A Multiple Method Of Discriminating And Correcting Abnormal Operation Data Of Power Grid Based On Point-To-Point Ratio

Guangyao Yu[1],Tianhao Wang[1],Liang ZHang[2],Xianxu Liu[3]

[1]State Grid Tianjin Electric Power Research Institute, Tianjin 300384, China
[2]State Grid Tianjin Chengnan Power Supply Company, Tianjin 300000, China
[3]State Grid Tianjin Chengxi Power Supply Company, Tianjin 300000, China

Keywords: point-to-point multiple ratio method; abnormal operation data; data screening; data modification.

Abstract

By mining and analyzing the operation data of power grid, it is possible to obtain the regularity, characteristics and status of power network. The data may be distorted or lost in the process of acquisition, transmission and storage, which may cause abnormal data and bring negative influences on analysis work. The screening and revising method of the operation data which has been proposed in this paper, can filter out and correct the abnormal data. The application case shows that the revised operation data can closely match the actual operation data.

1 Introduction

The grid dispatching centre is also the centre of grid operation data. SCADA, EMS, CPS and other systems generate a large amount of data every day, which contains all the information of grid operation and safe production [1,2]. It is possible to get the regularity, characteristics and status of grid operation through relative data deep excavation, and realize the lean operation management, improve the safety, quality and economic operation level of power grid.

Grid operation data has already played an important role in the operation analysis of power enterprises. From 2011, State Grid Corporation of China, China's largest power grid company, has collected and applied operation data to carry out grid development diagnosis analysis. These data contains main network historical factors (above 220kV) like equipment active power P, equipment reactive power Q, current I, voltage U, and has been collected from the power grid terminal acquisition equipment on the hour for 24 points of operating data. The diagnosis and analysis work includes the annual maximum load rate distribution of transformers and lines, the annual average equivalent load rate distribution and other power grid dispatching operation indicators. Finding weak links and potential safety hazards is the purpose of the analysis work. Operation data is the most basic and important data for power enterprises grid operation diagnosis and analysis. The authenticity and accuracy of data are directly related to real results of diagnostic analysis.

Grid operation data is divided into 2 categories [2], one is normal data, which conforms to the statistical characteristics value and can be normally used; the other is abnormal data, which is in violation of business logic or deviating from its statistical characteristics. Abnormal data comes from many reasons[3], such as harsh environment interference on terminal acquisition equipment operation, leading to data acquisition, transmission, storage, information loss or distortion. Abnormal data values can often reach several times or even several dozens of times than the equipment safety threshold, so it is necessary to identify and correct abnormal data, otherwise there will be a significant negative impact on the power grid diagnosis.

2 Basic theory basis

The power grid short-term load has the following five characteristics [4,5]:

(1) the similarity of the whole 24 hours variation rules between different days;

(2) the similarity of same typical weekdays in different weeks;

(3) the respective similarity on workdays and holidays;
(4) the similarity of the major holidays load curve in different years;
(5) the short-term load is also significantly affected by environmental factors, such as seasonal change, equipment accidents and maintenance. Based on the characteristics which are mentioned above, the real operation data and the adjacent acquisition time operation data collected by monitoring equipments also have the same obvious periodicity at the abnormal data acquisition moment, and affected by the operation environment.

The point-to-point multiple ratio method mentioned in this paper is a normal daily forecasting method based on time series analysis. It is commonly applied in short-term load forecasting. The basic idea of this method [4,6].is to obtain an exponential smoothing result about certain moment load forecasting value from other same period values in recent days. This one by one moment procedure is so-called point-to-point method.

3 Abnoramal data screening and correction method

In this paper, the grid abnormal data screening and correction method is mainly divided into three steps: Firstly, set up operating data packets as the basis of data processing, the data contains operation data and grid operation information collected by monitoring equipment; Secondly, filter abnormal data from the whole operation packet; Finally, based on the normal point-to-point multiple ratio method, the operation data modification method is proposed. Each step will be described in detail in the following paragraph.

Assume that a device element, the data collected on the certain day i and at the certain moment t of the day is D_{it} , the device component limit threshold is A_{im}.

3.1 Set up power grid equipment operation monitoring data packet

Using power grid equipment operation data collected by information acquisition system, and storing data into the database of dispatching control system. The data storage is based on the unique ID number of each device element, and then each ID number corresponds to a series of operation data for entire year of certain device. At the same time, each device in the system has been set a limit threshold value, lines (above 220kV) stable control limit value and, transformers rated capacity are the main factors of equipment element limit threshold value. In addition, maintenance, failure, disable period of standby equipments are supposed be marked in this mechanism.

3.2 Filter abnormal data from the whole operation packet

When 0, if in the period of normal outage, D_{it} is normal data, if in the period of abnormal outage, D_{it} is abnormal data; if $D_{it}>A_{im}$, D_{it} is abnormal data, if $D_{it}<A_{im}$, D_{it} is normal data.

3.3 Use Point-to-Point ratio multiple method to modify abnormal data to correct value

Point-to-point ratio multiple method[4,6]: used for load forecasting is based on the prediction before running data calculations predict daily operation data, and operation data before and after the abnormal data point are known. The operating data before and after the abnormal data points are taken into account based on the method, and an improved point-to-point ratio multiple method is proposed to deal with the abnormal data and obtain higher accuracy.

Point-to-point ratio multiple data processing methods mainly include the following steps:

Based on the cyclical feature of grid operation, researchers can pick up 7 days in 1 week as a related load period and set up the day when abnormal data D_{it} happens as a boundary, the 7 days or a week before the moment as the is 1st period, the 7 days or a week before the moment as the is 2nd period. If there are other abnormal data in the two selected periods, these data are supposed to be removed. If the abnormal data is in 1st period, then pick up same time point data of same day type in the week before 1st period. If the abnormal data in 2nd period, then pick up same time point data of same day type in the week after 2nd period, and the rest may be deduced by analogy.

When select relative loads in Day 14(n=14), take the abnormal data collected at moment t:

Pick up values at t moment in different type of days in a week of 1st period and figure out a smooth curve between

each point.

$$D_{it}^{1} = \alpha D_{3it} + \alpha（1-\alpha）D_{4it} + \cdots + \alpha（1-\alpha）^{5} D_{8it} \quad (1)$$

$D_{3ij} \ldots D_{8ij}$ is operation data at j moment in different type of days in a week of 1st period.

Pick up values at j moment in different type of days in a week of 2nd period and figure out a smooth curve between each point.

$$D_{it}^{2} = \alpha D_{9it} + \alpha（1-\alpha）D_{10it} + \cdots + \alpha（1-\alpha）^{5} D_{14it} \quad (2)$$

$D_{9it} \ldots D_{14it}$ is operation data at t moment in different type of days in a week of 2nd period.

In this sense, α is a smooth coefficient of continuous operation data points and could be selected from 0 to 1.

In addition, choose D_{1it} as a operation data at t moment in same type of days in a week of 1st period.

$$\frac{D_{it}^{'}}{D_{it}^{1}} = \frac{D_{1it}}{D_{it}^{2}} \quad (3)$$

Consequently, it shows that modification operation data at t point is $D_{it}^{'}$:

$$D_{it}^{'} = \frac{D_{1it}}{D_{it}^{2}} * D_{it}^{1} \quad (4)$$

4 Application Examples

In order to prove the application effect of this method, the transformer operation data collected by SCADA of Tianjin Power Grid is taken as an example.

Table Ⅰ is the format of the grid equipment operation data content collected by the grid information collecting system, which collect and save data into the database of the dispatching control system. In which ID is the equipment number, DDATE is the acquisition date, V00 ~ V23 are 24 integral points of a day for operation data it collected (MW) in segment from July, 23th, 2014 to July, 29th, 2014.

Table Ⅰ Original Transformer Operation Data In Grid dispatching Control Center

ID	DDATE	V00	V01	V02	…	V20	V21	V22	V23
hx2202P	2014-6-17	15.26	12.25	13.59		20.62	21.96	19.95	15.26
hx2202P	2014-6-18	10.24	7.22	9.87		23.30	23.30	17.27	11.58
hx2202P	2014-6-19	12.58	10.24	10.91		16.60	16.27	13.92	11.91
hx2202P	2014-6-20	11.24	14.59	13.92		21.96	21.29	14.93	11.58
hx2202P	2014-6-21	11.91	11.56	11.89		16.27	14.59	14.59	9.23
hx2202P	2014-6-22	10.23	12.14	12.81		15.93	17.61	24.98	14.93
hx2202P	2014-6-23	19.95	8.23	12.58		18.95	19.62	12.25	06.55
hx2202P	2014-6-24	5.54	15.60	12.92		111.91	16.27	8.90	5.88
hx2202P	2014-6-25	10.57	5.88	9.23		23.30	18.28	13.92	16.27
hx2202P	2014-6-26	7.55	10.57	0		11.56	12.69	13.25	12.58
hx2202P	2014-6-27	14.93	6.21	8.26		14.93	0	12.92	10.91
hx2202P	2014-6-28	17.27	6.55	104.54		15.93	21.96	18.28	11.91
hx2202P	2014-6-29	7.55	12.58	6.21		13.25	9.23	7.22	13.92
hx2202P	2014-6-30	3.5342	5.8796	7.22		19.28	15.93	12.25	6.89

The following process steps are carried out according to the method proposed by the present invention:

4.1 Set up power grid equipment operation monitoring data packet

Add two pieces of data after the device with the ID hx2202P, one is the device limiting threshold value, which means that the threshold is line stability control limit If the device is a line and the threshold is transformer rated capacity if the device is a transformer. Obviously Hx2202P is a transformer ID with certain rated capacity. The other is to mark the device whether it is in a known outage time period, and if so, it is marked as 1, and if not, it is set to the default value 0. After the process above, the content format of the monitoring data packet of the completed network equipment operation is shown in Table Ⅱ.

Table Ⅱ Network Equipment Operation Monitoring Packet

ID	DDATE	V00	V01	V02	…	V20	V21	V22	V23	Threshold Value	Device Status
hx2202P	2014-6-17	15.26	12.25	13.59		20.62	21.96	19.95	15.26	24	0
hx2202P	2014-6-18	10.24	7.22	9.87		23.30	23.30	17.27	11.58	24	0
hx2202P	2014-6-19	12.58	10.24	10.91		16.60	16.27	13.92	11.91	24	0
hx2202P	2014-6-20	11.24	14.59	13.92		21.96	21.29	14.93	11.58	24	0
hx2202P	2014-6-21	11.91	11.56	11.89		16.27	14.59	14.59	9.23	24	0
hx2202P	2014-6-22	10.23	12.14	12.81		15.93	17.61	24.98	14.93	24	0
hx2202P	2014-6-23	19.95	8.23	12.58		18.95	19.62	12.25	06.55	24	0
hx2202P	2014-6-24	5.54	15.60	12.92		111.91	16.27	8.90	5.88	24	0
hx2202P	2014-6-25	10.57	5.88	9.23		23.30	18.28	13.92	16.27	24	0
hx2202P	2014-6-26	7.55	10.57	0		11.56	12.69	13.25	12.58	24	1
hx2202P	2014-6-27	14.93	6.21	8.26		14.93	0	12.92	10.91	24	0
hx2202P	2014-6-28	17.27	6.55	104.54		15.93	21.96	18.28	11.91	24	0
hx2202P	2014-6-29	7.55	12.58	6.21		13.25	9.23	7.22	13.92	24	0
hx2202P	2014-6-30	3.5342	5.8796	7.22		19.28	15.93	12.25	6.89	24	0

4.2 Filter abnormal data from the whole operation packet

When some operation data is larger than the limit value after 2 comparison, it is abnormal data and showed in Table Ⅱ with red mark number. If the operation data is 0, outage tag should be marked and if the tag is 1, it is normal data like V02 on day of July, 26th, 2014 showed in Table Ⅱ which is known as outage time period. If the tag is 0, then it is the abnormal data, such as V21 data in Table 2 on the day of

July, 27th, 2014.

4.3 Use Point-to-Point ratio multiple method to modify abnormal data to correct value

Take the V20 data of June 24th as an example, the week before the working day is regarded as the 1st period and the following week is the 2nd week;

Then choose V20 data in different type of days in a week of 1st period to figure out a smooth curve and get the following formula:

$$D^1_{i20} = \alpha D_{3i20} + \alpha（1-\alpha）D_{4i20} + \cdots + (1-a)^5 * D_{8i20} \quad (5)$$

is operation data at V20 moment in different type of days in a week of 1st period.

$$D^1_{i20} = \alpha * 18.95 + \alpha(1-\alpha) * 15.93 + \cdots + \alpha(1-\alpha)^5 * 23.3 \quad (6)$$

Similarly, choose V20 data in different type of days in a week of 2nd period to figure out a smooth curve and get the following formula:

$$D^2_{i20} = \alpha D_{9i20} + \alpha（1-\alpha）D_{10i20} + \cdots + (1-a)^5 * D_{14i20} \quad (7)$$

$D_{9i20} \ldots D_{14i20}$ is operation data at V20 moment in different type of days in a week of 2nd period.

$$D^2_{i20} = \alpha * 23.3 + \alpha(1-\alpha) * 11.56 + \cdots + \alpha(1-\alpha)^5 * 19.28 \quad (8)$$

In this sense, α=1/2 is a smooth coefficient of continuous operation data points.

Therefore, D^1_{i20}= 17.75 D^2_{i20}=18.12

Another factorD_{1i20}=20.62 is known as t moment value in different type of days in a week of 1st period. $\frac{D'_{i20}}{D^1_{i20}} = \frac{D_{1i20}}{D^2_{i20}}$;

And modified V20 moment value D'_{i20} is,

$$D'_{i20} = \frac{D_{1i20}}{D^2_{i20}} * D^1_{i20} = 20.62/18.12*17.75 = 20.20 \quad (9)$$

It shows that modification value is more closer to actual situation, this method performs pretty well in data correction.

5 Summary

In this paper, based on the equipment operation data obtained by the grid operation monitoring system, researchers had set up foundation operation data packets and finally proposed an abnormal data correction mode based on the point-to-point ratio multiple method, in which all the abnormal dada can be screened scientifically .The method can accurately filter the abnormal data in entire operation data and correct the abnormal data accurately, and the revised operation data can closely match the actual operation data.

References

[1] Shao Lizheng,Liu Dunnan, Zeng Ming.Operating Analysis and Data Mining System for Power Grid Dispatching[C].2011.p.140-143

[2] Zhang Liang. Researchon Data Quality for Electric Power Dispatch Datacenters[J].East China Electric Power,2009,37(3): 0403-0405.

[3] Li Canbing, Liu Xiaoguang, Zhao Hongjun, etal.Analysis and Processing Methods of Bad Load Data in Medium-voltage Distribution Network[J].Automation of Electric Power ystems , 2008, 32(20):97-100.

[4] Kang Chongqing,Xia qing,Li umei.Power System Load Forecasting[M].Beijing:China Electric Power Press,2007.

[5] Kang Chongqing , XIA Qing , ZHANG Boming. Review of power system load forecasting and its development[J]. Automation of Electric Power Systems, 2004, 28(17) : 1-11.

[6] Liu Mei,Xu Hongqiang,Li Zhen,etal.Short-Term Load Forecasting Based on Periodic Analysis[C].China Universities Annual Conference on Electric Power Systems & Automation,Zheng zhou,China.2004.p.1633-1636