Analyzing problems of distribution transformer service areas based on customer-side data mining

Shanshan Xiao*, Liancheng Wang
School of Electrical Engineering, Shandong University, Jinan, China

*Corresponding author e-mail: 1141159667@qq.com

Abstract. With the development of economy and the continuous improvement of living standard, the contradiction between the rapid growth of electric load and the untimely renewal of power grid equipment is increasingly prominent. Power supply reliability issues or frequent power outages and customer complaints exist in many areas. Due to the limited human and material resources of power supply companies, the problems existing in the distribution transformer service areas cannot be resolved timely and thoroughly. As the direct service objective of electric power, customer’s experience of using electricity will reflect the power quality of the distribution transformer service areas to a great extent. Therefore, this paper will start with work orders from customer sides, find out the existing problems in distribution transformer service areas based on the data association rules, and help make informed decisions. The proposed method not only reduces the investment of capital and manpower, but also reduces the number of power outages, and improves the customer satisfaction.

1. Introduction

In modern life, people require better quality and more reliable power supply. Improving power supply reliability is not only the needs of customers, but also the guarantee for electric companies to establish a good image [1]. Distribution transformer service areas is the last kilometer of power supply, its health will directly affect the power quality of customers. Building a healthy distribution transformer service area is the core of scientific management for electric companies.

In recent years, a lot of research work has been carried out around the power supply reliability of distribution transformer service areas. These studies show that meteorological indicators, industry categories, power consumption categories, wind chill index and air pollution index have a great impact on the probability of power failures [2,3]. When evaluating the reliability of power supply in one area, it is generally based on the number of repeated power failures and the interval between two power failures [4]. Judging from the current study, the analysis of distribution transformer service areas is mostly carried out from the aspects of electrical equipment, operations, load and weather conditions, without considering the reliability of power supply from the perspective of customers [5]. However, the reliability of power supply is more concerned about whether customers can use electricity reliably. Therefore, this paper starts with the customer's historical work orders, finds association rules from all kinds of work orders, and digs out the problems of the distribution transformer service areas, which can effectively analyze the problems of power grid, so as to improve the quality of power supply and reduce customer complaints.
2. Problems in improving power supply reliability based on customer side data

The work orders on the customer side include fault repair, complaints and comments. The reasons for these work orders include the customer's own equipment failure, long power supply distance, small distribution transformer capacity, low distribution gear, three-phase imbalance, etc. Therefore, measures can be taken to increase the capacity of distribution transformer, change the line, add distribution transformer, etc. However, these measures are troublesome and resource consuming. Therefore, when the power supply company receives the call from the customer, it also temporarily restores the power supply to the customer, and fails to solve the fundamental problem. In the future, the same problem will occur, resulting in the increase of the customer's complaint rate.

This paper uses the method of data association to analyze all kinds of work orders, mining out the problems that need to be solved in the distribution transformer service area. To some extent, it is helpful to upgrade the power grid.

3. Data sources

As a comprehensive work, the management of the distribution transformer service area needs the cooperation of multi departments and multi systems. Therefore, the acquisition and synthesis of work order data from multiple systems is the basis of analysis. By analyzing data from various systems, it is found that the customer's fault repair work order data is stored in Rush Repair System, which records the customer ID and the reason of the fault repair. The Business Support System stores customer complaint work orders and comment work orders, which records the ID of dissatisfied customer and the reasons for dissatisfaction. The Power Supply Customer Service System stores the data of the active operation inspection work orders, which records the monitoring data of low voltage and heavy overload. The Marketing System stores the customer's measurement data and the relationship between customers and distribution transformer service areas.

The above data from all systems are massive, and the field names from all systems are not uniform, which makes the data more complex. Therefore, data preprocessing is needed before analysis.

4. Data preprocessing

According to the original data, there are three preprocessing stages:

(1) Remove unrelated work orders

Through the analysis of data, it is found that there are some fault repair work orders caused by the customer's own reasons, such as improper operation of the customer and misinformation of work order. In addition, there are some customer complaint work orders caused by poor attitude of the staff and malicious complaints. These work orders do not reflect the problems of electrical equipment and power lines, so these data should be removed before analysis.

(2) Remove irrelevant fields

The data from various work orders on the customer side is large and disorderly. There are many field types in each type of work orders. Take the fault repair work orders as an example, a piece of data contains more than 30 field types, such as work order ID, customer ID, acceptance time, contact address, acceptance content, cause of power failure, etc. These fields are not all required by us, so we need to remove unimportant fields, keep important fields such as customer ID, address and fault causes, etc.

(3) Standardize field content

These work orders on the customer side are usually formed by the customer calling the electrical company. Due to various factors such as call quality and telephone operator's simple record, some fields such as customer ID and area ID in these work orders are missing. In addition, the content of the field describing the reason is not uniform and standard, which makes data analysis more difficult. Therefore, it is necessary to use the relationship model between customers and distribution transformer service areas to find the ID of the distribution transformer service area where the customer is located. And it is necessary to unify and standardize the different statements that represent the same reason, and then incorporate them into the database.
5. Analysis of the problems in the distribution areas based on association rules

5.1. Association rules

Association rules reflect the correlation between several things and can find valuable information from a large amount of data. If there is a certain repetition rate between several things, then there is a certain correlation between them, and the association rule is to find this correlation.

Association rule is an implication:

\[ R : M \Rightarrow N \]  

(1)

In this rule, M is the previous occurrence of this rule, and N is the occurrence of this rule after M. This rule means that if M happens first, then N is likely to happen next.

Support and confidence can be used as two more important indexes to measure association rules. We only need to know the items with high support and confidence among a large number of items to find the rules. In addition, it should be pointed out that the value of this high support degree and confidence degree must be greater than the threshold values of minimum support degree and minimum confidence degree.

The support degree of association rule is the percentage of transactions \( M \cup N \) in database D. This indicator reflects the possibility that event M and event N will occur in the same time period.

\[
\text{support}(M \Rightarrow N) = \frac{\text{count}(M \cup N)}{|D|} 
\]  

(2)

The confidence level reflects the possibility of the occurrence of event N after the occurrence of event M.

\[
\text{confidence}(M \Rightarrow N) = \frac{\text{support}(M \Rightarrow N)}{\text{support}(M)} 
\]  

(3)

The higher the degree of support, the greater the proportion of occurrence of N after occurrence of M in all the occurrence of events. The higher the confidence level, the greater the proportion of the number of occurrences of N after M in the number of all events containing M. The valuable association rules we want to mine also have high support and confidence[7].

5.2. Apriori algorithm

The essence of this algorithm is to find out all frequency sets. The frequency of these item sets is at least the same as the predefined minimum support, and then the frequency sets produce association rules. These rules must meet the threshold (min-confidence and min-support). The core idea of the algorithm is based on the frequency of the recursive method of set theory, used layer by layer search iterative method. This method is helpful to find out the hidden relationship between the item sets in the database. This algorithm was used in the analysis of shopping basket data at first. Through the analysis of shopping basket data, we can find the shopping tendency of users, which is helpful for the merchants to sell better. At present, the algorithm is widely used in various fields, such as industry, agriculture and business. These potential data and information will play a guiding role in decision-making to some extent.

The figure 1 shows the steps of the algorithm. In the flow chart, \( C_k \) represents the candidate item set containing k items, and \( L_k \) represents the frequent item set containing k items. Only those item sets that are greater than the minimum threshold will be retained. We use recursion to find this potential rule. This method is simple and easy to implement, the recursive process is shown in Figure 1.

The Apriori algorithm is used to analyze the association rules in kinds of work orders, which is helpful to find the problems in one area. This process can be divided into five parts, figure 2 is a detailed description of the algorithm.
Figure 1. The process of Apriori algorithm

Figure 2. Algorithm steps
6. Example analysis

Preprocessing the data of a quarter's fault repair work orders, complaint work orders, comment work orders and active operation inspection work orders in a city, and then integrate these data to form database D. There are 1470 pieces of fault repair data, 156 pieces of complaint data, 1214 pieces of comment data and 2069 pieces of active operation inspection data. These work order data contain the situation of power failures, customer complaints and customer’s comments. These data have overlapping part in the same distribution transformer service area. The Apriori association algorithm can find the association rules in these complex work order data. In a certain area, if there are related reasons for power failure or customer’s dissatisfaction in various work orders, the problems in the area can be found by integrating these related reasons.

This article focuses on mining frequent itemsets with less than 5 items. Set the minimum number of support to 30 and the minimum confidence to 0.6. The following table shows the association results.

| area  | Work orders situation                                      | Frequent count | Confidence degree |
|-------|------------------------------------------------------------|----------------|-------------------|
| A     | (1) Customer's fault repairs: outgoing line of the transformer is damaged by external force (2) Customer's comments: insufficient safety distance (3) Customer's comments: transformer noise (4) Customer complaints: equipment relocation plan is not implemented | 34             | 0.79              |
| B     | (1) Customer's fault repairs: the transformer has been overloaded for a long time (2) Customer's fault repairs: the primary leakage protector trips repeatedly (2) Customer complaints: frequent power outage during agricultural irrigation. (3) Active operation inspection: three-phase imbalance | 47             | 0.78              |
| C     | (1) Customer's comments: the terminal customer's voltage is low for a long time (2) Customer complaints: low voltage has not been improved. (3) Active operation inspection: transformer gear is low | 42             | 0.75              |
| D     | (1) Customer's fault repairs: external force to hang up the power lines (2) Customer's comments: the power line is too low, which has potential safety hazards (3) Customer's complaints: line modification plan is not implemented | 36             | 0.70              |
| B     | (1) Customer complaints: transformer noise (2) Customer's fault repairs: tripping caused by loose outgoing line of secondary leakage protector (3) Customer's comments: the insulation of the lower line is damaged | 32             | 0.69              |
| E     | (1) Customer's fault repairs: theft (2) Active operation inspection: abnormal collection | 31             | 0.61              |

From the above correlation results, we can find the following conclusions:

(1) The location of transformer in area A is unreasonable, the safety distance from residents is insufficient. Residents explained the situation to the electrical company, but relevant departments did not carry out corresponding rectification measures. (2) In area B, the distribution transformer is often overloaded, especially in the period of centralized irrigation. There is a problem that the power supply capacity cannot meet the customer's power demand. It is necessary to increase the capacity of the
transformer or change the phase in time. In addition, there are aging of equipment and damage of insulation layer in the area. The manager of the area should strengthen the routine maintenance and management. (3) Customers living at the end of the power lines in area C are too far away from the transformer. When the gear of distribution transformer is low, these customers will always have low voltage. The electrical company should timely adjust the gear of distribution transformer or add a transformer in the middle of the line. (4) The distance between the low-voltage lines and the ground in area D is insufficient, so there is a potential safety hazard. The electrical company should put the line reconstruction plan on the agenda. (5) In area E, there are more active operation inspection data showing low voltage, but other work orders do not reflect power failure or low voltage, so it can be inferred that the acquisition of electricity meter is abnormal, and the electricity meter or acquisition module needs to be replaced.

7. Conclusion
At present, the reliability of distribution transformer service areas is assessed and thus corresponding correction measures are taken mostly based on the electrical equipment, load conditions and operation conditions. Keeping customer satisfaction as the focus point, this paper has proposed a new method to analyze the reliability of distribution transformer service areas from the perspective of customer's electricity experience. Based on the algorithm of data association rules, the association relationship among the customer's data such as fault repairs, complaints, comments and active operation inspection data is found out, therefore, the root-causes of the problems are identified and fixed accordingly. This method verifies the feasibility of improving the reliable operation of the distribution network from the perspective of the customer side. It provides an effective support for improving customer's satisfaction with electricity consumption and has certain guiding values.

References
[1] Chen Lingjun, Zhang Ye, Wu Yinghui, et al. Evaluation of health status of low-voltage distribution network based on K-means algorithm[J]. Distribution & Utilization, 2019, 36(2): 42-47
[2] He Jianzhang, Wang Haibo, Ji Zhixiang, et al. Analysis of factors affecting distribution transformer overload in smart grid [J]. Power System Technology, 2017, 41(1): 279-284
[3] 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 (PES), 2013 IEEE, IEEE, 2013.
[4] Lu Tao, Liu Jie, Qiu Zhangquan, et al. Judgment of repeated power failures in power supply region based on big data technology [J]. Power & energy, 2019, 40(01): 52-54
[5] Yin Xiuyan, Zhang Hong, Tian Xincheng, et al. Monitoring and analysis of transformer districts with frequent power outage based on big data. [J]. Distribution & Utilization, 2017, 34(1): 38-41
[6] Huang Xu, Wang Weiheng, Wu Shuang, et al. An intelligent key feature selection method of power grid based on artificial intelligence technology [J]. Distribution & Utilization, 2019, 36(10): 22-29
[7] Zhang Guobin, Wang Xiaorong, Deng Chunyu. Prediction method in distribution transformer heavy and overload supply areas with relevance analyze and machine learning[J]. Big data, 2018