Direct Power Supply Identification Method of PV Power Based on Affinity Propagation Clustering

Hanjun Deng[1,2], Xing He[1,2(✉)], Rui Huang[1,2], Yuping Su[1,2], Suihan Zhang[1,2], and Wenwei Zeng[1,2]

[1] State Grid Hunan Electric Power Company Limited, Changsha 410000, Hu'nan, China
2065057002@qq.com

[2] Hunan Province Key Laboratory of Intelligent Electrical Measurement and Application Technology, Changsha 410000, Hunan, China

Abstract. At present, the marketization of distributed photovoltaic power generation faces problems such as the trading model is still immature and the trading mechanism is complicated, which makes the management of distributed energy trading very difficult. In this paper, a direct power supply identification method based on affinity propagation (AP) clustering is proposed. First, the historical PV output data of the station or neighboring stations are used to obtain the output data of the "same type" PV arrays for reference. Then, the AP clustering algorithm is used to cluster the power generation data in the same temperature segment, and the PV power sources suspected of direct power supply are identified according to the corresponding relationship between the clustering results and the electricity sales state. Finally, the proposed method is verified by the actual operation data in a certain area. The simulation results verify the effectiveness of the proposed method for the identification of direct PV power supply, and provide a reference for the subsequent on-site inspection of power grid companies.

Keywords: Photovoltaic Power · Direct Power Supply Identification · Affinity Propagation Clustering · Correlation Coefficient

1 Introduction

The direct power supply (DPS) is conducive to the local consumption of distributed photovoltaic (PV) [1–3]. It is also conducive to the implementation of "double carbon" and energy conservation and emission reduction policies. The Notice on the Pilot of Distributed Generation Market Trading (2017) No. 1901 also explicitly supports the relevant mode of electricity trading between distributed generation project units and

Fund project: This work was supported in part by the State Grid Hunan Electric Power Co., Ltd. 2022 "New Research Control Model" Pilot Project (5216AG220007); in part by the Changsha Science and Technology Project of Hunan Province under Grant (kq2202213).

H. Yang et al. (Eds.): SmartGIFT 2022, LNICST 483, pp. 218–228, 2023.
https://doi.org/10.1007/978-3-031-31733-0_19

nearby power users within the distribution network. According to a series of policies, it is obvious that the state encourages distributed PV projects to participate in DPS. However, due to the volatility and randomness of distributed PV output, it usually cannot fully meet the electricity demand of DPS users. It still needs the power system to provide full capacity investment, but it does not bear the corresponding transmission and distribution price. This means that customers who do not participate in the DPS will share more of the transmission and distribution costs. In addition, the DPS site is prone to the occurrence of electricity theft [4–6]. The mode of DPS also has some influence on power quality management [7–9]. In order to standardize the trading behavior of distributed PV power market, it is necessary to provide a DPS identification method of distributed PV power as an auxiliary means for the implementation of new energy policy.

The classification and abnormal identification of load data have been studied in the literature, but the abnormal identification of PV power generation data is less studied. In [10], a detection method with low false alarm rate based on identification of production and operation status is proposed. The false alarm rate of abnormal load identification is reduced by studying the corresponding relationship between various production and operation states of load and three-phase power. In [11], a power data anomaly detection method based on Holt-winters model and DBSCAN clustering is proposed. It can improve the accuracy of detection by clustering the residual term obtained by subtracting the predicted value and the real value.

Unlike load power consumption, the DPS of PV behavior is more difficult to identify and classify due to the randomness and uncertainty of PV. The existing research method is to predict the PV output and identify whether there is any abnormality by comparing the predicted data with the actual measured data. In [12], a comparison method for PV day-ahead prediction models based on deep learning neural networks is proposed. The method shows that the hybrid prediction model based on convolutional neural network and long short-term memory network has the best effect. In [13], a PV prediction method based on deep convolutional neural network and meta-learning is proposed, which achieves high accuracy and reliability of day-ahead prediction. In [14], a graphical modeling method for PV prediction based on multiple meteorological factors is proposed, which improves the accuracy of PV day-ahead forecasting. However, there are many factors affecting the prediction of PV, including climate, temperature, weather and geographical environment. The existing methods cannot comprehensively consider all the influencing factors.

There are many factors that affect the output of the PV array, such as climate, temperature, weather, solar radiation, and shade. In [15], a prediction method based on regional similarity is proposed, but the factors such as shading, dust and attenuation during the use of PV arrays are not considered. In this paper, a DPS identification method of PV output in the same area based on AP clustering is proposed. Firstly, through the basic data such as the model and installation date of PV arrays in the same area, the PV arrays of the same type and area are found. Through the correlation analysis of historical output data, the clustered PV power supply set is found. Then, the AP clustering algorithm is used to cluster the power generation data in the same temperature segment to find out the corresponding relationship between the clustering results and the state of PV power supply, so as to identify the PV power supply suspected of DPS. Finally, 10 distributed

PV power sources in a region are taken as an example to verify the effectiveness of the proposed method.

2 Output Correlation Analysis of Photovoltaic Power Supply

Since the capacity of each PV power supply is different, the power generation and grid-connected power P used in the calculation of this paper are calculated by taking the power value corresponding to the 100 kW capacity PV power supply.

$$P=\frac{P_{act}}{P_e} \times 100 \tag{1}$$

where P_{act} represents the actual value of power. P_e represents the rated power of the corresponding PV power supply.

The climatic conditions, solar radiation amount and other conditions of PV power in the same area are similar, but the shading, dust, attenuation and other factors in the process of use are different. Through the basic data such as the model and installation date of PV arrays in the same area, the interference of these factors can be eliminated by finding out the PV arrays in the same type and area as a reference. However, due to the different factors such as solar cell module tilt Angle, energy conversion efficiency and maximum power point tracking (MPPT), the output data of these PV power sources do not necessarily have sufficient similarity.

As an evaluation index to measure the degree of linear correlation between two variables, the correlation coefficient can be used to measure the output similarity between PV power sources. In this paper, Pearson Correlation Coefficient (PCC) is used to measure and screen out the PV power sources suitable for analysis together. Assuming that the corresponding variables of PV power supply a and b are P_a and P_b. The generation data of PV power supply is selected as a variable at an interval of 15min, and the correlation coefficient of PV power supply a and b is:

$$r\left(P^a, P^b\right) = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(P_i^a - \overline{P}^a\right)\left(P_i^b - \overline{P}^b\right)}{\sqrt{\sum_{i=1}^{n}\left(P_i^a - \overline{P}^a\right)^2}\sqrt{\sum_{i=1}^{n}\left(P_i^b - \overline{P}^b\right)^2}} \tag{2}$$

where n represents the number of time intervals in a day. ${P_i}^a$ and ${P_i}^b$ represent the output of PV power supply a and b at the i-th moment, respectively. $\overline{P}^a$ and $\overline{P}^b$ represent the average output of PV power supply in a day, respectively. The range of PCC is [0, 1]. PCC > 0 represents positive correlation between variables, and larger PCC represents the stronger correlation.

Assuming that the total number of PV power sources is N_a in a region. The correlation coefficient of each power source can be calculated to form the correlation coefficient matrix of the region.

$$R = \left[r\left(P^a, P^b\right)\right]_{N_a \times N_a} \tag{3}$$

In order to improve the differentiation, the data of each PV power source in the same area for one day and one hour are counted separately as variables to calculate the correlation coefficient matrix. According to the correlation coefficient less than 0.5 in the correlation coefficient matrix, the PV power sources with low correlation are screened out. Thus, the PV power population calculated together is obtained.

3 The Identification of PV Power Sales Status

For the PV power supply with strong correlation in the same station area, the power generation data at the same temperature is selected as the research sample with the resolution of 15 min. Assume that there is no DPS for PV power supply in a certain station, so the grid-connected power is the generation power. The three-phase grid-connected power curves of two typical days at 25–35 °C are shown in Fig. 1 and Fig. 2, respectively. The PV power supply had stable weather on the first day. The power fluctuation is small, and the curve presented a single peak shape. The next day is rainy and there was a temporary shortage of power generation. The curve shows a double peak shape in Fig. 2.

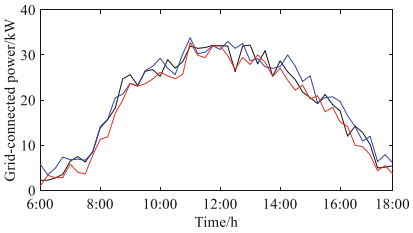

Fig. 1. The grid-connected power on the first day

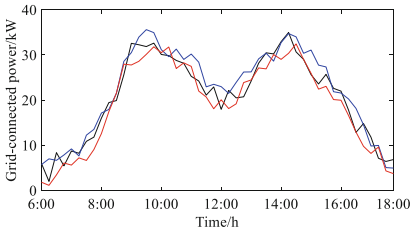

Fig. 2. The grid-connected power on the second day

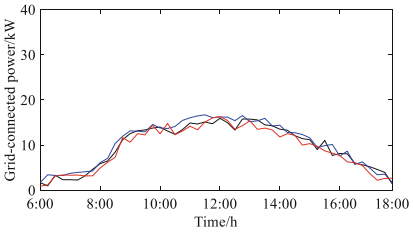

Fig. 3. The grid-connected power under "factory power sale"

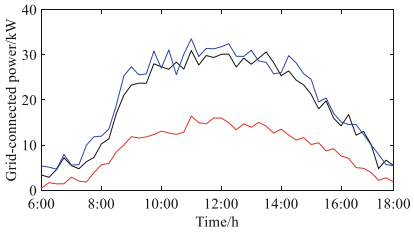

Fig. 4. The grid-connected power under "residential customers power sale"

DPS from PV power sources is mainly divided into two types of power sales to residential customers and power sales to factories. The power consumption of the factory is mainly high-power three-phase equipment, and the three-phase power is basically balanced. The power consumption is large, which is easy to cause the continuous large power shortage of PV power supply. It is shown in Fig. 3. The residential power consumption is mainly single-phase equipment with small power, which is easy to cause three-phase power imbalance of PV power supply. It is shown in Fig. 4.

According to Figs. 1, 2, 3 and 4: Under the same temperature conditions, the maximum daily power generated by PV power will not be significantly different. The power of the three phases is basically balanced. No matter what form of direct power supply will have a certain impact on the grid-connected power. In the case of "factory power sale", the three-phase grid-connected power of PV power supply will be significantly reduced. In the case of "residential customers power sale", the grid-connected power of PV power will appear three-phase imbalance. These differences can be used as the basis for the identification of direct power supply of PV power.

4 Direct Supply Power of PV Identification Based on AP Clustering Algorithm

The daily grid-connection curves of PV power sources with high correlation in the same station area under the same temperature are taken as samples, and the corresponding clustering algorithm is designed to identify whether there is DPS. Due to the random and changeable weather conditions, the number of clusters is difficult to determine. In this paper, the AP clustering algorithm [16, 17], which does not require setting the number of class families in advance, is used for clustering. The core of AP clustering algorithm is to find the set of optimal samples to maximize the sum of similarity between all samples and the nearest optimal sample. The specific algorithm flow is as follows:

The number of samples involved in the identification of PV power supply is N, and the similarity between samples is calculated using the Euclidean distance as the standard. For samples x_i and x_j, the similarity $s(i,j)$ is defined as

$$s(i,j) = -d_{i,j} = -\left\|x_i - x_j\right\| \tag{4}$$

where $d_{i,j}$ denotes the Euclidean distance between samples x_i and x_j. By calculating the similarity between all samples, the similarity matrix $S \in \mathrm{R}^{N \times N}$ can be formed. The goal of AP clustering algorithm is to calculate the corresponding class representative of each sample, that is, the clustering center of each cluster. The diagonal element of the similarity matrix is taken as the bias parameter p_i of the clustering algorithm, which represents the possibility of each sample being selected as a class representative. All diagonal elements p_i are equal. The larger the bias parameter is, the more class representatives are selected and the more class clusters are selected. On the contrary, the fewer class representatives are selected. In order to make the final number of class clusters reasonable, p_i is taken as the median of all elements in the similarity matrix.

The class representation is calculated by the attractiveness index and the attribution index. The attractiveness index $b(i,j)$ represents x_j as the class representative fit of x_i and is a parameter of x_i pointing to x_j. All the attractiveness indices form the attractiveness matrix $B \in \mathrm{R}^{N \times N}$. The attribution index $a(i,j)$ represents the probability that x_i chooses x_j as a class representative and is a parameter of x_j pointing to x_i. All the attribution indices form the attribution matrix $A \in \mathrm{R}^{N \times N}$. The initial values of the attractiveness matrix B and the attribution matrix A are taken as zero matrices.

$$b_0(i,j) = 0 \tag{5}$$

$$a_0(i,j) = 0 \tag{6}$$

The latter one attractiveness matrix element $b_1(i,j)$ and the attribution matrix element $a_1(i,j)$ are calculated as follows.

$$b_1(i,j) = \begin{cases} s(i,j) - \max\limits_{k \neq j}\{a_0(i,k) + s(i,k)\}, & i \neq j \\ s(i,j) - \max\limits_{k \neq j}\{s(i,k)\}, & i = j \end{cases} \tag{7}$$

$$a_1(i,j) = \begin{cases} \min\left\{0, b_0(j,j) + \sum_{k \neq j} \max\{b(k,j), 0\}\right\}, & i \neq j \\ \sum_{k \neq j} \max\{b(k,j), 0\}, & i = j \end{cases} \tag{8}$$

The number of iterations is denoted by t. The attractiveness matrix element $b_{t+1}(i,j)$ and the attribution matrix element $a_{t+1}(i,j)$ are calculated by iterating the following equation.

$$b_{t+1}(i,j) = \lambda \times b_{t-1}(i,j) + (1-\lambda) \times b_t(i,j) \tag{9}$$

$$a_{t+1}(i,j) = \lambda \times a_{t-1}(i,j) + (1-\lambda) \times a_t(i,j) \tag{10}$$

where λ represents the attenuation coefficient of the AP clustering algorithm, also known as the damping factor, which mainly affects the iterative convergence speed.

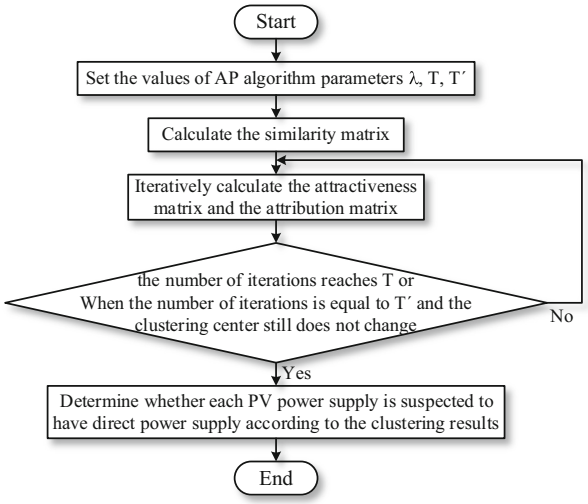

Fig. 5. Flowchart of direct power supply identification for PV power based on AP clustering algorithm

The sum of the attractiveness matrix elements $b_t(i,j)$ and the attribution matrix elements $a_t(i,j)$ obtained at each iteration step is calculated, and the sample x_j corresponding

to the largest sum is the class representative of sample x_i.

$$sum = b_t(i,j) + a_t(i,j) \tag{11}$$

Find all the class representatives of each iteration and determine the cluster center. The iteration process ends when the number of iterations reaches T or when the number of iterations is equal to T′ and the clustering center still does not change. According to the final clustering results, the direct power supply corresponding to each sample can be identified. The whole AP clustering process can be expressed as Fig. 5.

5 Case Study

5.1 Output Correlation Analysis

The case study was conducted with 10 PV power sources of the same type in a certain station. These PV arrays are similar in terms of shading, dust, and attenuation during use. The hourly generation data of each photovoltaic power supply in the same time period is shown in Fig. 6.

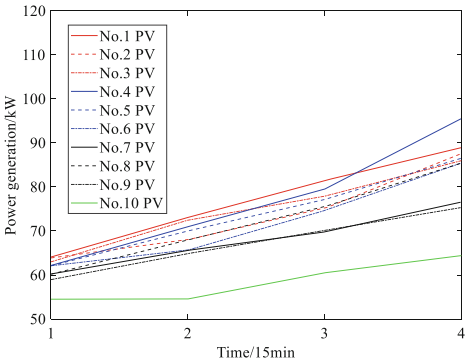

Fig. 6. The hourly generation data of each photovoltaic power supply in the same time period

PCC is used to measure the correlation of the power output of each PV power source. The hourly generation data shown in Fig. 6 is used as the variable to calculate the correlation coefficient matrix recorded in Table 1.

As can be seen from Table 1, items less than 0.5 in the correlation coefficient matrix mainly appear in the corresponding correlation coefficients of No. 7 and No. 10 photovoltaic power sources. Therefore, No. 1–6 and No. 8–9 are selected as the photovoltaic power supply groups clustered together.

5.2 The Direct Power Supply Identification of Photovoltaic Power Supply

All daily three-phase power generated by PV power sources 1–6 and 8–9 at 25–35 °C in the past year are selected as samples for analysis. The total number of samples is 163.

Table 1. The correlation coefficient matrix of each photovoltaic power source.

PV	1	2	3	4	5	6	7	8	9	10
1	1	0.9812	0.9693	0.9672	0.9125	0.9813	0.5795	0.9524	0.7324	0.2538
2	0.9812	1	0.9005	0.9854	0.9558	0.9955	0.4757	0.9132	0.8364	0.2826
3	0.9693	0.9005	1	0.9104	0.9846	0.9871	0.5566	0.9980	0.7912	0.3178
4	0.9672	0.9854	0.9104	1	0.9864	0.8842	0.5815	0.8983	0.7725	0.2805
5	0.9125	0.9558	0.9846	0.9864	1	0.9440	0.5505	0.9936	0.7810	0.4014
6	0.9813	0.9955	0.9871	0.8842	0.9440	1	0.5397	0.9752	0.7847	0.1926
7	0.5795	0.4757	0.5566	0.5815	0.5505	0.5397	1	0.5560	0.6737	0.6138
8	0.9524	0.9132	0.9980	0.8983	0.9936	0.9752	0.5560	1	0.7906	0.3334
9	0.7324	0.8364	0.7912	0.7725	0.7810	0.7847	0.6737	0.7906	1	0.6706
10	0.2538	0.2826	0.3178	0.2805	0.4014	0.1926	0.6138	0.3334	0.6706	1

The AP clustering algorithm is used to cluster these samples. The damping factor λ is 0.5. The bias parameter p is taken as the median of all terms of the similarity matrix S ranked from largest to smallest. The maximum number of iterations T is 400. The maximum number of iterations T′ is 40 if the clustering center does not change. The three-phase grid-connected power curves corresponding to each cluster center are shown in Figs. 7, 8, 9, 10, 11 and 12.

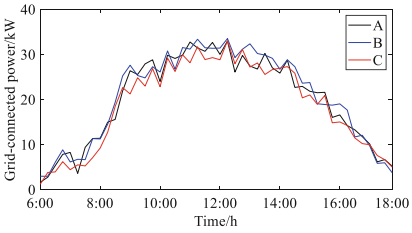

Fig. 7. The grid-connected power of class cluster center 1.

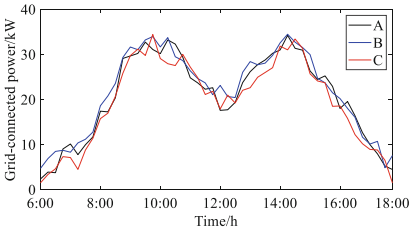

Fig. 8. The grid-connected power of class cluster center 2.

Cluster 1 represents the three-phase grid-connected power without direct power supply. The power generation is high, and the power of the three phases is basically balanced. The weather conditions corresponding to this kind of cluster are good, and the power generation first rises and then declines with the advance of time, presenting a "single peak" shape.

Cluster 2 shows a "double peak" shape, which corresponds to the brief rainy days during the day. The power generation power showed a temporary drop during the day and then rebounded to a higher state. The highest grid-connected power of photovoltaic power is close to the cluster 1, and the three-phase power is basically balanced. This type of cluster also corresponds to the case of no DPS.

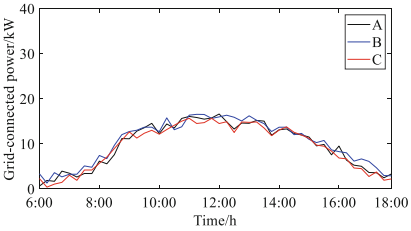

Fig. 9. The grid-connected power of class cluster center 3.

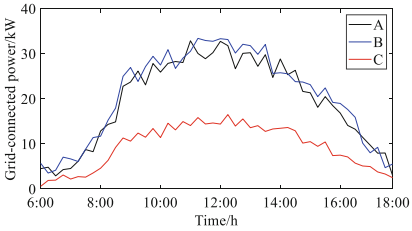

Fig. 10. The grid-connected power of class cluster center 4.

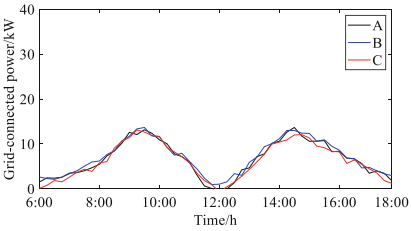

Fig. 11. The grid-connected power of class cluster center 5.

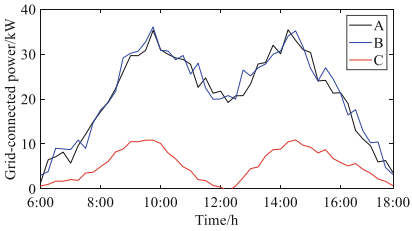

Fig. 12. The grid-connected power of class cluster center 6.

The three-phase power of cluster 3 and cluster 5 is basically balanced, but there is a large shortage of three-phase power. This indicates that the PV power supply corresponding to this kind of cluster directly supplies power to the factory. The different weather conditions affect the power generation, so the cluster 3 and cluster 5 are "single peak" shape and "double peak" shape, respectively. Samples contained in these two clusters belong to No. 8 PV power supply, indicating that No. 8 PV power supply is suspected of supplying direct power to factory users.

The highest power of cluster 4 and cluster 6 is higher, but there is a large power shortage in one phase. This corresponds to single-phase electrical equipment such as household appliances. The PV power supply corresponding to this kind of cluster directly supplies power to residential users. The different curve shapes of cluster 4 and cluster 6 are also caused by different weather states. The samples contained in these two clusters belong to No. 9 PV power supply, indicating that No. 9 PV power supply is suspected of supplying direct power to residential users.

In summary, the suspicion of DPS of No. 8 and No. 9 PV power supply is identified by AP clustering algorithm.

6 Conclusion

The identification method of DPS based on AP clustering algorithm proposed in this paper can effectively identify the DPS of PV power. The differences in climate, light intensity and other factors can be reduced by selecting the PV power source in the same

station for analysis. Choosing the same type of PV power supply can reduce the shading, dust, attenuation and other factors. The output correlation of each PV power supply can be measured by the correlation coefficient, and the PV power supply with low correlation can be screened out. The AP clustering algorithm is used to divide the three-phase grid-connected power curves of each PV power supply at the same temperature into different clusters. The characteristics of the grid-connected power of each type of cluster can be used to know whether the corresponding sample is suspected of having direct power supply. Further, it can judge which type of DPS exists and provide reference for the on-site inspection of the power supply company.

References

1. Singh, A.K., Parida, S.K.: A review on distributed generation allocation and planning in deregulated electricity market. Renew. Sustain. Energy Rev. **82**, 4132–4141 (2018)
2. Nguyen, S., Peng, W., Sokolowski, P., et al.: Optimizing rooftop photovoltaic distributed generation with battery storage for peer-to-peer energy trading. Appl. Energy **228**, 2567–2580 (2018)
3. Georgilakis, P.S., Hatziargyriou, N.D.: Optimal distributed generation placement in power distribution networks: models, methods, and future research. IEEE Trans. Power Syst. **28**(3), 3420–3428 (2013)
4. Jokar, P., Arianpoo, N., Leung, V.C.M.: Electricity theft detection in AMI using customers' consumption patterns. IEEE Trans. Smart Grid **7**(1), 216–226 (2015)
5. Shaaban, M., Tariq, U., Ismail, M., et al.: Data-driven detection of electricity theft cyberattacks in PV generation. IEEE Syst. J. (2021)
6. Ismail, M., Shaaban, M.F., Naidu, M., et al.: Deep learning detection of electricity theft cyber-attacks in renewable distributed generation. IEEE Trans. Smart Grid **11**(4), 3428–3437 (2020)
7. Campanhol, L.B.G., Da Silva, S.A.O., De Oliveira, A.A., et al.: Power flow and stability analyses of a multifunctional distributed generation system integrating a photovoltaic system with unified power quality conditioner. IEEE Trans. Power Electron. **34**(7), 6241–6256 (2018)
8. Miller, W., Liu, A., Amin, Z., et al.: Power quality and rooftop-photovoltaic households: an examination of measured data at point of customer connection. Sustainability **10**(4), 1224 (2018)
9. Kharrazi, A., Sreeram, V., Mishra, Y.: Assessment techniques of the impact of grid-tied rooftop photovoltaic generation on the power quality of low voltage distribution network-a review. Renew. Sustain. Energy Rev. **120**, 109643 (2020)
10. Du, Z., Su, S., Liu, Z., et al.: Second inspection method for electricity theft detection with low false alarm rate based on identification of production and operation status. Autom. Electr. Power Syst. **45**(02), 97–104 (2021)
11. Xiao, Y., Zheng, K., Yu, Z., et al.: Power data anomaly detection based on holt-winters model and DBSCAN clustering. Power Syst. Technol. **44**(03), 0320 (2020)
12. Wang, K., Qi, X., Liu, H.: A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. Appl. Energy **251**, 113315 (2019)
13. Zang, H., Cheng, L., Ding, T., et al.: Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning. Int. J. Electr. Power Energy Syst. **118**, 105790 (2020)
14. Cheng, L., Zang, H., Ding, T., et al.: Multi-meteorological-factor-based graph modeling for photovoltaic power forecasting. IEEE Trans. Sustain. Energy **12**(3), 1593–1603 (2021)

15. Lu, S., Peng, S., Yang, Y., et al.: Identification method of abnormal photovoltaic users based on mean impact value and heuristic forward searching. Electr. Power Autom. Equip. **42**(02), 106–111 (2022)
16. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. Science **315**(5814), 972–976 (2007)
17. Wang, K., Zhang, J., Li, D., et al.: Adaptive affinity propagation clustering. arXiv preprint arXiv:0805.1096 (2008)