems. We know that the actual network attack is still not managed by the network. Caused by the presence or negligence, many companies and companies' computer systems are not even armed. Some even have professional security service systems, but after using it for a period of time, they find that they are not as free and respectful as before, and they change the security policy without permission. Poor management raises cyber security threats. These are all caused by a lack of strong cyber security awareness, and this situation is still widespread in today's society.

3. Concepts and principles of deep learning

3.1. The Concept of Deep Learning

Deep learning is a very important part of machine learning algorithms. It plays an important role in artificial neural networks by constructing deep feature learning on the data by constructing neural tissue structures similar to those recognized and understood by the human brain. The structure of early artificial neural networks is a perceptron network with multiple hierarchies [4]. The essence of deep learning is to extract the feature layer by layer from the input data, so that low-level linear features become high-level combined features after deep neural network learning, thereby improving the data's deep understanding. Deep learning has achieved great achievements with the help of neural network technology, and neurons are regarded as the most basic components of neural networks. Early neural network structures were generally multi-layered perceptron networks, but it could only solve simple linear classification tasks or could not even handle simple non-linear tasks. However, when an additional layer of network becomes a network with a hidden layer, it can not only handle simple XOR scenes, but also can approximate arbitrary non-linear functions. The back-propagation optimization method of early neural networks was proposed by Rumelhar and Hinton in 1986. The problem of how to optimize the neural network is solved [5]. Although the BP algorithm can be used for the inverse solution in the neural network, it takes too long to perform a training of the network, and the solution can only converge to the local optimal value. The problem makes neural networks difficult to solve. In 2006, the use of pre-training and fine-tuning the network to train deep neural networks was first proposed. The process of pre-training made the parameters of the neural network converge to a solution close to the optimal value, and then through the fine-tuning method to adjust the parameters of the entire network, these two processes make the speed of training a multi-layer neural network greatly improved. Due to the constraints of hardware conditions, the scale of the neural network structure used at that time was generally small [6]. In recent years, graphics processors have made huge breakthroughs in the field of parallel computing and are widely used to train large-scale deep neural networks. At the same time, with the introduction of the deep learning framework TensorFlow, the network structure can be quickly implemented. Deep learning technology has greatly promoted the progress of machine learning which has attracted the attention of many related researchers and many technology companies. Learning has gained great achievements in the field of complex text processing, image, audio, video and the like.

3.2. Principles of Deep Learning Algorithms

Deep learning is a non-linear neural network structure with multiple hidden layers. The deep neural network consists of an input layer, several hidden layers, and an output layer. Each layer has several neurons, and there are connection weights between the neurons. Each neuron mimics a human neuron cell, and the connection between nodes mimics the connection between neural cells [7].

The first layer (input layer): You can directly receive the raw data, not just the features.

The middle layer: contains several neurons, each neuron contains two functions, one is the weighted superposition of the input, and the other is the transformation of the superimposed value and the transformed value is passed to the next layer.

Output layer: Make decisions based on the final results of the obtained multi-layer operations.
4. Application of deep learning in computer network information security

4.1. Application Status of Deep Learning in Information Security

Deep learning is not a panacea for all information security problems because it requires extensive labeled datasets. Unfortunately, there are no such labeled datasets available. However, there are significant improvements in information security cases by deep learning networks that do an existing solution. Malware detection and network intrusion detection are just two of these areas, and deep learning has shown more significant improvements than rules-based and classic machine learning solutions.

Network intrusion detection systems are usually rule-based and signature-based controls that are deployed on the perimeter to detect known threats. Attackers can easily avoid traditional network intrusion detection systems by changing the signature of malware. Systems based on deep neural networks have been used to solve traditional security application problems, such as detecting malware and spyware [8]. Compared with traditional machine learning methods, deep learning-based technologies have better generalization capabilities. Now, deep learning-based neural networks are being used by users and entity behavior analysis. Traditionally, UEBA uses anomaly detection and machine learning algorithms. These algorithms extract security events to analyze and baseline every user and network element in the enterprise IT environment. Any major deviation from the baseline will be triggered for anomalies, further alarms from security analyst investigations were triggered. UEBA enhanced detection of insider threats, albeit to a limited extent. Deep learning-based systems are used to detect many other types of anomalies. Pawel Kobojeck, University of Warsaw, Poland uses keystroke dynamics verify that the user is whether using an LSTM network.

4.2. Application of Deep Learning in Detecting TOR Traffic

The main purpose of a cyber attack is to steal corporate user data, sales data, intellectual property files, source code, and software keys. Attackers use encrypted traffic to mix stolen data in regular traffic and transfer it to remote servers. Most attackers often use anonymous networks, making it difficult for security guards to track traffic. In addition, stolen data is often encrypted, which makes rule-based network intrusion tools and firewalls ineffective. Recently, anonymous networks have changed ransomware / malware changes. The body form is used for C & C. For example, onion ransomware uses the TOR network to communicate with its C & C server.

Anonymous network / traffic can be done in a variety of ways, of which TOR is one of the more popular choices. TOR is a free software that enables anonymous communication on the internet through a dedicated routing protocol called the onion routing protocol. This protocol relies on redirecting Internet traffic from multiple free hosted relays worldwide. During the relay, just like the layer of onion skins, each http packet is encrypted using the public key of the receiver [9]. At each receiving point, use the private key to decrypt the data packet. After decryption, the next target relay address will be disclosed. This process will continue until the exit node of the tor network is found. Here the packet decryption ends, a simple http data packet will be forwarded to the original target server. The original purpose of starting TOR was to protect user privacy. However, attackers used it instead of other illegal methods to intimidate good people. As of 2016, about 20% of TOR traffic involved illegal traffic activity. In a corporate network, TOR traffic is blocked by not allowing the TOR client to be installed or blocking the IP addresses of the protection or ingress nodes. However, there are many ways for attackers and Italian TOR software to access the network to transmit data and information. IP interception strategy is not a sensible strategy. According to the data, it shows that in 2017, 70 percent of automated attacks using multiple IP, 20% of automated attacks using more than 100 IP.

The TOR traffic can be detected by analyzing the traffic packets. This analysis can be performed on the TOR node, or between the client and the ingress node. The analysis is done on a single packet flow, and each packet flow constitutes one tuple includes source address, source port, destination address, and destination port. It analyzes network flows by extracting them at different time intervals. Most of the mainstream work in this field uses time features and other features such as size, port information to detect TOR traffic.
4.3. Application of Deep Learning in Pornographic Image Detection
Online pornography is a serious online crime, and the development of the Internet and communication technology has made it easier and faster and more concealed to spread pornographic information, which has severely affected the physical and mental health of adolescents. Manual template matching and other methods commonly used in image recognition generally used artificial. The methods of feature extraction and machine learning face some difficulties in identifying pornographic images, which has led many pornographic websites to repeatedly ban. In image recognition, the extraction of image features is the key to image recognition performance. Pornographic recognition based on skin detection detects parameters such as the proportion of exposed skin, such as the HSV color model, and then uses relevant methods to determine, but it is not ideal for complex textures and lighting effects; Porn recognition based on manual feature extraction, such as visual bag-of-words model, obtains classification results through feature extraction and classifiers of related parts, and the speed and accuracy can not fully meet the requirements. The advantage of generalization capabilities and robustness of deep learning in porn image detection is obvious. During the detection of pornographic images based on deep learning, the images to be tested are input into a convolutional neural network CNN, such as ResNet, VGGNet, AlexNet, or GoogleNet, to determine whether the picture is pornographic.

4.4. Application of Deep Learning in Illegal Text Information Detection
The use of the Internet to disseminate illegal text information will have a bad impact. This information includes false information, reactionary information, fraud information, etc. The use of homics, splitting, and pinyin can evade the detection of sensitive words that are currently commonly used, which requires natural language processing to be more complete and efficient. Natural language processing usually uses a shallow model based on statistics. Most researches use separate parts of speech tagging, semantically related words, named entity recognition, semantic role tagging, etc., and the feature concatenation of several separate tasks increases the error, while ignoring the integrity of language. But with the development of deep learning technology, the research on natural language processing has become a hot spot, and the application of network models such as convolution, circulation, and recursion to natural language processing has made illegal text information detection more accurate.

4.5. Application of Deep Learning in Public Network Voice Regulation
Due to the current increase in voice crimes, on the one hand, the amount of speech is huge and difficult to troubleshoot; on the other hand, speech is different from text analysis and can rely on sensitive keywords for detection, so the error rate of existing speech recognition systems is too high. In this case, public network voice regulation has played its huge advantages. The principle of public network voice supervision is as follows:
First: a large number of speech samples are input into a deep learning neural network, and the abstract features of the speech samples are obtained through the samples to obtain a speech feature information database.
Then: input the speech information into the deep learning neural network to get an abstract representation of the speech information and compare it with the speech feature database.
Finally: A classifier can distinguish normal information from suspicious information. Normal information is directly ignored, and suspicious information is recognized as text information by voice and screened manually, so that analysis and early warning of voice information are realized.

4.6. Application of Deep Learning in Smart Phone Intrusion Detection
With the rapid development of science and technology, the number of mobile smart devices and user data traffic on the Android platform has grown exponentially, and at the same time, the damage caused by malicious intrusion programs to mobile phone users has become greater. Intrusion detection is a prevention security mechanism, through the monitoring of smartphone status, network behavior, etc., to
discover whether user unauthorized behavior or intrusion behavior has occurred. Android malicious application detection system based on deep learning, breaking the traditional technical barrier of low efficiency, not only the theory of feasibility certificate has been obtained on the Internet, and the actual verification has also achieved better detection results.

There are two main directions for deep learning intrusion detection:

The first is to find intrusion rules and patterns, and compare them with the training model;

The second is for anomaly detection, to find out the normal behavior of users, and to create a normal behavior library for users.

4.6.1. Detection principle. The principle of deep learning intrusion detection is as follows: First use static code analysis technology to extract multiple types of behavioral feature data for Android applications, then convert the feature data into sample feature matrices, and then use the convolutional neural network algorithm file to train the sample feature matrices [10]. The last batch download did not participate in training the Andriod application of the deep neural network, and then perform system steps on its APK to get the relevant prediction report of the unknown sample APK.

4.6.2. Detection model. Mobile phone intrusion detection models based on deep belief networks are mainly composed of a restricted Boltzmann machine model (RBM) and a BP neural network.

First process the input data, and then use RBM for unsupervised training to make the features output by each layer more significant. Ensure that when the features are mapped inward to different feature spaces, retain as much feature information as possible to form a training model. To the more obvious feature information; the last layer uses the BP neural network for classification. The BP network layer receives the feature vector output by the RBM layer as its input data, and the training process of this layer is supervised training. After training with a fixed number of layers, set the classification parameter to 2 to get the final error value and calculate the corresponding accuracy rate.

4.7. Application of Deep Learning in HTTP Malicious Feature Analysis

With the development of web applications, the use of the http protocol has further expanded, and it has also become the main carrier of malicious behavior on the network. Therefore, many malicious behavior characteristics are reflected in the request data. There are many web attacks such as SQL injection, cross-site scripting attacks, the malicious behaviors such as cookie tampering are reflected in the http request, and the attack method of the request is changeable. The malicious characteristics are not only reflected in a specific place, but also many malicious characteristics are concentrated in the path or other parts.

The http security monitoring model is as follows:

First, analyze the http request format and malicious characteristics, and design a large number of features from the aspects of structure, length, and characters according to the characteristics of the data. Based on the automatic generation of a sensitive lexicon, count the number of sensitive words in the request content, and use it as one of the characteristics of the request. Then use the feature selection algorithm based on information entropy to select the distinguishing features from the designed features and use these selected features to vectorize the request. After use the classification algorithm to train the security detection model, it can be used to detect the category of the data.

4.8. Other Applications of Deep Learning

Deep learning is also gradually being widely used in various other areas of network information security. Face recognition based on deep learning, as an efficient means of identity authentication, is more secure than account passwords; its application in phishing website detection avoids malicious information risk of spreading and browsing users; deep learning technology analysis of malicious features based on the http protocol can avoid malicious operations in the web application using the http protocol; applying deep learning technology in information retrieval can prevent malicious links from appearing in search results of search engines.
5. Conclusion
In recent years, the use of computer networks has become more and more popular, which has made a huge push in the digital economy era, and network information security has also attracted everyone's attention. At present, there are still great hidden dangers in network information security that cannot be solved in a short time. Network information security is not a simple technical issue, but also involves management issues. Its development requires the joint efforts of the whole society through various preventive measures and publicity education to improve network security. In the increasingly technological age, network security technology will gradually progress and develop, those hidden dangers and problems will gradually be solved, and the network environment will become more and more complete and secure. At present, the application of deep learning in information security is still in its infancy, but it provides new ideas for the current field of information security. With the development of deep learning, the application of deep learning in information security will become more mature and more and more wide. With the development of cloud computing-related technologies, cloud computing and deep learning technologies can be combined to study information security prevention technologies.

References
[1] Jiang Luning. Machine learning-deep learning and network security technology[J]. Web Industry Forum. 2016 (05): 122-124.
[2] Shao Yan, Zhang Fanzhong. Research on the application of deep learning in public network security management [J]. Network Security Technology and Application. 2015 (6): 89-90.
[3] Chen Chen. Malicious code detection method based on deep learning of opcode behavior[J]. Harbin Institute of Technology. 2013 (12): 106-108.
[4] He Miao. Research on mobile data security detection technology based on machine learning[J]. Beijing University of Posts and Telecommunications. 2015 (01): 209-212.
[5] Zhang Haijian. Andriod malicious application detection system based on deep learning[J]. Electronic Technology and Network Security. 2017 (01): 89-91.
[6] Zhang Meijuan. Research on smart phone intrusion detection system based on deep learning [J]. Beijing Jiaotong University. 2016 (06): 256-258.
[7] Xue Xuefeng, Zhou Guodong. Research on deep learning for natural language processing[J]. Acta Automatica Sinica. 2016, 42 (10): 1445-1465.
[8] Zhang Qiaoli, Zhao Di, Chi Xuebin. Review of medical image diagnosis based on deep learning[J]. Computer Science. 2017, 44 (11a): 1-7.
[9] Su Zhida, Zhu Yuefèi, Liu Long. Android malicious application detection based on deep learning[J]. Computer Applications. 2017 (6): 1650-1656.
[10] SZEGEDY C,LIU W,JIA Y.et al. Going deeper with convolutions[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. IEEE,2015: 1-9.