Research on the Combination Technology of Smart Meter Status Evaluation and Acquisition Function Based on Big Data Technology

Guorong Xie, Xiaohui Zheng, Dahai Zhuang, Hong Zheng, Yuanting Chen, Qiangqiang Gao, Weishan Zhen
State Grid Info-Telecom Great Power Science and Technology Co., LTD, Fuzhou Fujian 350003, China

Abstract. In order to study the key technologies suitable for large-scale intelligent electric energy meter automatic verification, the discrete verification mode is discussed from three aspects of data interaction, scheduling control, and intelligent diagnosis. First, build an application architecture for automated verification data from the perspective of multi-dimensional perception, use WCF and MQ communication modes to build a data interaction architecture, and propose a feedforward feedback integrated control mechanism; second, based on the automated verification control architecture, establish a space-time vertical and horizontal scheduling control strategy; finally, according to the category of automatic verification equipment, classify abnormal / fault categories and establish a fault knowledge base. Based on the intelligent diagnosis process of “alarm-discovery-diagnosis-processing-learning” mechanism, build an intelligent diagnosis application framework. The proposed method effectively guarantees the accuracy and reliability of the interactive data, realizes the automatic verification of space-time cooperative operation, and can effectively improve the daily operation and maintenance efficiency of large-scale automated verification.

1. Introduction
With the development of smart grid, smart electric energy meters, which are important support for power companies to strengthen lean management, improve quality service levels, expand power markets, and innovate trading methods, have been widely used in China. For this reason, the existing manual smart electric energy meter on-site fixed-period inspection determines whether the meter is out of tolerance. It not only consumes a lot of manpower and material resources, but also has a certain risk of on-site operation, and it is impossible to find the abnormal operation of the smart electric energy meter in a timely and effective manner. It has been unable to meet the lean operation and maintenance requirements of smart energy meters. Therefore, based on the technical characteristics of smart electric energy meters and making full use of data from information systems such as marketing business application systems, electricity information collection systems, and metering and production scheduling systems, an online inspection and evaluation indicator for smart grid status and its online platform are designed.
To this end, a smart grid state online inspection evaluation index and its online platform design scheme are proposed to provide a strong basis for the online smart energy meter defect inspection and operation and maintenance. First, it analyses the influencing factors of smart energy meter status verification, elaborate on the smart energy meter status verification evaluation index system; secondly, designs the system architecture and business process of the smart energy meter online detection and evaluation platform; finally, through practical cases, verification The appraisal index and the applicability of the online system are expected to provide a certain reference value for the lean operation and maintenance of smart energy meters [1].

2. Overview of Big Data Technology

Big data in the smart grid has the characteristics of "4V": large scale, many types, low value density and fast change. The processing of big data includes: acquisition-storage-analysis processing-display. From the four stages of the big data management life cycle, data storage and processing are the key to big data processing.

2.1. Big data storage technology

Big data storage mainly includes relational database cluster, distributed database, distributed file system, NoSQL database and distributed cache. These technical methods are suitable for different scenarios: relational database clusters and distributed databases are suitable for structured and transactional data; distributed file system (HDFS) is suitable for large-scale unstructured archive data; NoSQL database for large-scale unstructured, semi-structured streaming data. HDFS is the storage foundation of distributed computing. It can be deployed on decentralized and inexpensive hardware devices to store massive data sets, and provides high throughput for data reading and writing. HDFS architecture diagram is shown in 1.

![HDFS Architecture diagram](image)

**Figure 1.** HDFS architecture of distributed file system

2.2. Big data analysis and processing technology

Big data analysis and processing technology mainly uses specific analysis algorithms combined with appropriate distributed computing programming models and parallel execution engines to perform real-time analysis and mining and batch parallel calculation on massive data stored in the big data platform to meet different purposes and performance Analysis needs. On current parallel machines, the more popular parallel programming environments can be divided into three categories: shared storage, message passing, and data parallelism. The analysis logic diagram is shown in Figure 2.
As can be seen from logic diagram 2, the Agent component on each computing node in the distributed computing environment collects static resource information (such as CPU, memory, disk, network, and other performance parameters) and dynamic resource information on the corresponding node at any time. Such as CPU utilization, used memory, etc.). The agent can also perform necessary monitoring on the specified software module on the node where it is located, and transmit the running information corresponding to the module process in real time. When the system receives one or more analysis and processing tasks designed and defined by the user in accordance with the appropriate distributed computing model, the distributed task scheduler cluster can organize all the node information collected in real time, combined with the predefined resource organization Strategy to schedule and allocate resources according to application requirements, and achieve flexible, efficient and reliable task execution and result aggregation reports through parallel execution engines.

2.3. Data analysis algorithm

It is known that the measured data satisfies the Wei Gaussian distribution, then the squared Mahalanobis distance of the measured data approximately satisfies the $\chi$ distribution. Based on this conclusion, the minimum volume ellipse method compares the Mahalanobis distance and the square root value of $\chi$, namely:

$$
\varepsilon(\mathbf{x}_{n}, C_{n}, p) = \left\{ \mathbf{x}; d_{2}(x) \leq \sqrt{\chi_{p}^{2}} \right\}
$$

Thus, the boundary of the smallest volume ellipse is determined. With this boundary as the boundary, the measurement data falling outside the boundary is diagnosed as abnormal measurement data.

The convex shell peeling method is to construct a convex shell using the measurement data of the outermost layer. After deleting the measurement data that fell on the convex hull, continue to build the second layer of convex hull inwards, and so on, you can build a multilayer onion-like convex hull. The convex shell peeling method gives different weights to the measurement data falling on each layer of convex shells, namely the depth. The depth is positively related to the distance between the convex hull and the data center, and the measured data at the outermost side has the smallest depth. The convex shell peeling method determines the outermost data that should be removed, that is, abnormal data, by setting the depth or the amount of peeling data. The stripping quantization index can also be determined by least square method, least median square method or least trimming square method. The advantage of the convex shell peeling method is that there is no need to classify and set conditions to eliminate abnormal data. The disadvantage is that improper parameter setting may cause too many normal measurement data to be misdiagnosed as abnormal measurement data.
In the low dimension, the minimum volume ellipse method and convex shell peeling method can diagnose abnormal measurement data more efficiently, but when the measurement data dimension is high, they will show greater limitations. The principal component analysis method can better solve the problem of diagnosis efficiency under high-dimensional data.

The basic idea of the principal component analysis method is to map high-dimensional measurement data to low-dimensional subspaces according to the relevant attributes of the data, to achieve dimensionality reduction analysis of high-dimensional measurement data. Mathematically, the principal component analysis method uses an orthogonal linear transformation to transform the data into a new coordinate system so that the maximum variance of any projection of the data is on the first coordinate; the second largest variance is on the second Coordinates, and so on. As such, the equation for solving the largest principal component is:

$$w_1 = \arg_{\|w\|_2} \max E\{(w^T X)^2\}$$  
(2)

The Gaussian mixture distribution is a typical mixed probability distribution algorithm. The basic idea is to divide the measurement data into subsets, such as the mean method, and then use the data in each subset as independent Gaussian distribution measurement data with different parameters., Ie the overall measurement data is expressed as

$$p(t \mid x) = \sum_{i=1}^{k} \alpha_i(x) \phi_i(t \mid x)$$  
(3)

Among them, $\alpha$ is the weight of the Gaussian distribution; $\phi$ is the Gaussian probability function, namely:

$$\phi_i(t \mid x) = \frac{1}{(2\pi)^{d/2} \sigma_i^d(x)} \exp \left[-\frac{\|t - \mu_i\|^2}{2\sigma_i^2(x)}\right]$$  
(4)

The theory of extreme value distribution only focuses on the extreme points of the measurement data that are far from the probability distribution function, and determines whether the measurement data is abnormal by observing the probability distribution functions of these extreme points. For the selection of extreme points, you can simply take the limit of the normal distribution, or you can segment the data and take the maximum and minimum values in each segment. The probability distribution function satisfied by the extreme point is:

$$p = \exp \left[-\exp\left(\frac{x - \mu}{\sigma}\right)\right]$$  
(5)

Compared with determining the abnormal measurement data directly by the probability distribution function, the extreme value distribution theory has a wider application range, and in some special cases (such as rare abnormal data) has a higher detection resolution, Squeeze.

The goal of the support vector machine method is to establish a maximum-interval hyperplane, which is to measure the data group while maximizing the distance between the subspaces corresponding to the grouped subset. Generally speaking, the measurement data is distinguished in a finite-dimensional space, and the set used for distinguishing is usually linearly inseparable in this space. To this end, the support vector machine method maps the original finite-dimensional space to a much higher-dimensional space, and separates subsets in the high-dimensional space. Support vector functions can be expressed as:
$$SV = \text{sign} \left[ \sum_{i=1}^{K} \alpha_i L_i K(x_i, z) + b \right]$$

The support vector machine method judges whether the data is abnormal by the density of the support vector function in the region. Generally, high-density area data is normal, while sparse areas indicate measurement outliers. Compared with the traditional statistical data diagnosis method, the support vector machine method has higher diagnostic accuracy; the disadvantage is that the input data needs to be completely marked, and after classification, the model parameters are not intuitive [2].

3. Development of an online inspection and evaluation platform for the status of smart electric energy meters

The data sources of the online inspection and evaluation of the status of smart energy meters are mainly marketing business application systems, electricity information collection systems and metering production scheduling systems. Their physical infrastructure and front-end devices do not need to be repeatedly constructed. Demand analysis, design methods, detailed design, etc. have been described in more similar documents, and will not be described here. According to the idea of modularization, the system architecture of its platform is designed according to the data source layer, data management layer and business application layer, as shown in Figure 3.

![Platform system architecture diagram](image)
3.1. Data source layer
The data source layer is the calculation data source of the online platform system, which mainly includes the marketing business application system SG186, electricity information collection systems and the measurement production scheduling system.

3.2. Data management
The data management layer is mainly divided into two parts: data collection and data application; among them, data collection mainly realizes the smooth extraction of data required for calculation from the three basic platforms, including collection interface management, collection quality management, data storage management and data supplementary recording management. The data application mainly realizes the basic processing of the extracted original data and provides the data foundation for subsequent platform business applications, including online monitoring management, data mining analysis, evaluation model management and status index management.

3.3. Business application layer
The business application layer mainly includes two parts: business function and business decision-making; among them, the business function mainly includes trigger condition management, status inspection process, status score evaluation and status judgment management; business decision-making mainly includes status evaluation analysis, monitoring alarm statistics and inspection strategy report. And test closed-loop tracking [3].

According to the above platform system architecture, the script Java language is applied, and the online platform for smart energy meter status inspection and evaluation is developed according to the B/S development framework, and it is deployed on the server side; The business process of the online platform for smart energy meter status detection and evaluation has been designed in four ways, including family defects, online monitoring, on-site detection, and timing trigger, as shown in Figure 4.

![Figure 4. Flow chart of automated business of meter monitoring](image)

4. Analysis of key technologies of smart meter collection function

4.1. Development of network side anomaly analysis model
Power supply radius, device model, user's power consumption characteristics and other reasons will also lead to abnormal line loss, collectively referred to as grid-side abnormality. Considering the lack of user load data, the traditional theoretical line loss calculation method cannot be applied in actual
production. Therefore, based on limited data (such as line diameter, power supply radius, user power consumption characteristics, etc.), it is necessary to calculate the line loss reference value of the station area and identify the network-side cause of the abnormal line loss.

First of all, according to different data bases, appropriate selection of appropriate models to calculate the line loss reference value of the station area. In view of the fact that the line topology of the station area and the data of the equipment area are insufficient, but the load data can be obtained, the equivalent load calculation method is used to calculate the equivalent impedance data of the distribution network using the limited load data, and based on the equivalent data and load data Reference value of line loss in Fujian province area. If none of the load data can be obtained, the stable and reliable station area in the power supply area is selected as the sample data, and the model data is trained on the station area in the power supply area with limited power data and station feature data to obtain a line based on the limited feature. Reference value. Then analyse each user according to the calculated line loss reference value, calculate the degree of influence of the large amount of electricity users on the line loss, and identify the existing network side problems according to the typical characteristics of the network side. [4].

4.2. Analysis of intelligent diagnosis of anomalies Fujian province

Analyse the stations with abnormal line loss every day, and use the data collection and measurement function of the collection system to automatically analyse the users with abnormal data to obtain the abnormal metering caused by the communication reasons. For abnormal problems caused by non-communication reasons, further use big data abnormal analysis algorithm to identify and manage abnormal data. Its business process is shown in Figure 6.

![Figure 5. Intelligent diagnosis and analysis of anomalies in Fujian province](image)

4.3. Status evaluation of smart energy meters and closed-loop transformation of replacement operation and maintenance

The related design of the deepening application project of the main station software of the electricity information collection system (smart energy meter status evaluation and replacement of the collection and operation closed-loop transformation) mainly involves the development of the support function of the operation error of the collection and operation closed-loop management smart meter; closed-loop work order processing flow and measurement Abnormal implementation monitoring, measurement abnormal historical monitoring, on-site operation terminal APP transformation; abnormal data acquisition interface interaction with the acquisition system and operation and maintenance closed-
loop management module, marketing system power inspection interface transformation; system data
governance-line loss abnormality, acquisition communication Abnormal diagnosis, load abnormal
analysis, power abnormal analysis, photovoltaic abnormal analysis, grid-side abnormal analysis and
other work collection operation and maintenance closed-loop management smart meter operation error
support function development [5].

5. Experimental testing
According to the above-designed energy meter, the data collection experiment was carried out, and
three measured energy meters and one data acquisition device were prepared. The data collected
includes the table number, peak-hour power, normal power, valley power and total power. The data of
the three measured energy meters are shown in Table 1.

Table 1. Energy Meter Data Information

| The main parameters | Power meter under test 1 | Energy meter under test 2 | Energy meter under test 3 |
|---------------------|--------------------------|---------------------------|---------------------------|
| Table number        | 151023000688             | 151023000697              | 151023000653              |
| Peak hour electricity (kWh) | 0.51                    | 0.27                      | 1.33                      |
| Normal power consumption (kWh) | 1.08                    | 0.32                      | 2.47                      |
| Valley power (kWh)   | 0.00                     | 0.11                      | 0.32                      |
| Total power (kWh)    | 1.59                     | 0.70                      | 4.12                      |

Collect the data of the three measured energy meters under the power-on state and the power-off
state, and collect 100 times at different distances (0.2m, 0.5m, 1.0m, 1.5m, 2.0m), Record the number
of successful data collection, as shown in Table 2 and Table 3.

Table 2. Electric energy meter data collection experiment

| Collection distance (m) | Successful times of collecting energy meter data |
|-------------------------|-----------------------------------------------|
|                         | Power meter under test 1 | Energy meter under test 2 | Energy meter under test 3 |
| 0.2                     | 100                          | 100                          | 100                          |
| 0.5                     | 100                          | 100                          | 100                          |
| 1.0                     | 100                          | 100                          | 100                          |
| 1.5                     | 99                           | 100                          | 100                          |
| 2.0                     | 97                           | 100                          | 99                           |

Table 3. Data collection experiment of power meter in power-off state

| Collection distance (m) | Successful times of collecting energy meter data |
|-------------------------|-----------------------------------------------|
|                         | Power meter under test 1 | Energy meter under test 2 | Energy meter under test 3 |
| 0.2                     | 100                          | 100                          | 100                          |
| 0.5                     | 100                          | 100                          | 100                          |
| 1.0                     | 98                           | 100                          | 98                           |
| 1.5                     | 98                           | 99                           | 99                           |
| 2.0                     | 95                           | 95                           | 100                          |

It can be obtained from the above data that in the range of 2m, the data collection success rate of
the power meter in the power-on state is 99.7%, and the data collection success rate of the power
meter in the power-off state is 98.8%. Analyse the reason why the success rate in the power-off state is
lower than the power-on state. When the power is turned off, the energy sent by the RF antenna of the
energy meter all comes from the energy transmitted by the reader. The part of the transmitter also
includes the energy provided by the power meter's own power supply. The greater the energy, the longer the communication distance, and the better the effect. That is, at the same distance, the success rate of data collection is higher [6].

6. Conclusion
In view of the shortcomings of the current smart meter manual management mode, which is time-consuming and wasteful of resources, a system model of energy meter operation status analysis based on big data technology is proposed. The concept, key technology and data mining technology of big data, the platform and clustering algorithm based on big data mining are mainly introduced. Big data distributed storage and distributed analysis and processing have solved the problems of wide processor distribution, large data volume, and low practicality in calculation. Aiming at the status of data in power consumption information collection system, metering production scheduling platform, marketing business system and other systems, a hierarchical structure big data platform system is constructed to realize the analysis of the operation status of electric energy meters, and the massive data of multiple systems are mined. And comprehensive analysis, in the same system to achieve the docking application of energy meter data from different information sources, more timely, accurate, and truly carry out operational status analysis and evaluation. Through the data mining platform, the massive data of the energy meter information is converted into a smart meter operation status report, and the staff is guided to calibrate or rotate the meter, which solves the problem of wasting human and material resources in the current meter management method.

References
[1] María Rodríguez Fernández, Adolfo Cortés García, Ignacio González Alonso, & Casanova, E. Z. Using the big data generated by the smart home to improve energy efficiency management. Energy Efficiency, 9(1) (2015) 1-12.
[2] Chung, S. J. Sohn, S. H., Kim, C. H., & Seo, I. Y. A study on technology trend and future directions of smart meter. Journal of the Korean institute of illuminating & electrical installation engineers, 32(2) (2018) 98-103.
[3] Tosi, D. Cell phone big data to compute mobility scenarios for future smart cities. International Journal of Data ence & Analytics, 4(4) (2017)265-284.
[4] Guzik, P. Matiolanski, A., & Dziech, A. Real data performance evaluation of caiss watermarking scheme. Multimedia Tools & Applications, 74(12) (2015) 1-15.
[5] Guenduez, A. A. Mettler, T., & Schedler, K. Smart government — partizipation und empowerment der bürger im zeitalter von big data und personalisierter algorithmen. hmd praxis der wirtschaftsinformatik, 54(4) (2017) 477-487.
[6] Godwin-Jones, R. Scaling up and zooming in: big data and personalization in language learning. Language Learning & Technology, 21(1) (2017)4-15.