ISSN: 2664-9640

An Intrusion Detection System Based on Entropy and Neural Networks

Minjue Ma

Wenzhou Polytechnic, Wenzhou, China

Abstract

Information entropy is a very important concept in machine learning. In machine learning, we often need to use the idea of information entropy to describe the probability distribution of data, use information entropy to find the expected information related to random events, and use information entropy to quantify the similarity between different data probability distributions. This article proposes an intelligent detection algorithm for network attacks based on neural networks, which extracts features from data through information entropy calculation, trains data using neural networks, and optimizes the classification process. Generally speaking, the higher the quality of data contained in training data, the better the detection performance of the model. This method can more effectively respond to various network attacks, improve the accuracy of network attack detection, and reduce false positives.

Keywords

Intrusion Detection, Entropy, Neural Networks.

1. Introduction

Researchers have conducted extensive research on the main threats faced by networks in reality, including four main aspects: network intrusion detection including DDoS attacks, malware detection, phishing email detection, and webpage tampering. Among them, the problem of network intrusion is the most serious. At present, in the field of network intrusion detection, it is mainly divided into two categories based on the deployment location of detection mechanisms and the detected targets: one is network abnormal traffic detection based on forwarding devices. This type of detection mechanism is usually deployed in forwarding units in the network, mainly using pre-defined rules or fixed patterns trained based on machine learning methods to detect forwarding traffic and determine whether there is malicious traffic; The second is intrusion detection based on end systems. This type of detection mechanism is deployed in the end system, usually PC or other access devices, and is generally detected through two methods: "attack mode matching" or "abnormal behavior discovery". Before machine learning technology was applied to the field of network security, network anomaly detection typically used static rules predefined by security experts based on experience or analysis of attack behavior for malicious traffic detection. However, with the increasing complexity of network operations and the increasing richness and concealment of attack behaviors, traditional network anomaly detection mechanisms face two major problems: firstly, the growth rate of new attack behaviors is much faster than the speed of security experts discovering and solving them. Secondly, the complexity of attack behavior leads to increasingly complex rule updates and may conflict with historical rules.

2. Related research

In 1990, Heberlein et al. [1] developed a network security monitoring system, which for the first time directly used the network as a source of information for intrusion detection. Since

ISSN: 2664-9640

then, network-based intrusion detection has been a hot topic in intrusion detection research with the development of networks. In the early days, there were some algorithms based on probability and statistical ideas, for example, Qayyum et al. [2] discussed the impact of statistical parameters such as mean and variance on anomaly detection, and proposed a distance measurement method based on chi square distribution. Ashfaq et al. [3] proposed an accuracy measurement method based on standard deviation regularization entropy for multiple classifiers, which achieved good performance improvement.

Wenke Lee et al. [4] proposed in 1999 to establish an anomaly detection model by analyzing network data flow, and introduced machine learning algorithms into intrusion detection systems. Subsequently, various algorithms related to data mining, machine learning, and deep learning were applied, and researchers studied anomaly based network intrusion detection from various aspects. At present, "intrusion detection" usually refers to anomaly based network intrusion detection, so in subsequent descriptions, "intrusion detection" is used to describe "anomaly based network intrusion detection" without causing confusion.

Big data and machine learning are playing an increasingly important role in this field, including the AI2 analyst in the loop security system proposed by MIT (Massachusetts Institute of Technology), the B-Profile network data analysis system jointly proposed by UNB (University of New Brunswick) and the Canadian Institute of Cybersecurity, and Cisco's DNA security [5]

The Network Information Security Laboratory (NISL) of Tsinghua University has proposed solutions to a series of vulnerabilities in network and system security, enhancing the security of basic internet software and services; The "High Speed Network Real Time Diversified Monitoring Technology and System" completed by the University of National Defense Technology uses the "Network Hawkeye" to detect network information in real time, and uses big data to ensure network information security; Kelai Company has proposed the "Kelai Big Data Security Situation Awareness Platform (BAP)" [6], which uses technologies such as boundary defense, network analysis, and backtracking analysis to provide comprehensive security situation awareness; 360 Company's Tianyan system utilizes big data to drive network security and achieve comprehensive threat detection.

3. A Feature Extraction Method Based on Information Entropy

Information theory and information entropy are very important concepts in AI and machine learning. In machine learning, we often need to use the idea of information entropy to describe the probability distribution of data, use information entropy to find the expected information related to random events, and use information entropy to quantify the similarity between different data probability distributions. Information theory and information entropy have extensive applications in Bayesian learning, decision tree learning, and classification systems. For example, in decision tree algorithms, information entropy can be used as the basis for tree branching to construct decision trees.

By extracting features from network traffic, we can obtain a large number of features that describe a connection record, such as connection duration, destination address, source address, protocol type, etc. In the KDD CUP dataset, a record is described using 41 dimensional features, while in the CICIDS2017 dataset, a traffic record is described using 84 dimensional features. A large number of features provide a detailed description of network traffic, but the effectiveness of each feature in detecting network attacks varies. There are a large number of redundant features in the entire feature space, which often have a low contribution to attack detection and even serve as noise interference in the attack detection process. Meanwhile, selecting the most useful features to reduce feature dimensionality can effectively reduce the system's computational burden and improve algorithm performance. Therefore, we can conduct feature correlation studies to determine useful features for attack detection. To address the issue of a

ISSN: 2664-9640

large number of redundant features in IDS data, this chapter proposes a feature selection method based on information gain. In the project, we first analyze and calculate the impact of different features in traffic data on the changes in system entropy, and then select the feature that causes the greatest change in system information entropy as the input for the classification system. By using this method, the impact of irrelevant features in attack detection can be reduced, system noise can be removed, and feature dimensionality reduction can reduce the consumption of algorithm computing resources and improve attack detection efficiency.

In the classification system, assuming that the set of network traffic categories we collect is C, which represents the labels corresponding to different records in the dataset. The number of traffic categories in the entire dataset is N, and we use c_i to represent the i-th class ($c_i \in C$) in the entire dataset. So the entropy H（C）of this classification system can be expressed by the following formula

$$H(C) = -\sum_{i=0}^{N} p(c_i) * \log_2 p(c_i). \quad (1)$$

$p(c_i)$ is the probability of occurrence c_i in the entire dataset. Because we can obtain information on the entire system and the distribution of traffic categories, therefore we can use the probability of their occurrence in the dataset to represent the total number of class occurrences in the dataset

$$p(c_i) = \frac{num(c_i)}{num(C)}. \quad (2)$$

The feature space in the data we studied is $F = \{f_0, f_1, \ldots, f_n\}$. To calculate the information gain of f_n for system classification, it is necessary to study the changes in the amount of system information when the feature f_i is in the system and not in the system, that is, the information gain brought by f_i for the classification system.

The information gain $IG(f_i)$ brought by feature fi to the system can be expressed as the difference between the original entropy of the system and the entropy value of the system when f_i is fixed. The larger the difference, the greater the change in entropy. That is, the more information this feature fi brings to the system, the more conducive it is to classification.

$$IG(f_i) = H(C) - H(C|f_i) \quad (3)$$

$$H(C|f_i) = \sum_{j=1}^{m} p(x_j) H\left(C\middle|f_i = x_j\right) \quad (4)$$

The method of feature selection based on information gain determines the importance of a particular feature in our study to the system, and uses all samples in the entire system to participate in the calculation, effectively reducing noise interference, comprehensively considering the changes in system information entropy, and accurately grasping the internal connections between the entire system samples. The method of information gain is relatively simple to calculate and can be operated on larger datasets, making it easy to capture the relationship between features and results.

Based on the above method, we selected 19 dimensions of features in the dataset, with significantly higher entropy increase than other features, including:

protocol type ,service ,flag ,source bytes ,destination bytes ,logged in ,count ,srv count ,serror rate ,srv serror rate ,same srv rate ,diff srv rate ,dst host srv count ,dst host same srv rate ,dst host diff srv rate ,dst host same src port rate ,dst host srv diff host rate ,dst host serror rate ,dst host srv serror rate.

4. Intrusion Detection Algorithm Based on Neural Networks

The most typical neural network structure consists of three parts, namely the Input Layer, where the number of neurons is the data feature dimension; The Hidden Layer, as the intermediate layer of the neural network, accepts the output from the previous network as the current input value and calculates the current result to be passed on to the next layer. The

ISSN: 2664-9640

number of neurons in the hidden layer will directly affect the fitting ability of the model; The Output Layer is the network layer that outputs the final result, and its number of neurons is the number of classification categories. The classic three-layer neural network model is shown in the figure. When the number of neurons in the input and output layers of the network is determined, it is necessary to consider the depth and width of the hidden layer to achieve better fitting results.

Using the KDD99 dataset to complete classification, it is divided into two categories: attacked and not attacked. After scrambling the data, 80% of the total data is selected as the training set, and the remaining 20% is used as the test set. A simple four layer neural network is used with RELU as the activation function, see Table 1

Table 1: Hierarchical Division

Layer	Node	Weight
Input layer	19	w1=19*30
Hidden layer one	30	w2=30*20
Hidden layer two	20	w3=20*2
Output layer	2	

5. Effect analysis

Using information entropy increment for feature selection, selecting the most typical neural network model, dividing the training set and test set in KDD99 data, using entropy increment algorithm to achieve feature selection and network attack detection. After learning the training set 100 times, predicting and comparing the test set, the prediction accuracy is 94.8%. The prediction accuracy is shown in Figure 1.

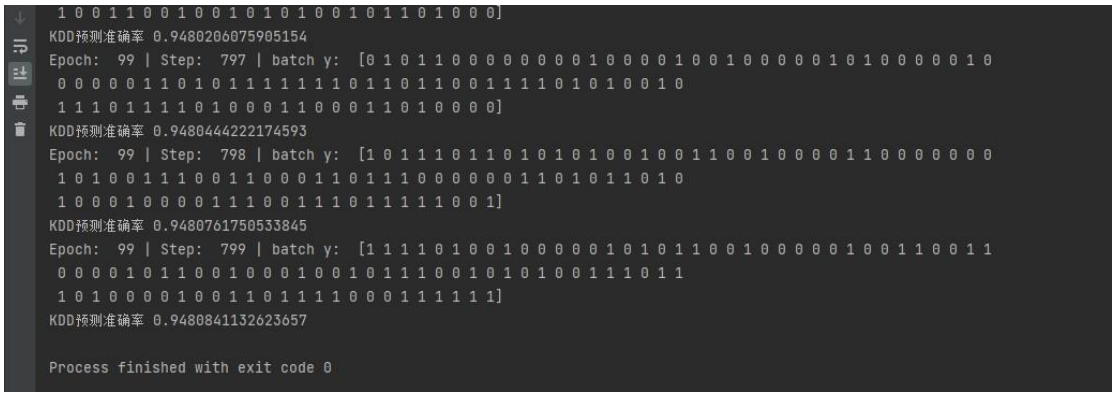

Figure 1: The prediction accuracy

Acknowledgements

Artificial Intelligence Academy, Wenzhou Polytechnic.

References

[1] Heberlein LT, Dias GV,Levitt KN,et al.A network security monitor[C].//1990 IEEE Computer Society Symposium on Research in Security and Privacy,Oakland, CA, USA, 7-9 May 1990.

[2] Qayyum A, Islam M H,Jamil M.Taxonomy of statistical based anomalydetection techniques for intrusion detection[C].// Emerging Technologies, 2005.Proceedings of the IEEE Symposium on.IEEE,2005.

ISSN: 2664-9640

[3] Ashfaq A B, Javed M, Khayam S A, et al. An Information-TheoreticCombining Method for Multi-Classifier Anomaly Detection Systems[C].// IEEEInternational Conference on Communications (ICC), Cape Town, South Africa,2010.

[4] Lee W, Stolfo S J.Mok K WA Data Mining Framework for Building IntrusionDetection Models[C].// Proceedings of the 1999 IEEE Symposium on Security andPrivacy. Oakland:IEEE Press. 1999:120-132.

[5] Cisco Enterprise Network Security[Z]. Cisco public,2018.

[6] http://www.colasoft.com.cn/products/bap.php.