Research on Intellectualized Network Fault Diagnosis System based on Big Data Analysis Technology

Kong Lu

Jiangxi Modern Polytechnic College, Nanchang, 330095, China

Keywords: big data analysis technology; intellectualization; network fault diagnosis system

Abstract: With the continuous progress of China's science and technology, network technology has been widely applied in all walks of life. Various new types of network business expand the entire scale of network. The network environment is becoming complex, and users put forward higher requirements for the reliability and service quality of network system, thus traditional network system maintenance mode cannot satisfy the basic demand of current network application. This paper studies network fault diagnosis based on big data analysis technology, and designs a diagnosis system with high intellectualization, which can comprehensively analyze and process different network faults. It has good application effect and deserves popularization.

At present, people increasingly depend on network, the scale of network is expanding, various new types of network business and equipment are widely applied, and the probability of network fault is simultaneously increasing, thus, network management faces great challenges. Scientific and efficient network management can constantly monitor the operation of various equipment in the system, timely find out fault problems, and remind network management staff of taking measures to ensure the stable operation of network system. Big data analysis technology, with good application effect, can monitor various types of fault, improve the overall performance and reliability of network, at the same time, enhance the work efficiency of network management staff and avoid serious fault problems.

1. Network Fault Diagnosis

1.1 Basic model and main background

The basic task of network management is monitor network fault so as to comprehensively improve the reliability of network. Scientific and effective network fault monitoring can perform necessary diagnosis on fault in network system before users finding network fault, solve network problems and put forward suggestions and basic planning about the overall transformation of network system [1]. Network fault diagnosis is mainly to examine carefully relevant parameters such as basic performance of network equipment and network flow to judge the basic operation of network. Various problems in network performance and network system fault can bring big loss to users. Therefore, it is urgent for operators to take reasonable measures to realize accurate and reliable network system fault prediction by network system parameters. Traditional network fault diagnosis is mainly to perform visualized analysis and judgment based on causality. Many enterprises proposed solutions through this method. However, this method requires effective model built on propagation mode and specific causes of network fault, which requires enterprises to invest a large amount of capital and time, with higher cost and lower efficiency. Thus, it cannot adapt to current complex network business and network environment.

With the progress of science and technology, big data is widely applied. The fault diagnosis and analysis based on this technology has been used in product testing, quality control and production process. It is not necessary for this method to model fault propagation and causes, which just requires the model of statistical analysis to determine the basic types of fault. The most commonly used method is threshold fault monitoring, which realizes set threshold value through specific network parameters and judge whether specific parameter is within the threshold scope. Main problems of this method is that efficient processing of data below threshold value is lack, and many detailed problems related to network status are easily ignored. At the same time, the determination of threshold value always depends on the work experience of technicians, there is a lack of standard, and the uncertainty is too much high. In order to solve these problems, it is required to deeply study network anomaly and improve the utilization of monitored information of network system. Using statistical theory, the quantitative analysis is conducted on the network status. It can be assumed that the parameter set Xt monitored by the network parameter at time t is the same as the parameter set Xt+1 monitored at time t+1, and the final threshold is determined asE（X）=ΣXt. Once there is deviation in the network monitoring parameters, the statistics of the parameters will be obviously abnormal from the normal conditions. By analyzing the significance of this abnormality, the basic situation of the network fault can be determined [2].

1.2 Basic method of network behavior fault diagnosis

At present, network fault system can extensively collect network information, and various types of parameters will be stored in time sequence to form complete time series, which belongs to statistical method to explore inherent law of dynamic data resources and can analyze basic law existing in the system according to history of a certain period. To obtain information valuable for maintaining network safety from mass information is the research focus currently. If the abnormal network system parameters are detected, it indicates that there will be fault in network or related equipment. Using the random interference term and the data of a specific time period, a mathematical model can be established, and the network parameter set Xt is obtained according to the white noise and the response intensity. Once there is fault in network system, the mathematical model will generate a relatively large deviation. In this way, network fault can be effectively diagnosed and processed. This method requires simple calculation and small data volume, and it can realize the rapid diagnosis on fault based on stable operation of network system. Meanwhile, it can realize the pre-judgment on various faults, and perform comprehensive transformation and upgrading of network management system. Besides, operators can improve the processing efficiency of network fault and increase current utilization of network system [3].

1.3 The performance verification of scheme

The network fault diagnosis using time series analysis can help network managers find various anomalies in time and improve work efficiency. In the data collection of network equipment, the network anomaly early-warning system can comprehensively analyze the time series and realize the advance prediction of various faults. In the specific network anomaly early-warning, it is necessary to determine the basic parameters of the mathematical model reasonably, and establish a model in normal situation. Once there is deviation in network parameters, the system will alarm the first time [4].

In the actual operation process, the AR model can be used for mathematical modeling to determine whether the data result satisfies the basic requirements of stationarity, and analyze the residuals to finally determine the network fault. In order to verify the accuracy of the results, the Q-Q statistical graph can be used to perform detailed residual numerical analysis to determine whether it conforms to the normal distribution. If the residual has a strong linear characteristic after data analysis and processing, and the basic situation of the normal distribution is satisfied, it can be determined that there is no network abnormality in the data.

At present, network operators require to determine various types of faults that may exist in the system in a short term, and also need to reduce false non-fault judgments to a certain extent. The use of big data analysis technology in fault diagnosis is correct significantly than traditional threshold detection mode. The false alarm rate is also lower than traditional detection method. Although the difficulty of the overall calculation has increased, the fault analysis effect is remarkable, and the work efficiency can be improved comprehensively, which is worthy of further study.

2. The Intelligent Network Analysis Positioning System

The intelligent network analysis positioning system belongs to the auxiliary network operation system, and has the functions of data visualization, data analysis, data fault diagnosis, data acquisition, etc. According to the design scheme of the open source software, the system has a relatively strong expansion capability, and the system can offline or online detect various abnormalities in the network system, provide an accurate display of the network health status, and support data collected in different formats and from different data sources, including configuration files, system log, and mixed data files of indicator data [5]. The intelligent network analysis positioning system has a variety of fault diagnosis technologies, which can judge network element and network faults, comprehensively detect numerical and event type abnormal conditions, determine the basic cause of abnormal network status, and provide an effective solution for repairing faults. The algorithm of the intelligent network analysis positioning system is highly targeted and can effectively detect different types of faults. Users can also use various algorithms provided by the algorithm library to select according to the actual situation.

2.1 The basic system structure

The intelligent network analysis positioning system mainly includes a data application layer, a data analysis layer, and a data acquisition layer, and these three levels are distributed from top to bottom. The data application layer can provide different applications for the system. According to the actual needs, the user can use the basic graphical interface to perform reasonable configuration of the application in the system, and finally realize the custom processing of the network management capability. The system can also analyze basic algorithms customized in data analysis layer and the data resources developed by other types to be used reasonably in the specific application layer. The data analysis layer can provide users with specific diagnosis methods for different kinds of faults, and comprehensively analyze the data resources to determine the fault types and provide suggestions for fault repair. The data collection layer can collect various types of data resources through the bottom network equipment, and perform preliminary processing to store data in a structured basic mode for use to other layers [6].

2.2 Data analysis layer

The data analysis layer has the ability of network fault analysis and diagnosis and comprehensive prediction and prevention, including fault prevention and determination, fault analysis and abnormal condition detection. Fault analysis and anomaly detection mainly include performance index analysis, protocol analysis, configuration analysis, alarm analysis, log analysis, etc. These functions are based on the big data analysis mode to find network faults by comparing different dimensions such as space and time, and can effectively solve various problems in the manual analysis mode. Among them, the anomaly detection in spatial dimension is to carry out horizontal comparative analysis on equipment. The anomaly detection of the time dimension compares the historical data with the real-time data, and uses the characteristics of various data analysis to establish a basic matrix of network behavior, including KPI time series data.

Log anomaly diagnosis is the main content of network operation and maintenance. The software and hardware running information of the equipment is recorded in the system log. The analysis of these information can be used to obtain the specific running status of the equipment, and comprehensive optimization and troubleshooting. It can also be used for similarity detection, regression analysis, random forests, neural networks, etc., to achieve timely fault diagnosis through the scientific detection mode. During the actual operation, the abnormality of the equipment can be clearly detected according to the change of the log time characteristics. The specific change is mainly the change of the time interval. There is a specific precedence relationship between the new event and the old event, but the overall time interval remains within the characteristic range. Once there is an abnormal situation in equipment, the interval between the two will be significantly beyond the normal range, and if the time interval between the events occurred exceeds a certain range, it can be considered that there is fault in equipment. In the normal operation of the equipment, a certain amount of logs will be generated from time to time. In the case of anomaly, the amount of logs will increase significantly, and a new type of log will appear when there is hardware or software fault in equipment. The detection of the similarity of the two phases before and after the log can detect the abnormality of the equipment. When it is normal, the overall log status is stable. The logs in each time period have a high similarity. When it is abnormal, the logs will be different [7].

2.3 Data acquisition layer

The data acquisition layer can collect information such as interface utilization, memory usage, processor, notification message data, and system log data, which are online and offline. The cluster equipment such as Flume and third-party servers can be applied to analyze the actual operation of data. The offline data can be imported into the analysis system by various means to realize data analysis. The data acquisition layer connects the whole system, supports the docking of multiple equipment, and can also adapt the format to achieve standardized storage of data resources, avoiding excessive differences before the format.

2.4 KPI anomaly detection

The time series of the same type of indicators of different ports or equipment constitutes a specific matrix of key performance indicator data. The detection of ports or time points with abnormal KPI conditions can be performed by multivariate analysis. There is KPI time series of specific equipment in each column of KPI matrix, and it also can be composed by specific KPI density or frequency in the port of equipment. Since the number of equipment and ports in a network system is large, the characteristic matrix as a high-dimensional matrix cannot be directly analyzed, so it is necessary to use a subspace method to solve such problem. The subspace method divides the multivariate time series into an abnormal modal space and a normal modal space, and completes fault diagnosis through effective analysis of the principal components. The subspace segmentation can maximize the variance, determine the normal vector with different numbers, and use the 90% orthogonal vector to form the normal state space, and the other part is the abnormal space.

2.5 The monitoring of anomaly in configuration

The main cause for the abnormality in the network system is that the system configuration has an abnormal situation. Therefore, all the commands in the configuration file need to be extracted during the monitoring, and the keywords included in these commands are counted and classified, and the command with the highest frequency is set to main configuration template, to be compared with other configurations to determine the fault occurrence. This detection method easily leads to the formation of additional templates, thus decreasing the accuracy of statistical results and leading to unnecessary calculations [8]. Thus, it is feasible to first determine the fault, through the Bayesian network, decision tree, fault tree for diagnosis, determine the main cause of the fault, and adopt the knowledge of the system operation fault to effectively query the performance and data of various network components, determine the basic cause of the fault and achieve effective processing.

3. Conclusion

Big data analysis technology can promote the intellectualization of network fault diagnosis, improve fault processing efficiency and solve various problems in manual detection mode, playing a significant role in the development of China’s network system. Relevant technicians shall conduct in-depth research to comprehensively improve the application.

References

[1] Yu Ying, Zhu Zhengguo, Huang Chao, Deng Kun. Fault trend judgment for distribution network based on big data analysis [J]. Chinese Journal of Power Sources, 2018,42(01):132-134+146.

[2] Sun Shisheng, Sun Qing. Analysis on Effect and Measures of “Information Cocoons” of New Media in the Era of Big Data [J/OL]. New Media Research, 2018(22):7-10.

[3] Di Kejin. Research on Development Trend of Artificial Intelligence in Big Data [A]. Beijing Science & Technology Information Society. The Symposium of Information First in Smart Technology Development Forum –Academic annual meeting of Beijing Science & Technology Information Society in 2018 [C]. Beijing Science & Technology Information Society: Beijing Science & Technology Information Society, 2018: 7.

[4] Li Shuqin. The Design and Realization of Big Data Platform in Colleges and Universities Faced with Intelligent Decision [A]. Network Application of China Computer Users Association. The Symposium of annual meeting of The 22th Network New Technology and Application in 2018 - Network Application of China Computer Users Association [C]. Network Application of China Computer Users Association: Beijing Key Laboratory of Information Service Engineering of Beijing Union University, 2018: 4.

[5] Qin Yanping, Wang Nan, Rui Xiaohua. Challenges and Countermeasures in the Construction of Ideology Discourse Power in China’s Cyberspace in Big Data [J]. Police Research & Exploration, 2018(11):15-16.

[6] Chen Haoying. The Innovative Application of Communication Technology of Computer Remote Network in the Era of Big Data [J]. Electronic Technology & Software Engineering, 2017(20):33.

[7] Li Shuifang. Clustering Analysis of Abnormal Network Traffic Based on Spark Platform [A]. China Computer Federation. The Symposium of the 33th National Computer Security Conference [C]. China Computer Federation: CCF TCCS, 2018: 5.

[8] Zhang Situo, Wu Liu, Cai Yaoguang, Cheng Xiaorong, Wei Chang. Research on Data Network Fault Management Model based on Alarm Correlation [J]. Electric Power Information and Communication Technology, 2014,12(04):114-118.