Network Log Anomaly Detection Model Construction and Performance Improvement Method Based on Artificial Intelligence

Shaopei Su[1], Yan Zhang[2]

[1]Department of Information and Communication, Sichuan University of Science and Technology, Meishan, 620000, China

[2]Chengdu Chuanke Foreign Language School, Chengdu, 611730, China

Keywords: Artificial Intelligence; Network Log; Anomaly Detection Model; Performance Improvement; Network Security

Abstract: This article focuses on anomaly detection of network logs, aiming at building an efficient anomaly detection model and improving its performance under the background of increasingly complex network systems and limited traditional detection methods. Through in-depth analysis of related concepts of network logs, artificial intelligence (AI) technology and anomaly detection principle, a hierarchical architecture design model is adopted, and the long-term and short-term memory network (LSTM) algorithm based on deep learning (DL) is applied, combined with data cleaning, over-sampling and parameter optimization strategies. The experimental results show that the model with improved performance is significantly superior to the unoptimized model in terms of accuracy, precision, recall and F1 value, and the detection accuracy of different types of abnormal logs is also greatly improved. The research shows that the model and performance improvement method are effective and provide a reliable scheme for anomaly detection of network logs, but it still needs to be continuously optimized to adapt to the dynamic network environment.

1. Introduction

At present, the scale and complexity of network system are rising. As the record carrier of network activities, network logs contain rich information about system running status and user behavior. It is very important to analyze them to detect abnormal conditions to ensure the safe and stable operation of network systems [1]. However, the traditional rule-based or statistics-based anomaly detection methods are often stretched in the face of massive, complex and changeable network log data [2].

In this context, AI technology, with its powerful data analysis and pattern recognition capabilities, has brought new ideas and methods for anomaly detection of network logs [3]. The anomaly detection model of network logs based on AI is expected to identify potential abnormal behaviors more accurately and efficiently, and provide strong support for network security protection [4]. At present, although a lot of research has been carried out on anomaly detection of network logs based on AI, there are still some problems to be solved urgently [5]. Some models have insufficient generalization ability in complex network environment, and it is difficult to adapt to diverse network scenarios. At the same time, the optimization method of model performance is not perfect, and there is still much room for improvement in detection accuracy, false alarm rate and detection efficiency [6].

This article focuses on the AI-based network log anomaly detection model construction and performance improvement methods. The purpose of this article is to design an anomaly detection model with strong adaptability and superior performance, which can effectively meet the challenges faced by the current anomaly detection of network logs. By systematically analyzing the characteristics of network logs and the advantages of AI technology, a scientific and reasonable model architecture is constructed, and targeted performance improvement strategies are put forward, hoping to provide useful reference for related research and practice in the field of network security.

  DOI: 10.25236/icmmct.2025.030

2. Related theoretical and technical basis

Network log is a record file generated by network devices, servers and applications during operation, which records all kinds of network activity information in detail, such as login and logout time of users, access resource paths and system operation instructions [7]. Formally, it can be divided into text log and binary log, among which text log is widely used because of its readability and easy analysis. The content of network log is rich, which can reflect the running state of network system and user behavior patterns, and provide a key data source for subsequent anomaly detection.

AI aims to make machines simulate human intelligence and realize abilities such as learning, reasoning and decision-making. In the field of network log anomaly detection, commonly used AI technologies include machine learning (ML) and DL. ML covers supervised learning, unsupervised learning and semi-supervised learning [8]. Supervised learning needs to train models on labeled data. For example, support vector machines can construct classification models based on labeled normal and abnormal log data. Unsupervised learning aims at unlabeled data, and clustering algorithm can classify similar network logs into one category and find potential abnormal patterns [9]. DL is a branch of ML. By constructing a deep neural network, it can automatically learn the hierarchical features of data, such as the normal feature representation of network logs by a deep self-encoder, so as to detect abnormal logs deviating from the representation.

Anomaly detection is based on the assumption that normal behavior pattern has certain regularity and stability, while abnormal behavior deviates from this pattern. Its core is to build a normal behavior model, and when the deviation of data points from the model exceeds the set threshold, it is judged as abnormal [10]. Common methods include statistics-based methods, which set thresholds according to statistical characteristics of data such as mean and variance; Based on the distance method, the distance between data points and normal data sets is calculated to identify anomalies. These principles provide theoretical support for the anomaly detection model of network logs based on AI.

3. Construction of anomaly detection model of network log based on AI

3.1. Model requirements analysis

In the complex and changeable network environment, the network log anomaly detection model needs to meet various requirements. The model should have high accuracy, can accurately identify all kinds of abnormal behaviors, and reduce the rate of false positives and false negatives as much as possible. With the increasingly covert and complex network attack methods, the model needs to capture subtle abnormal features keenly. In addition, the model should have good real-time, and it can quickly analyze and detect when massive log data are constantly generated, find anomalies in time and give early warning, so that operation and maintenance personnel can respond quickly. In addition, the model should have strong adaptability, which can cope with the logs generated by different types of network systems and adapt to the dynamic changes of network environment.

3.2. Model architecture design

This model adopts hierarchical architecture design, which is divided into data preprocessing layer, feature extraction layer and model training and detection layer. The data preprocessing layer is responsible for cleaning and converting the original network log, removing noise data, filling in missing values, and converting the log data into a format suitable for subsequent processing. The feature extraction layer uses natural language processing technology or statistical methods to extract representative features from the preprocessed logs. In the model training and detection layer, the appropriate AI algorithm is selected to construct an anomaly detection model, and the model is trained according to the extracted features, and then the new log data is detected abnormally.

3.3. Key algorithms and processes of the model

The core algorithm of this model adopts LSTM based on DL. LSTM can effectively deal with the long-term dependence in time series data, and is suitable for network logs, which have time

series characteristics. In the LSTM, the forgetting gate determines the proportion of the characteristic information about the network log transmitted from the memory unit of the previous moment to the current moment, and the calculation formula is as follows:

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{1}$$

Where f_t is the forgetting gate output at t time; σ is a sigmoid activation function, which compresses the value between 0 and 1 to control the degree of transmission of characteristic information of weblogs. W_f is the weight matrix of forgetting gate; $\left[h_{t-1}, x_t\right]$ means splicing the hidden state h_{t-1} of the previous moment (which contains the characteristic information of the previous weblog sequence) with the weblog input x_t of the current moment; b_f is the offset of the forgetting gate.

In the data preprocessing stage, the original network logs are divided into log sequences with fixed length in time sequence. Each log record is segmented and vectorized, and the text information is converted into a numerical vector x_t. Then, in the process of feature extraction, the vectorized log sequence is input to the LSTM network in turn in time step. The memory unit in the LSTM network will capture the long-term dependence in the network log sequence according to the above formula and learn the normal feature pattern of the network log. In the training process, a large quantity of labeled normal and abnormal network log sequence data are used to minimize the cross entropy loss function L between the predicted value and the real label. The formula of cross entropy loss function is:

$$L = -\sum_{i=1}^{N}\left[y_i \log\left(\hat{y}_i\right) + \left(1 - y_i\right)\log\left(1 - \hat{y}_i\right)\right] \tag{2}$$

Where: N is the quantity of samples; y_i is the real tag (normal log is 0, abnormal log is 1); $\hat{y}_i$ is the predicted value of the model.

By adjusting the parameters of LSTM network through back propagation algorithm, the model can accurately learn the characteristic patterns of normal and abnormal network logs. When new network log data arrives, it goes through the same preprocessing and feature extraction steps, and is input into the trained LSTM model. The model outputs the prediction result, and judges whether the log sequence is abnormal according to the preset threshold. If the prediction result exceeds the threshold, it is judged as an abnormal log and the corresponding alarm mechanism is triggered. In this way, the model based on LSTM can effectively detect the anomaly of network logs, and meet the demand of network security for real-time and accurate detection of abnormal behavior. In order to solve the problem of data quality, data cleaning technology can be used to improve data quality by removing duplicate records, correcting wrong data and filling in missing values. For the problem of data imbalance, over-sampling or under-sampling methods can be used.

4. Experiment and result analysis

The experiment was carried out on a server equipped with IntelCorei7-10700K processor and 32GB memory. The operating system is Ubuntu20.04. The programming environment adopts Python3.8 and related DL framework. The data set used comes from the log records of several actual network systems, covering different types of network activities. After data preprocessing, removing irrelevant fields and cleaning noise data, a data set containing 10000 log records is finally obtained. Among them, there are 8000 normal logs and 2000 abnormal logs. The data set is divided into training set and test set according to the ratio of 7:3.

Based on the model architecture and algorithm designed above, a network log anomaly detection model based on LSTM is constructed. At the same time, the performance improvement method

mentioned above is applied to oversample the abnormal samples in the training set, and the LSTM model parameters are optimized by random search. In order to compare and verify the performance improvement effect of the model, two groups of experiments are set up: one is the basic model without performance improvement method, and the other is the optimized model with performance improvement method.

The basic model and the optimized model were tested several times on the test set, and various performance indicators were recorded. The results are shown in the following Table 1:

Table 1: Comparison of Performance Metrics between the Basic Model and the Optimized Model

Model	Accuracy (%)	Precision (%)	Recall (%)
Basic Model	82.5	78.3	75.6
Optimized Model	90.2	87.4	85.8

It can be clearly seen from Table 1 that the optimized model with performance improvement method is superior to the basic model in all indexes. The accuracy of the optimized model is improved to 90.2%, which is nearly 8 percentage points higher than that of the basic model, which shows that the optimized model can distinguish normal and abnormal logs more accurately. The accuracy rate and recall rate reached 87.4% and 85.8%, respectively, indicating that the optimized model can better detect the actual abnormal logs while reducing false positives, and then the F1 value was increased from 76.9 to 86.6, and the comprehensive performance was significantly improved.

Further analysis of the detection of different types of abnormal logs leads to the results shown in Table 2 below:

Table 2: Comparison of Detection Results for Different Types of Anomaly Logs

Anomaly Type	Basic Model Detection Accuracy (%)	Optimized Model Detection Accuracy (%)
Login Anomaly	75.0	88.0
Operation Permission Anomaly	78.5	85.5
Traffic Anomaly	80.0	90.0

It can be seen from Table 2 that the detection accuracy of the optimized model is also higher than that of the basic model for different types of abnormal logs. Especially in the aspect of traffic anomaly detection, the accuracy of the optimized model reaches 90.0%, which is greatly improved compared with the 80.0% of the basic model. This shows that the performance improvement method not only enhances the detection ability of the model as a whole, but also has better recognition effect for various specific abnormal situations.

The results show that the anomaly detection model and performance improvement method of network log based on AI proposed in this article have achieved good results. The optimized model has significantly improved the key indicators such as accuracy, precision, recall and F1 value, and the ability to detect different types of abnormal logs has also been significantly enhanced. However, the network environment is constantly changing, and it is needed to continue to pay attention to the emerging abnormal behavior patterns in the future, further optimize the model and improve its adaptability to the complex and changeable network environment.

5. Conclusions

This article focuses on the construction of AI-based network log anomaly detection model and the research on performance improvement methods, and has achieved a series of valuable results. By systematically analyzing the challenges faced by network log anomaly detection, and deeply discussing the relevant theoretical and technical basis, a layered anomaly detection model is successfully designed, and the LSTM algorithm is the core detection mechanism. In the aspect of performance improvement, the performance of the model is significantly improved through the application of a series of methods such as improving data quality, dealing with data imbalance and optimizing model parameters and structure. The results clearly show that the optimized model has made remarkable progress in several key performance indicators, which not only greatly improves

the overall detection accuracy, but also significantly improves the detection accuracy for various specific types of anomalies, effectively verifying the effectiveness and practicability of the proposed method.

However, the field of network security is always in a dynamic development, and new network attack methods and abnormal behavior patterns are constantly emerging. Therefore, future research still faces many challenges. On the one hand, it is needed to continuously pay attention to the changes of network environment and update and optimize the model in time to ensure its ability to detect new anomalies. On the other hand, we can explore the integration of more cutting-edge technologies, such as reinforcement learning and transfer learning, to further enhance the adaptability and generalization ability of the model and provide a more solid guarantee for the safe and stable operation of the network system.

Acknowledgements

The 2nd Huang Yanpei Vocational Education Thought Research Plan Project in 2024- Research on the Targeted Spiral Talent Training Model Based on Huang Yanpei's "Doing and Learning in One" Teaching Concept (ZJS2024ZN027)

References

[1] Jiang X, Luo H, Sun Y, et al. Fast Anomaly Detection for IoT Services Based on Multisource Log Fusion[J]. IEEE Internet of Things Journal, 2023, 11(6): 9405-9419.

[2] Chen S, Liao H. BERT-Log: Anomaly Detection for System Logs Based on Pre-trained Language Model[J]. Applied Artificial Intelligence. 2022, 36(1):24.

[3] Jia T, Li Y, Yang Y, et al. Hilogx: noise-aware log-based anomaly detection with human feedback[J]. The VLDB Journal, 2024, 33(3):883-900.

[4] Lu S, Wang M, Wang D, et al. Black-box attacks against log anomaly detection with adversarial examples[J]. Information Sciences, 2023, 619:249-262.

[5] Zhao Z, Xu C, Li B. A LSTM-based anomaly detection model for log analysis[J]. Journal of Signal Processing Systems, 2021, 93(7): 745-751.

[6] Ghafoori Z, Erfani S M, Bezdek J C, et al. LN-SNE: Log-normal distributed stochastic neighbor embedding for anomaly detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 32(4): 815-820.

[7] Guan W, Cao J, Gu Y, et al. GRASPED: A GRU-AE network based multi-perspective business process anomaly detection model[J]. IEEE Transactions on Services Computing, 2023, 16(5): 3412-3424.

[8] Liu X, Liu W, Di X, et al. LogNADS: Network anomaly detection scheme based on log semantics representation[J]. Future Generation Computer Systems, 2021, 124: 390-405.

[9] Wei X, Sun C, Zhang X Y. KAD: a knowledge formalization-based anomaly detection approach for distributed systems[J]. Software Quality Journal, 2024, 32(2): 821-845.

[10] Elsayed M A, Zulkernine M. PredictDeep: security analytics as a service for anomaly detection and prediction[J]. IEEE Access, 2020, 8: 45184-45197.