Power grid data fusion based on block chain technology

LZ Yin, XW Han, CJ Gao, J Zhang and Z Li
Anshan Power Supply Company of the Liaoning Electricity Power Company, Liaoning, 132012, China
*corresponding author’s email: cliuping@36haojie.com

Abstract. In order to improve the efficiency of power grid data processing, a new power grid data fusion method based on block chain model is proposed in this paper. The results show that the power grid data fusion method based on block chain technology has higher efficiency.

Keywords. Blockchain technology; Power grid data fusion; Big data processing

1.Introduction
With the continuous advancement of enterprise informatization, the value of data information is particularly important in the process of enterprise development [1]. At present, the data management research of China's power grid enterprises is in the exploratory stage. The smart grid, new power equipment and power grid information system are updated frequently, and massive data are waiting to be screened and analyzed, which has created unprecedented challenges for the traditional data analysis system. The data information of the new power grid is quite different from that before. On the one hand, the data information comes from a wide range of sources, such as equipment operation, automatic feedback of intelligent system, etc. On the other hand, the data structure of power grid enterprises is quite different, and the data contains power grid data information, text information, picture information, etc. [2-3]. Blockchain technology is just suitable for the analysis and processing of complex data. It is considered as a subversive innovation after big data and cloud computing, which is highly concerned by all industries. The development of blockchain technology has become an irreversible trend [4-5].

This paper proposes an analysis method of power grid data fusion based on block chain technology to improve the ability to deal with power grid data problems.

2.Demand analysis of smart grid data platform to introduce block chain technology
2.1 Characteristic analysis of smart grid data information
Smart grid system mainly includes: data collection and monitoring control system, electric power production management system, the wide area monitoring system, power distribution network management system, power grid dispatching management system, smart meter system, energy management system of geographic information system and so on more than a smart grid management system, the intelligent CNC system is used to deal with different types of grid data. As smart grid upgrades, the value of these data has been recognized by grid experts. However, the development time of these systems is different and they are relatively independent in the process of operation, which ultimately determines the multi-source nature of grid data. Besides the smart grid is made up of several parts, including power transmission lines, electronic equipment components, terminal sensor, data acquisition terminal of the smart grid will to power grid enterprises transmission, substation, distribution, sale of electricity and so on each link of real-time data acquisition, the electric net data contains information such as the sort is more, such as Numbers, text, images, sound, essential attribute of the smart grid data, in turn,
led to the heterogeneity of grid data information. By analyzing the multi-source heterogeneity of smart grid data, the differences can be divided into four types: system differences, attribute differences, pattern differences and semantic differences. These problems will lead to poor sharing of data within power grid enterprises, which will easily lead to the phenomenon of "information island" of power data, which is not conducive to optimal management of power grid data and mining of potential value.

2.2 The fusion analysis of block chain technology and power grid data management platform

Blockchain technology is an extension of "Internet +" thinking, and its technological advantages can be applied to all aspects of smart grid data platform. The grid data requires high security and reliability of its platform framework, and the management platform should process and analyze the data in the process of data collection and storage. Block chain technology is a decentralized data processing technology, with the advantages of distributed storage, information traceability, and not easy to tamper with. The application of this technology to the technical framework of smart grid data system can effectively solve the problems of multi-source heterogeneity of power grid data and information storage. To realize the seamless connection between blockchain technology and smart grid data platform, it is necessary to demonstrate the compatibility between the two.

2.3 The selection of block chain technology by power grid data management platform

Block chain technology can be divided into three types according to the access mechanism: public chain, proprietary chain and alliance chain. The public chain is open to the public network, and all personnel can access and process data through block nodes. Proprietary chain is a network chain structure controlled by individuals. It can only be regarded as a block chain in a broad sense without the main features of decentralization and distribution of the block chain. The alliance chain, on the other hand, can set login access rights, retains some functions of centralized control, and has powerful information processing functions and strong scalability, mainly aiming at the establishment of network framework for specific companies or organizations. The technicians of the power grid company can set the node permissions of the alliance chain, and those with special permissions can read, write and modify the data information on the node, and meanwhile ensure the security of the data. In general, the alliance chain has low operating cost, strong degree of trust, high security coefficient of data storage, and can customize system access rights, which is very suitable for the construction of data platform framework of power grid companies.

2.4 Collection and extraction of smart grid data management platform

There are multiple data sources for power grid data, and different data sources come from different power grid data acquisition systems. The poor coordination of these systems leads to diverse data formats and large differences in data types of power grid, and the phenomenon of repeated collection and multiple extraction of information occurs from time to time. These factors bring many difficulties to the data analysis and application work of power grid enterprises. In order to eliminate the phenomenon of "information island" of power grid data, the collected intelligent data should be preprocessed. When preprocessing power grid operation parameters and data information, corresponding summary tables should be generated and data statistics should be conducted according to the data demand type of block chain power grid data management platform.

2.5 Security analysis of power grid data platform under alliance chain technology

Under the block alliance chain technology, all nodes of the smart grid data platform can participate in the data processing process, and the network nodes jointly verify the legitimacy of the grid data transaction. The operation of the grid data platform will no longer rely on any trust agency or tripartite intermediary, and even if some nodes in the system are damaged, the security of the whole smart grid data platform will
not be affected. At the same time, the grid data platform relies on the distributed storage, P2P transmission, asymmetric encryption, intelligent contract and other features of the block alliance chain technology to ensure the untamperability and traceability of the grid information.

3. Power grid data fusion processing based on block chain technology

The goal of building smart power grid is coordination, efficiency, integration, energy conservation and environmental protection. Therefore, in order to cope with the development of smart power grid and build efficient and energy saving power grid information processing algorithm, this paper proposes the power grid data fusion analysis method based on block chain technology, so as to improve the fusion algorithm rate and reduce energy consumption. The accuracy of power grid data analysis and processing and prediction has been improved by the method of parallel computing. The parallel fusion process algorithm of the generation power prediction method is shown in Figure 1.

Figure 1  Parallel fusion processing algorithm

(1) Objective function

Discretize the data. In selecting discretization method considering the correlation between data, so the unsupervised discretization methods, unsupervised discretization method at the present stage mainly isometric, such as frequency, and the method based on clustering, according to the data processing in this paper have a temperature, humidity, power generation, history, power, temperature, etc., so the isometric discretization method is applied, is too simple and crude, in order to improve discretization of the K-means clustering algorithm through the eigenvalue distribution characteristics, and the characteristics of the proximity cluster, and then USES the division principle of top-down division of the clusters, The bottom-up merging principle is adopted to merge the clusters until they are divided into a specified number. Where the switching variables and enumerators are discrete variables, There is no need for discretization. Other data should be discretized before being used.

The monitoring data is matrixed, and the vector is set as the collection and monitoring matrix of a certain fan at a certain time:

$$E = (d_1, d_2, \cdots, d_m, t)$$  (3-1)

Where, $m$ represents the dimension, $d_i$ represents the value of the $i$ dimensional data, and $t$ represents the acquisition time.

Let the monitoring data matrix of a certain fan at a certain time be $E_q$.
Here the $M$ Represents the value returned by multiple devices in a certain period of time, in the form as follows:

$$
E_q = \begin{bmatrix}
E_1 \\
E_2 \\
\vdots \\
E_n
\end{bmatrix} = \begin{bmatrix}
E_{11}E_{12}\ldots E_{1n} t_1 \\
E_{21}E_{22}\ldots E_{2n} t_2 \\
\vdots \\
E_{n1}E_{n2}\ldots E_{nn} t_n
\end{bmatrix}
$$

(3-2)

4. Simulation

Through the simulation experiment of the calculation example, the pre-results of different algorithm models on power are obtained from the aspect of prediction accuracy, as shown in Figure 2.

![Figure 2](image_url)

**Figure 2** Power prediction comparison of different models

As can be seen from the trend of the curve in Fig 2, the input variables of PCA-BP and "MI-BP-MP" are few and the training time is short. The algorithm in this paper and the actual power is close, "MI - BP - MP method under the premise of considering the variable correlation, to select variables training, so as to achieve the large amount of data fusion, the purpose of removing redundant variables, and can keep the original data most characteristic information, achieve big data fusion, and power demand forecast.
Fig. 3 reflects the RMSE, MAPE and test of the two prediction models from a quantitative perspective. From the numerical point of view, the error values of the algorithm presented in this paper are all reduced; from the result curve, it can be seen that the algorithm presented in this paper has a good performance with a weak and reduced absolute error fluctuation. From the perspective of computational timeliness, the data in this experiment is divided into four groups. After data expansion, 1GB, 12GB, 24GB, and 120GB are respectively taken for experimental comparison between the traditional method and the algorithm of the model in this paper. The comparison results of time consumed by parallel operation is show in Table 1.

Table 1  Comparison results of time consumed by parallel operation

| Model algorithm      | The amount of data (GB) | Time (s)  |
|----------------------|-------------------------|-----------|
| BP network           | 1                       | 278.54    |
| BP network           | 12                      | 384.96    |
| BP network           | 24                      | 576.77    |
| BP network           | 120                     | 1477.40   |
| algorithm in paper   | 1                       | 253.21    |
| algorithm in paper   | 12                      | 374.45    |

Data sets of different sizes were set and run on clusters of nodes of different sizes. The experimental results are shown in Figure 4.
The experimental results show that when the data set is small, the difference in data processing efficiency is not significant, but with the increasing data volume, the platform shows an efficient processing rate. The processing time of the algorithm presented in this paper increases linearly, and the cluster acceleration ratio is nearly 3 times compared with traditional processing. This shows that the power grid data fusion analysis method based on block chain technology can meet the performance requirements of smart power grid big data fusion processing.

5. Conclusions

In this article, through analyzing the characteristic of the smart grid data information and chain block, the block chain technology embedded in the grid data management framework, based on grid data fusion processing, not only to integrate multi-source data centrally, found in large amounts of complex data with grid, operating and managing the production control information knowledge, in one of the decisive power grid data fusion method based on block chain model, and the BP network prediction algorithm, PCA - BP prediction algorithm with in this paper, based on the technology of block chain grid data fusion method, prove the method's ability to deal with large data problems.

6. References

[1] Wu H and Wang ZH 2018 Multi-source fusion-based security detection method for heterogeneous networks[J] Computers & Security 74 pp 55-70
[2] Andoni M, Robu V and Flynn D 2017 Blockchains: Crypto-control your own energy supply[J] Nature 548(7666) p 158
[3] Sutanto D 1999 Power management solutions for energy management, power quality and environment using battery energy storage systems[C] IEEE 1999 International Conference on Power Electronics and Drive Systems Hong Kong 1 pp 143-157
[4] Jeroen T, Frederik G and Daah S 2013 Multiobjective battery storage to improve PV integration in residential distribution grids[J] IEEE Transactions on Sustainable Energy 4(1) pp 182-192
[5] Li GY, Zhou M and Liu B 2006 An Alternative Approach for Coordinating Dispatching Day-Ahead and Real-time Energy Markets[J] Power Systems Conference and Exposition 52 pp 1162-1167
[6] Mahamadou AT, Mamadou BC and Brayima D 2013 Use of Ultracapacitors and Batteries for Efficient Energy Management in Wind-Diesel Hybrid System[J] IEEE Transactions on Sustainable Energy 4(2) pp 414-424
[7] Ahmed SA, Awad T and Magdy MA 2015 Optimal ESS Allocation for Load Management Application[J] IEEE Transactions on Power systems 30(1) pp 327-336