International Conference on Applied Energy 2019
Aug 12-15, 2019, Västerås, Sweden
Paper ID: 0219

LINE LOSS ANALYSIS OF LOW-VOLTAGE DISTRIBUTION SYSTEM: A BIG-DATA METHOD

Haiwen Chen[1], Shouxiang Wang[1*], Yingjie Tian[2]

1 Key Laboratory of Smart Grid of Ministry of Education, Tianjin University, Tianjin 300072, China (*Corresponding Author)

2 State Grid Shanghai Electrical Power Research Institute, Shanghai 200437, China

ABSTRACT

Line loss in power distribution system occupies one fifth to one quarter of the total power loss in power systems. Different from the line loss in transmission system, distribution line loss is closely related to the load characteristics and economic types of power users. And the proportion of line loss caused by management problems is significant. Hence the traditional physical model of line loss is not suitable for distribution line loss analysis. To analyze line loss of low-voltage distribution system accurately, a big-data method based on MSApriori association rule mining algorithm is proposed in this paper. Based on the operational data of more than 20,000 distribution stations and the detailed data of 1.8 million power users in Shanghai, a line loss analysis method based on big data is proposed in this paper. The relationship between the line loss and its influencing factors is analyzed from the perspectives of user type, user number, power supply volume, reactive power shortage, and transformer type. The association factors obtained from the analysis are helpful to further guide the reduction of line loss.

Keywords: distribution line loss, big data, association rules

1. INTRODUCTION

Line loss has always been the focus of power suppliers, because it means a meaningless waste of energy. Reducing line loss can effectively improve the efficiency of energy transmission and promote the economic operation of the whole energy system.

In power system, the line loss of distribution system is the most serious. On the one hand, compared with the transmission system (110 kV or more), the voltage level of distribution system is lower (35 kV, 10 kV and 0.4kV in China), and the cable impedance is relatively high, which leads to higher energy loss. On the other hand, the distribution system has complex lines, which are prone to cause grounding short circuit and electricity theft. At the same time, the reliability and accuracy of the sensor in distribution system are much lower than that of transmission system, and the measurement results are not accurate enough. Line loss caused by these factors is called management line loss. For power suppliers, the line loss of distribution system is assessed and managed according to the power distribution area. A power distribution area refers to the 10kV/0.4kV transformer and its connected lines and power users. The line loss value can be obtained by calculating the sold electricity and the input electricity in the distribution area. The result is often divided by the power supply, and the line loss rate is obtained for assessment. Therefore, the analysis of line loss of low-voltage distribution system in this paper is based on 10kV/0.4kV distribution area.

At present, researches about line loss are mainly focusing on the transmission system based on its mechanism model. For example, ref [1] proposed an online transmission line loss calculation and analysis method based on the SCADA system. Ref [2] proposed an optimal allocation method of solar PhotoVoltaic (PV) systems based on the physical model of transmission line. Ref [3] studied the impact of renewable energy access in line loss of microgrid from the perspective of the physical model. The conventional researches start from the point of view of the physical model of line loss, which is not suitable for line loss in low-voltage distribution system. Since the management line loss accounts for a large proportion of the line loss in distribution system, only analyzing from the perspective of the physical model is not comprehensive enough to reflect the relationship of complex factors (economic type, number of users, number of measurements, etc.) in actual operation. Big-data analysis method can effectively overcome the insufficient consideration of influencing factors in line loss mechanism modeling. It

Selection and peer-review under responsibility of the scientific committee of the 11th Int. Conf. on Applied Energy (ICAE2019).

can more effectively mine the internal relationship between various types of data, so as to discover new knowledge and direct the production activities [4].

In this paper, an analysis method for the influence factors of line loss in low-voltage distribution system based on big-data analysis is proposed. Compared with the traditional physical model-based analysis method, the big-data method is entirely data-driven and model-free, which can mine the related factors more comprehensively and guide the line loss reduction.

2. ANALYSIS PROCESS

The analysis process is shown in Fig 1.

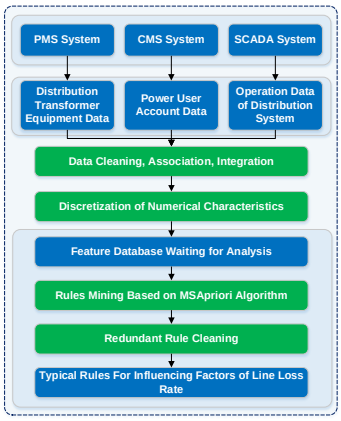

Fig 1 Analysis process

The main data sources include power production management system (PMS), customer management system (CMS), supervisory control and data acquisition (SCADA) system. Firstly, the equipment ID of distribution transformer in the PMS database is extracted and matched with the connection relationship data in the CMS database. All users belonging to the same distribution area are obtained. The number of users, economic type, industry attributes, contract power supply capacity and quantity of smart meters in each distribution area are counted. Then, detailed operation data of each distribution area are obtained through the SCADA system, including daily average power supply, peak-valley ratio, reactive power shortage, line loss (by weekly statistics), success rate of information acquisition, etc. The missing and abnormal data are screened and cleaned, and the continuous variables are discretized by the box-dividing method to form the data set to be mined. Finally, the MSApriori association rule mining algorithm is used to analyze the relationship between line loss rate and complex factors by setting a confidence threshold and filtering repetitive rules.

3. DATA PROCESS

3.1 Data description

The data source used in this paper comes from Shanghai Electric Power Big Data Application System, which connects PMS, CMS, SCADA system, meteorological and economic data. The object of this paper is the operation data of 25,673 low-voltage distribution areas in Pudong area in 2017.

3.2 Multi-source data association and discretization

The database association relationship of the big data system is shown in Fig 2.

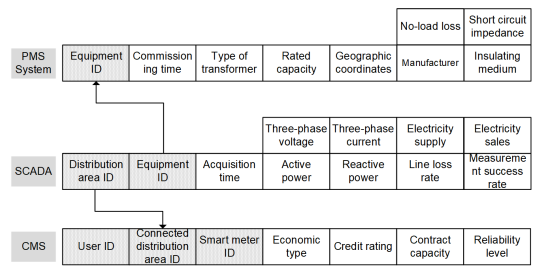

Fig 2 Database association

After reading the related data from the database, the characteristics of distribution area are sorted out as follows. (1) Statistics the number of all users in each distribution area. (2) By calculating the proportion of contract capacity of users of different economic types in each distribution area, the highest proportion of economic types is designated as the dominant economic type of the distribution area. (3) The peak-valley ratio of power supply and three-phase unbalance are calculated. (4) Calculate the average user power supply in each distribution area. Association rule learning method requires discretization input, so the ChiMerge algorithm is used to discretize continuous features [5]. Parts of the discretized features are shown in Table 1.

Table 1

Feature name	Discretized level	Range
Line loss rate	1 -> 10	0% -> 16%
Credit rating	1 -> 6	Poor -> Excellent
Smart meter number	1 -> 4	1 -> 1157
Power supply	1 -> 5	110.4MWh -> 55483.3MWh
Reactive shortage	1 -> 3	-215.51kVar -> 178kVar
User number	1 -> 4	1 -> 1139

3.3 Overview of line loss data in Pudong area

Firstly, the transformer is marked on the map, and the line loss rate in each distribution area is expressed by the color shade. It can be seen from Fig 3 that the transformers in Pudong area are mainly distributed along the economic development zone (CBD of Shanghai). The distribution networks in economic development zone

are dense, and the line loss rate is higher than that in other areas.

Furthermore, the proportion of distribution area with different dominant economic types is briefly analyzed. The total number is 2,5673 and the result is shown in Fig 4.

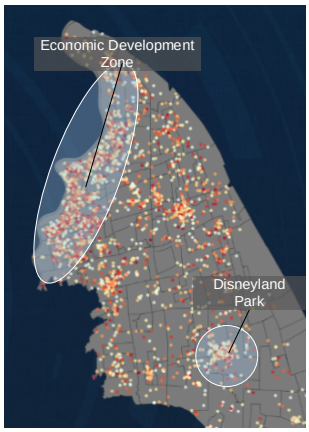

Fig 3 Analysis process

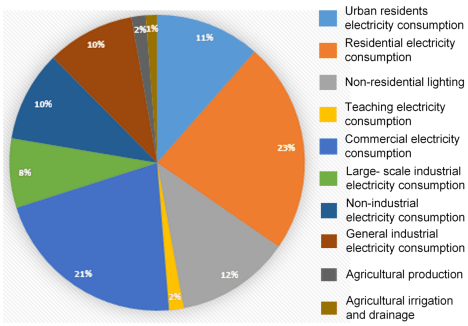

Fig 4 Portion of different economic types

Consistent with the results of Fig 3, the proportion of distribution area with commercial electricity consumption is the second highest, reaching 21%. In Pudong, distribution area with agricultural power consumption accounts for a very low proportion, less than 1%.

4. ASSOCIATION RULES MINING

4.1 MSApriori algorithm

Apriori is a frequent item mining method and is quite useful for association rule mining on big data problem. It identifies the frequent items and extending them to a larger item set. Considering the distribution of economic types of distribution area is unbalanced, we adopted MSApriori algorithm to overcome it. MSApriori algorithm can set multiple support degrees for different items to guarantee the discovery of some rare item sets [6]. Details can be found in [6].

4.2 Useful rule concepts

Here we explain some concepts that will be used in the following section. Let *X,Y* be itemsets, $X \Rightarrow Y$ an association rule and *T* a set of transactions. Support is an indication of how frequently the itemset appears in the dataset [7]. The support of *X* with respect to *T* is defined as the proportion of transactions *t* in the dataset which contains the itemset *X*.

$$\text{support} = \frac{|\{t \in T; X \subseteq t\}|}{|T|} \tag{1}$$

The lift of a rule is defined as:

$$\text{lift} = \frac{\text{support}(X \cup Y)}{\text{support}(X) \times \text{support}(Y)} \tag{2}$$

If the lift is > 1, that lets us know the degree to which those two occurrences are dependent on one another, and makes those rules potentially useful for predicting the consequent in future data sets [7].

5. CASE STUDY

Restricted by the number of pages required for conference papers, only the partial related factors of the highest and lowest line loss are analyzed here.

5.1 Rules for the lowest line loss rate distribution area

First, we limit the right rule to the lowest level of line loss rate. Only the factors related to the result of "the lowest line loss rate" were considered. There are 157 effective rules. After ranking according to lifting degree, this paper only takes 10 rules with the highest lifting degree as an example to analyze. The digraph of rules is shown in Fig 5. The size of the circle represents the degree of support, and the color represents the degree of lift.

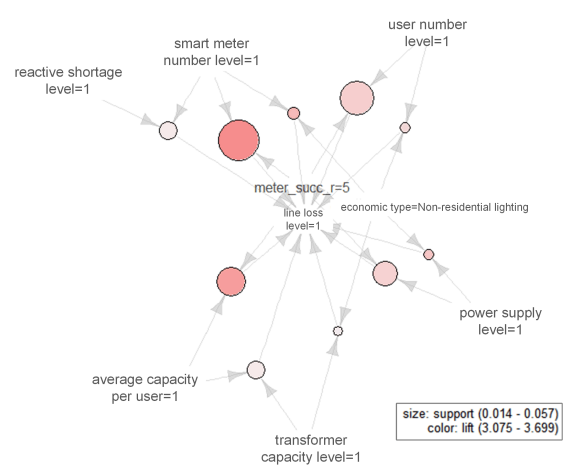

Fig 5 Rules for low-line-loss distribution area

Line loss rate level 1 (0-2.1%) has obvious characteristics: 1) the number of meters (level 1), reactive power, three-phase unbalance are all at the lowest level; 2) measurement success rate is at the

highest level (level 5); 3) non-resident lighting is the economic type. From the principle analysis, the non-resident lighting power consumption in Pudong area mainly refers to the power consumption of public roads, bridges, wharfs, public toilets, traffic command lights of public security departments, public security indicator lights, police booths and streetlights in parks. This type of power consumption is almost all resistive equipment, and the use of time is relatively short. Therefore, the line loss rate of this type of station area is at the lowest level.

5.2 Rules for the highest line loss rate distribution area

There are 169 effective rules for high line loss rate distribution area. Carry out the same process as in Section 5.1. The digraph of rules is shown in Fig 6.

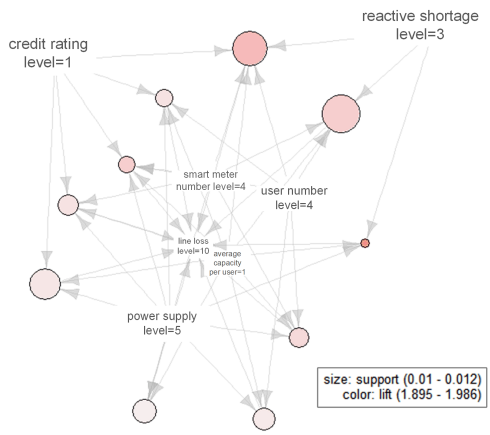

Fig 6 Rules for high-line-loss distribution area

Line loss rate level 10 (8%-16%) has the following characteristics. Factors such as the highest level (level 4) of user number and smart meter number, the highest level (level 5) of power supply have an obvious support degree. That is to say, the distribution area with high power supply, more power users and more meters tend to have a higher line loss rate. This shows that the main factor of high line loss rate is the management line loss, that is, the line loss caused by the error of measuring equipment and the line loss caused by the mismanagement and mistake. For example, the line loss caused by missed copy and miscalculation in the process of meter reading. In the process of data acquisition, it is difficult to accurately unify the date of meter reading. Because of the large number of users and meters, the error caused by meter reading calculation is magnified, which makes the line loss rate deviate from the reasonable range in theory. Besides, users in high line loss areas have the lowest credit rating level, which means there is more risk of electricity theft. Electricity theft also increases the line loss of distribution system.

Through big data analysis, it can be found that the low-voltage distribution areas with high line loss rate are mainly caused by management factors. So in the area with high statistical line loss rate, it is necessary to focus on the management of line loss.

6. CONCLUSION

In this paper, a big data method for the analysis of low-voltage distribution line loss is proposed. Compared with the traditional physical model of line loss, the big data method can take more relevant factors into consideration, such as economic types, user number, credit rating, etc., which is helpful to guide the reduction of line loss. We take distribution areas in Pudong, Shanghai as an example, and prove that management line loss is a key concern in the treatment of line loss. Limited by the number of pages, there are other valuable rules to be discussed, and specific measures can also be proposed combined with the effective rules, which is the future research priority.

ACKNOWLEDGEMENT

Supported by the State Grid Technology Project and National High-Tech Research and Development Program of China (863 Program) (2015AA050203).

REFERENCE

[1] Zhang Y, Zhu Y, Bai X, Wei H. CIM-based data-sharing scheme for online calculation of theoretical line loss. Energy Procedia 2011;16:1619–26.

[2] Dhanalakshmi B, Rajasekar N. Dominance square based array reconfiguration scheme for power loss reduction in solar PhotoVoltaic (PV) systems. Energy Convers Manag 2018;156:84–102.

[3] Zehir MA, Batman A, Sonmez MA, Font A, Tsiamitros D, Stimoniaris D, et al. Impact of Renewable Based Microgrid Supply/Demand Profiles on Low Voltage Distribution Networks. Energy Procedia 2016;103:231–6.

[4] Tabesh P, Mousavidin E, Hasani S. Implementing big data strategies: A managerial perspective. Bus Horiz 2019;62:347–58.

[5] Kerber R. Chimerge: Discretization of numeric attributes. Proc. tenth Natl. Conf. Artif. Intell., 1992.

[6] Yen-Liang Chen, Ya-Han Hu. Mining association rules with multiple minimum supports: a new mining algorithm and a support tuning mechanism. Decis Support Syst 2006.

[7] Hahsler M, Grün B, Hornik K, Buchta C. Introduction to arules – A computational environment for mining association rules and frequent item sets. J Stat Softw 2005.