Research on Error Correction Method of Massive Data Based on Arbitrary Partition Area

Runqing Bai1, a, Yong Xin2, Xin He1, Wensi Huang2, Xin Lu2 and Jing Chen2

1Electric Power Research Institute, State Grid Gansu Electric Power Company, Lanzhou 730050, China
2SGIT-GreatPower, FuJian, FuZhou, 350003, China

Abstract. With the continuous construction of intelligent power grids, a large amount of data on power grid operation, production management and market operation have been accumulated in power systems, which are affected by factors such as collection, communication, and environment. There are quality problems in terms of integrity, consistency, and effectiveness of data. In this paper, by studying the characteristics of massive data and data anomaly identification method, a method of error correction for massive data in power grid based on arbitrary partition area is proposed. Taking abnormal data as the center and dividing the area based on the network topology structure, the abnormal data of the network is corrected layer by layer to achieve the purpose of improving the data quality.

1. Introduction

With the rapid development of modern society and science and technology, electric energy plays an increasingly irreplaceable role in today's social life. At present, all countries in the world are accelerating investment and building large-scale smart grids. The installation and deployment of a large number of smart meters and smart terminals enable the integration of communication, computer and information technology on the basis of physical grid to form a new kind of Power composition network [9]. Digital technology is widely used in intelligent power systems, which will inevitably generate massive of data, involving various links such as power generation, transmission, substation, power distribution, power consumption, and dispatching [2]. However, due to various reasons such as remote terminal unit acquisition, energy meter acquisition, channel status, parameter setting, system operation, hardware platform, provincial and municipal forwarding, and human factors [1], quality problems such as incomplete data, inconsistent cross-system data and abnormal data generally exist in some business system data. With the accelerating pace of grid informationization, these data problems have seriously affected the integration and highly integrated application of system data. Therefore, it is necessary to strengthen the monitoring of power grid data quality and the error correction processing of abnormal data, so as to fundamentally solve the data quality problem and provide a strong guarantee for improving the operation management, development planning and strategic decision-making of power grid enterprises.

At present, the data error correction processing method mainly focuses on the error correction when an abnormality occurs in the data transmission process, and lacks of error correction of the data itself...
anomalies. In [3], a Forward Error Correction Method (AFEC) with adaptive characteristics based on Real-Time Transport Protocol (RTP) that can repair single packet loss is proposed. By dynamically adjusting the error coding parameters (error coding packet length), the network congestion change can be tracked to ensure the quality of transmission service. In [4], a method of error correction by secondary packaging is proposed. In view of the failure of data synchronization, the data packets that failed to synchronize were supplied again by adopting the method of secondary packaging, so as to solve the problem of data inconsistency caused by errors such as packet loss and unpacking failure due to unexpected faults in the process of database synchronization. Data inconsistency caused by errors. In [5], relevant strategies for data error correction are proposed based on the verification information of EEC check bits. By introducing a filtering algorithm, the correct bits are eliminated and a set of suspect bits is obtained, which contains most of the erroneous bits. These suspicious bits are then further tested using a random flip algorithm with polynomial complexity to find the most likely erroneous bits, thereby minimizing the number of erroneous bits in the data packet.

Aiming at the anomaly of data itself, this paper puts forward an error-correcting processing method of massive data of power grid based on arbitrary segmentation region. By using the detection results of power grid data quality and power grid topology structure, the power grid is segmented to achieve error-correcting processing of abnormal data in different regions and improve data quality.

### 2. Research on error correction processing based on massive data of arbitrary segmentation area power grid

![Diagram](image.png)

**Figure 1.** Abnormal data identification and process of error correction processing of power grid.

The massive data of the power grid is formed by the interconnection or interaction of data of systems such as equipment (asset) operation and maintenance lean management system (PMS2.0), dispatching OMS/SCADA system, marketing business application system, electric energy information collection system, electricity information collection system, marketing basic data platform, grid GIS platform. For
these structured, semi-structured and unstructured massive data, in order to improve the data quality and improve the data availability rate, for identification of abnormal data and error correction. This paper proposes a multi-source data error correction processing method based on arbitrary segmentation region to realize multi-source data reliability management of power grids and improve grid data quality. The following is the process of abnormal data identification and error correction processing:

2.1. Research on abnormal identification method of massive data in power grid

The quality of the massive data of the power grid is an important basis for managers to make scientific decisions. The low quality of accuracy, integrity and consistency of data will affect the quality of decision analysis and even produce wrong decisions and suggestions. Therefore, the data quality of each business system must be improved to ensure the accuracy, integrity and consistency of the data source, improve the data availability and credibility, reduce the difficulty and risk of decision-making, and better provide services for decision analysis.

The data quality includes the data itself quality, the quality of data storage, the quality of data usage and the quality of data transmission, etc. This paper studies the quality problem of the data itself, based on the detection methods of database constraint rules, business rules, data mining methods, state estimation methods, etc., the data quality of each power business system is tested.

(1) Data anomaly identification based on database constraint rules

Database constraint rules are mainly data integrity constraints, and some data validation rules are enforced on the data table, including Not Null, Unique, Primary Key, Foreign Key, and Check, using the constraints conditions of the database itself to achieve the detection of data quality of each business system, such as data type, length, format, precision, etc., the constraints conditions are as follows:

| The serial number | The constraints conditions | Constraint rule specification |
|-------------------|---------------------------|-----------------------------|
| 1                 | Not Null                  | Used to ensure that the value of the current column is not null |
| 2                 | Unique                    | Used to ensure that any two columns of data in the protected data column are different |
| 3                 | Primary Key               | Used to uniquely identify a record in the table |
| 4                 | Foreign Key               | The mechanism used to maintain data integrity and consistency has a good validation effect on business processing |
| 5                 | Check                     | Used to limit the range of values in a column |

(2) Data anomaly identification based on business rules

The data anomaly identification based on business rules mainly uses the knowledge of the power system business, combined with the relevant design specifications of the power grid, to detect the data that violates the power system business rules appearing in each business system.

1) Testing the rationality of archival data. In the equipment (asset) operation and maintenance lean management system, check the reasonableness of wire length, cable length, transformer capacity, etc., for example:

| The serial number | Equipment type | Detection item | Detection rules |
|-------------------|----------------|----------------|-----------------|
| 1                 | Wire           | Length         | Have a value, greater than 0, can not seriously exceed the design specifications |
| 2                 | Cable          | Length         | Have a value, greater than 0, can not seriously exceed the design specifications |
| 3                 | Transformer    | Capacity       | Have a value, greater than 0, can not seriously exceed the design specifications |
2) Detecting the reasonableness of data collection. Under normal operating conditions, the data collected by the power grid has certain continuity. When the irrational values such as sudden changes occur in the electric energy collection system and the electricity information collection system, for example, the 10kVXX distribution load fluctuation analysis shown in the figure below can be known, there is an abnormal jump in the active power at "7:30".

![Analysis of 10kVXX distribution load fluctuation](image)

Figure 2. Detecting the reasonableness of collected data.

3) Power grid topology relationship detection. There is a certain topological connection relationship between "station - line - transformer - household" in the power grid, and the topology connection relationship in the GIS platform of the power grid is detected, including free equipment, topological connection disconnection, abnormal loop, and so on.

4) Busbar balance detection. The unbalance degree of busbars of different voltage levels in the power network has a reasonable interval. Based on the results of multi-source data fusion, combined with the topology structure of the power grid, the unbalance degree of busbars is detected.

\[
\text{Busbar imbalance degree} = \frac{A_{\text{input power}} - A_{\text{output power}}}{A_{\text{input power}}} \times 100\% \quad (1)
\]

(3) Data anomaly identification based on data mining method
1) Data anomaly identification based on K-MEANS clustering analysis

Using K-MEANS clustering analysis method, clustering algorithm is used to cluster the analysis objects, and then the distance between each analysis object and the cluster center is judged, thus the method of judging data quality [6]. Data seriously deviating from the clustering center is often the data with quality problems.

Modeling with a certain data as the analysis object, assuming that the sample data is \( \{\chi^{(1)}, \chi^{(2)}, \ldots \chi^{(m)}\} \), \( \chi^{(i)} \in \mathbb{R}^n \), perform K-MEANS cluster analysis:

a) Randomly select \( k \) data points and divide them into \( k \) object groups. They are: \( \mu_1, \mu_2, \ldots, \mu_k \subseteq \mathbb{R}^n \).

b) According to the center of each cluster object, calculate the distance between each object and these center objects, and re-divide the corresponding objects according to the minimum distance until all data are allocated. The minimum distance function is:

\[
c^{(i)} = \arg \min_j \| \chi^{(i)} - \mu_j \|^2 \quad (2)
\]
c) Recalculate the center of each cluster.

\[ \mu_j = \frac{\sum_{i=1}^{m} 1\{c^{(a)} = j\} x^{(i)}}{\sum_{i=1}^{m} 1\{c^{(i)} = j\}} \] (3)

d) Repeat "b) and c)" until the function converges, then the algorithm terminates. Focus on the analysis of data that is severely out of the cluster center and achieve abnormal data detection.

2) Data anomaly identification based on correlation analysis

The measurement value of the power system changes at every moment, but the intrinsic relationship does not change regardless of how the measured data changes. For example \( P = U I \cos \phi \), the proportional relationship between \( P \) and \( I \) does not change. Data quality detection based on correlation analysis is to use this relationship to detect the quality of bad data. When abnormal data, sudden load changes or system structure changes, it is bound to break this relationship. Therefore, based on the constraint relationship of these laws, the correlation of measured values in the power system is used to realize the detection of bad data.

The variance can reflect the case where the data deviates from the average. In the case of system stability, the measured value of each point deviates from the average value within a certain range. If it exceeds the range, it indicates that the data of the point changes greatly. Therefore, the variance is used to detect abnormal data [11].

The measurement variance of the measurement point \( i \) is:

\[ \sigma_{i,t}^2 = \frac{\sum_{j=1}^{N} (x_{i,j} - \mu_t)^2}{N} \quad (i = 1,2,\ldots,N) \] (4)

Where: \( x_{i,j} \) is the measurement value of \( i \)-point \( t \)-time, \( \mu_t \) is the expectation of the measurement value of \( i \)-point \( t \)-time.

\[ \mu_t = \frac{\sum_{j=1}^{N} x_{i,j}}{N} \quad (i = 1,2,\ldots,N) \] (5)

Take \( N=1 \) to get the self-variance of measurement point \( i \):

\[ \sigma^2 = (x_{i,j} - \mu_t)^2 \quad (i = 1,2,\ldots,N) \] (6)

Self-variance reaction \( t \) time \( i \) point measurement value deviates from the range of average value.

(4) Data anomaly identification based on state estimation

State estimation is a method of estimating the internal state of a dynamic system based on available measurement data. The measurement data of the power system includes both telemetry and remote signaling. However, due to system failure and human factors, these measurement values may deviate from the true value, resulting in the occurrence of bad data and affecting subsequent state estimation. The method of bad data detection based on state estimation mainly includes estimation identification method, non-quadratic criterion method and residual search method. The flow of these methods is mainly as follows: ① Assume that the weighted residual or standard residual as the eigenvalue obeys a certain probability distribution. ② Set a threshold based on a certain confidence level to perform a hypothesis test. ③ Found index bad data [11].
2.2. Research on grid area segmentation method

(1) Classification identification of abnormal data

Based on the mass data quality detection results of the power grid, combined with the main network, distribution network, and low-voltage network power flow calculation, the data requirements and abnormal data are identified, classify the abnormal data that affects the power flow calculation into one category, and do not affect the abnormal data of the power flow calculation into another class.

Table 3. Classification identification of abnormal data

| The serial number | Classification          | Acquisition device type                      | Whether it affects power flow calculation |
|-------------------|-------------------------|-----------------------------------------------|------------------------------------------|
| 1                 | Main network            | Transmission line                             | No                                       |
| 2                 |                         | High pressure side of main transformer        |                                          |
| 3                 |                         | Medium pressure side of main transformer      |                                          |
|                   |                         | (triple coil transformer)                     |                                          |
| 4                 |                         | Low pressure side of main transformer         |                                          |
| 5                 |                         | The generator                                 | Yes                                      |
| 6                 |                         | External equivalent load                     | Yes                                      |
| 7                 |                         | Compensation equipment                        |                                          |
| 8                 | Distribution network    | Line                                          |                                          |
| 9                 |                         | Distribution transformer                      | Yes                                      |
| 10                |                         | On-column transformer                         |                                          |
| 11                |                         | High voltage user                             |                                          |
| 12                |                         | Compensation equipment                        |                                          |
| 13                | Low voltage network     | Distribution transformer                      | Yes                                      |
| 14                |                         | On-column transformer                         |                                          |
| 15                |                         | Low voltage user                              |                                          |

(2) Power grid area segmentation

Based on the detection and identification results of abnormal data, the abnormal data is located by using the topology of the power grid. As shown in Figure 4, the red station or line is marked with the abnormal data as the center to segment the power grid. The principle of segmentation is as follows:

1) The internal load of the main network, main transformer, and reactive power compensation equipment fails to be collected, and the plant station is divided into one area, as shown in area 3 in Figure 3.

2) If the transmission line, external equivalent system and generator acquisition fail, the plant stations connected with these devices are divided into the same area, as shown in area 1 and area 3 in Figure 3.

3) The distribution network or the station area fails to collect equipment and the distribution line or station area shall be used as the area for division.

4) If the divided area cannot completely correct all the abnormal data in the area, the area division range can be expanded.
3) The power grid segmentation area boundary equivalent
According to the area where the power grid is divided, the boundary of the divided area is equivalent, as follows:
1) Power injection point, such as generators, external equivalent systems, power input terminal of transmission line, etc., equivalent to external equivalence systems.
2) The power output point, such as the transmission line power output, load and other equipment, is equivalent to the load.

2.3. Research on abnormal error correction methods for data in grid segmentation areas
(1) Power grid division regional power flow calculation
According to the division area and boundary equivalent value, the power flow calculation model is constructed, and the Newton-Raphson method is used to calculate the regional power flow. The algorithm is as follows:
The nonlinear equations are set as follows:

\[
\begin{align*}
    f_1(x_1, x_2, \ldots, x_n) &= 0 \\
    f_2(x_1, x_2, \ldots, x_n) &= 0 \\
    \vdots \\
    f_n(x_1, x_2, \ldots, x_n) &= 0
\end{align*}
\]  

(7)

Its approximate solution is \(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}\). Let the approximate solution and the exact solution differ by \(\Delta x_1^{(0)}, \Delta x_2^{(0)}, \ldots, \Delta x_n^{(0)}\), respectively.
Each equation in the above equation is expanded by Taylor series. Taking the i-th equation as an example, we can get:

\[
f_i(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)} + \Delta x_i^{(0)}) = 0
\]

\[
f_i(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)} + \Delta x_i^{(0)}) = f_i(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}) + \frac{\partial f_i}{\partial x_1} \Delta x_1 + \frac{\partial f_i}{\partial x_2} \Delta x_2 + \ldots + \frac{\partial f_i}{\partial x_n} \Delta x_n + \Phi_i
\]

(8)

Where \(\Phi_i\) is a function containing the product of the higher order of \(\Delta x_j (j = 1, 2, \ldots, n)\) and the higher order partial derivative of \(f_i\). If the approximate solution \(x_j^{(0)}\) is not much different from the exact solution, the higher power of \(\Delta x_j\) can be omitted, so that \(\Phi_i\) can be omitted. In addition, \(\frac{\partial f_i}{\partial x_i}\) represents the value of the partial derivative of the function \(f_i\) to the independent variable \(x_i\) at the point \(x_j^{(0)} (j = 1, 2, \ldots, n)\). Thus it follows:

\[
\begin{align*}
&f_i(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}) + \frac{\partial f_i}{\partial x_1} \Delta x_1 + \frac{\partial f_i}{\partial x_2} \Delta x_2 + \ldots + \frac{\partial f_i}{\partial x_n} \Delta x_n \\
&f_2(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}) + \frac{\partial f_2}{\partial x_1} \Delta x_1 + \frac{\partial f_2}{\partial x_2} \Delta x_2 + \ldots + \frac{\partial f_2}{\partial x_n} \Delta x_n \\
&\vdots \\
&f_n(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}) + \frac{\partial f_n}{\partial x_1} \Delta x_1 + \frac{\partial f_n}{\partial x_2} \Delta x_2 + \ldots + \frac{\partial f_n}{\partial x_n} \Delta x_n
\end{align*}
\]

(9)

Written in matrix form as follows:

\[
\begin{bmatrix}
  f_1(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}) \\
  f_2(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}) \\
  \vdots \\
  f_n(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})
\end{bmatrix} =
\begin{bmatrix}
  \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \ldots & \frac{\partial f_1}{\partial x_n} \\
  \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \ldots & \frac{\partial f_2}{\partial x_n} \\
  \vdots & \vdots & \ddots & \vdots \\
  \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \ldots & \frac{\partial f_n}{\partial x_n}
\end{bmatrix}
\begin{bmatrix}
  \Delta x_1 \\
  \Delta x_2 \\
  \vdots \\
  \Delta x_n
\end{bmatrix}
\]

(10)
At this time, $x_j^{(1)}$ is not a true solution of the equations, but a step closer to the true solution $x_j$, which can be iterated and corrected repeatedly. The iterative process continues until the convergence judgment is met:

$$\max \left\{ f_i(x_1^k, x_2^k, \ldots, x_n^k) \right\} < \varepsilon_1$$  \hspace{1cm} (12)

Or

$$\max \left\{ \Delta x_j^k \right\} < \varepsilon_2$$  \hspace{1cm} (13)

K is the number of iterations, $\varepsilon_1, \varepsilon_2$ are pre-set small positive numbers.

(2) Power grid segmentation area abnormal data error correction

- Use power business rules, such as busbar balance, power supply/sales relationship, etc., to correct data affecting power flow calculation.
- Based on the results of power flow calculation, the abnormal data that does not affect the power flow calculation are corrected, such as the collected values of the high and low voltage side of transformer, the collected values of the first and end of transmission line, etc.

3. Case analysis

In this paper, the local power grid data of a company is taken as an example to conduct simulation. The simulation data comes from the collection of smart meters. The data of November 5, 2017 in this region is taken as the source data of data governance, combining with the error correction method of mass data of power grid based on arbitrary segmentation region listed above, the power grid data is governed. Local power grid geographical wiring is shown in Figure 4:

![Geographical wiring diagram of a company's local power grid.](image)

Figure 4. Geographical wiring diagram of a company's local power grid.

According to the data quality detection method introduced above, the acquisition of electricity at the lower hanging load 1 and load 2 of substation 3 fails, and the acquisition of electricity at the 2-3 gateway of 110kV line fails, and the grid power flow calculation does not converge.
Substation 3 includes: 110kV, 35kV, 10kV 3 voltage levels, 3 110kV in and out lines, 2 35kV in and out lines, 2 110kV busbars, 2 35kV busbars, 2 10kV busbars, 2 sets of 20000kVA main transformers, 2 loads, as shown in Figure 5. The abnormal data error correction processing is as follows:

1) Abnormal data classification. Based on the analysis of data quality detection results, it can be seen that the abnormal data affecting the power flow calculation include: electricity collection data of load 1 and load 2, and does not affect the power flow calculation include: 110kV line 2-3 gate power collection data, as shown in the red virtual box mark in Figure 5.

2) Divide the area. With load 1 and load 2 as the center, substation 3 is divided into one area, as shown in Figure 5.

3) Using the collected data to construct the calculation model to correct the abnormal data. According to the electric quantity collected by the low-voltage side of the main transformer 1# and the main transformer 2# of substation 3, consider the 10kV mother level of the substation 3, and calculate the load 1 and load 2 in the station. The details of the correction result are as follows:

| The serial number | Device name                  | Active power (MWh) | Reactive power (Mvarh) |
|-------------------|------------------------------|--------------------|------------------------|
| 1                 | Main transformer 1#- low pressure side | 29.6352            | 11.52                  |
| 2                 | Main transformer 2#- low pressure side | 47.736             | 20.64                  |
| 3                 | Load 1#                       | 29.6263            | 11.5165                |
| 4                 | Load 2#                       | 47.7217            | 20.6338                |

4) Correct the data using the power flow calculation results. After the load data correction is completed, the power flow calculation is performed on the local power grid shown in Figure 5, and the power flow calculation result is used to correct the collected power of the “110kV line 2-3 gateway”, and the correction result is as follows:
### Table 5. Correct the data using the power flow calculation results

| The serial number | Device name                        | Active power (MWh) | Reactive power (Mvarh) |
|-------------------|------------------------------------|--------------------|------------------------|
| 1                 | Substation 2-110kV line 2-3 gateway| 190.092            | 770.5872               |
| 2                 | Substation 3-110kV line 2-3 gateway| 185.3736           | 723.8208               |
| 3                 | 110kV line 2-3 line loss           | 4.7184             | 46.7664                |

5) After the data correction is completed, power flow calculation was carried out for the whole network, and the gap between the calculation result and the collected data is within a reasonable range, and the abnormal data is successfully corrected by using the method of arbitrary segmentation.

### 4. Summary

To ensure the data quality of the power grid is not only the guarantee for dispatching and operation and maintenance personnel to accurately and comprehensively grasp the operation status of the power grid, but also the important basis for management personnel to make scientific decisions. At present, the data error correction methods adopted by the power grid are generally global error correction or point-to-point error correction, on the data error correction effect can not meet the requirement of the power grid of the lean management, this paper combines the characteristics of the massive data of the power grid and the data quality inspection method. The multi-source data error correction processing method in the segmentation area corrects the surface and points of the anomaly data, and realizes the power grid data management from course to fine, step by step a correction, which improves the accuracy of data error correction and improves the data quality of the grid.

### Acknowledgments

Funded by the state grid corporation of headquarters science and technology project “research and application demonstration of key technologies of power grid mass data loss reduction analysis oriented to the same period line loss management” (Fund Project: 522722180003).

### References

[1] Huang Hui, Zhu Qiliang. Analysis and Prospect of the Development of Data Quality Control in Smart Grid [J]. Science and Technology Information, 2012 (07): 92-93.

[2] Lei Yu. Design and implementation of data cleaning subsystem for grid big data [D]. Beijing University of Posts and Telecommunications: 1, 2015.

[3] Li Shiyin, Liu Yuying, Qian Jiansheng. Research on adaptive forward error correction method for real-time data transmission [N]. Journal of China University of Mining and Technology, 2015-11 (6).

[4] Wang Weiwei. Research on database synchronization mechanism based on quadratic packet error correction [D]. Southwest University: 31-35, 2010.

[5] Wei Xingshen. Research on data error correction mechanism based on error-correcting code [D]. Nanjing University: 37-39, 2013.

[6] Yan Yuping, Xiao Zhanhui. Application of Lie group detection method based on cluster analysis in power grid data quality management [J]. Modern Electronic Technology, 2016, 39 (15): 138-139.

[7] Yang Yi, Yang Jie, Nie Tian. Research on intelligent analysis of power grid data based on data mining [J]. Electronic Technology and Software Engineering, 218, 224.

[8] Zou Jie. Research on Data Cleaning and Evaluation Model Based on Data Mining [D]. Beijing University of Posts and Telecommunications: 22, 2017.

[9] Lei Yu. Design and implementation of data cleaning subsystem for grid big data [D]. Beijing University of Posts and Telecommunications: 21, 2015.

[10] Wan Shun, Gui Ning, Chen Jiaying, Chen Xiaolong. Research on Power Grid Data Mining Technology under Big Data [J]. Information Communication, 2019, 193 (1): 182-183.

[11] Zhou Xiaobao. Research and application of fault detection and identification methods for power
system state estimation [D]. Hunan University: 14, 2013.

[12] Qin Xuan. Research on quality assessment of power statistics and anomaly detection methods, 2013: 2-3

[13] Yao Jinsong, Xin Yong, Huang Wensi, Lu Xin, Chen Jing, Huo Chengjun. Research on Quality Control Technology of Power Line Loss Data. Industrial Instrumentation and Automation Device, 2018 (21).

[14] Wang Weizhou, Lu Peijing, Xu Wei. Research on reactive power optimization analysis and calculation of arbitrary segmentation area [J]. Electric Power Equipment, 2007, (8).

[15] Yin Rui, Yu Yangqi, Wang Manyi, et al. Research on power data quality evaluation model and system in big data environment [J]. Automation Technology and Application, 2017, 36 (04): 137-141.

[16] Zhang Bo, Zhu Zeting, Lu Mingkai, et al. Discussion on the method of data loss anomaly data in the camping station [J]. Zhejiang Electric Power, 2015 (02): 9-11+19.

[17] Peng Xiaosheng, Deng Diyuan, Cheng Shijie, Wen Jinyu, Li Zhaohui, Niu Lin. Key Technologies of Power Big Data for Smart Grid Applications [J]. Proceedings of the CSEE, 2015 (03).

[18] Song Yaqi, Zhou Guoliang, Zhu Yongli. Current Status and Challenges of Big Data Processing Technology in Smart Grid [J]. Power System Technology, 2013 (04).