Overload Analysis of Distribution Transformers Based on Data Mining

Nan Chen 1, Taotao Dai 2, Liancheng Wang 1, Wei Zhao 1, Ke Lu 1

1 School of Electrical Engineering, Shandong University, Jing Shi Road No. 17923, Jinan, Shandong, China
2 Shandong Hising Electric Power Tech Co., Ltd., Orson Mansion, Jinan, Shandong, China
805430679@qq.com

Abstract. Overload problems of distribution transformers frequently occur in distribution networks. To avoid the inadvertent effect on the networks and take corresponding measures, the association rules are used to analyze the heavy overload phenomenon of distribution transformers. For the operation of the distribution network, it is very important to study the strong association rules between the heavy overload phenomenon of the transformer in different areas and the seasons, weather and holidays. In this paper, the data preprocessing of heavy overload data and other data of transformer network is first processed, and then a data mining model is established. Finally, the strong association rules of heavy overload are found by using Apriori algorithm. The strong association rules can be used to guide the operation of regional distribution network and avoid the influence of heavy load overload on power supply reliability.

1. Introduction

As the last level power supply unit for low-voltage users, distribution network transformer directly affects the quality of power supply. The heavy overload operation of the equipment will speed up the abnormal loss of the internal components, reduce the service life of the equipment, and bring troubles and risks to the power grid. In some important periods, such as during the period of high temperature or low temperature in one year, and during the major holidays, the load rate of equipment will fluctuate greatly. At present, the load conditions of the distribution change mainly include the detection and post processing in the event, and the lack of prejudgement on the load. At present, reference [2] and reference [3] are used to analyze the correlation and probability of variation of heavy overload with the factors such as meteorological indexes and electricity categories. Finally, the decision tree model based on the Stochastic Forest theory is used to predict the overload status of the distribution variable heavy load. Reference [4] studies the fast developing region with rapid load growth. Based on user, meteorological and historical load data, a medium and long term prediction method for heavy overload based on logistic regression is proposed. A heavy overload prediction method based on BP neural network and grey model is proposed in reference [5]. The load change of distribution transformers before and after the Spring Festival as a model input is used to predict the load rate, and then the overload condition of the equipment is judged. Reference [6] takes one week as a cycle to study 7 types of load curves of 6000 residents' stations in the same day a day, and add meteorological and economic data to predict the load curve by different machine learning methods.
However, this method does not take into account the relationship between load and holidays, as well as the quarterly annual rule. In reference [7], the relationship between meteorological index, wind cold index, air index, human comfort index and the heavy overload of transformer network transformer is studied, and the wavelet neural network model is used to predict the load of the transformer.

The phenomenon of heavy overload essentially reflects that the power supply capacity of transformer in distribution network cannot meet the needs of users in specific scenarios. The electricity consumption in the power supply area is affected by many natural conditions such as season, weather, holidays, economy, and social environment. In addition, the defect of the device itself and the unreasonable topology of the power grid will also affect the occurrence of heavy overload. In this paper, according to the heavy overload data of the distribution network substation, considering the temperature, weather, holiday and the periodicity of the seasonal and daily cycles, and the periodicity of one week, the data was found by the association rules to explore the law of the occurrence of heavy overload in distribution network.

2. Data source selection
The main factors affecting the heavy overload of distribution network transformer include 1) temperature, weather conditions, air quality and other weather factors; 2) human behavior factors such as holidays, working days, and other human behavior factors; 3) social factors such as economic development level, industrial structure and other factors of power supply area; 4) its own characteristics such as capacity, protection mode, etc.

After the actual investigation and the full consideration of the way and possibility of data acquisition, this paper mainly selects transformer load data, transformer account data, transformer topology, meteorological data and holiday data as data sources. Load data of transformer in distribution network include assignment name, sampling time, load rate and other fields. Meteorological data include daily mean temperature, precipitation, air quality and maximum wind power. The properties of distribution transformers include equipment capacity, commissioning time, cooling mode and production date.

3. Data preprocessing
In distribution network transformer load data, the distribution transformers with load ratio greater than 100% is defined as overload transformer, the definition between 80%-100% is heavy load, and less than 80% is considered normal.

In terms of time dimension, the influence of multidimensional time characteristics on overload of distribution transformers is considered. The seasonal and daily periodicity and cycle period characteristics are taken into account. March, April and May are divided into spring, and June, July and August are divided into summer, September, October and November are divided into autumn, December, January and February divided into winter. On a 24 hour cycle, 0 o'clock to 6 o'clock is in the morning, 6 to 12 o'clock is in the morning, 12 to 18 is afternoon, 18 to 24 o'clock is evening. Seven days a week will be set from Monday to Sunday.

The influence of social production activities on heavy overload of distribution transformers is mainly based on the arrangement of national holidays.

The precipitation in the meteorological data includes clear, small rain, thunderstorm, heavy rain and snow. The air quality grade includes excellent, good, mild and severe pollution, and the maximum wind grade includes 1, 2, 3, 4, and more, and the daily average temperature is a continuous numerical range.

Association rules mining algorithms deal with discrete data, so the continuous daily mean temperature data need to be discretized. The typical discretization methods are divided into unequal learning method, equal width method, equal frequency method, supervised learning entropy method and card method. According to the regularity of heavy overload with time distribution and the requirement of correlation analysis, the method of equidistant dispersion is used to discretize the temperature data. The method can keep the original distribution of data. The minimum length of the
The fault is Min, and the maximum value is Max. There are two main methods to make the equal width discretization. One method is to set the discrete width $\delta$ first and find the discrete point number $k$. The other method is to calculate the discrete width $\delta$ by the discrete points. The calculation expression is as follows.

$$\delta = \frac{(Max - Min)}{k}$$  \hspace{1cm} (1)

The break point can be obtained by this method.

$$t_i = Min + \delta \times i \quad i = 0,1,2,\ldots,k$$  \hspace{1cm} (2)

According to actual demand, the daily average temperature is divided into 5 intervals, namely, low temperature, low temperature, medium temperature, partial high temperature and high temperature, as shown in the following table.

| Name          | Section     |
|---------------|-------------|
| low temperature | $[Min, t_1]$ |
| Hypothermia   | $[t_1, t_2]$ |
| Middle temperature | $[t_2, t_3]$ |
| Partial high temperature | $[t_3, t_4]$ |
| high temperature | $[t_4, Max]$ |

4. Analysis of overload regulation of variable weight based on association rules

4.1. Association rules

Association rule is a data mining method to find the relationship between items and items in a given data set. The core of its algorithm is to solve the problem of quickly selecting the two or more group variables with the strongest correlation from many variables. If there is a certain regularity between the values of several variables, it can be said that there is a correlation between these variables. The definition of association rules is: the data set of association rule mining is recorded as $D$, $D = \{t_1, t_2, t_3, \ldots, t_k, \ldots, t_n\}$, $t_k = \{i_1, i_2, \ldots, i_m, \ldots, i_p\}$, $t_k (k = 1, 2, \ldots, n)$ is called a transaction, $i_m (m = 1, 2, \ldots, p)$ is called an item. Each transaction has a identifier called T. A set of items is called an item set.

Let A be an item set, and the transaction T contains A when and only if $A \subseteq T$. Association rules like $A \rightarrow B$ among $A \subset I, B \subset I$, also $A \cap B = \emptyset$, A is called the preceding part of association rules, and B is called the back part of association rules.

There are two basic measures to measure association rules: support and confidence. The purpose of association rules is to find association rules with high confidence and high support between variables. In the database D, the requirement of association rule $A \rightarrow B$ is to meet the requirement of minimum support s and minimum confidence C. Support s is the percentage of transactions $A \cup B$ (i.e. A and B) in database D.

$$\text{sup} \text{port}(A \rightarrow B) = P(A \cup B)$$  \hspace{1cm} (3)

The confidence level C is the ratio of transactions containing both A and B in database D to transactions that contain only A.

$$\text{confidence} \ (A \rightarrow B) = \frac{P(B/A)}{P(A)}$$  \hspace{1cm} (4)
The support and confidence of association rules are values between 0~100%. The support and confidence of association rules are values between 0~100%. The higher the confidence level, the more likely the occurrence of B is caused by A. The most common view is that only association rules that satisfy minimum support degree (Min-sup) and minimum reliability (Min-conf) can be called strong association rules.

4.2. Apriori algorithm

Apriori algorithm belongs to Boolean association rule mining algorithm. For quantitative association rules can also be transformed into Boolean association rules, and then we use Apriori algorithm for mining. The Apriori algorithm is a priori probability algorithm, which takes advantage of the prior knowledge of the frequency set characteristics and uses a cyclic method of hierarchical sequence search to complete the mining of frequent itemsets.

The Apriori algorithm is applied to analyze the heavy overload law of transformer in distribution network. The specific process is as follows:

START
The data of weather, distribution, overload and holiday data are integrated to form a mining database
The minimum support threshold is determined, and Apriori algorithm is applied to mining association rules.
Obtain strong association rules and generate association rules database.
Combining strong association rules with reality
END

Figure 1. Data mining process

5. Example analysis

The data of distribution network transformer heavy overload data, weather data, holiday data and other data of one year of a regional power company are integrated to form a database D, with a total of 1190 data. Apriori algorithm is used to find the regularity of data overload and weather, holidays and other data.

This paper mainly digs out frequent 4- frequent itemsets, selects 3% as the minimum support threshold, and the minimum support count is 40. After getting frequent itemsets, association rules can be generated based on frequent itemsets. Select association rules with a confidence level of more than 60%, so that strong association rules can be obtained. As shown below.

Table 2. Strong association rules

| Consequent | Aforesaid                          | Support count | Confidence degree |
|------------|------------------------------------|---------------|-------------------|
| A          | Mild pollution, summer and clear   | 44            | 75%               |
| B          | Low temperature, grade 4, evening  | 48            | 70.8%             |
| C          | High temperature, clear and good   | 41            | 65.7%             |
| B          | Saturday, winter, evening          | 42            | 64%               |
Based on the above strong association rules, we can draw the following conclusions.

1. Under the mild and sunny weather in summer, regional A needs to make a plan to deal with heavy overload of transformer in distribution network. (2) On the contrary, regional B should pay more attention to low temperature weather, especially in the evening and holidays with the maximum wind power of 4. (3) The regional C needs to pay attention to the high temperature weather with good weather and clear air. On Friday, the heavy air quality and the maximum wind 4 area C are more likely to be overloaded.

6. Summary
Through mining strong association rules, we can guide the operation of distribution network and effectively deal with the phenomenon of heavy overload in distribution network. According to the high accuracy of weather forecast and the pre-arrangement of holidays, we can predict the heavy overload of the transformer in the distribution network, and therefore reasonably arrange the operation of the distribution network.

References
[1] Wang Dewen, Sun Zhiwei. Big data analysis and parallel load forecasting for power user side [J]. Chinese Journal of electrical engineering, 2015,35 (03): 527-537.
[2] He Jian Zhang, Wang Haibo, Ji Zhi Xiang. Analysis of factors affecting the heavy overload of distribution transformers facing smart grid [J]. power grid technology, 2017,41 (01): 279-284.
[3] He Jian Zhang, Wang Haibo, Ji Zhi Xiang. Prediction of heavy overload of distribution transformers based on random forest theory, [J]. power grid technology, 2017,41 (08): 2593-2597.
[4] Li M, Zhou Q. Distribution transformers mid-term heavy load and overload pre-warning based on logistic regression[C]// PowerTech, 2015 IEEE Eindhoven. IEEE, 2015:1-5.
[5] Shi Chang, Yan Wen Qi, Zhang Xiaohui. Based on the BP network and grey model, the Spring Festival distribution overload prediction [J]. Journal of power science and technology, 2016,31 (03): 140-145.
[6] Singh M J, Agarwal P, Padmanabah K. Load forecasting at distribution transformers using IoT based smart meter data from 6000 Irish homes[C]// International Conference on Contemporary Computing and Informatics. IEEE, 2017:758-763.
[7] Sun X, Luh P B, Michel L D, et al. An efficient approach for short-term substation load forecasting[C]// Power and Energy Society General Meeting. IEEE, 2013:1-5.
[8] Zhang Guobin, Wang Xiaorong, Deng Chunyu. Heavy load forecasting method for distribution network area based on correlation analysis and machine learning [J]. big data, 2018,4 (01): 105-116.
[9] Wang Xing. Big data analysis: methods and applications [M]. Beijing: Tsinghua University press, 2013:97-104.