Reliability Analysis of Distribution Network Operation Based On Short-Term Future Big Data Technology

Zhe Jing\textsuperscript{1,a}, Cong Yu\textsuperscript{2,b}, Feiruo Xi\textsuperscript{3}, Fengzhi Wu\textsuperscript{4,d}, Zhixiang Tao\textsuperscript{5,e} and Peihao Yang\textsuperscript{6,f}

\textsuperscript{1}State Grid Ningxia Electric Power Co. Ltd. Yinchuan Power Supply Company, Yinchuan, China
\textsuperscript{2}Ashikaga University Tochigi-Ken, Japan
\textsuperscript{3}Hunan Institute Of Science and Technology University, Yue Yang, China
\textsuperscript{4}Massachusetts Institute Of Technology, Commonwealth Of Massachusetts, US
\textsuperscript{5}Tianjin Chengjian University, China, Tianjin China
\textsuperscript{6}Xi’an Thermal Power Research Institute Co., Ltd., Xi’an, China

\textsuperscript{a}small0406@sina.com; \textsuperscript{b}yucong9503@gmail.com; \textsuperscript{c}1064317970@qq.com; \textsuperscript{d}gbsc999@gmail.com; \textsuperscript{e}bruce789123@163.com; \textsuperscript{f}44960280@qq.com

Abstract. This paper firstly establishes a four-dimensional index system for distribution network operation reliability. Using principal component analysis method, it extracts the main evaluation indicators from a large amount of data, and analyzes the influencing factors of the indicators according to the main evaluation indicators. According to the parallel association rules, the method establishes the relevant model, and extracts the main indicators of operational reliability and the strong correlation rules between each influencing factors, so as to obtain the main influencing factors. The artificial neural network is proposed to predict, based on historical data and real-time data. The main influencing factors are used as the input and output of the forecast, and the output is expressed by the main evaluation index, and the operational reliability index for a period of time is judged. The calculation of the proposed strategy can quickly and effectively predict the operational reliability of the distribution network.

1. Introduction

As a large and complex data set, large data can not be processed in a period of time using conventional tools and software, and has the characteristics of high capacity, fast, diverse and low density [1-2]. The ability to collect and process data of various states is the basis of the implementation of smart grid, which refers to a series of data generated during the operation of the grid. With the rapid development of distribution network and the increasing demand for data, data in various distribution network management systems are collected and applied. For a long time, these data include not only business information such as operation, monitoring, dispatching and maintenance of distribution system, but also a large amount of social and economic information[3-6]. In the field of distribution network research, this kind of data provides a basis for the intellectualization of distribution network.

In this paper, the large data processing method is applied to the reliability analysis of distribution network operation. The large data in the distribution network are collected, stored, calculated and processed. Combining this kind of data with prediction technology, the short-term future nodes are
predicted, and the operation reliability of the distribution network at that time is obtained.

2. Data Base of Distribution Network
To study the operation reliability of distribution network, a large amount of data in distribution network should be obtained first. Due to the different information of the structure, operation status, environment and maintenance plan of each component in the distribution network system, real-time monitoring and acquisition of these data are carried out. When the system is in normal operation, the operation reliability of the distribution network in a period of time is analyzed, and the operation reliability index of the system is obtained [7]. Overall, the data related to reliability indicators are divided into four categories as shown in Table 1.

Table 1: The related data and its sources of operational reliability

| Type          | Data sources                                      | Data classification                                |
|---------------|--------------------------------------------------|---------------------------------------------------|
| Recording data| Distribution Network Monitoring System and Production Management Software Database | Equipment installation account, equipment nameplate parameters |
| Real-time data| Production Management Software Database, Information Acquisition Platform | Voltage, current, active power, reactive power, power factor, frequency and other electricity data, etc. |
| Load data     | Distribution network transformer load monitoring platform, load dispatching and distribution control system, etc. | Real-time load data and load forecasting data of each load node, etc. |

3. Large Data Analysis of Distribution Network

3.1. Reliability analysis model and evaluation index system of distribution network operation
In different scenarios, large data technology is applied to analyze the operational reliability schematic as shown in Figure 1. The purpose of analysis and evaluation of distribution network operation reliability is to put forward operation control strategy, help decision-making, guide dispatching operation, early warning of possible faults, ensure the reliability of distribution network, and improve the security and stability level of power grid in the future [8]. The flow chart of reliability analysis for large data operation is shown in Figure 2.
1) Establishment of distribution network operation reliability index system based on large data processing technology. 2) Principal Component Analysis (PCA) is used to extract the characteristic data of each index by combining with the historical operation data, and the relevant indexes affecting the operation reliability are obtained. 3) The parallel association rule mining method is used to analyze the operation reliability of distribution network, and the main factors affecting the operation reliability are obtained. 4) Input the results of the previous step into the artificial neural network, and the output of step 2 is regarded as the main reliability index of the artificial neural network. The reliability prediction model of the artificial neural network is obtained through continuous training.

The state index is an index describing the state reliability of a system. It focuses on three aspects: health state, critical state and risk state. Degree index is an index to quantitatively describe the reliability of the system, indicating the reliability of the system under normal operation. Hierarchical dimension indicators include system layer, region layer, node layer and element layer. Time dimension indicators reflect the reliability of different time constraints, such as minutes, hours, days, months and so on. According to the different prediction time, the calculation results of time dimension index can be obtained.

3.2. Extraction of Main Indicators Based on Principal Component Analysis

The number of index variables is very large, the calculation is very high, and there is some redundant information between them. Principal Component Analysis (PCA) is a statistical method for large data analysis. Its purpose is to simplify the object model, collect key information and reduce the dimension of variables. It can use the square of variance and variance of analysis data to calculate the information of each index. By deleting the same correlation index, the correlation can be reduced, and the reliability index in most of the original information can be included, and the key index which has great influence on the system can be selected.

1) Standardize evaluation indicators

The feasibility of each index varies in scale and use, so it is necessary to unify the parameters. The index variables obtained from large data analysis are processed by normal distribution and transformed into corresponding normal distribution variables. The transformation process is shown in Formula (1).

\[ Z = (X - \overline{X})/\sigma \]  \hspace{1cm} (1)

In the formula: \( Z \) is the normal distribution variable, \( X \) is the index variable, \( \overline{X} \) is the average value of the index variable, and \( \sigma \) is the standard deviation.

2) Establishing correlation matrix and calculating eigenvalues

The correlation between index variables means that when we know that a variable group changes one of the variables, we can determine the value of other variables. Pearson correlation coefficient can be used to measure the correlation strength between two random variables. Pearson correlation coefficient \( \sigma_{XY} \) is generally defined by variables \( X \) and \( Y \).

\[ \sigma_{XY} = \frac{\text{cov}(X,Y)}{\sigma(X)\sigma(Y)} \]  \hspace{1cm} (2)

3) Obtain the main evaluation index

\( U = [\sigma(f_i, z_j)] \) is the load matrix of the main component factors. The different values of \( \sigma(f_i, z_j) \) in \( U \) correspond to the related factors of \( z_j \) in the \( j \) evaluation index of the main component \( f_i \) in \( i \). Its value range is \([-1,1]\]. With the increasing of absolute value, the correlation becomes larger. Not all information is necessary in evaluating system reliability. Therefore, only some indicators representing the main information can be selected. In the load matrix, for the \( p \) principal component, the index \( z_j \) corresponding to the \( \sigma(f_i, z_j) \) maximum value is selected as the main evaluation index, and the \( q \) indices obtained are the main index of the system reliability evaluation.
3.3. Prediction Model Based on Artificial Neural Network
Using historical and real-time data to analyze the reliability index of distribution network in the future is called reliability prediction of distribution network system. At present, the existing methods include traditional forecasting methods and artificial intelligence methods. Traditional forecasting methods are complicated, and the influencing factors and operation methods are difficult to determine. This paper uses the method of artificial neural network to accurately analyze the relationship between various evaluation indicators.

According to the method of artificial neural network, the technical route of analyzing the operation reliability of distribution network system is as follows: from the collected historical data, according to the parallel association rules mining method proposed above, the main influencing factors index is obtained. The input \( I(T) \) of the artificial neural network model is the main influencing factors index, and then the evaluation time \( t \) is selected by the demand for electricity. The main evaluation indexes obtained by the 2.2-section principal component analysis method use \( T+t \) time as output \( O(T+t) \). At this time, the continuous training model can get the complete model based on the artificial neural network prediction method. The method proposed in this paper can be used to evaluate the reliability index value of the system.

According to different prediction and calculation evaluation indexes, corresponding time indices can be obtained, and system components can be evaluated in different occasions to determine whether the voltage or current inside and outside the system will mutate in the next few minutes or hours. Selecting time scales, such as dates, months and years, can predict the future time of system load reduction probability expectations, available electricity, decision-making control, dispatch and other indicators or power grid planning.

4. Case analysis
The text of your paper should be formatted as follows:
This paper chooses a medium-sized city distribution system as an example, the actual data collection time is 10 years, every 15 minutes to collect and statistics data, each sample as a sample, verify the method, a total of 350,400 data samples. There are nine system-level indicators, including voltage and current expectations and other major indicators. These main indicators are called variables \( X_i (i = 1,2,...,9) \). The distribution of each variable \( X_i \) can be obtained from the historical data collected and the method proposed in this paper. The normal distribution process can get the normal distribution variable \( Z_i \). According to formula (2) and formula (3), the correlation matrix \( R \) of nine index variables \( Z = (Z_1,..., Z_9) \) is calculated. Finally, their eigenvalues \( \lambda_i \) are calculated by formula (4), as shown in table 2. The variance and cumulative variance of each index are obtained by formula (5) and formula (6), as shown in table 2.

| Components            | Characteristic value | Variance/% | Cumulative variance /% |
|----------------------|----------------------|------------|------------------------|
| Principal Component 1 | 6.5882               | 73.202     | 73.202                 |
| Principal Component 2 | 1.2074               | 13.415     | 86.615                 |
| Principal Component 3 | 0.9567               | 10.632     | 97.258                 |
| Principal Component 4 | 0.1375               | 1.525      | 98.774                 |
| Principal Component 5 | 0.0617               | 0.688      | 99.465                 |
| Principal Component 6 | 0.0425               | 0.474      | 99.933                 |
| Principal Component 7 | 0.0057               | 0.063      | 99.994                 |
| Principal Component 8 | 0.0005               | 0.005      | 100.00                 |
| Principal Component 9 | 0.0000               | 0.0000     | 100.00                 |

Index of the system is the one with the largest load rate, and the operation reliability of the distribution network can be evaluated by the main evaluation index. The three main abnormal
evaluation indicators are: load drop rate, power shortage and expected voltage exceeding the standard.

Through the three main anomaly evaluation indicators as samples, based on a certain historical data analysis, a sample is a "matter", called $T$, each transaction contains factors such as influencing factors and evaluation indicators. Using the above method, we get frequent item sets according to the basic database, and divide each item in the frequent item sets into two small sets: the main evaluation index set $A$ and the main evaluation index influencing factor set $B$. According to the method proposed above, we get association rules $a \Rightarrow b$ ($a \subseteq A, b \subseteq B$), calculate the support $s(a \Rightarrow b)$ and confidence $c(a \Rightarrow b)$ of association rules, and then calculate the support degree. Thirty percent and 85 percent confidence are the main factors influencing the evaluation index. According to the artificial neural network prediction method and parallel association rule mining technology proposed in this paper, a prediction model for evaluating the reliability of distribution network operation is established.

Fourteen samples are selected as 14 conditions of this case. Using the scheme proposed in this paper, the specific prediction results of distribution network operation reliability are obtained, as shown in Figure 3-5. Figure 3 shows the load reduction probability under 14 operating conditions. Curve 0 is the condition under normal operation and curve 1 is the condition of load shedding. Curve 1 is obtained by using the method proposed in this paper. Curve 2 is the result of artificial neural network prediction without association rule mining.

Figure 4 is the result of the probability of voltage exceeding limit. Curve 1 is the operating condition of voltage exceeding limit, and Curve 0 is the normal working condition. Curve 1 is the result of voltage overrun obtained by the method proposed in this paper. Curve 2 is the result of artificial neural network prediction without association rule mining.

Figure 5 is the result of the probability of power shortage. Curve 1 is the operation condition of power shortage and Curve 0 is the normal operation condition. Curve 1 is the result of power shortage obtained by the method presented in this paper. Curve 2 is the result of artificial neural network prediction without association rule mining.

According to the above experimental results, the parallel association rules mining technology and artificial neural network prediction technology proposed in this paper can be used to predict the operational reliability of distribution network system. It can predict the running state, scheduling plan and maintenance plan in real-time running environment. The method proposed in this paper is more accurate than the method without association rule mining technology. In addition, the main evaluation indicators are directly used in artificial neural network prediction. According to the method proposed in this paper, the prediction can be completed in 6s. It can easily track the latest socio-economic developments and new equipment operation information. According to these data, the original prediction model can be readjusted to achieve the highest prediction accuracy.

5. Conclusion
The data of distribution network system is huge and complex. The key data of the system can be extracted by the method proposed in this paper. This paper studies the reliability analysis method of distribution network system based on big data technology. It can analyze and predict the distribution network in real time and accurately. The main conclusions of this paper are as follows:
1) The four-dimensional evaluation index system of distribution network operation reliability includes state dimension, degree dimension, level dimension and time dimension. The operation reliability level in different areas such as equipment layer, node layer and system layer is explained in different degrees by normal working conditions, voltage overrun and load shedding.

2) With the increase of the number of equipments, more and more system states are used to evaluate the reliability of the system, and the dimension disaster may be encountered. The principal component analysis method proposed in this paper can reduce the correlation among the indicators, so it can reduce the evaluation dimension, reduce the evaluation workload, improve the efficiency, achieve rapid evaluation, and make real-time decision-making for scheduling.

3) Association rule mining technology and artificial neural network prediction technology can extract useful information from a large number of data, accurately locate the indicators that affect the main factors, reduce the dimension of the input data of the evaluation model, reduce the difficulty of modeling, and realize the rapid and accurate evaluation of the operation reliability of distribution network.

References
[1] Yuan, Jiann-Shiun, Jie Lin, Qutaiba Alasad, and Shayan Taheri. "Ultra-Low-Power Design and Hardware Security Using Emerging Technologies for Internet of Things." Electronics 6, no. 3 (2017): 67.
[2] Sun Baohua, Chen Lei, Xia Dong, etc. Design and application of intelligent operation and maintenance control platform for distribution network based on large data platform [J]. Electrical automation, 2018,40 (06): 81-84.
[3] He Anhong, Xiao Xubing. Distribution network operation and maintenance management and control platform based on intelligent technology and big data technology [J]. Automation application, 2018 (11): 97-98.
[4] Taheri, Shayan, Jie Lin, and Jiann-Shiun Yuan. Security interrogation and defense for SAR analog to digital converter. Electronics 6, no. 2(2017): 48.
[5] Hu Dinglei. Application and Prospect Analysis of Big Data Technology in Intelligent Distribution Network [J]. Computer Products and Circulation, 2018 (09): 67.
[6] Jie Lin and Jiann-Shiun Yuan, A 300 mV, 6-bit ultra-low power SAR ADC.2016 13th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT), Hangzhou, 2016, pp. 713-715.
[7] Hu dinglei. Application and Prospect Analysis of big data technology in intelligent distribution network [J]. Computer products and circulation, 2018 (09): 67.
[8] Feng Lei. Application of big data technology in distribution network [J]. Shandong industrial technology, 2018 (16): 138.