Operation and Maintenance Analysis for Power Communication Networks based on Big Data

Zhan SHI[1] Ying ZENG[1] Li-qian SUN [2]
[1]Power Grid Control Center of Guangdong China South Power Grid
[2]State Key Laboratory of Network and Switching Technology, Beijing University of Posts and Telecommunications

Abstract： To support the reliable operation and reasonable construction demands of power communication networks, this paper uses Big Data technologies to analyze the equipment operation and maintenance status of power communications networks based on the equipment ledger, failure and maintenance data. The actual network data is mined using this method, and according to the data mining result, the equipment operation and maintenance situation of power communication networks is analyzed from the points of maintenance effect, maintenance cost and time distribution of maintenance.

Key words: Power、telecommunication、network, equipment operation and maintenance, Big Data, data mining

1. Introduction

Currently, how to get value from unstructured, heterogeneous and diversified data by using the big data technologies has become a hot topic in academia and industry [1-3]. For example, Internet service providers use the consumers' behavior data by data mining to find out their personalized preferences, which provides the basis for targeting advertising. Meanwhile, network operators diagnose the potential network anomalies and its corresponding factors through comprehensive analysis. It studies the parameter configuration data, alarm data and KPI big data, to provide effective decisions to guarantee the quality of service (QoS) [4-5]. However, data analysis of power communication network operation and maintenance is still based on the traditional experience-based judgment [6]. And it's lack of efficient mining and analysis means in the era of big data [7-8].

To address the problem above, we will use big data technologies to analyze and process the data of equipment account, operation, maintenance and construction in electric power communication networks. We accomplish the correlation between the different factors by combining Map Reduce and FP-growth technologies based on factors' characteristics. Finally, we give effective analysis results for the features of equipment operation and maintenance.

2. Indicators and influencing factors of equipment operation and maintenance

2.1. Maintenance effect

Maintenance effect is mainly used to test the failures occurred after overhaul for equipment of type *E*. We set $K(e)$ to represent a certain equipment $e \in E$ is faulty, and $T_J(e)$ to represent the latest overhaul time of the equipment *e*. The probability function $\Pr(K(e) \mid t - T_J(e)) = 1$ indicates that the equipment *e* fails at time $t - T_J(e)$ after its latest overhaul, therefore the length of time duration $t - T_J(e)$ determines the effect of maintenance for equipment *e*.

2.2. Maintenance costs

2.2.1. Time overhead

Time cost is the time duration between the start time and the completion time of maintenance. We suppose the maintenance of a certain equipment e($e \in E$) starts at time $T_S(e)$, and completes at time $T_O(e)$. So the average time cost of equipment *e* is defined by the following function:

$$T_{Cost}(E) = \frac{\sum_{N(E)} \left(T_S(e) - T_O(e)\right)}{N(E)} = \frac{\sum_{N(E)} \left(T_{Fix}(e) + T_{Non-fix}(e)\right)}{N(E)} \quad (1)$$

where $N(E)$ represents the total number of maintenance for equipment of type E. $T_{Fix}(e)$ is the fixed time cost, which represents the actual effective time of project implementation. Besides, $T_{Non-fix}(e)$ is the rest time cost and mainly affected by the remoteness of the equipment that needs maintenance.

2.2.2.Integrated cost

Besides the time cost, the integrated cost also covers the labor cost $M_{Cost}(e)$. So, the integrated cost $Cost(e)$ of equipment e can be measured by:

$$Cost(e) = (M_{Cost}(e))^a (T_{Cost}(e))^b \qquad (2)$$

where a and b respectively represent the coefficients of influence degree of labor and time.

2.2.3.Time distribution of maintenance

After determining the equipment type, time distribution of maintenance describes the dispersion of workload on different time scales (yearly, monthly, and weekly). It is defined by:

$$D(E) = \sum_{N(E)} \Pr(J(e) \mid \Delta T_{obs}) \qquad (3)$$

In this formula, $\Pr(J(e) \mid \Delta T_{obs})$ indicates the maintenance state. If the equipment e is under maintenance, $\Pr(J(e) \mid \Delta T_{obs}) = 1$, otherwise $\Pr(J(e) \mid \Delta T_{obs}) = 0$.

3. Data preprocessing

3.1. Data cleaning

1) Considering their special meanings, we ignore the missing values in power grid communication network maintenance database.

2) For noise data, we apply the outlier analysis.Through this method, we can identify the data which are significantly deviate from the rest, and then, put them aside.

3.2. Data integration

To reduce the dimensions of database, here we apply the correlation analysis to identify the useful data indexes among all the indexes given in power grid communication networks maintenance database.

For the nominal data of index *A* and *B*, we apply the Chi-square test

$$\chi^2 = \sum_i \sum_j \frac{(o_{ij} - d_{ij})^2}{d_{ij}} \qquad (4)$$

In the formula above, o_{ij} represents the observed frequency of joint event (A_i, B_j), and d_{ij} stands for the expected frequency of (A_i, B_j). We can compute d_{ij} by:

$$d_{ij} = \frac{count(A = a_i) \times count(B = b_j)}{n} \qquad (5)$$

Here, n indicates the number of data tuples and $count(A = a_i)$ is the number of data tuples that having index A with value a_i. While $count(B = b_j)$ is the number of data tuples that having index *B* with value b_j.

For numeric attribute, we apply the correlation coefficient test:

$$r_{A,B} = \frac{\sum_{i=1}^{n}(a_i - \bar{A})(b_i - \bar{B})}{n\sigma_A\sigma_B} = \frac{\sum_{i=1}^{n}(a_i b_i) - n\bar{A}\bar{B}}{n\sigma_A\sigma_B} \qquad (6)$$

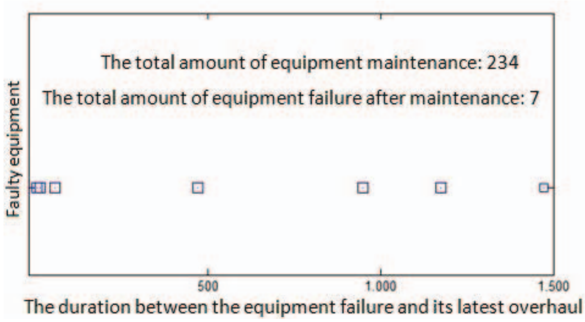

Figure 1 Distribution of maintenance effect for data network equipments

3.3. Data reduction

In this paper, we realize data reduction by the principle components analysis. The data set that needs reduction consists of n attributes or data vectors. The intention of principal components analysis is to find k n-dimension orthogonal vectors which can mostly represent original data set with $k \le n$. Compare to the method of correlation analysis that reduce the dimension of data by identifying the most correlated indexes, the principal components analysis projects the original data set into a smaller one by constructing an alternative and smaller index set.

By data preprocessing, the accuracy, consistency and integrity of data are increased, and the dimension of date is reduced meanwhile. Furthermore, it reduces the difficulty of data mining in power grid communications network equipment maintenance analysis, and improves the efficiency of mining algorithm.

4. MR FP-Growth algorithm for mining

By FP-Growth method, the problem of looking for long-frequency mode is converted into recursively searching for some short modes in smaller condition database, and then connects suffix. The most infrequent entries are used for suffix, which offers better selectivity.

When the database is too large, such as the power grid communication network herein, it is hard to construct FP-Tree based on the main memory. To solve this problem, the MapReduce technology and FP-Growth algorithm are combined in this paper. Parallel computing characteristic of Hadoop cluster is used to accelerate the convergence speed. Therefore, the enhanced FP-Growthi algorithm (MR FP-Growth algorithm) is proposed with specific ideas as follows:

(1) implementation methods:

a) Dividing object databases, and distributing them to different nodes in the cluster.

b) Computing support count for each node, aggregating these data to the same node, and then constructing frequent item sets.

c) Obtaining local frequent item sets by using FP-Growth algorithm.

(2) implementation process:

a) TDB will be evenly distributed to different working nodes.

b) Distributedly Computing F_List.

c) Parallelly computing local frequent item sets. The affairs set will be distributed to the same node to build FT_i by Map. Reduce constructs SubFP-Tree.

5. Result analysis

We select fault data, equipment maintenance data and related resource account data as the input source. All data is between 2010 and 2014 from a power company. We will analyze the data in terms of maintenance effect, maintenance costs and time distribution of the maintenance workload.

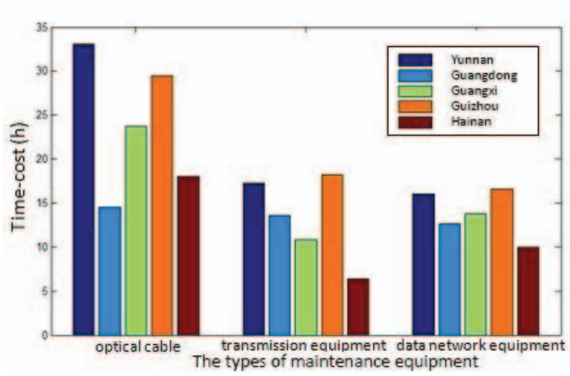

Figure 2 Distribution of average maintenance time-cost

5.1. Maintenance effect

Figure 1 presents the maintenance effect for different types of data network equipments, in which X-axis represents the duration between the equipment failure and its latest overhaul. From this figure we can find that, 234 overhauls of data network equipment have been carried out from 2010 to 2014. Among those overhauled equipments, 7 equipments failed later. Moreover, the time duration of three of them between equipment failure and the latest overhaul is within two months, showing a certain correlation..

5.2. Maintenance time cost

Figure 2 presents the average maintenance time–cost of different equipments in different regions. The average maintenance time cost is represented by Y-axis, while the types of maintenance equipment and their provinces are shown on the X-axis. As shown in figure 2, the cables have longer average maintenance time than transmission equipments and data network equipments. We can find that, maintenance work in Guangdong Province has better enforcement in terms of maintenance time cost than other provinces.

5.3. Time distribution of maintenance workload

From a macro perspective, figure 3 reflects the monthly changes of the average maintenance workload for various equipments in different years. As shown in this figure, the major overhauled equipments like optical cables, transmission networks and data networks will embrace maintenance peak after the Spring Festival.

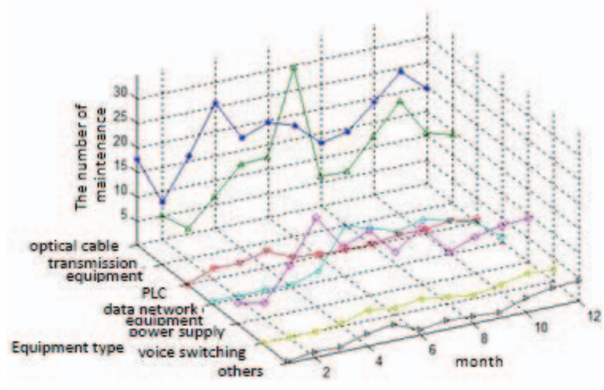

Figure 3 The monthly changes of the average maintenance workload for different equipments

Figure 4 reflects the weekly changes and monthly changes of the average maintenance workload for different equipments. From the perspective of week, the peak periods for all kinds of equipment concerned in this analysis occurred in weekdays. In general, the maintenance workload in weekend is less than in weekdays, even though the cables and transmission equipments that have more maintenance requirements need a certain workload in the Saturdays.

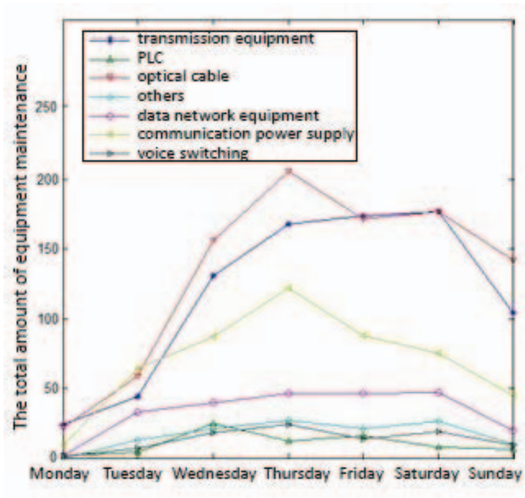

Figure 4 The weekly changes of the average maintenance workload for different equipments

6. Conclusion

In this paper, big data processing technique is used to analyze operation and maintenance of power communication network equipment. Firstly, the data reduction for desired attributes is adopted to pretreatment the original data set, and to cluster the important attributes. Then, indicators and influencing factors of equipment operation and maintenance are proposed. Finally, fault data, equipment maintenance data and related resources ledger data are analyzed by the proposed method. Analysis results can provide effective support for operation and maintenance of network.

Reference

[1] Li J, Lin L J. The research on vulnerability of power grid communication networks. INFORMATION & COMMUNICATIONS, 2014 (11): 57-58.

[2] Zhang P, Cui Q M, Hou Y Z, et al. Opportunities and challenges of wireless networks in the era of big data (in Chinese). Chin Sci Bull, 2015 (60):433-438.

[3] Zhong Fan, Parag Kulkarni, Sedat Gormus, et al. Smart Grid Communications: Overview of Research Challenges, Solutions, and Standardization Activities [J]. IEEE Communications Surveys and Tutorials, 2013, 15(1): 21-38.

[4] SHI Jian, LIU Yi-chang, CHAO Yu-jian, LIU Jun. Research on Power Communication Index System. Electric Power Information and Communication Technology, 2013, 11(12): 20-25.

[5] Yuan Li-na. The design and implementation of the analysis system of the operation index of electric power communication networks [D], Beijing University of Posts and Telecommunications, 2013.

[6] ZENG Ying. Power Communication Network Reliability Analysis and Evaluation Methods, Telecommunications for Electric Power System, 2011, 32(226): 13-16.

[7] GAO Hui-sheng, RAN Jing-xue, SUN Yi-qun. Risk evaluation of communication network of electric power based on improved FAHP. SYSTEMS ENGINEERING-THEORY & PRACTICE, 2008(3): 133-138.

[8] BAI Yi-long. Research on the reliability of electric power communication system, Information & Communications, 2015(3): 185-185.

Author's brief introduction and contact information

Zhan SHI, is an engineer of Power Grid Control Center of Guangdong China South Power Grid, Guangzhou, China.
His current research interests include Communication and Information Systems. Tel.: 15813383500. Address: Meihua Road No. 75 Yuexiu District Guangdong Province. Email: rhett.shi@139.com.

Li-qian SUN, is a master student of State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, China. Her current research interests include Smart Grid and Network Management. Email: lqsun@bupt.edu.cn.