Research on Computer Network Fault Prediction and Proactive Maintenance Strategy Based on Data Mining

Liang Pang

Yinchuan University of Energy, Yinchuan, Ningxia, 750001, China

Keywords: Data mining; Computer network; Fault prediction; Proactive maintenance strategy

Abstract: This paper focuses on the field of computer network fault prediction and proactive maintenance, exploring the application of data mining technology. By analyzing the types and causes of network faults, it expounds on the key data mining technologies, constructs a fault prediction model, formulates proactive maintenance strategies, and evaluates the implementation effects. The aim is to enhance network reliability and stability, providing theoretical support and practical guidance for network operation and maintenance.

1. Introduction

In the digital age, computer networks serve as critical infrastructure for social operation and are widely applied across various fields. However, with the expansion of network scale and increasing complexity, faults occur frequently, causing inconvenience to users and even significant economic losses, such as interruptions in financial transactions and delays in medical diagnoses. Therefore, accurately predicting network faults and conducting proactive maintenance to improve network reliability and stability is an urgent task in the field of network operation and maintenance. Data mining technology offers a new approach by collecting and analyzing data such as device status, traffic, and logs during network operation. It can mine potential patterns and rules to predict faults in advance and formulate proactive maintenance strategies, effectively reducing the fault occurrence rate and enhancing network service quality.

2. Literature Review

2.1 Research Status of Network Fault Prediction

Scholars at home and abroad have conducted extensive research on network fault prediction. Early studies mainly relied on expert systems and rule-based reasoning methods, formulating fault diagnosis rules by summarizing expert experience to predict and diagnose network faults. However, this approach faced challenges such as difficulties in knowledge acquisition, complex rule maintenance, and poor adaptability, making it difficult to meet the fault prediction needs of large-scale and complex networks. With the development of machine learning technology, fault prediction methods based on statistical learning and neural networks have gradually become research hotspots. Algorithms such as Support Vector Machines (SVMs), decision trees, and neural networks have been widely applied in the field of network fault prediction. By training models to learn patterns and rules in network data, they can classify and predict network faults. However, these methods encounter issues such as high computational complexity and insufficient model generalization ability when dealing with large-scale, high-dimensional, and nonlinear network data[1].

2.2 Research Status of Data Mining Application in Network Fault Prediction

The application of data mining technology in the realm of network fault prediction has emerged as a groundbreaking approach to tackle the long - standing challenges in this field. Network systems are becoming increasingly complex, and traditional methods of fault detection and prevention often fall short in dealing with the dynamic and multifaceted nature of modern networks. Data mining technology, with its powerful analytical capabilities, offers a new and promising avenue to address

  DOI: 10.25236/iwmecs.2025.002

these issues.

Association rule mining, a key technique within data mining, plays a crucial role in uncovering the intricate relationships between network faults and a multitude of influencing factors. By sifting through vast amounts of network data, it can discover potential causes and patterns of fault occurrence. For instance, it can identify which specific network configurations, usage patterns, or external environmental factors are more likely to trigger certain types of faults. This knowledge enables network administrators to take proactive measures to prevent faults before they occur, thereby enhancing the overall reliability and stability of the network.

Cluster analysis is another valuable data mining method that can group network data based on similarity. In the context of network fault prediction, it can effectively identify abnormal data points that deviate from the normal patterns. These abnormal data may indicate the presence of potential fault patterns or even emerging network attacks. By clustering network traffic data, for example, cluster analysis can distinguish between normal traffic flows and those that exhibit unusual characteristics, such as sudden spikes in data volume or irregular communication patterns. This early identification of abnormal traffic allows for timely intervention and the prevention of potential network failures or security breaches.

Classification algorithms, on the other hand, are designed to classify and predict network faults. They can analyze historical network data, learn from past fault occurrences, and then use this knowledge to detect potential fault hazards in advance. By assigning network states or events to different predefined categories, classification algorithms can provide clear warnings about the likelihood of a fault happening. For example, they can predict whether a particular network component is at risk of failure based on its current performance metrics and historical failure records.

There are numerous real - world examples that demonstrate the effectiveness of these data mining techniques in network fault prediction. Some studies have successfully employed association rule mining technology to analyze network alarm data. By examining the relationships between different alarms, they can predict possible faults that may follow a sequence of specific alarm signals. This enables network operators to respond quickly and prevent the escalation of faults. Others have adopted cluster analysis methods to cluster network traffic data. Through this approach, they can identify abnormal traffic patterns, such as those associated with distributed denial - of - service (DDoS) attacks or malware infections, and issue early warnings to mitigate the impact of these threats on the network.

However, despite the significant progress made in applying data mining technology to network fault prediction, existing research still has several shortcomings. In the area of data preprocessing, the quality and consistency of network data can vary greatly, and ineffective preprocessing methods may lead to inaccurate analysis results. Feature extraction is another challenge, as selecting the most relevant and informative features from the vast amount of network data is a complex task. Moreover, model optimization is crucial for improving the accuracy and efficiency of fault prediction models, but current research in this aspect is still not sufficient. Therefore, further in - depth research is urgently needed to overcome these limitations and fully unlock the potential of data mining technology in network fault prediction.

2.3 Research Status of Proactive Maintenance Strategies

Proactive maintenance strategies aim to take maintenance measures in advance based on network fault prediction results to prevent faults or reduce their impact. Currently, proactive maintenance strategies mainly include preventive maintenance and predictive maintenance. Preventive maintenance involves regularly inspecting and replacing equipment based on its operating time and historical fault data to reduce the fault occurrence rate. However, this method lacks targeting and may lead to over-maintenance or insufficient maintenance. Predictive maintenance, on the other hand, takes maintenance measures before fault occurrence based on network fault prediction results, offering higher targeting and effectiveness. However, existing proactive maintenance strategies still lack scientific and reasonable decision-making methods in terms of maintenance timing selection, maintenance resource allocation, and maintenance plan formulation, requiring further improvement[2].

3. Analysis of Network Fault Types and Causes

3.1 Network Fault Types

Network faults can be broadly classified into two categories: physical faults and logical faults. Physical faults refer to network connection interruptions or performance degradation caused by hardware damage to network devices, line failures, or loose interfaces. For instance, hardware malfunctions in network devices such as routers and switches may result in the inability of the network to communicate normally. Damaged network cables or loose interfaces can cause intermittent network connections. On the other hand, logical faults are network anomalies arising from incorrect network configurations, software issues, virus infections, and other such causes. For example, network configuration problems like IP address conflicts or incorrect subnet mask settings can lead to poor network communication. Exploitation of vulnerabilities in operating systems or applications may subject the network to attacks, triggering faults.

3.2 Network Fault Causes

The causes of network faults are complex and diverse, mainly including device factors, environmental factors, human factors, and management factors. Device factors are one of the main causes of network faults, as the quality, performance, and service life of network devices directly affect network stability. For example, aging network devices are prone to hardware failures, leading to network interruptions. Environmental factors such as temperature, humidity, and electromagnetic interference can also affect network devices and trigger faults. Human factors include operational errors and unauthorized operations, such as accidentally deleting network configuration files or arbitrarily changing network parameters, which may cause network faults. Management factors mainly include unreasonable network planning, inadequate monitoring, and untimely maintenance, such as an unreasonable network topology design leading to network congestion and affecting network performance, and a lack of effective network monitoring means making it difficult to detect and handle network faults in a timely manner[3].

4. Key Data Mining Technologies

4.1 Data Preprocessing

Data preprocessing is an important step in data mining, aiming to improve data quality and provide a reliable data foundation for subsequent data analysis and mining. Data preprocessing mainly includes data cleaning, data integration, data transformation, and data reduction. Data cleaning is used to remove noise, outliers, and duplicate data from the data, improving data accuracy and consistency. For example, for missing values, methods such as mean, median, or mode can be used for filling, while for outliers, statistical or visualization methods can be used for identification and processing. Data integration involves integrating data from different data sources into a unified dataset. For example, integrating device status data, traffic data, and log data of network devices for comprehensive analysis. Data transformation converts data from the original format into a format suitable for data mining algorithm processing, such as numericalization, standardization, and discretization. Data reduction reduces data volume through methods such as dimensionality reduction and clustering to improve data mining efficiency.

4.2 Feature Extraction and Selection

Feature extraction and selection involve extracting effective information from raw data that can reflect network fault characteristics to improve the accuracy and efficiency of fault prediction. Feature extraction methods include Principal Component Analysis (PCA), Factor Analysis, and Independent Component Analysis (ICA), which can convert high-dimensional data into low-dimensional data while retaining the main features of the data. Feature selection methods include algorithms such as Information Gain, Chi-Square Test, and ReliefF, which can select the most discriminative features from the extracted features, removing features that are irrelevant or have low correlation with fault prediction, reducing data dimensionality, and improving model training speed and prediction accuracy[4].

4.3 Data Mining Algorithms

Commonly used data mining algorithms include Support Vector Machines (SVMs), decision trees, neural networks, and random forests. SVMs find the optimal hyperplane to distinguish different classes of data, offering high generalization ability and being suitable for classification and regression problems with small samples and high-dimensional data. Decision trees classify or regress data through a tree structure, having the advantages of strong interpretability and ease of understanding, and can visually display fault prediction rules. Neural networks simulate the human brain's information processing process through connections between multiple layers of neurons, having powerful nonlinear mapping capabilities and being able to handle complex network fault prediction problems. Random forests are an ensemble learning method that builds multiple decision tree models and combines their prediction results to improve model prediction accuracy and stability.

5. Construction of a Network Fault Prediction Model Based on Data Mining

5.1 Model Construction Framework

The framework for constructing a network fault prediction model based on data mining mainly includes data collection, data preprocessing, feature extraction and selection, and model training and evaluation. Data collection involves collecting data related to network faults from various data sources such as network devices, log files, and monitoring systems, including fault types, fault occurrence times, fault durations, and fault impact ranges. Data preprocessing performs operations such as cleaning, integration, transformation, and reduction on the collected data to improve data quality. Feature extraction and selection extract valuable features from the preprocessed data and select the most representative feature subsets. Model training and evaluation use training datasets to train fault prediction models and adopt methods such as cross-validation and confusion matrices to evaluate the models. Based on the evaluation results, model parameters are adjusted to optimize model performance.

5.2 Model Selection and Optimization

As shown in Figure 1, comparing the accuracy of algorithms such as support vector machines, decision trees, neural networks, and random forests on the test set provides a visual reference for selecting models tailored to specific network data characteristics.

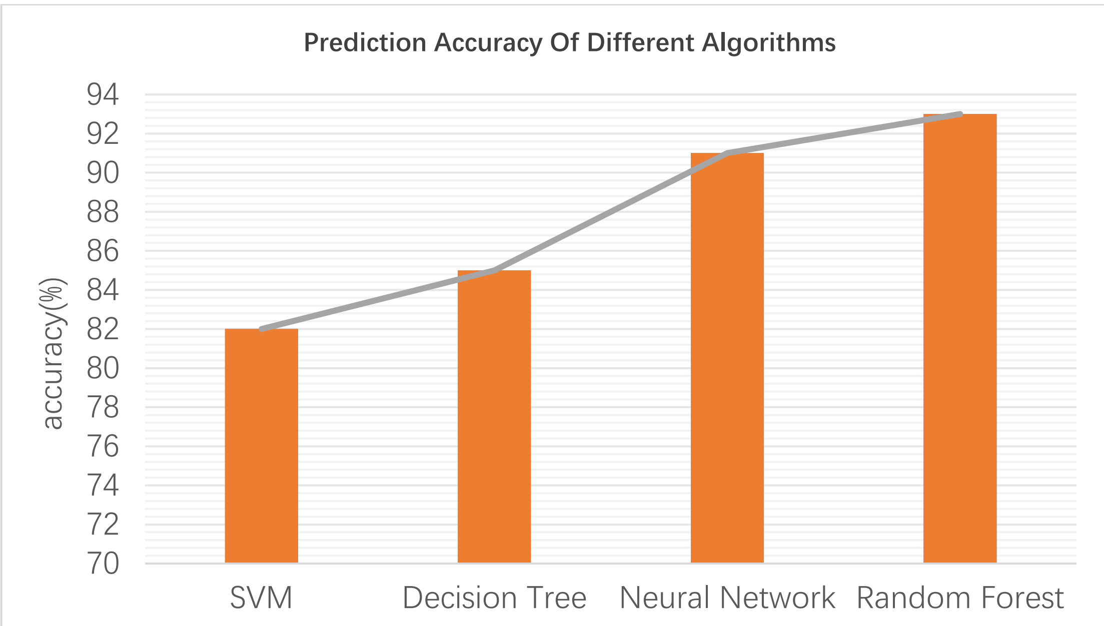

Figure 1. Accuracy Comparison of Data Mining Algorithms

In network fault prediction work, accurately selecting an appropriate prediction model is a critical prerequisite, which requires full consideration of the characteristics of network fault data and prediction needs. If the data exhibits linear separability, the Support Vector Machine model often demonstrates good prediction performance, effectively distinguishing between normal and fault states by finding the optimal classification hyperplane in high-dimensional space. When the data is nonlinear, neural network models or random forest models are more advantageous. Neural networks can handle complex nonlinear relationships by simulating the structure of human brain neurons, while random forests enhance prediction stability and accuracy by constructing multiple decision trees and combining their results. In the model training process, methods such as grid search and genetic algorithms are indispensable for optimizing model parameters, significantly improving the model's prediction accuracy and generalization ability. Additionally, as the network environment constantly changes and new fault types and data continuously emerge, it is necessary to regularly update and iterate the model, adjusting its structure and parameters according to new data and fault situations to ensure the model's effectiveness and accuracy and provide reliable guarantees for the stable operation of the network.

6. Formulation of Proactive Maintenance Strategies Based on Fault Prediction

6.1 Maintenance Timing Selection

Accurately grasping the maintenance timing is crucial in network operation and maintenance. Based on network fault prediction results, the importance of network devices and the potential impact of faults should be considered to reasonably determine the maintenance time. Key network devices are the "lifelines" of network operation, and once a fault occurs, the consequences can be disastrous. When a high probability of fault occurrence and potential serious losses are predicted, maintenance must be arranged promptly to nip the fault in the bud. For non-key network devices, the availability of maintenance resources and maintenance costs need to be comprehensively weighed. When resources are tight or maintenance costs are high, maintenance can be appropriately postponed to avoid resource waste caused by over-maintenance. Reasonable selection of maintenance timing can not only ensure stable network operation but also improve resource utilization efficiency, achieving a win-win situation between economic benefits and network reliability[5].

6.2 Maintenance Resource Allocation

Scientific and reasonable allocation of maintenance resources is the key to ensuring timely repair of network faults. It is essential to precisely allocate resources such as maintenance personnel, tools, and spare parts based on the type, scale, and distribution of network faults. For network equipment or areas with frequent faults, it is advisable to increase resource investment. This includes adding more maintenance personnel, equipping them with advanced repair tools, and maintaining an adequate stock of spare parts. These measures can enhance maintenance efficiency and facilitate the rapid restoration of normal network operation. Meanwhile, establishing a maintenance resource-sharing mechanism is crucial. This mechanism helps break down resource barriers and enables the optimized allocation and efficient utilization of maintenance resources across different regions and departments. Through resource sharing, idle resources can be avoided, and wastage can be minimized. This, in turn, improves the overall maintenance capability, ensuring a swift response and effective handling when network faults occur.

6.3 Maintenance Plan Formulation

Formulating a scientific and rational maintenance solution is the core step in resolving network faults, necessitating clear definition of maintenance steps, methods, and scheduling based on specific network fault conditions and predictive results. For hardware faults, promptly replacing damaged equipment or components is crucial to restore normal hardware performance, such as swapping out a failed network switch or hard drive. In case of software faults, implementing upgrades or vulnerability patches enhances software stability and security by introducing new features and closing security gaps. If configuration errors cause faults, reconfiguring network parameters ensures operations comply with standards and resolves issues like connectivity problems or inefficient routing. When formulating the solution, full consideration must be given to the impact on network services, scheduling maintenance during off-peak hours when traffic is low to minimize user disruptions. Continuous and stable network operation is ensured by adopting methods with minimal or no downtime, such as hot-swapping, redundant systems, and rolling updates, which significantly improves user satisfaction by reducing interruptions in network-dependent tasks. In summary, a well-crafted maintenance solution is vital for effectively addressing network faults, ensuring seamless and reliable network operation, and ultimately enhancing user experience.

7. Conclusion

This paper has studied computer network fault prediction and proactive maintenance strategies based on data mining. By analyzing the types and causes of network faults, it has expounded on key data mining technologies, constructed a network fault prediction model based on data mining, and formulated corresponding proactive maintenance strategies. The research results show that data mining technology can effectively mine potential patterns and rules in network data to achieve accurate prediction of network faults, and proactive maintenance strategies based on fault prediction can take maintenance measures in advance to reduce the network fault occurrence rate and improve network reliability and stability.

Although this study has achieved certain results, there are still some shortcomings. For example, in the data preprocessing process, the handling of data quality issues in complex network environments is not yet perfect; in terms of model construction, the generalization ability and adaptability of the model still need to be further improved; in terms of formulating proactive maintenance strategies, the scientificity and rationality of maintenance decisions need to be further optimized. Future research can be carried out from the following aspects: first, in-depth research on data preprocessing methods in complex network environments to improve data quality; second, explore more advanced model construction methods and algorithms to improve the accuracy and generalization ability of fault prediction; third, combine artificial intelligence technology to optimize the process of formulating proactive maintenance strategies to achieve more intelligent and automated network fault prediction and proactive maintenance.

References

[1] Zhang Tianze. Design and Verification of a Reconfigurable Accelerator for Vibration Signal Fault Prediction [D]. University of Electronic Science and Technology of China, 2025.

[2] Zhu Jianbin. Exploration of the Application of AI in Computer Network Fault Prediction and Automatic Repair [J]. China Management Informationization, 2025, 28 (04): 154-156.

[3] Zhang Dacheng. Analysis of a Computer Network Fault Prediction and Adaptive Recovery System Based on Cloud Computing [J]. Electronics Technology, 2024, 53 (09): 314-315.

[4] Lang L, Yonghong H, Xingming L. Study on the mining method for specific fault data of multimedia sensor networks in cloud computing environment [J]. Multimedia Tools and Applications, 2017, 76 (16): 17113-17128.

[5] Liu Zhen. Mining of Fuzzy Association Rules for Alarms in Multi-Domain Distributed Networks [D]. University of Electronic Science and Technology of China, 2015.