Identification of low-voltage connection relation in distribution platform based on similarity of voltage curve and grey correlation degree of entropy weight

Yuehao Yan¹, Wei Bao¹, Lijiao Li² and Zongjie Wang*³

¹ Zhengzhou Power Supply Company, State Grid Henan Electric Power Company, Zhengzhou 450000, China
² Zhengzhou Songyao Information Technology Limited Company, Zhengzhou 450000, China
³ School of Electrical Engineering, Zhengzhou University, Zhengzhou 450000, China
*Corresponding author’s e-mail: wangzongjie97@163.com

Abstract: The correct low-voltage connection relationship in distribution area is of great significance to the safe operation and efficient management of power grid. As there are many reconstruction projects in the low-voltage platform area, the assets change frequently, and the interconnection perception ability of the low-voltage platform area is weak, which brings great difficulties to the identification of the user connection relationship. Traditional identification methods are heavy and inefficient. This paper proposes a method based on trend similarity and distance measure to identify low-voltage connection relation in distribution platform area. Firstly, Pearson correlation coefficient and discrete Fréchet distance are calculated to measure the trend similarity of voltage curve, and abnormal users are found out. GIS is used to search for adjacent stations, and finally, the correct station area of users is determined by analyzing the entropy weight grey correlation degree. The applicability and correctness of the proposed algorithm are verified by the application results of an example.

1.Introduction
The safety and stability of power distribution system is directly related to the benefit, quality and reliability of power consumption of each user. [1] The current distribution network has a series of problems. To solve these problems, the most important thing is to obtain accurate low-voltage connection relation of distribution platform area. Only by obtaining the correct low-voltage connection relationship in real time, can the internal information of the stations be comprehensively obtained and provided to the operation and maintenance personnel, so as to fundamentally solve the problem of inaccurate file data of the stations, so as to realize the intelligent control and lean operation of the low-voltage distribution network.

With the expansion of the number and scale of smart electricity meter users, power grid companies can master massive and high-density data through the feedback of smart electricity meter. The increasing scale of data gives birth to the research soil of using data analysis methods for topological identification of stations. [2]

Literature [3] uses the correlation coefficient between voltage sequence data as the correlation measure in the algorithm to analyze the degree of outlier in the user ring domain, so as to verify the
wrong users in the platform region. But the method is not universal. In literature [4], the verification platform area is first modeled, and then the user voltage and current data obtained are put into the model to verify the accuracy of the connection relation. Although the effect is better, but the process is more cumbersome.

To solve the above problems, in order to identify the low-voltage connection relationship in distribution platform area more efficiently and accurately, this paper proposes an identification method based on voltage curve similarity and entropy weight grey correlation degree. Considering the characteristics of the variable household relationship and the characteristics of the collected data, the Pearson correlation coefficient and the discrete Fréchet distance of the user voltage and the distribution voltage of the station were calculated to screen out the abnormal users and generate the user set with inconsistent distribution. According to GIS data, the adjacent distribution station area of abnormal users is obtained, and the correct belonging station area of users is judged by using entropy weight grey correlation degree. The result of case analysis proves that the identification effect of this method has higher recall rate and precision rate, and compared with the traditional data-driven method, it is more accurate and efficient, which helps to improve the service quality and management efficiency of low-voltage power supply area better.

2. Algorithm

2.1 Voltage curve similarity measure

In the distribution network, the voltage often fluctuates because of the uncertainty of loads everywhere. The voltage fluctuation curves of the loads with relatively close electrical distance are similar, while the voltage fluctuation curves of the loads with relatively far electrical distance are less similar. Voltage curve trend similarity adopts Pearson correlation coefficient, Pearson correlation coefficient is also called Pearson product moment correlation coefficient, is measure the linear correlation between two variables $X$ and $Y$, with a value between -1 and 1.

Pearson correlation coefficient is defined as formula (1).

$$
\rho = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}
$$

In the formula, $\bar{X}$, $\bar{Y}$ is the mean of $X$ and $Y$; $\rho \in [-1,1]$, if $\rho > 0$, it indicates that the two variables are positively correlated; if $\rho < 0$, it indicates that the two variables are negatively correlated; The greater the absolute value of $\rho$, the stronger the correlation.

It can be seen from Equation (1) that Pearson coefficient ignores the distance between voltage curves. This is an omission for under-voltage power stealing users or users with large amplitude difference but similar trend of connection information error. Therefore, on the basis of Pearson correlation coefficient calculated by using the user's intelligent voltmeter voltage and three-phase voltage at the transformer's low-voltage side, discrete Fréchet distance is used for review.

2.2 Distance measure of voltage curve of smart electricity meter

The Fréchet distance was first proposed by Fréchet, and its calculation object was continuous curve. Eiter and Mannila proposed discrete Fréchet distance on the basis of continuous Fréchet distance, and took discrete Fréchet distance as the distance measure of curve:

Suppose a polygon line with m endpoints $P$:

$$
P = \left\{ P_1, P_2, ..., P_m \right\}
$$

A polygon line with n endpoints $Q$:
The link sequence $L$ formed by each endpoint on $P$ and $Q$ is as follows:

$$L = \{p_{u1}, q_{v1}, p_{u2}, q_{v2}, \ldots, p_{uk}, q_{vk}\}$$

Length $L_{\text{max}}$ is defined as the maximum value in sequence $L$, that is, the longest connection length:

$$L_{\text{max}} = \max_{i=1,2,\ldots,k} \text{dist}(p_{ui}, q_{vi})$$

Then the discrete Fréchet distance between the two sequences is:

$$\delta_{\text{DF}}(P, Q) = \min \{L_{\text{max}}\}$$

The discrete Fréchet distance has the following properties:

1. Calculating the discrete Fréchet distance does not require the same number of endpoints in each curve;
2. The discrete Fréchet distance is the minimum distance between the links that represents the maximum distance required between the curves, which is better than calculating the average distance between all the links. Therefore, the discrete Fréchet distance is selected as the distance measure between curves.

### 2.3 Search method of adjacent platform area based on GIS

In the GIS platform of smart power grid, the problem of low-voltage connection relation of distribution network is usually that users in a certain station area are incorrectly recorded to neighboring stations. Therefore, when correcting users with incorrect station area and phase information, it is only necessary to consider their physical location in neighboring stations.

According to technical guidelines for distribution network planning and design, 220/380 V lines should have a clear power supply range. In principle, the power supply radius of class $A+$ and $A$ power supply area should not exceed 150m, class $B$ should not exceed 250m, class $C$ should not exceed 400m, and class $D$ should not exceed 500m. Class $E$ and $F$ should be calculated and determined as required.

Calculate the physical distance between two distribution transformers according to the latitude and longitude coordinates in GIS:

$$d_{AB} = R \times \cos^{-1}\left[ \cos \omega A \times \cos \omega B \times \cos(jA - jB) + \sin \omega A \times \sin \omega B \right]$$

Where, the longitude and latitude of transformer A and B are respectively $(jA, \omega A)$ and $(jB, \omega B)$; the radius of the earth is $R$.

![Figure. 1. Two physically similar platform regions](image)

### 2.4 Identification of low voltage connection relation based on entropy weight grey correlation

Suppose that there are altogether $m$ total table voltage sequences of the original station area of the user to be verified and the station area adjacent to the physical location in GIS, and there are altogether $T$ users to be verified, then the reference sequence is denoted as $y_1, y_2, \ldots, y_m$, and the comparison sequence

$$Q = \{q_1, q_2, \ldots, q_m\}$$

...
is denoted as \(x_i, x_2, \ldots, x_t\), where:
\[
y_i = \{y_i(1), y_i(2), \ldots, y_i(n)\}, i = 1, 2, \ldots, m
\]
\[
x_j = \{x_j(1), x_j(2), \ldots, x_j(n)\}, j = 1, 2, \ldots, t
\]
(8)

Where, \(n\) is the voltage sequence indicator used for analysis. When \(n\) is taken at a certain moment \(k\), \(y_i(k)\) and \(x_j(k)\) represent the \(k\) indicator of \(y_i\) and \(x_j\) respectively.

Calculate the correlation coefficient between the total meter of known station and the voltage sequence of the watt-hour meter to be identified \(\xi_y(k)\):
\[
\xi_y = \frac{D_{\min} + D_{\max}}{D_y(k) + \rho D_{\max}}
\]
(9)

\[
D_{\min} = \min_i \min_k |y_i(k) - x_j(k)|
\]
(10)

\[
D_{\max} = \max_i \max_k |y_i(k) - x_j(k)|
\]
(11)

\[
D_y(k) = |y_i(k) - x_j(k)|
\]
(12)

Where, \(\rho\) is the resolution coefficient, and its value range is \([0,1]\), usually 0.5.

Therefore, the average correlation between reference sequence and comparison sequence is:
\[
\gamma_{ij} = \frac{1}{n} \sum_{k=1}^{n} \xi_y(k)
\]
(13)

Since the weight of the average correlation degree \(\gamma_{ij}\) in Formula (8) is \(1/n\), it cannot reflect the influence of different indicators on the reference indicators, so the concept of information entropy is introduced. Put forward using the entropy weight method and gray correlation analysis for low voltage articulated identification, entropy weight method in check change relation, each point voltage indicator in the area of low pressure articulated identification has different importance and characteristics of must give the corresponding weights of different indexes, weights reflect the important degree of each point voltage in sequence.

It is assumed that there are a total of \(m\) total table voltage sequences of the original station to which the user to be verified belongs and adjacent stations at their physical locations, and each sequence takes the voltage values of \(n\) points to form \(m \times n\) moments \(A'\):
\[
A' = [a'_{ik}]_{m \times n}
\]
(14)

Where, \(a'_{ik}\) represents the value of the \(k\) indicator in the voltage sequence of the \(i\) station.

After averaging:
\[
A = [a_{ik}]_{m \times n}
\]
(15)

The entropy of the \(j\) indicator is:
\[
E_k = \alpha \sum_{i=1}^{m} b_{ik} \ln b_{ik}, k = 1, 2, \ldots, n
\]
(16)

\[
\alpha = \frac{1}{\ln m}
\]
(17)

\[
b_{ik} = \frac{a_{ik}}{\sum_{i=1}^{m} a_{ik}}
\]
(18)
When \( b_{ik} = 0 \), let \( \sum_{i=1}^{m} b_{ik} \ln b_{ik} = 0 \) then the entropy weight of the \( k \) indicator is:

\[
\omega_k = \frac{1 - E_k}{\sum_{k=1}^{n} (1 - E_k)}
\]

(19)

Where, \( \omega_k \in [0,1] \), \( \sum_{j=1}^{n} \omega_k = 1 \), The weight of each voltage index is obtained.

\[
W = [\omega_1, \omega_2, \ldots, \omega_n]
\]

(20)

Replace fixed weight \( 1/n \) in Formula (8) with variable weight \( \omega_k \):

\[
\gamma_{ij} = \sum_{k=1}^{n} \omega_k \tilde{x}_{ij}(k)
\]

(21)

3. The specific algorithm

User identification of low-voltage connection errors in the electricity information acquisition system is realized on the basis of calculating the Pearson correlation coefficient and discrete Fréchet distance of the total table of the platform and the users under the platform, which includes the following steps:

A. Obtain one-day voltage data of the master meter of the station and the user's smart meter from the electricity consumption information collection system.

B. Calculate Pearson correlation coefficient and discrete Fréchet distance between each user and the phase voltage series connected to the electricity information acquisition system within one day in the embodiment area. If Pearson correlation coefficient is less than 0.6, but the discrete Fréchet distance is greater than the threshold, the user is suspected of stealing electricity by undervoltage method; If Pearson correlation coefficient is greater than or equal to 0.6, and the discrete Fréchet distance is less than the threshold, the user is judged to belong to this region. Otherwise, the user low-voltage connection information is incorrect, and all users with incorrect low-voltage connection relationships are obtained.

C. The transformer area adjacent to the user to be verified can be obtained according to GIS data, and the three-phase voltage of the low-voltage side of the transformer in the distribution platform area is taken as the reference sequence. On this basis, the entropy weight of each voltage indicator is obtained, and the voltage of the user to be verified is taken as the comparison sequence.

D. The voltage sequence of the low-voltage side of the adjacent station area on the original station area and physical location of the user to be verified is obtained as the reference sequence, and the voltage value of 24 points is taken for each sequence. The weight of voltage index is obtained by entropy weight method. The larger the difference between the reference voltages, the greater the entropy weight. The stage and phase of the voltage curve with the largest correlation degree belong to the user to be verified.
4. Application instance

The data of some low-voltage stations under the jurisdiction of some municipal power supply company are taken as the input for example analysis. All the selected stations are "problem stations" where the line loss index in the system is abnormal due to the low voltage connection relation inconsistent with the actual situation on site. The abnormal data in the stations have been verified by the maintenance personnel.

There are 72 users in the implementation sample area. Pearson correlation coefficient and discrete Fréchet distance of each user in the implementation sample area and the phase voltage sequence connected to it in the electricity information acquisition system are calculated. It is found that the correlation coefficient between low-voltage users 13, 14, 15, 22 and phase A is significantly lower than that between other users and phase A. The correlation coefficient between user 40 and B phase on the low-voltage side is significantly lower than that between other users and B phase. Therefore, users 13, 14, 15, 22, and 40 do not belong to the zone. The correlation coefficient between the voltage of each phase on the low-voltage side of the distribution platform area and other users is high, and the discrete distance is lower than the threshold value, so it can be judged that they belong to this transformer area. Two adjacent transformer regions of TA2 can be obtained according to GIS data, namely TA1 and TA3 in Figure 3.
A total of 6 low-voltage side voltage sequences of TA2 and TA1 and TA3 were obtained from the five stations to which users originally belonged and the adjacent stations in physical locations as reference sequences. Voltage values of 24 points were taken from each sequence to form a 246-order data matrix $\mathcal{A}$, obtained after standardized treatment. The weight of voltage index is obtained by entropy weight method. The larger the difference between the reference voltages, the greater the entropy weight. The entropy weight of the ten indicators is as follows, accounting for 86.4%.

$$\omega = [\omega_{10}, \omega_{11}, \omega_{12}, \omega_{16}, \omega_{17}, \omega_{18}, \omega_{19}, \omega_{20}, \omega_{21}, \omega_{23}] = [0.0977, 0.0644, 0.049, 0.0511, 0.0181, 0.0341, 0.0848, 0.2417, 0.0229, 0.0375]$$

Calculate the weighted correlation degree between the low-voltage side voltage of the transformer at the known station and the voltage sequence of the user’s electricity meter to be identified. The weighted correlation degree in the embodiment is listed as follows:

| Courts | The user | m13 | m14 | m15 | m22 | m40 |
|--------|----------|-----|-----|-----|-----|-----|
| TA1 A phase | 0.619 | 0.652 | 0.630 | 0.718 | 0.617 |
| TA1 B phase | 0.593 | 0.625 | 0.619 | 0.756 | 0.589 |
| TA1 C phase | 0.587 | 0.618 | 0.618 | 0.776 | 0.783 |
| TA3 A phase | 0.780 | 0.752 | 0.740 | 0.598 | 0.573 |
| TA3 B phase | 0.575 | 0.616 | 0.590 | 0.876 | 0.583 |
| TA3 C phase | 0.560 | 0.587 | 0.567 | 0.674 | 0.558 |

It can be seen that users 13, 14, and 15 are highly correlated with PHASE A of TA3, user 22 with phase B of TA3, and user 40 with phase C of TA1. Thus, the correct low-voltage connection relationship of each user can be determined, and the correctness of identification can be determined through on-site confirmation.

5. conclusion

The positive and beneficial effects of the identification algorithm of low-voltage connection relation in distribution platform area based on voltage curve similarity and entropy weight grey correlation are as follows:

1. Pearson correlation coefficient with discrete Fréchet distance is used to find the wrong user, avoiding the single consideration of sequence trend similarity and ignoring the distance factor. Meanwhile, the calculation times are reduced to improve the efficiency of the algorithm.

2. When correcting and connecting wrong users, the entropy weight grey correlation method is used to distinguish the sampling index of reference sequence, so as to improve the identification accuracy and avoid the problem of equal weight in the grey correlation method.

3. The reference voltage sequence adopted is the voltage at the low-voltage outlet side of the transformer in the adjacent station area rather than the voltage of all users in the adjacent station area,
which has high reliability and improves the detection efficiency.

References
[1] Li Kun, Zhou Zhou, Zhang Yongjun, Liu Xuan, Li Guichang, Huang Guoquan. Reliability evaluation of topology identification results of low voltage station based on measurement data quality [J]. Automation of Electric Power Systems, 2014, 45(17):99-107.
[2] Chen Zhaoran, Huang Chun, Zhang Zhidan, Jiang Yaqun. Topology Identification method of Low-voltage Distribution Network Based on T-type grey Correlation degree and KNN Algorithm [J]. Power System Protection and Control, 2014, 49(01):163-169.
[3] Shirong Zou, Hui Lu, Ziming Hong, Jiedong Huang, Lintao Chen, Zhiwei Liao. Topology checking of low-voltage distribution network based on voltage correlation analysis of dynamic time series [J]. Guangdong electric power, 2020, 33(05):52-59.
[4] Guo Shen-Lin, Lin Jia-ying, WANG Peng, Zhang Jichuan, Chen Lei, TANG Guo-jing. Topology verification of low-voltage distribution network based on ROF outgroup detection [J]. Control engineering of China, 2020, 27(01):201-206.
[5] Yong Xiao, Yun Zhao, Zhidong Tu, Bin Qian, Runmian Chang. Topology Verification Method of Low-voltage Distribution Network Based on Improved Pearson Correlation Coefficient [J]. Power system protection and control, 2019, 47(11):37-43.
[6] Zhang Pan, Gao Qiangwei, Huang Xu, Zhu Cong, Sun Weiqing. Identification Method of Low-voltage Distribution Network Topology Based on Power Carrier Communication [J]. Electronic Devices, 2014, 44(01):162-167.