Cloud-Edge Collaboration Based Data Mining for Power Distribution Networks

Li An and Xin Su(✉)

College of IOT Engineering, Hohai University, Changzhou 213022, Jiangsu, China
anjiali0302@163.com, leosu8622@163.com

Abstract. The automation rapid development of the power distribution network have not been fully utilized with the terminal coverage rate increment. The demand and complexity of the power distribution network applications are also fast updated leading a huge calculation pressure from cloud service. This paper does data mining from power distribution network in three aspects, including delay, complexity and power. It defines them with respective weights according to the application requirements, and propose a cloud-edge collaborative communication scheme to effectively reduce the computing complexity of the system.

Keywords: Power distribution network · Edge computing · Cloud-Edge collaboration · Data mining · Computing complexity

1 Introduction

1.1 A Subsection Sample

According to the statistics of the National Energy Administration, China's electricity consumption has reached 7.2255 trillion kilowatts in 2019, an increase of 4.5% over the same period last year [1]. The construction of the power distribution network, which sends electricity to thousands of households in the last kilometer, is worthy of attention and research. At the beginning of the 20th century, the data collection of power distribution network terminal is quite simple, which can only provide data support for basic tasks such as power grid dispatching, maintenance and planning. With a large number of monitoring devices [2, 3], electric vehicle charging piles [4, 5], new energy [6] and other widely connected to the power distribution network, there are higher requirements for data acquisition and analysis of power distribution network terminals.

In [7], it studies the impact factors of cloud virtual machine startup time for users plan, making resource manage decisions in power grid. The power grid uses cloud computing technology, which makes the grid lack the ability to process data for real-time application. As a result, some scholars have introduced edge computing into the power grid. In [8], it introduces the edge computing to power grid, and takes smart measurement as an example to analyze the power grid edge computation in terms of efficiency and security. In addition, by the end of 2018, Chinese State Grid Co., Ltd. has been connected to 540

H. Gao et al. (Eds.): ChinaCom 2021, LNICST 433, pp. 438–451, 2022.
https://doi.org/10.1007/978-3-030-99200-2_33

m terminals of various types, and the daily data collection has exceeded that of 60TB [9]. Two aspects, including the limited storage and real-time processing capacity of the cloud center and the fast update of terminal data, complex types, poor quality and urgent need to tap its value, have been restrict to the construction process of power distribution network. In [10], a cloud-edge collaboration was proposed to exploit the advantages of both edge computing and cloud computing. The cloud-edge collaborative aims to address the problem of excessive inference delay caused by heavy computing tasks in the power distribution networks cloud server. Moreover, a multi-node collaborative computing management model to reduce the processing delay of cloud service providers was also suggested in the same work. A cloud-edge collaborative architecture for power distribution networks was proposed in [11] to solve the problems of network resource redundancy and overload caused by the widespread use of chimney-type independent service access.

Based on above, this paper proposes an effective mining method about the data value of power distribution network, which is a method of data cleaning, classification and fusion based on delay, complexity and power. The proposed method can monitor the abnormal data from the power distribution network terminal, reduce data redundancy and integrate heterogeneous data. It also processes data not only improve the accuracy and effectiveness, but also optimize the data quality. The threshold value of the processed data attribute value is compared and discriminated, and the transmission scheme of the user data of the power distribution network terminal is determined, so that the delay and power in transmission are reduced.

2 Power Distribution Network Terminal Data Analysis

2.1 Terminal Data Type

Load data and environmental monitoring data are the main parts of terminal data of power distribution network, among which static data and dynamic data constitute load data. In the power grid, the static data mainly refers to the parameters of equipment attributes and the topology data of the static network. The dynamic data mainly comes from the voltage monitoring system and the power system monitoring network. The dynamic data of the power grid can be monitored from all angles by the terminal equipment, so as to obtain the formation of real-time status data. The environment monitoring data includes both the operation of the internal device of the distribution plate, such as temperature, humidity, water level, noise, vibration, etc., it also includes external operating environment of the power distribution network equipment, which monitors access control, anti-theft, equipment, small animal detection, etc. The effective monitoring and control of environmental monitoring data and environmental equipment can prevent accidents, prevent the loss of control of the operating environment, extend the service life of equipment, and reduce the high cost caused by the extensive management of power distribution.

2.2 Data Preprocessing Based on DCP

As the development of smart grids has led to the rapid growth of multi-source heterogeneous data, people have begun to mine all kinds of information contained in terminal data. For example, people start using clustering technology to characterize different types of users, and help power grid companies to achieve business activities, like precision marketing.

However, looking at the terminal data application development process, whether it is services such as scheduling planning, resource allocation, status alarms to maintain the safe and economic operation of the power grid, or services such as user portraits and precision marketing for customers, or achieving stable operation/energy saving and emission reduction, etc. multi-objective energy efficiency management, power consumption optimization and other integrated services can mine the value of data from the three aspects of delay, complexity and power. In addition, these three categories do not exist in isolation, as shown in Fig. 1.

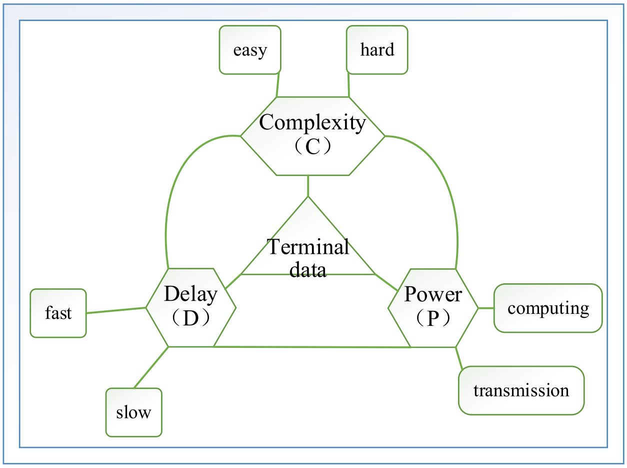

Fig. 1. Internal interaction of terminal data.

From the perspective of demand in delay, terminal data can be divided into 'fast' and 'slow'. When the terminal data belongs to the 'fast' attribute, it will be processed in time at the edge; otherwise, it will be processed in the cloud far from the terminal. The complexity of data also contains two attributes, namely 'hard' and 'easy'. When the terminal data is judged to be 'easy', it will be transmitted directly to the edge to reduce the computing load of the cloud; otherwise, it will be transmitted to the cloud. When analyzing the power attributes of terminal data, key factors such as power in computing and transmission need to be considered. When the power is high, it is dealt with in the edge to achieve low cost, environmentally friendly, and promotion of the good development of the power grid; otherwise, it is processed in the cloud. The three dimensions of terminal data analysis are complementary to each other and are also the foundation for the generation of higher level business.

Based on the DCP power distribution network terminal data analysis, in theory, when the demand in delay of the data is "slow", the complexity demand is "hard", and the demand in power is low, the task data is suggested to transmit to the cloud for processing. When the demand in delay of the data is "fast", the complexity demand is "easy", and the

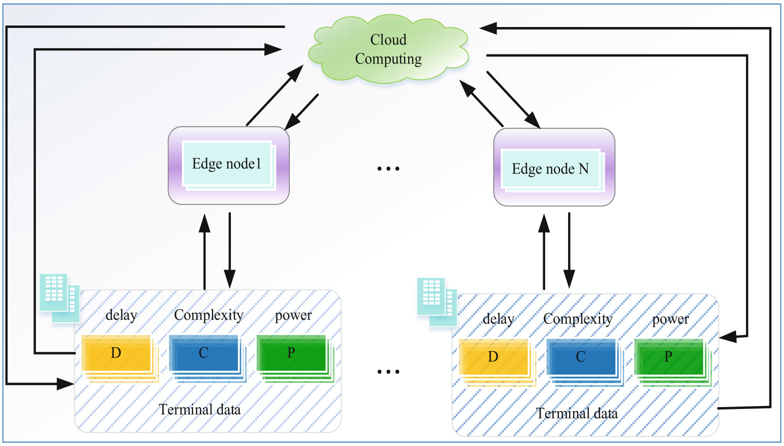

Fig. 2. Schematic diagram of cloud-edge-end data interaction.

demand in power is high, the task data is suggested to transmit to the edge for processing, as shown in Fig. 2. However, this is not in line with the actual situation analysis, from the point of view of the terminal, it may produce data with "slow" delay requirements, "easy" complexity requirements and high power requirements, so the above single division is no longer applicable. Therefore, in order to solve this problem, this article combines the actual needs, and the DCP sets different weights according to business requirements, and sum them in Eq. 1. The weighting function is as follows:

$$Q = \alpha \times t_i + \beta \times e_i + \lambda \times c_i \tag{1}$$

Here, t_i, e_i and c_i correspond to the delay attribute, power attribute and complexity attribute of data uploaded by terminal equipment respectively. α, β and λ respectively correspond to the weight distribution values of delay attribute, power attribute and complexity attribute of data uploaded by terminal equipment, α, β and λ satisfy $\alpha, \beta, \lambda \in [0, 1]$, $\alpha + \beta + \lambda = 1$. Moreover, α, β and λ can be self-adjusting according to the own conditions of the terminal equipment. For example, when the user has a higher demand for the delay, it will pay more attention to the cost of delay, thereby setting a relatively higher α value. When the terminal equipment is at a low power, the power saving as much as possible will be considered when the transmission plan is determined, so the user will select a relatively higher β value. In addition, when the complexity of the task data uploaded by the terminal equipment is "hard", it will pay attention to the calculation resource costs caused by task data processing, thereby setting relatively high λ values.

2.3 Data Preprocessing Process

Due to a large number of terminal task data, it is easy to enlarge data incompleteness and inaccuracy, which affects subsequent data mining and analysis. According to the characteristics of power grid data, data errors include three classes of repeated data, abnormal data, and missing data.

Repeated data has two types of repeated data attributes and repeated data instances. Due to the fluctuation of the grid power status, the problem of repeated data attributes is mainly derived from business definition or code repetition, and the repeated data instance is mainly derived from repeated records generated by the detection system. Abnormal data is divided into irregular data and invalid data. Irregular data mainly refers to instance data that violates rules due to data transmission or manual error operations. Invalid data mainly refers to the reserved fields of a specific data table in the power grid system that are mostly empty or default data. The main reason for missing data is that the power grid data input system fluctuates, data transmission is lost or manually missed, and the number of lost instances is relatively small from the perspective of the overall data. These repeated data, abnormal data, and missing data need to be preprocessed by DCP to avoid the waste of storage and computing resources in the power distribution network.

In this paper, we have mined the task data of the terminal from three aspects of delay, complexity and power consumption, so as to improve the value density of task data and reduce the resource waste of the power distribution network. The data mining process is shown in **Algorithm 1**. The core contains the following four situations:

Case1: The delay, complexity and power consumption of the task data are mined, and the weighted sum is carried out. If the value of the weighting function is greater than the upper limit of the given threshold, it indicates that the storage and computing resources used to process the task data are large after multi-objective measurement. The system of power distribution network then transmits the task data of the terminal to the cloud server for processing.

Algorithm 1: DCP

Input: data of terminal task set $D=\{task_1, task_2, ..., task_n\}$

Output: the transmission plan of terminal task data set D

1. **for** special data $task_i$ in D **do**
2. the task data is directly transmitted to the cloud server
3. **end**
4. carry on task data mining, analyze delay, complexity, power consumption.
5. calculate the weighted model $Q=\alpha\times t_i+\beta\times e_i+\lambda\times c_i$
5. $th_{\min}$ = the lower limit of the given threshold
6. $th_{\max}$ = the upper limit of the given threshold
7. **if** Q < $th_{\min}$
8. the task data is not processed and discarded.
9. **else if** Q > $th_{\max}$
10. the task data is transmitted to the cloud server.
11. **else**
12. the task data is transmitted to the edge server.
13. output the transmission plan of terminal task data set D

Case2: The delay, complexity and power consumption of the task data are mined, and the weighted sum is carried out. If the value of the weighted function is less than the lower limit of the given threshold, it indicates that the contribution value of the task data

to the power distribution network approaches zero or none after the task data is analyzed and processed after multi-objective measurement. Therefore, the system of power distribution network will consider whether it is a data error. Moreover, it is also necessary to avoid waste of storage and computing resources in power distribution network, the system will not process and abandon this task data.
Case3: The delay, complexity and power consumption of the task data are mined, and the weighted sum is carried out. If the value of the weighted function is between the upper limit and lower limit of the given threshold, it indicates that the storage and computing resources used to process the task data are small after multi-objective measurement. The system of power distribution network then transmits the task data of the terminal to the edge server for processing.
Case4: After the task data is input into the system of power distribution network, if the task data is special, such as a large venue event should have direct communication channels with the cloud server, and the venue uploads the future activities arrangement to the cloud server. The cloud server will arrange the power supply plan after accurately predicting its daily load, and guarantee the venue is going smoothly.

3 System Model

In the power distribution network cloud-edge collaborative system model, it is assumed that there are M servers in this system, H is used to represent the set of servers, and $H_b \in H, b \in \{1, 2, ..., M\}$ represents the b server in the system model. At the same time, it is assumed that there are N tasks of the power distribution network terminal, T is used to represent the tasks set of the power distribution network terminal, and $task_i \in T, i \in \{1, 2, ..., N\}$ represents the $task_i$ uploaded by the power distribution network terminal.

Define the variable $y_{i,b}$ to indicate whether $task_i$ will be uploaded to server H_b, that is,

$$y_{i,b} = \begin{cases} 1, \ task_i \quad upload\ to\ server \quad H_b \\ 0, \ other \end{cases} \tag{2}$$

Taking into account the continuity of tasks, we assume that a task can only be handled by one server if it is to be unloaded, that is,

$$\sum_b y_{i,b} \in \{0, 1\}, \forall i \in T, \forall b \in B_e \cup B_c \tag{3}$$

Where, B_e represents the collection of servers at the edge, and B_c represents the collection of servers in the cloud, and $H = B_e \cup B_c$.

3.1 Data Complexity

According to the chaotic theory, the complexity of the pseudo-random sequence refers to the degree of similarity to the random sequence, and is a measure of the extent of the overall extent to which the sequence is restored [12]. In this paper, the calculation of task data complexity is expressed by fuzzy entropy, that is,

$$FuzzyEn(m, r, N) = \ln \phi^m(r) - \ln \phi^{m+1}(r) \tag{4}$$

Where,

$$\phi^m(r) = (N - m - 1)\sum_{i=1}^{N-m-1}\left[(N - m)^{-1}\sum_{i=1,j\neq 1}^{N-m-1} A_{ij}^m\right], \forall i, j \in [0, N]$$

$$A_{ij}^m = \exp[-In(2) * (\frac{d_{ij}^m}{r})^2, \forall i, j \in [0, N]$$

Where, m is the dimension of the reconstructed phase space, r is the similarity tolerance limit and N is the sequence length. d_{ij}^m is the distance between the window vectors after the phase space is reconstructed from the original sequence, and the phase space dimension m is generally based on the correlation dimension (fractal dimension) [13]. The calculated fuzzy entropy is mapped between [0, 1]. If the fuzzy entropy of task data is located [0, 0.5], the task data of complexity is considered to be "easy". If the fuzzy entropy of the task data is located (0.5, 1], the task data is similar to the random data, and its complexity is "hard".

3.2 Delay and Power of Data

In terms of terminal task data analysis, we mine the data from three aspects: delay, complexity and power consumption, which can improve the value density of task data and reduce the overhead cost of the system. On this basis, a cloud-edge collaborative task scheduling algorithm for power distribution network based on DCP is proposed in this paper, which includes the following steps:

Step 1: calculating the delay, complexity and power consumption of task data through the DCP algorithm.
Step 2: setting the weighted coefficient of delay, complexity and power consumption based on business requirements and the conditions of the business equipment.
Step 3: making a decision on the weighted DCP value to determine the unloading party of the task data.

Local Calculation Cost. Suppose the power consumption of the local terminal is modeled as $pl = k \cdot (f_i^l)^3$, where f_i^l and k are the computing power of the terminal CPU and the coefficients of the related processor chip structure, respectively [14]. When the local terminal processes the task data, the execution time and energy consumption are respectively.

$$t_i^l = y_{i,b} \cdot \frac{c_i D_i}{f_i^l}, \forall i \in T, \forall b \in B_e \cup B_c \tag{5}$$

And

$$e_i^l = y_{i,b} \cdot c_i \cdot D_i \cdot k \cdot (f_i^l)^2, \forall i \in T, \forall b \in B_e \cup B_c \tag{6}$$

Where, D_i- the amount of data of terminal $task_i$; c_i-complexity of terminal $task_i$;t_i^l-local terminal computing delay; e_i^l-the energy consumption calculated by the local terminal.

Unloading Cost. In this section, the unloading cost mainly includes time cost and energy consumption. When analyzing the time cost and energy consumption, we focus on the delay and energy consumption of the uplink transmission of tasks to edge servers and cloud servers, as well as the delay and energy consumption of edge servers and cloud servers in processing tasks. This is because there is a big gap between the amount of data of the calculation result and the amount of data of the calculation task. For example, when the business needs to monitor whether the equipment is in good condition, a large amount of electrical data needs to be processed, while the calculation result is only whether or not, the amount of data is very small, so it can be ignored, so the computing delay of sending the calculation results back to the terminal equipment from the edge or cloud is not considered.

When $task_i$ is uplink transferred to an edge server or cloud server, the transfer rates are assumed to be,

$$R_i^e = y_{i,b} \cdot W_i^e \log_2(1 + \frac{p_i^e g_i^e}{\sigma_{ie}^2 + \sum_{j \in N \setminus \{i\}} p_j^e g_j^e}), \forall i \in T, \forall b \in B_e \tag{7}$$

and

$$R_i^c = y_{i,b} \cdot W_i^c \log_2(1 + \frac{p_i^c g_i^c}{\sigma_{ic}^2 + \sum_{j \in N \setminus \{i\}} p_j^c g_j^c}), \forall i \in T, \forall b \in B_c \tag{8}$$

Here, $W_i^e, W_i^c, \sigma_{ie}^2, \sigma_{ic}^2, g_i^e, g_i^c, p_i^e$ and p_i^c denote the bandwidth of the $task_i$ transferred to the edge server and the cloud server, the white Gaussian noise in the channel, the gain on the channel, and the transmission power of the terminal equipment, while $p_j^e \cdot g_j^e$ and $p_j^c \cdot g_j^c$ represent the interference to the terminal $task_i$ when other terminal tasks are unloaded to the edge server and the cloud server, respectively.

Suppose the computing resources allocated to the $task_i$ uploaded to the edge server and the cloud server are represented by f_i^e and f_i^c respectively, that is, the number of cycles allocated to the CPU on the edge server and cloud server. Here, for the convenience of analysis, we assume that computing resources are evenly distributed. Therefore, when performing a computing task, if the uninstall decision is to upload to the edge server, the time cost and energy consumption of unloading the $task_i$ are these,

$$t_i^e = y_{i,b} \cdot \left(\frac{c_i D_i}{R_i^e} + \frac{c_i D_i}{f_i^e} \right), \forall i \in T, \forall b \in B_e \tag{9}$$

and

$$e_i^e = y_{i,b} \cdot \left(\frac{c_i D_i}{R_i^e} * p_i^e + \frac{c_i D_i}{f_i^e} * pl_{ie} \right), \forall i \in T, \forall b \in B_e \tag{10}$$

Where, pl_{ie} represents the calculated power of the edge server. If the $task_i$ is to upload to the cloud server, the time cost and energy consumption of the $task_i$ are as follows:

$$t_i^c = y_{i,b} \cdot \left(\frac{c_i D_i}{R_i^c} + \frac{c_i D_i}{f_i^c} \right), \forall i \in T, \forall b \in B_c \tag{11}$$

and

$$e_i^c = y_{i,b} \cdot \left(\frac{c_i D_i^c}{R_i^c} * p_i^c + \frac{c_i D_i^c}{f_i^c} * pl_{ic} \right), \forall i \in T, \forall b \in B_c \tag{12}$$

Where,pl_{ic} represents the calculated power of the cloud server.

System Overhead. In this paper, the aggregation function method is adopted in the cloud-edge collaborative task scheduling algorithm of power distribution network based on DCP, which combines delay, energy consumption and complexity into a cost function, that is,

$$F(y) = \alpha \cdot \frac{L(y) - L_{\min}}{L_{\max} - L_{\min}} + \beta \cdot \frac{E(y) - E_{\min}}{E_{\max} - E_{\min}} + \lambda \cdot \frac{C(y) - C_{\min}}{C_{\max} - C_{\min}} \tag{13}$$

Where,α, β and λ are the weights of delays, energy consumption and complexity functions, $\alpha, \beta, \lambda \in [0, 1], \alpha + \beta + \lambda = 1$, $L_{\max}, L_{\min}, E_{\max}, E_{\min}, C_{\max}, C_{\min}$ represents the maximum value of average delay, minimum value of average delay, maximum value of average power consumption, minimum value of average power consumption, maximum value of average complexity and minimum value of average complexity, respectively. By standardizing the three objective functions respectively, the aggregation function reduces the numerical difference caused by different target types. α, β and λ values can quantify the emphasis of the system on the optimization objective. It is a single-objective optimization of the delay when $\alpha = 1$. Similarly, when $\beta = 1$, it is a single objective optimization of power consumption, when $\lambda = 1$, it is a single objective optimization of complexity.

Before exploring the system overhead, we need to analyze and calculate the parameters $L_{\max}, L_{\min}, E_{\max}, E_{\min}, C_{\max}$ and $C_{\min}$.

$L_{\max}$: maximum value of average delay. Delay reaches the maximum value when all tasks are offloaded to the cloud.
$L_{\min}$: minimum value of average delay. Assuming that the edge is rich in computing resources, it will not cause queuing. So delay reaches the minimum value when all tasks are offloaded to the edge.
$E_{\max}$: maximum value of average power consumption. Power consumption reaches the maximum value when all tasks are offloaded to the cloud.
$E_{\min}$: minimum value of average power consumption. All tasks are assigned to the nodes with the best computing capability in the edge, and the power consumption reaches the minimum.
$C_{\max}$: maximum value of average complexity. Scheduling the task with missing data to the server will increase the complexity cost dramatically. Here, it is assumed that the data of all tasks are missing, which is the extreme value.
$C_{\min}$: minimum value of average complexity. Because it is a non-deterministic polynomial problem, it is unrealistic to get the absolute minimum. In this problem, the result of single objective optimization of complexity is taken instead of the exact value.

4 Simulation Result

In this paper, the DCP algorithm is used to determine the optimal strategy of service uploading at the terminal. Without special declaration, the parameter settings are summarized in Table 1.

Table 1. Simulation parameter information.

Parameter	Meaning	Value
W	Channel bandwidth	10 MHz
n_0	Noise unilateral power spectral density	− 174 dBm/Hz
L	Terminal data	50 kbit
N	Time series length	3000
m	Reconstruct the dimension of phase space	3
r	Similarity tolerance limit	3.5
${f^l}_k$	terminal computing capacity	5*10^8 cycle/s
${f^e}_k$	edge computing capacity	3*10^9 cycle/s
${f^c}_k$	could computing capacity	10*10^9 cycle/s
C	CPU per bit	100 cycles/bit
D	Task data quantity	[1000,2000] kB

4.1 Verification of Analysis

There are many businesses in the power distribution network, such as water level alarm, power distribution equipment intelligent diagnosis, large customer load management, etc. In order to facilitate simulation, this paper selects the electrical fault service to simulate the cloud-edge collaboration based on DCP, the traditional cloud computing architecture and the cloud-side cooperation algorithm based on decentralized “cloud”. First of all, 1000 simulations are carried out on the complexity of the electrical fault service, in which Fig. 3 is only part of our iterative experiment on the complexity of the electrical fault service.

In the power distribution Internet of things, the greater the complexity of the services data, the greater its randomness, and the more difficult it is for the services data to be restored. As can be seen from Fig. 3, the complexity of electrical fault service is within a stable range, and its maximum value is 0.4709, the minimum is 0.4533. After many iterative numbers, the average value is 0.4654, belongs to between 0–0.5, therefore electrical fault data complexity is considered to be “easy”. If delay and power consumption is not considered, it is worth suggested, and the electrical fault is preferably transferred to the edge.

In Fig. 4 and Fig. 5, “TCC” represents traditional cloud computing architecture, that is, its data is directly uploaded to the cloud; “DE” represents the cloud-side collaboration based on decentralized “cloud”, that is, data is uploaded to the cloud and edge

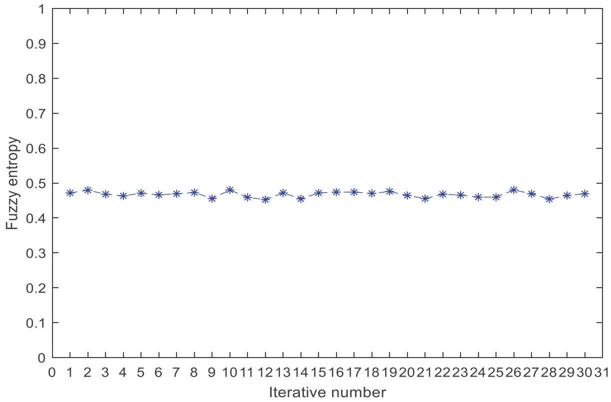

Fig. 3. Estimation of electrical faults complexity based on fuzzy entropy

side through random allocation; "DCP" represents cloud-edge collaboration based on DCP, which means that terminal data is mined from three perspectives of delay, complexity and power, and then uploaded to the cloud or edge servers as required. It can be seen from Fig. 4 and Fig. 5 that in the case of certain edge service providers and cloud service providers, the total time consumption and power consumption of cloud edge collaborative architecture based on DCP algorithm are always lower than those of traditional cloud computing architecture and cloud-side cooperation based on decentralized "cloud", because the algorithm based on DCP preprocesses data at the terminal and eliminates redundant data, and increasing the value density of terminal data. In addition, scientific and reasonable scheduling of the resources of the cloud and the edge has been carried out to optimize the time and power consumption of business processing, which further verifies the availability and efficiency of the DCP algorithm for mining terminal data.

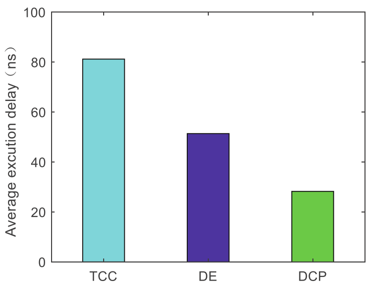

Fig. 4. Time delay observation of electrical faults.

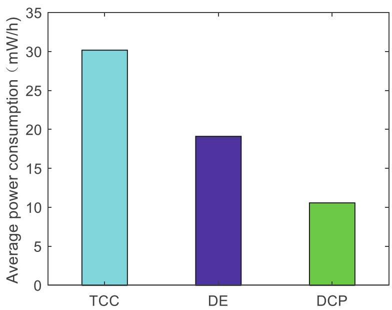

Fig. 5. Power consumption observation of electrical faults.

4.2 Influence Factors

In this paper, 560 terminals in the power distribution network are randomly selected for simulating. In addition, based on the three algorithms including the cloud-edge collaboration based on DCP, the traditional cloud computing architecture and the cloud-side cooperation algorithm based on decentralized "cloud", the task data of 560 terminals is analyzed from the delay and power consumption. The simulation results Fig. 6 and Fig. 7 are as follows.

In Fig. 6 and Fig. 7, the advantages of the algorithm in the power distribution grid are horizontally compared from the perspective of time delay and power consumption. The red node line represents the situation of the traditional cloud computing architecture, and the data is directly uploaded to cloud, that is, "TCC"; the blue node line represents the situation of cloud-side collaboration based on decentralized "cloud", and the data is uploaded to the cloud and the edge through random allocation, that is, "DE"; the green node line represents the cloud-edge collaboration based on DCP, and the terminal data is mined from the perspectives of delay, complexity and power, and then uploaded to the cloud or edge server as required, that is, "DCP".

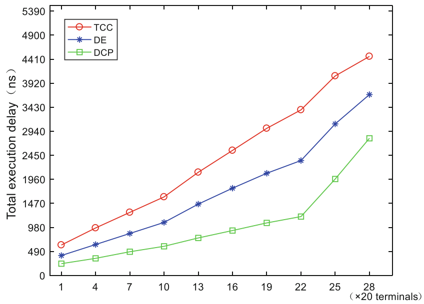

Fig. 6. Delay observation of 560 terminals.

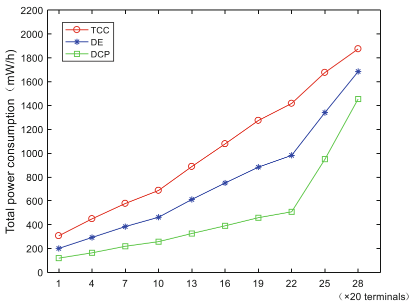

Fig. 7. Power consumption observation of 560 terminals.

It can be seen from Fig. 6 and Fig. 7 that when the number of edge nodes and cloud center is constant, with the increase of terminal tasks, the total delay and power consumption of processing tasks are increasing, whether it is traditional cloud computing architecture, the cloud-side cooperation algorithm based on decentralized "cloud" or cloud-edge collaboration based on DCP. Under the same conditions, the total delay and power consumption of cloud-edge collaborative based on DCP algorithm is always lower than that of traditional cloud computing architecture and cloud-side cooperation algorithm based on decentralized "cloud". which verifies the availability and efficiency of cloud-edge collaborative based on DCP architecture. However, it is worth noting that when the processing terminal task exceeds about 440, the execution delay and power consumption of the processing task of DE and the proposed scheme are sharply increased. This is because the calculation and storage resources of the edge service are limited, not enough to support multiple task processing, which may result in queuing phenomena. The computing and storage resources of the cloud are larger than the edge, so TCC is the same as usual.

5 Conclusion

This paper proposes an effective method for mining the value of the terminal data of the power distribution network. This method uses the function model of the data attributes established on the terminal to mine the value of the terminal user data in terms of complexity, time delay and power. According to the output function value, the data transmission direction is judged to determine the transmission plan of the power distribution network terminal user. This method not only reduces the time and power of server processing tasks, but also reduces the transmission and storage pressure of the cloud servers, making the obtained power distribution network terminal business scheduling plan scientific and effective, and can be applied in the actual engineering field.

Acknowledgment. This research work is supported by the National Key Research and Development Program of China (2021YFE0105500); and the National Natural Science Foundation of China (61801166).

References

1. Zheng, L., et al.: Research on power demand forecasting based on the relationship Between economic development and power demand. In: 2018 China International Conference on Electricity Distribution (CICED), pp. 2710–2713, IEEE (2018)
2. Chen, W., Guo, M., Jin, Q., Yao, Z.: Reliability analysis method of power quality monitoring device based on non-parametric estimation. In: 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), pp. 1526–1529. IEEE (2019)
3. Zhou, Z., et al.: Validity evaluation method of DGA monitoring sensor in power transformer based on chaos theory. In: 2018 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP), pp. 402–405. IEEE (2018)
4. Choi, W., Lee, W., Sarlioglu, B.: Reactive power control of grid-connected inverter in vehicle-to-grid application for voltage regulation. In: 2016 IEEE Transportation Electrification Conference and Expo (ITEC), pp. 1–7. IEEE (2016)
5. Cai, A., Yu, Y., Xu, L., Niu, Y., Yan, J.: Review on reactive power compensation of electric vehicle charging piles. In: 2019 22nd International Conference on Electrical Machines and Systems (ICEMS), pp. 1–4. IEEE (2019)
6. Yuan, W., Wang, C., Lei, X., Li, Q., Shi, Z., Yu, Y.: Multi-area scheduling model and strategy for power systems with large-scale new energy and energy storage. In: 2018 Chinese Automation Congress (CAC), pp. 2419–2424. IEEE (2018)
7. Mao, M., Humphrey, M.: A performance study on the VM startup time in the cloud. In: 2012 IEEE Fifth International Conference on Cloud Computing, pp. 423–430. IEEE (2012)
8. Shi, W., Sun, H., Cao, J.: Edge computing-an emerging computing model for the internet of everything era. J. Comput. Res. Develop. **54**(5), 129–133 (2017)
9. Xue, M., Shi, K., Chen, X., Wu, Q., Li, B., Qi, B.: A summary of research on converged communication model of power network and information network. In: 2017 2nd International Conference on Power and Renewable Energy (ICPRE), pp. 1062–1066. IEEE (2017)
10. Zhu, W., Qiang, F., Liu, N., Wu, X., Li, K., Zhang, K.: Research on multi-node collaborative computing management model for distribution Internet of Things. In: International Conference on Artificial Intelligence and Computer Applications, pp. 1019–1022. IEEE (2020)

11. Fan, C., Lu, Y., Leng, X., Luan, W., Gu, J., Yang, W.: Data classification processing method for the power IoT based on cloud-side collaborative architecture. In: IEEE 9th Joint International Information Technology and Artificial Intelligence Conference, pp. 684–687. IEEE (2020)
12. Chen, X.J., Li, Z., Bai, B.M., Cai, J.P.: A certain chaotic pseudo-random sequence complexity entropy measure of fuzzy relations. **60**(06), 379–388 (2011)
13. Wen, B.H., Yuan, M., Hou, L.: The study of CSI 300 index's complexity and comparison of model efficiency based on entropy algorithm. J. Quant. Econ. **32**(1), 19–25 (2015)
14. Zhang, W., Wen, Y., Guan, K., Kilper, D., Luo, H., Wu, D.O.: Energy-optimal mobile cloud computing under stochastic wireless channel. IEEE Trans. Wireless Commun. 12(9), 4569–4581(2013)