Research on Analysis Method of Information System Operation Discipline

Yintie Zhang, Lijun Wang*
State Grid Electric Power Research Institute, Nanjing, China

*Corresponding author e-mail: wanglijun@sgepri.sgcc.com.cn

Abstract. Compared with traditional operation and maintenance, intelligent operation and maintenance can better meet the needs of the big data era. At present, the intelligent analysis technology of operation and maintenance data is not perfect enough, and still has a large research space. Intelligent operation and maintenance technology can help operation and maintenance personnel to find out the cause of abnormal operation of system and prevent system failure. The discipline of a large number of operation and maintenance data can help the operation and maintenance personnel to discover the potential problems of the system, and transform passive maintenance into active prevention. But most systems only consider the trend of a certain operation and maintenance data over time, instead of the discipline of the data collected at the same time. According to the characteristics of operation and maintenance data, this paper selects unary linear regression model to simulate the discipline of operation and maintenance data obtained at the same time, and mines the regularity of data from another perspective. The results of experiments verify reasonableness and effectiveness of this method.

1. Introduction
With the wide and deepening of IT applications, more and more business needs require the support of IT systems, and each enterprise attaches great importance to the maintenance and management of computer hardware and software resources. A large number of operation and maintenance technology tools and operation and maintenance management tools record the generated data at all times, and each information system operation also generates a large amount of process data. The performance monitoring data generated during system operation and maintenance, as well as the log files of various systems, constitute the company’s huge information system operation and maintenance data [1]. This includes asset data, performance data, log data, document data, video data, paper data, etc. Operation and maintenance personnel can know the operating status of IT equipment based on these operation and maintenance data. When the information system runs abnormally or fails, the source of the fault can be traced back from the error log to determine whether the fault is caused by the system itself or by the application, so as to help the operation and maintenance personnel deal with the fault in time and make the system return to normal operation quickly [2].

There are two major problems in the traditional operation and maintenance system: first, it requires a lot of manual operations, and the operation and maintenance efficiency is low. Second, the failure warning mechanism is imperfect, and the operation and maintenance personnel are in a passive
working state. With the wide and deep application of IT applications, operation and maintenance data has grown exponentially, and there is an urgent need to automatically maintain large amounts of data. In the era of big data, how to effectively manage IT operation and maintenance is a key issue for every IT operation and maintenance personnel [6]. However, traditional operation and maintenance system has not been able to meet the growing business needs. The intelligent operation and maintenance system introduces data analysis and machine learning technologies on the basis of the traditional platform, which improves the automation degree of the system and can proactively predict system anomalies and failures. Operation and maintenance personnel can be freed from some tedious and repetitive work, focus on risk warning and system fault monitoring, which can effectively improving the operation and maintenance level of IT systems [9].

The existing intelligent operation and maintenance system is not enough to mine the data. For example, it is rare to distinguish between workdays and non-workday data or collect data by moment, thereby further mining the operation and maintenance data. Therefore, there is still a large research space for intelligent operation and maintenance technology.

2. Operation and maintenance intelligent transformation opportunities

2.1. Traditional operation and maintenance

With the continuous deepening and improvement of IT construction, enterprises attach great importance to the operation maintenance and management of computer hardware and software resources. In traditional operation and maintenance system, most of the update management work requires manual operation. The increase in the number of hardware and software devices makes IT operation and maintenance more complex, and a large amount of human resources need to be invested. In addition, due to the lack of a fault warning mechanism, the operation and maintenance personnel can perform maintenance after the abnormality or failure occurs. They work in low efficiency and are in a passive operation and maintenance state for a long time, the system troubles cannot be detected and processed in time [5, 10].

Because of the wide and deep application of IT applications, the operation and maintenance data has grown exponentially. It is urgent to automatically maintain a large amount of data. Traditional operation and maintenance cannot meet the needs of existing business development. The system needs to introduce data analysis, machine learning and other technologies to discover the operation and maintenance rules, improve the automation and intelligence degree, realize the proactive prediction of future faults, effectively avoid risks, and enhance the work efficiency of maintenance personnel and the capability of failure detection and risk warning. Thereby effectively improving the operation and maintenance level of IT systems and enhancing the competitiveness of enterprises [8].

2.2. Intelligent operation and maintenance

Compared with traditional operation and maintenance, intelligent operation and maintenance can improve the system automation, and meet the needs of existing business development. In the traditional system, the operation and maintenance personnel passively handle the faults after the monitoring system triggers the threshold to generate a warning. While the intelligent operation and maintenance platform can automatically handle system malfunction. The operation and maintenance personnel can be freed from the cumbersome and repetitive work, and devote themselves to risk analysis and fault handling, which can reduce the IT maintenance cost to a certain extent, and improve the processing speed of faults [5]. In addition, the intelligent operation and maintenance system can establish predictive models through data analysis, machine learning and other technologies, predict certain problems in advance, objectively and accurately capture the symptoms before the failure occurs, proactively predict the faults, effectively avoid risks, and improve the efficiency of operation and maintenance personnel and the utilization of data [4]. Operations and maintenance personnel's working status has changed from passive maintenance to active prevention, which can prevent system
failure according to a large number of data, thus improving the ability of fault monitoring and risk warning of system.

In existing intelligent operation and maintenance system, the utilization of data is still not sufficient and the prediction method for discovering data regularity is not perfect. The operational status of some businesses is completely different on weekdays and non-working days, while the system rarely distinguish between data generated on weekdays and non-workdays in the process of analyzing the past operation and maintenance data. In addition, when analyzing data, the system rarely collects data by moment and better discover the operation and maintenance rules according to the data law obtained at a certain moment. Therefore, there is still a large research space for intelligent operation and maintenance technology. This paper will further explore the law of operation and maintenance data from a new perspective.

3. Research on analysis method of information system operation discipline

3.1. The innovation ideas of operation and maintenance data intelligent analysis technology

At present, the research on intelligent analysis technology of operation and maintenance data is still not perfect. In many enterprises, information operation and maintenance is not regarded as a business. It does not have the same development perspective as other businesses. It relies too much on the technical capabilities of operation and maintenance personnel to handle informational events and problems. The data analysis from the perspective of business development is struggling. In addition, enterprises spend a lot of time and energy to build their own operation and maintenance service system, seriously neglecting the accumulation of operation and maintenance data, which is extremely unfavorable for the precipitation of operation and maintenance service knowledge, so that many informational decisions are separated from the data foundation, making data analysis impossible. However, the traditional system cannot meet the requirements of the big data era, and it is necessary to subvert the pre-defined operation and maintenance mode of traditional operation and maintenance. Therefore, the research of intelligent operation and maintenance technology is particularly important.

In general, in the process of analyzing operation and maintenance data, the trend of a certain operation and maintenance index (such as CPU total load rate) with time is only considered horizontally to study the law of data. This paper analyzes the operation and maintenance data from a new perspective, extracts the data at the same moment and considers the change law of data from the vertical angle, and then obtains the normal value (this value is called the model value in this paper) of the total CPU load rate at that time according to the data law, so that it can predict whether the total CPU load rate is abnormal at some time in the future, which can help the operation and maintenance personnel to find out the cause of the abnormal operation of the system or prevent the occurrence of system failure. Passive maintenance becomes active prevention, shifting the focus of operation and maintenance personnel to risk analysis and fault handling, effectively avoiding risks, improving the work efficiency of operation and maintenance personnel, and improving the fault monitoring and risk warning capabilities of operation and maintenance systems. In order to more objectively characterize the law of data, we collect data in a steady state of the system. Then, the obtained operation and maintenance data will be cleaned, and extracting the valid data at a certain moment after the abnormal data removed. Since the system is stable, the data collect at the same time is also in a stable state. Then, the linear regression model is used to predict the model value of a certain operation and maintenance data at a certain time, and the model value is used as the standard value of the system. The operation and maintenance personnel can judge whether the system runs abnormally by the deviation between the actual value of the data and the model value, which helps the operation and maintenance personnel to better complete the maintenance work.

3.2. Calculation method model values

Linear regression [7] is an important method in statistical analysis and is widely used in the research of society, economy, technology and many natural sciences. In statistics, linear regression is a
regression analysis that models the relationship between one or more independent variables and dependent variables using a least squares function called a linear regression equation. This function is a linear combination of one or more model parameters called regression coefficients (the independent variables are all one power). The case of only one independent variable is called simple regression, and the case of more than one independent variable is called multiple regression.

We usually use $X$ to represent the data matrix,

$$X = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^{n \times p} \quad (1)$$

Where $x_i \in \mathbb{R}^p$ represents a data sample with a p-dimension.

The label of the data is usually represented by $Y$, and only consider the case where each sample belongs to one class here.

$$Y = (y_1, y_2, \ldots, y_n)^T \in \mathbb{R}^n \quad (2)$$

In a linear regression model, for a sample $x_i$, its output value is a linear combination of its characteristics:

$$f(x_i) = \sum_{m=1}^{p} w_m x_{im} + w_0 = w^T x_i \quad (3)$$

Where $w_0$ is called the intercept and $x_{i0} = 1$.

In the regression analysis, only one independent variable and one dependent variable are included, and the relationship between the two can be approximately represented by a straight line. This regression analysis is called unary linear regression analysis. According to the characteristics of the acquired operation and maintenance data, this paper adopts the unary linear regression model. For a linear regression, it can be seen that the value of the dependent variable $Y$ varies with the value of the independent variable $X$. Each actual $X$ will have an actual $Y$ value, which we call $Y$ actual. Then we need to find out a straight line, each actual $X$ will have a $Y$ value calculated by the straight line, which we call the $Y$ prediction. The regression line minimizes the sum of the squares of the difference between the actual value and the predicted value of each $Y$.

The predictive model of the unary linear regression equation analysis method is:

$$Y = aX + b \quad (4)$$

Where $X$ represents the independent variable, $Y$ represents the dependent variable, and $a$ and $b$ represent the parameters of the unary linear regression equation.

Linear regression equations are usually solved using the least square method, but this method does not have the most appropriate and reasonable effect under various conditions. For example, when the amount of data collected is small and there are abnormal points in the data, the effect of using the least squares method is not satisfactory [3]. Firstly, there are many regression predictions in practice that require the sum of the absolute values of the total errors to be the minimum, not the sum of squares or the sum of absolute values of errors based on the sum of positive and negative errors to be zero. Secondly, the least squares method is very sensitive to outliers, and the regression line obtained by it is less robust and sensitive to some abnormal points. The method of minimizing the sum of the absolute values of the residuals reduces the sensitivity of outliers to regression coefficients and is stable. Therefore, in some respects, people have been biased to estimate regression coefficients by minimizing the absolute value of residuals. In addition, the data used in this paper is collected under
the condition of stable system. The data is relatively stable and the data volume is small. The regression line obtained by the method of the sum of the absolute values of the residuals with better robustness can better describe the steady state of the operation and maintenance data.

In this paper, we use the sum of the absolute values of the residuals to solve the unary linear regression equation, so that the sum $S$ of the absolute values of the fitting errors is minimized, and the corresponding linear regression equation is obtained.

$$S = \sum_{i=1}^{n} |y_i - \hat{y}_i| = \sum_{i=1}^{n} |Y_i - a - bx_i|$$  \hspace{1cm} (5)

4. Scene application verification and conclusion

4.1. Operation and maintenance source data

As the basis of the intelligent analysis framework of operation and maintenance data, the source data range should include infrastructure layer data, infrastructure layer data, platform layer data, application layer data, business layer data, and network security data for each layer and management scheduling data. According to the data type, the operation and maintenance data can be divided into structured data and unstructured data. According to the data time, the operation data can be divided into historical data and real time data. In order to support the data analysis to achieve the goal, the operation and maintenance data should be able to be acquired and analyzed. In addition, the support methods of operation and maintenance data are also divided into two types, one is to mine value on the basis of existing data, and the other is to prepare support data for a certain value target. The operation and maintenance data intelligent analysis source data support is shown in Figure 1.

4.2. Experimental results

The source data collected in this paper can be any running state of IT assets. Here, the total CPU load rate is taken as an example. Ideally, there are 144 collection points every day and there is one
collection point every 10 minutes. The data collection period is 30 days, and there is no distinction between working days and non-working days.

In a data collection cycle, the total CPU load rate at the same time in the original data is extracted and analyzed by unary linear regression model to predict the model value. When the amount of data collected at the same time is small, and it is more appropriate to use the method of minimizing the sum of the absolute values of the residuals with better robustness. Since the operation and maintenance data collected in a stable state of the system, the data at each moment tends to be stable, and the resulting prediction model is:

\[ Y = b \]  
(6)

The obtained model values are shown in Figure. 2.

![model values](image)

Fig. 2 model values

5. Conclusion
With the continuous development of information technology, the scale of IT architecture in data center is also expanding at the same rate. The number of servers, network devices and storage devices is also increasing, and the network architecture is becoming more and more complex, especially large and very large companies have many branches throughout the country and even around the world. The arrival of big data era makes traditional operation and maintenance system unable to meet the needs of data analysis. The operation and maintenance mode shifts to intelligent operation and maintenance, and intelligent operation and maintenance can make the operation and maintenance system more intelligent, efficient and reliable. At present, the data analysis technology applied in intelligent operation and maintenance system is not perfect enough. Most operation and maintenance systems only consider the trend of a certain operation and maintenance data over time, and do not extract the data at the same time to consider the change regularity of operation and maintenance data from a vertical perspective, and then predict the model value according to the regularity. According to the characteristics of operation and maintenance data, this paper selects unary linear regression model to simulate the regularity of operation and maintenance data collected at the same time, and analyzes the data from another perspective, which helps the operation and maintenance personnel to find the system operation abnormality or prevent the system from malfunctioning according to a large number of data, realizing real-time risk early warning and accurate trend prediction, achieving the operation and maintenance effect of changing passive maintenance to active prevention, the focus of operations personnel is shifted to risk analysis and troubleshooting. It can effectively avoid risks, enhance the
efficiency of operation and maintenance personnel, and improve the fault monitoring and risk warning capabilities of operation and maintenance systems.

6. Acknowledgment
This work was financially supported by the State Grid Corporation of Science and Technology (WBS number: 522723180006).

References
[1] Yingchao X U . Big Data Analysis and Forecast Based on IT Operation Data [J]. China Transportation Review, 2015.
[2] Tian J, Wenting C, Luo Z, et al. Application of Intelligent Platform in Operation and Maintenance of Data Center Infrastructure [J]. Modern Architecture Electric, 2015.
[3] Xie K G, Song Q K, Zhou J Q. A Linear Regress Model Based on Least-Absolute Criteria [J]. Acta Simulata Systematica Sinica, 2002, 14(2):189-192.
[4] Du Y, Corporation Z. Intelligent Operation and Maintenance, an Automatic Operation and Maintenance System Based on Self-learning [J]. Information & Communications Technologies, 2018.
[5] Guo Y, Cai H, Lei G E. Telecom intelligent operation system based on big data grid [J]. Telecommunications Science, 2018.
[6] Jian-hua Z, Center N. Efficient IT Operation and Maintenance Management in the Construction of Digital Campus in the Era of Big Data [J]. Journal of Changchun Normal University, 2015, 12: 012.
[7] Neter J, Wasserman W, Kutner M H. Applied linear regression models [J]. 1989.
[8] Huang T, Huang J. Data center operation and maintenance management measures in big data Era [J]. Electronic Test, 2017.
[9] Emily Misssner, Leslie J. Reynolds. Operations Management/Operations Research Web Sites [J]. Journal of Business & Finance Librarianship, 2003, 83-4:.
[10] Jiang H X, Yang Y Y. Study on comprehensive optimization model of multi parallel IT applications in IT systems management [J]. Computer Engineering & Applications, 2007, 43(27):240-242.