Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 8, No. 3, 2024

Research on Equipment Fault Identification Method and System based on Big Data Correlation Analysis

Dongxu Yang, Ran Huo, Jingquan Jia, Linxuan He

FAW Group Engineering and production logistics department - General Assembly Process Department, Changchun, Jilin Province, China

Abstract: With the rapid progress in the industrial field, equipment fault identification plays a crucial role in improving production efficiency and reducing operating costs. This article proposes a specialized and efficient equipment fault identification method and system based on big data correlation analysis. This system achieves precise identification of equipment faults by finely collecting and preprocessing equipment operation data, deeply exploring potential association rules between data. The method proposed in this article not only improves the accuracy of fault recognition, but also significantly enhances recognition efficiency, providing strong support for intelligent and refined management in the industrial field.

Keywords: Big Data; Correlation Analysis; Equipment Fault Identification.

1. Introduction

In the industrial field, equipment failures are one of the main reasons for decreased production efficiency and increased costs. Traditional equipment fault identification methods mainly rely on manual experience and regular maintenance, which have problems such as low identification efficiency and poor accuracy. With the continuous development of big data technology, more and more researchers are exploring equipment fault identification methods based on big data.

This article proposes a device fault identification method and system based on big data correlation analysis. Through the collection, preprocessing, correlation analysis, and fault identification of device operation data, efficient and accurate identification of device faults is achieved.

2. Principles of Big Data Correlation Analysis

Big data correlation analysis is a method of discovering data patterns by mining potential relationships between data. In equipment fault identification, correlation analysis can help us discover the inherent relationship between equipment operation data and faults, thereby achieving accurate identification of equipment faults. The basic principles of association analysis include data preprocessing, feature extraction, association rule mining, and result evaluation.

3. Equipment Fault Identification Method

The equipment fault identification method based on big data correlation analysis mainly includes the following steps:

Figure 1. System WEB Display

1. Data collection: Real time collection of various data during the operation of online devices through network monitoring devices, including key performance indicators such as CPU utilization and memory utilization. During the data collection process, the system pays special attention to the port status and flow situation of the device. By collecting real-time data on port traffic, incorrect frame rates, and broadcast frame rates, the system can accurately assess the quality and efficiency of network transmission.

2. Data preprocessing: Clean, denoise, and standardize the collected raw data to eliminate outliers and noise in the data.

3. Feature extraction: Extract key features from preprocessed data that can reflect the operating status of the equipment, and construct feature vectors.

4. Association rule mining: Using association analysis algorithms to mine feature vectors and discover association rules between equipment operation data and faults.

5. Fault identification and warning: Based on the mined association rules, real-time analysis of equipment operation status is carried out. When abnormal modes are found, fault warnings are issued in a timely manner.

The system is divided into four modules, namely network monitoring module, application monitoring module, analysis warning module, and simulation testing module.

Figure 1 shows the web end display of the system. The menu is divided into 9 items, including data quality monitoring, vehicle anomaly monitoring, big data analysis, and system monitoring, each of which can be analyzed.

4. Design and Implementation of Data Monitoring and Early Warning System

The design and implementation of a data monitoring and early warning system are described from two aspects: overall system functions and system algorithm interfaces.

4.1. Overall System Functionality

The data monitoring and early warning system is a monitoring platform that integrates data collection, data analysis, data calculation, data display, data monitoring, and early warning. The overall system functions are divided into two major parts: business modules and technical architecture.

(1) Business module

The business module is written and implemented in Java code, and its main function is to monitor, warn, and display the overall situation of the device.

Network monitoring: monitors the physical links of the network and identifies underlying physical faults for early warning.

Application monitoring: Monitor the data processing and flow process of the application system, and refine the monitoring granularity.

Analysis and warning: Integrating monitoring data from multiple links and devices, conducting fusion analysis, establishing a multi scenario warning model, and accurately and quickly locating faults.

Simulation testing: Simulate real environmental data, conduct system online simulation testing in advance, and automatically determine test results.

System management, including user management, role management, permission allocation, menu management, and log management.

(2) Technical architecture

After the successful construction of the high concurrency, high scalability, and high availability monitoring and early warning platform studied in this system, it can support the expansion of intelligent devices in various fields (SNMP, IPMI v2.0/v1.5 protocol), with only the adaptation of protocols and databases (Oracle, SQLServer, DM, Sybase, Informix, DB2, MySQL, PostgreSQL, etc.). The technical architecture is shown in Figure 2.

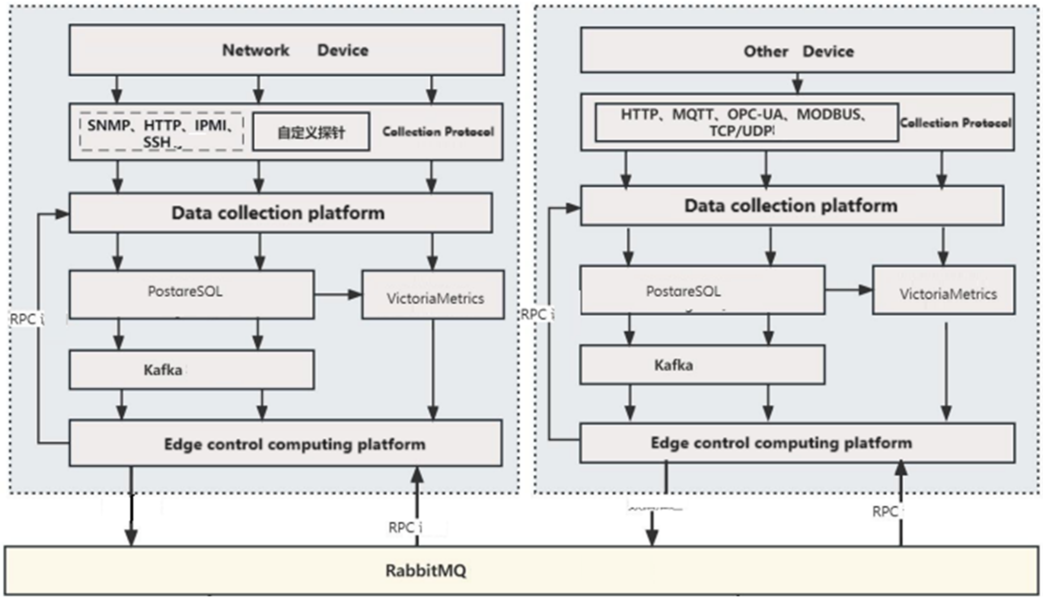

Figure 2. Overall Technical Architecture of the System

4.2. System Algorithm Interface

The algorithm platform is implemented using Java integrated Python script coding. The purpose of this design is to fully utilize the flexibility and scalability of Python coding in big data analysis, and to fully utilize the shell of the Java framework to make it an independent service. It can also integrate with the provided API and business platform to achieve the effect of integrating models. The integration solution adopts a technical approach of placing Python code as a folder in the Java framework's Resource folder, and calling and executing Python scripts through the command line in the code.

5. Predictive Model Algorithm

Realize advance prediction function through time series prediction algorithm. This algorithm predicts future data based on historical data, and the predicted results can be used to predict information such as IT infrastructure performance and resource usage. It also provides reminders, alarms, and other functions in case of performance degradation or resource shortages. Time series prediction algorithms usually use statistical analysis methods such as autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) to analyze the trend and periodicity characteristics of time series, respectively

Time series prediction algorithm is a method used to analyze time series data and predict future values. It usually predicts future data points based on patterns and trends of historical data. This system mainly applies three commonly used time series prediction algorithm models: autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA).

(1) Autoregressive model (AR):

The autoregressive model predicts future values based on the historical values of the time series itself. This model assumes a linear relationship between future values and past values.

The mathematical expression of the AR model is:

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$$

Among them, X_t is the value of the time series at time point t, C is a constant term, P is the order of autoregression, ϕ_i is the parameter of the model, and ε_t is the error term.

The parameters of the AR model can be estimated by minimizing errors, such as the least squares method.

(2) Moving Average Model (MA):

The moving average model is a method of predicting using the residuals of a time series (the difference between observed and predicted values). It assumes that future values are related to past error terms. The mathematical expression of the MA model is:

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i},$$

Among them, X_t is the value of the time series at time point t, μ It is the average value, q is the order of the moving average, θ_i is the model parameter, and ε_t is the error term.

Similar to the AR model, the parameters of the MA model can be estimated by minimizing errors.

(3) Autoregressive Moving Average Model (ARMA):

The ARMA model combines the AR model and the MA model, comprehensively utilizing the autoregressive and moving average properties of time series. The mathematical expression for the ARMA model is:

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Among them, p is the order of autoregression, q is the order of moving average, and ϕ_i and θ_i are the parameters of the model.

The parameters of the ARMA model can also be estimated by minimizing errors.

5.1. Methods for Implementing and Fitting Time Series Prediction Models:

Statistical software tools: Statistical software tools (such as R, Python's Statsmoodels library, etc.) can be used to implement and fit time series prediction models.

Parameter estimation: The parameters of a model are usually estimated by minimizing errors, including least squares methods.

Model diagnosis: After fitting the model, it is necessary to diagnose the model, check the fitting degree and residual properties of the model, in order to evaluate the effectiveness of the model.

5.2. Scenarios Applied to Equipment Fault Identification:

Performance prediction: Using historical performance data to predict future performance trends and help managers plan resources reasonably.

Resource usage prediction: Predict the future demand for resources, help managers adjust resource allocation in a timely manner, and avoid resource shortages.

Early warning and alarm: When the predicted results indicate performance degradation or resource shortage, the system can automatically trigger reminders, alarms, and other functions to remind managers to take corresponding measures.

The time series prediction algorithm is a powerful tool that can help IT managers discover problems and predict trends in a timely manner, thereby better managing and maintaining equipment. In practical applications, it is necessary to select appropriate prediction models based on specific situations, and make appropriate parameter adjustments and model diagnosis to ensure the accuracy of prediction results.

These models can be implemented and fitted through statistical software. For the performance prediction and resource utilization prediction of factory equipment, these models can be used to analyze historical performance data, predict future performance trends and resource requirements, so as to detect performance degradation or resource shortages in advance, and trigger functions such as reminders and alarms.

6. System Applications

In the era of big data, the application of algorithmic models is becoming increasingly widespread, especially in the field of device monitoring, where predictive models play a crucial role. The prediction model proposed in this article mainly focuses on monitoring equipment operation data, achieving efficient and accurate prediction of equipment status, and providing strong support for enterprise production and operation.

Equipment monitoring is an indispensable part of industrial production processes. Through real-time collection and analysis of equipment operation data, abnormal situations of equipment can be detected in a timely manner, preventing faults from occurring, and improving production efficiency. However, traditional equipment monitoring methods often rely on manual experience and regular inspections, leading to issues such as incomplete data collection and inaccurate analysis.

The prediction model proposed in this article utilizes big data technology and machine learning algorithms to deeply mine and analyze equipment operation data. By learning from historical data, the model can capture the changes in device performance and predict future operational status. This prediction not only includes the basic performance indicators of the equipment, such as temperature, pressure, flow rate, etc., but also predicts the probability of equipment failure, remaining life, etc.

7. Conclusion

This article integrates the business platform developed in Java and the algorithm platform developed in Python, and studies a fusion model optimized through technical solutions. Based on this, the design of a data monitoring and early warning system is implemented. This fusion model was proposed after conducting research on the current situation of the automotive industry, collecting user needs, and comparing and debugging Java and Python fusion solutions. Improvements and optimizations have been achieved in terms of functionality, technical architecture, technical operations, and business applications.

In summary, the predictive model proposed in this article has broad application prospects in the field of equipment monitoring. By achieving precise monitoring and prediction of device operation data, this model helps to improve production efficiency, reduce operating costs, and promote intelligent and refined management in the industrial field.

References

[1] Cheng Xueqi, Jin Xiaolong, Wang Yuanzhuo, et al. Overview of Big Data Systems and Analysis Technology [J]. Journal of Software, 2014,25 (9): 1889-1908.

[2] Zhao Li Design and Implementation of Vehicle Monitoring System Based on Big Data Processing [D]. Beijing: Beijing Jiaotong University, 2019.

[3] Zeng Xianyu Optimization and Implementation of Vehicle Monitoring System Based on Big Data Technology [D]. Changchun: Jilin University, 2016.

[4] Li Guojie, Cheng Xueqi. Big data research: a major strategic area for future science and technology and economic and social development - research status and scientific thinking of big data [J]. Journal of the Chinese Academy of Sciences, 2012, 27 (6): 647-657.

[5] Zeng Lei. Review of Big Data Research [J]. Software Guide, 2015,14 (8): 1-2.

[6] Fu Yan, Li Hongru. Research on the Digital Transformation System Architecture of Vehicle Enterprises [J]. Automotive Digest, 2022 (7): 20-26.

[7] Zhang Wenjie, Li Changlong, Hong Yu, et al. The Application of Remote Diagnosis System Based on Big Data Technology in Automobiles [J]. Automotive Digest, 2022 (4): 30-33.

[8] Chen Liang Research on Electric Vehicle Battery Monitoring System Based on Cloud Platform [D]. Shanghai: Shanghai University of Engineering and Technology, 2017.

[9] Wang Ruwei The Application of Java Programming Language in Big Data Development [J]. Electronic Technology, 2022, 51 (1): 160-161.

[10] Guo Yang, Chang Yingxian. On the Application of Java Language in Computer Software Development [J]. Digital Communication World, 2022 (1): 88-90+94.

[11] Zhang Jiaqi Design and Implementation of a Vehicle Test Management Information System Based on Java Web [D]. Beijing: Beijing Jiaotong University, 2016.

[12] Lu Shaobing. Research on Hybrid Language Programming and Its Implementation Based on Python [J]. Science and Technology Information, 2022, 20 (14): 31-33.

[13] Zhang Wei. Implementation of Vehicle Detection Algorithm Based on Python [J]. Mechanical Management Development, 2019,34 (12): 258-261.

[14] Cheng Yi, Cui Jiqiang, Wang Ruiying. Analysis of heavy-duty dump truck operating conditions based on Python [J]. Automotive Practical Technology, 2019 (20): 58-60.

[15] Zhao Zhisheng, Zhang Xiao, Ma Teng Research and Implementation of Java and Python Integration Scheme [J]. Network Security Technology and Application, 2015 (1): 146+148.

[16] Zhang Chunguang Design and Implementation of a Comprehensive Quality Assessment System for Java and Python Mixed Source Software [D]. Nanjing: Southeast University, 2021.

[17] Huang Jingxuan, Liu Kangjun, Yun Chentai, et al. Design and Implementation of an Assisted Bid Evaluation System Based on Java and Python [J]. Industrial Control Computer, 2021,34 (3): 18-20.

[18] Shen Ao, Wang Lei, Wang Xidian, etc Research on the Application of Jython Technology in Network Management Support Systems [J]. Telecommunications Engineering Technology and Standardization, 2013,26 (5): 52-56.

[19] Tan Yi, Wang Kai, Zhou Lanjiang. Rapid development of web applications using Jython [J]. Computer Knowledge and Technology, 2006 (35): 50+72.

[20] Zeng Hao Research on the Application of Python in Other Languages [J]. Popular Technology, 2012,14 (7): 13-14+19.

[21] Wang Wenyang, Chen Zheng, Chen Xiangwei. Design of New Energy Vehicle Monitoring Platform Software [J]. Equipment Manufacturing Technology, 2014 (12): 46-48.

[22] Zhang Yu, Che Xiaobo, Wang Yong, et al. Optimization Design of Remote Monitoring System for New Energy Vehicles [J]. Modern Electronic Technology, 2019,42 (3): 96-100.

[23] Aoyama K I. Universal Traffic Management System (UTMS) in Japan [C]//Proceedings of VNIS'94 1994 Vehicle Navigation and Information Systems Conference, Yokohama, Japan: IEEE, 1994.

[24] Ostrowski K, Birman K, Dolev D. Extensible Architecture for High Performance, Scalable, Reliable Publish Subscription Events and Notification [J] International Journal of Web Services Research, 2007, 4 (4): 18-58.

[25] YANG J. Web service componentization [J] Communications of the ACM, 2003, 46 (10): 35-40.