

# Research on Fault Prediction of Distribution Network Based on Large Data

Zhou Jinglong<sup>1,a</sup>, Han Changzhan<sup>1</sup>, Wu Xiangming<sup>1</sup>, An Jiakun<sup>2</sup>, He Chunguang<sup>2</sup> and Han Jinglin<sup>2</sup>

<sup>1</sup>State Grid Hebei Electric Power Company, No.32, fuqiang street, shijiazhuang, hebei province, China

<sup>2</sup>State Grid Hebei Economic Research Institute, fuqiang street, shijiazhuang, hebei province, China

**Abstract.** With the continuous development of information technology and the improvement of distribution automation level. Especially, the amount of on-line monitoring and statistical data is increasing, and large data is used data distribution system, describes the technology to collect, data analysis and data processing of the data distribution system. The artificial neural network mining algorithm and the large data are researched in the fault diagnosis and prediction of the distribution network.

Big data refers to the conventional software tools in a certain range of time capture, management and processing of data, is the need for new processing mode to have a stronger decision-making ability, insight discovery and process optimization capabilities of massive, high rates of growth and diversification of information assets.

Big data has four characteristics: massive data scale, fast data transfer, various data types and low value density. With the rapid development of computer information technology and intelligent distribution network technology, the application of large data in fault prediction and diagnosis of the distribution network is of great significance to improve the degree of automation of the distribution network.

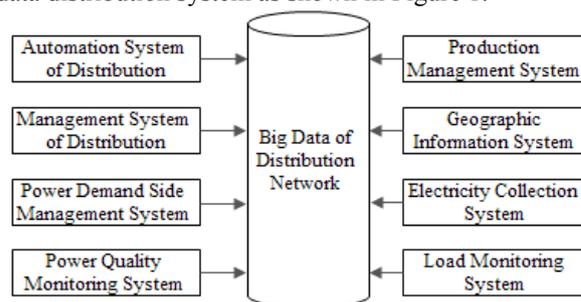
## 1 Big data of distribution

Distribution network is composed of overhead lines, cables, towers, distribution transformers, isolating switches, reactive power compensator, and some ancillary facilities, which play an important role in the distribution of electrical energy.

The research and application of the data distribution system can improve the network operation and management level. Therefore, the application of data distribution system has a broad market prospect. The distribution network system uses the methods of data collection, retrieval, analysis and storage to provide reliable information technology services for the distribution network.

The distribution network including big data distribution automation system, distribution management and scheduling system, power demand side management

system, power quality monitoring system, production management system, geographic information system, electricity information collection system and load monitoring system[1]. Among them, the distribution automation and management system is an important part of the distribution network monitoring, operation and management. The distribution automation and management system highly integrates information of dispatching automation system, production management information system, power marketing system and metering automation system. The main component of data distribution system as shown in Figure 1.



**Figure 1.** The main component of data distribution system

According to the data structure, data distribution system is mainly divided into structured data and unstructured data. Structured data stored in the database, most of the data in the distribution network is the kind of form, and with distributed energy, electric vehicles and its facilities appear in a large number of active distribution network, this type of data will continue to grow. Unstructured data is the data that cannot be expressed by two-dimensional logic table[2]. This part of the data mainly includes the monitoring of the lines and

<sup>a</sup> Corresponding author: zhoujinglong2017@sohu.com

equipment, pictures and video, equipment maintenance management, log information and so on.

## 2 Analysis and processing of large data

Large data analysis, including data analysis and interpretation. Large data analysis refers to the process of in-depth research and analysis of a large number of types of data, to find the implicit relationship between data and potentially useful data value. Big data interpretation refers to a deeper analysis of the process and results of data analysis and wood. Large data with multi dimension form of display, and big data analysis results will be converted to a specific industry issues.

Data visualization is the database of each data item as a single element, in order to construct a time dimension, space dimension and logic dimension of image data and other dimensions, so that from the different dimensions of the data were observed and analyzed deeply. Through from the time, space and logic and other dimensions of the data centralized management with visualization of multidimensional data, can get the stereo view[3].

Data mining is a work of extracting potentially useful information and knowledge from large amounts of incomplete data. Data mining consists of many steps, and each step is closely linked. In the process of data mining, data mining needs to be adjusted iteratively according to the knowledge acquired by mining data. Data mining can predict the unknown data according to the potential rules. The data mining process can be generalized into four parts: determination task, data preparation, mining modeling and result analysis application. The process of data mining is as shown in Figure 2.

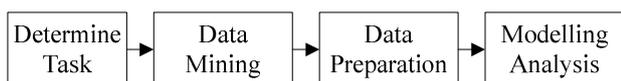


Figure 2. The process of data mining

Associated with a lot of data mining technology and data warehouse technology, data mining is processed on data mining, so data mining data preprocessing, data preprocessing, including data cleaning, data transformation and data integration process.

Because of the huge amount of large data, only after processing can we obtain useful information in these large amounts of data. At present, many fields of power system involve large data applications. The basic attributes of large data include quantity, speed, diversity, and so on, so the method of big data analysis is the key factor to determine whether information is valuable or not. Large data analysis methods generally have the following basic aspects:

Large data analysis techniques are applied to different types of users. Visual analysis is the basic requirement of users using large data analysis techniques. Visual analysis can display its characteristics through intuitive data graphics, and users can understand the

complex relationship existing behind the data by comparative information[4].

The core of large data analysis is data mining. The algorithms of data mining are usually founded on data types and formats and apply accepted algorithms to dig into the data and dig out useful information. Because of the large amount of data in large data, these data mining algorithms give large data processing faster solutions. If the algorithm takes a long time to reach a conclusion, then the big data is of no value and significance.

Data warehouse to complete data preprocessing provides a platform for data mining, thus eliminating the complicated process of data preparation. In addition, in the process of data warehouse construction, a comprehensive data processing and data analysis infrastructure including data access, integration, merging, database conversion and WEB access is set up around the data warehouse. The Structure of typical data mining system is as shown in Figure 3.

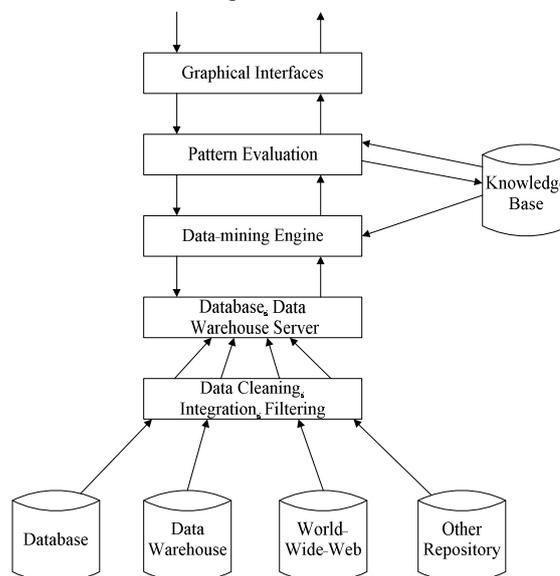


Figure 3. Structure of typical data mining system

## 3 Design of data mining algorithm

Large data mining methods are commonly used classification, regression analysis, neural network method, Web data mining, etc.. The methods of data mining from different angles, using the method of neural network technology to mine the data distribution system, through the analysis of data mining, prediction and fault diagnosis in the distribution network.

Artificial neural network (ANN) is a widely used technique in data mining. The data mining method of neural networks is to train the learning data sets repeatedly by imitating the human nervous system, and finds the model for prediction and classification from the data sets to be analyzed.

Neuron is the basic processing unit of artificial neural network. It is a nonlinear component with multiple inputs and single output. Besides the input signal, the output of neuron is also influenced by other factors in

neuron. In neural modeling, an additional input signal is often added to become a deviation or threshold.

Neurons are the foundation of work in a neural network. Its structure is quite simple, so the processing capacity is simple. A large number of neurons constitute a neuronal network that has many superior characteristics. The processing of information by neural networks is accomplished by a large number of neurons. It is a collection of functions that have access to information, distributed storage, and associative memory.

The neural network consists of different levels of node set, and each layer is output to the next level node. The output value is magnified, attenuated or suppressed due to the connection weights between nodes.

The structure of artificial neural networks is as shown Figure 4.

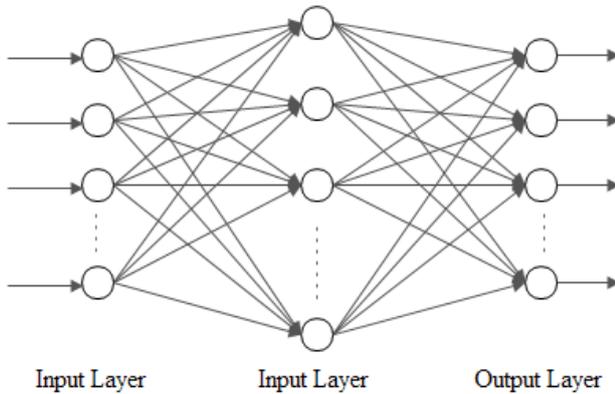


Figure 4. Structure of artificial neural networks

Feedforward networks have feedback signals in the training process, and in the classification process, data can only be transmitted forward until the output layer is reached, and there is no backward feedback signal between the layers. Perceptron and BP neural networks belong to feedforward networks. Figure 4 is a 3 layer feed forward neural network, in which the first layer is the input unit, the second layer is called the hidden layer, and the third layer is called the output layer[5].

The choice of activation function is an important step in the construction of neural networks. The activation function used in this paper is introduced below.

The main difference between the bipolar S shape function and the S shape function lies in the range of function, and the range of the bipolar S shape function is (-1,1), and the range of the S shape function is (0,1). Since the S shape functions and the bipolar S shape functions are derivable, the two activation functions are suitable for use in the BP neural network.

The double S type function is as formula (1)

$$f(x) = \frac{2}{(1 + e^{-ax})} - 1 \quad (-1 < f(X) < 1) \quad (1)$$

The derivatives of double S type function is as formula (2)

$$f'(x) = \frac{2ae^{-ax}}{(1 + e^{-ax})^2} = \frac{a[1 - f(x)^2]}{2} \quad (2)$$

For single neuron information, the information from other neurons is  $X_i$ , and their interaction strength with the neuron, that is, the connection weight is  $W_i$ ,  $i=0,1,\dots,n-1$ , the internal threshold of the processing unit is the  $\theta$ . The input of the neuron  $X$  is as formula (3)

$$X = \sum_{i=0}^{n-1} w_i x_i \quad (3)$$

The output of neurons is as formula (4)

$$Y = f\left(\sum_{i=0}^{n-1} w_i x_i - \theta\right) \quad (4)$$

#### 4 Fault prediction and of distribution network

Distribution network is the last link of the power system to supply power to users. Distribution network equipment is complex, many users, wide coverage, diverse geographical conditions, and affected by external conditions. Fault diagnosis of the distribution network is an important part of distribution automation. The realization of the high automation level of distribution network fault diagnosis is an essential guarantee for the economic development and the improvement of people's quality of life.

Distribution network fault prediction is based on distribution network fault data and fault influence factor data. Mining the relationship between fault and data of distribution network, constructing fault prediction model by using the correlation method, and forecasting the fault in a certain time area. The distribution network fault prediction is usually based on the actual demand of the power company for the fault prediction results, and the time scale and the region of the fault prediction of the distribution network are determined according to the optimization principle of the prediction model[6,7].

When a fault occurs in the distribution network, fault diagnosis system of large data analysis and judgment according to the real-time fault monitoring system to obtain the information on the distribution network based on the proposed regional, correct and effective power recovery strategy, fault isolation region. After years of operation, the distribution network has accumulated a large amount of sufficient fault data.

Data mining technology has powerful analysis function, can find out the cause of the malfunction analysis technique using its integrated, and formulate corresponding rules of decision-making, to take measures to reduce the failure probability, reliability and economic operation of the distribution network benefit.

## 4 Summary

With the continuous development of intelligent information system in distribution network, a rapid increase in the operation of the distribution network data collection, storage, analysis and processing of the effective data distribution system, provides important data for the stable operation of the distribution network. This paper mainly introduces the distribution network data, and the typical application of data distribution system analysis.

The application of large data in distribution network improves the management level of distribution network. Through the analysis of large data, it is of great significance to promote the construction of intelligent distribution network by quickly predicting and diagnosing the fault of distribution network.

## References

1. Zhang Dongxia, Miao Xin, Liu Liping, Zhang Yan, Liu Keyan. Research on Development Strategy for Smart Grid Bid Data[J]. Proceedings of the CSEE, 2015, 35(1): 2-11.
2. Tao Shun, Xiao Xiangning. Infrastructure of the power quality assessment system of power systems[J]. Transactions of China Electrotechnical Society, 2010, 25(4): 171 — 175.
3. Research of Distribution Network's Fault Diagnosis Based on Data[D]. Guangxi University, 2007.
4. Zhang Pei, Wu Xiaoyu, He Jinghan. Review on Big Data Technology Applied in Active Distribution Network[J]. Electric Power Construction, 2015, 36(1): 52-59.
5. Lu Yuzi. Research on Distribution Network Outage Prediction Based on Data Mining Technology[D]. Beijing Jiao Tong University, 2017.
6. Zhang Tiefeng, Liang Sibao, Gu Jiangwei. Overview of the distribution and utilization big data application[J]. Electrical Measurement & Instrument, 2017, 54(2): 92-98.
7. Wang Qian. Significance of the study methodology Big Data Analytics[D]. Wuhan University of Science and Technology, 2015.