Consumer-Transformer Relationship Identification Based on Two-scale similarity and SC Algorithm

Xuanping Lai2*, Siyangjie Liu1, Min Cao1, Yongjie Nie1, Tengfei Zhao2

1Electric Power Research Institute, Yunnan power Gird Co., Ltd., Kunming, Yunnan, 650217, China
2North China Electric Power University, Changping, Beijing, 102200, China

*Corresponding author’s e-mail: nieyongjie@stu.xjtu.edu.cn, 1060353878@qq.com

Abstract. In view of the chaotic relationship between user-transformer in the current low-voltage stations, manual identification consumes a long time period, low identification accuracy, and high input system costs. This paper proposes a method for identifying household change relations in low-voltage stations based on dual-scale similarity and spectral clustering algorithm. First, according to the different similarities of user voltage curves in different stations, consider Euclidean distance and cosine distance to establish a two-scale spectral clustering similarity matrix from the distance and shape of the voltage curve. Then based on the two-scale similarity matrix, perform spectral clustering on the voltage curves of the same station area to complete the clustering of users in the same station area. Examples show that the method proposed in the article has high recognition accuracy, low input cost, and has good application effects.

1. Introduction

For a long time, due to the large number of users in the low-voltage station area, the huge low-voltage equipment, and the complicated wiring form, the ownership relationship between users and the station area has basically relied on manual entry systems [1,2]. Relying on manual identification of user station zone attribution problems consumes a long time period, low recognition accuracy, and high input system cost [3]. The wrong household change relationship not only affects the user's power supply service quality, but also restricts the accurate calculation of line loss [4,5]. Therefore, there is an urgent need for a low-cost, high-efficiency, and widely-applicable method to identify the relationship between household changes in the station.

Aiming at the problems of long identification cycle time, low identification rate and high cost of current household change relationship.

Literature [6] proposed a method based on quantum genetics combined with fuzzy clustering, through fuzzy clustering of the magnitude of the zero-crossing voltage drift of each station area to classify the user's station area. Literature [7] proposed to judge abnormal household change relationship based on Pearson's correlation coefficient and K-means clustering. Literature [8] proposed a station topology recognition method based on gray correlation degree and KNN algorithm, by adding time series to the gray correlation degree to judge the degree of correlation between user voltage data and the station area, and then assign the station area to which the user belongs. Literature [9] proposed to recognize the household change relationship based on the derivative dynamic time bending distance and space density algorithm. Literature [10] proposes to identify the household
change relationship based on the similarity of voltage data. It is proposed that the above-mentioned documents only consider the distance similarity of a single voltage curve for clustering so as to classify the user's station area without considering the voltage curve trend [11].

The above literature only considers the distance similarity of a single voltage curve. The paper proposes a method of household change relationship recognition based on a two-scale similarity spectral clustering algorithm. First, according to the different similarities of user voltage curves in different stations, the Euclidean distance and cosine distance are considered to establish a two-scale similarity matrix from the voltage curve distance and shape; then based on the two-scale similarity matrix, the user voltage curve is subjected to multiple spectral clustering, so as to complete the identification of consumer-transformer relationship.

2. Voltage feature extraction method based on two-scale similarity

2.1 The basic principle of household-transformation relationship recognition and the preprocessing of voltage data

In the same station area, when the power grid fluctuates, the distribution side and the user side are also in a fluctuating state. The low-voltage station area has complex assets and numerous user accesses. Due to the relatively close electrical distance under the same line, the transformer gateway voltage and the end user voltage fluctuation time curve are the same; the transformer gateway voltage under different stations or under different lines The user voltage fluctuation time curve is quite different. The voltage similarity between users in the same station and users in different stations has the following magnitude relations:

\[
\begin{align*}
R_{xy} &> R_{xz}, R_{xy} > R_{yz} \\
x, y \in \{T_1\}, z \in \{T_2\}
\end{align*}
\]

(1)

In the formula: R is the voltage curve similarity measurement index, x and y belong to users in the same station, z is users in different stations from x and y, and T is the station area. In general, the similarity of voltage curves between users in the same station is much higher than the similarity of voltage curves between users in different stations and users in the same station.

Aiming at the current inconsistency of the voltage frozen data time axis due to the distance of transmission lines, collector clock drift, communication delay and other reasons, this paper uses data cleaning technology to deal with missing, abnormal, and invalid data to a certain extent. Select the daily voltage data of all users in the off-station area on the same date as input, and use interpolation to complement individual missing values to obtain the user's original voltage matrix V.

\[
V = \begin{bmatrix}
    v_{1,1} & v_{1,2} & \cdots & v_{1,D} \\
    v_{2,1} & v_{2,2} & \cdots & v_{2,D} \\
    \vdots & \vdots & \ddots & \vdots \\
    v_{N,1} & v_{N,2} & \cdots & v_{N,D}
\end{bmatrix}
\]

(2)

In the formula, N is the total number of users in the station area, D is the number of daily voltage collection points.

In order to ensure that the data is normalized to [0,1] and facilitate the subsequent voltage feature extraction and the calculation of the clustering algorithm, this paper uses the maximum and minimum normalization method to preprocess the original voltage matrix to obtain the user voltage normalization After the matrix, that is:
\[
\begin{align*}
V_i &= [v_{i,1}, v_{i,2}, \ldots, v_{i,D}]^T \\
V' &= \frac{V_i - \min(V_i)}{\max(V_i) - \min(V_i)} \quad i = 1, 2, \ldots, D \\
V' &= [V'_1, V'_2, \ldots, V'_D]
\end{align*}
\]

In the formula, \( V_i \) is the \( i \)-th column vector of the original voltage matrix \( V \), which represents the voltage data of all users at the \( i \)-th moment, \( \max(V_i) \) and \( \min(V_i) \) are the maximum and minimum voltage data of all users at the \( i \)-th moment, respectively. \( V' \) is the normalized column vector of \( V_i \).

### 2.2. Voltage feature extraction algorithm

(1) Similarity measure of voltage curve distance

Euclidean distance is a measure of the absolute distance between two points in \( n \)-dimensional space, and the distance between the user voltage curve and:

\[
x_{i,j} = \sqrt{\sum_{k=1}^{n} (v_{i,k} - v_{j,k})^2}
\]

(4)

(2) Similarity measure of trend characteristics

The paper considers the similarity of the voltage curve from the user voltage trend characteristics, and the cosine similarity between the user voltage curve and:

\[
y_{ij} = \frac{\sum_{k=1}^{n} (v_{i,k} - \overline{V_i})(v_{j,k} - \overline{V_j})}{\sqrt{\sum_{k=1}^{n} (v_{i,k} - \overline{V_i})^2} \sqrt{\sum_{k=1}^{n} (v_{j,k} - \overline{V_j})^2}}
\]

Then the cosine distance between the user voltage curve \( V_i \) and \( V_j \) is:

\[
y_{ij} = 1 - y_{ij}
\]

(5)

(6)

Compared with distance similarity to measure the similarity of user voltage curves, cosine similarity pays more attention to the difference in trend and direction of pairwise curves, which can reflect the trend characteristics of the curve in time. The result of the similarity matrix formed by the cosine distance is between 0 and 1, which is suitable for spectral clustering algorithm and can also describe the trend characteristics of each user's voltage curve.

(3) Two-scale similarity measure

The Euclidean distance can distinguish a certain degree of difference in user voltage, but it cannot effectively indicate the similarity of the trend change of the user voltage curve. Therefore, this article considers the comprehensive distance of the user voltage curve comprehensively considering the multi-factor similarity of the Euclidean distance and the cosine distance, which can fully reflect the distance and the degree of similarity of the trend change, which is defined as follows:

\[
W = \alpha X + \beta Y, \quad \alpha + \beta = 1
\]

(7)

In the formula, \( W \) is the spectral clustering comprehensive similarity matrix constructed by multiple factors; \( X \) is the similarity matrix obtained through Euclidean distance; \( Y \) is the similarity matrix obtained through cosine distance. And are the weights of the similarity matrix and respectively, in the text \( \alpha \) and \( \beta \) are both taken as 0.5.

### 3. Algorithm flow of consumer-transformer relationship identification

The specific steps of consumer-transformer relationship recognition based on two-scale similarity and SC algorithm are as follows.

Step 1: Import smart meters and substation transformers for users in low-voltage stations

Historical voltage time series.
Step 2: Calculate the voltage sequence of the low-voltage side of the substation and the user voltage.

Step 3: Set initial general clustering parameters $k$ and $\delta$, according to step 2. The two-scale similarity distance clusters station changes and users.

Step 4: Record the results of Step 3, the Taiwanese changes and users belong to the same category.

Step 5: Change the clustering parameters $k$ and $\delta$ and repeat step 4 until Reach the predetermined number of clusters.

Step 6: Calculate the probability that the station change and the user belong to the same station area.

Sexual results, to identify the household change relationship.

4. Case analysis

4.1 Low-voltage station area topology

The establishment of the topology model of the low-voltage distribution station area is shown in Figure 1. For the three transformers $T_1$, $T_2$ and $T_3$.

![Figure 1. Topology model of distribution station area](image)

According to the method in this paper, the two-scale distance between the station area and the user is shown in Table 1.

| V   | $U_{T1}$ | $U_1$ | $U_2$ | $U_3$ | $U_{T2}$ | $U_4$ | $U_5$ | $U_6$ | $U_{T3}$ | $U_7$ | $U_8$ | $U_9$ | $U_{10}$ |
|-----|----------|-------|-------|-------|----------|-------|-------|-------|----------|-------|-------|-------|---------|
| $U_{T1}$ | 0.00 | 0.21 | 1.03 | 1.38 | 1.28 | 1.29 | 1.29 | 1.32 | 2.02 | 1.93 | 1.95 | 1.97 | 1.96 |
| $U_1$ | 0.21 | 0.00 | 0.84 | 1.19 | 1.29 | 1.28 | 1.30 | 1.33 | 2.19 | 2.10 | 2.12 | 2.14 | 2.12 |
| $U_2$ | 1.03 | 0.84 | 0.00 | 0.38 | 1.78 | 1.68 | 1.75 | 1.78 | 2.93 | 2.86 | 2.88 | 2.88 | 2.88 |
| $U_3$ | 1.38 | 1.19 | 0.38 | 0.00 | 2.03 | 1.92 | 1.99 | 2.02 | 3.15 | 3.10 | 3.12 | 3.10 | 3.12 |
| $U_{T2}$ | 1.28 | 1.29 | 1.78 | 2.03 | 0.00 | 0.18 | 0.17 | 0.18 | 1.86 | 1.79 | 1.81 | 1.78 | 1.81 |
| $U_4$ | 1.29 | 1.28 | 1.68 | 1.92 | 0.18 | 0.00 | 0.17 | 0.21 | 2.02 | 1.94 | 1.97 | 1.94 | 1.97 |
| $U_5$ | 1.29 | 1.30 | 1.75 | 1.99 | 0.17 | 0.17 | 0.00 | 0.06 | 1.92 | 1.84 | 1.87 | 1.84 | 1.87 |
| $U_6$ | 1.32 | 1.33 | 1.78 | 2.02 | 0.18 | 0.21 | 0.06 | 0.00 | 1.89 | 1.82 | 1.84 | 1.82 | 1.84 |
| $U_{T3}$ | 2.02 | 2.19 | 2.93 | 3.15 | 1.86 | 2.02 | 1.92 | 1.89 | 0.00 | 0.13 | 0.10 | 0.13 | 0.10 |
| $U_7$ | 1.93 | 2.10 | 2.86 | 3.10 | 1.79 | 1.94 | 1.84 | 1.82 | 0.13 | 0.00 | 0.04 | 0.04 | 0.04 |
| $U_8$ | 1.95 | 2.12 | 2.88 | 3.12 | 1.81 | 1.97 | 1.87 | 1.84 | 0.10 | 0.04 | 0.00 | 0.16 | 0.01 |
| $U_9$ | 1.97 | 2.14 | 2.88 | 3.10 | 1.78 | 1.94 | 1.84 | 1.82 | 0.13 | 0.16 | 0.16 | 0.00 | 0.15 |
| $U_{10}$ | 1.96 | 2.12 | 2.88 | 3.12 | 1.81 | 1.97 | 1.87 | 1.84 | 0.10 | 0.04 | 0.01 | 0.15 | 0.00 |
It can be seen from Table 1 that the dual-scale distances between users 1, 2, and 3 and the station area T1 are smaller than the dual-scale distances between other users and the station area T1. Therefore, the clustering judges that users 1, 2, and 3 belong to the station area T1. In the same way, users 4, 5, and 6 belong to station area T2, and users 7, 8, 9, and 10 belong to station area T3. It is consistent with the original topology of the station area.

4.2 Engineering practical example
Collect the statistical tables of voltage penetration readings in a certain area, extract the distribution transformers of 3 neighboring stations and 30 users in each station, and use the algorithm in this paper to analyze the stations that 90 users belong to. According to the effective value of the three-phase voltage of the station substation, the station substation power of 3 station areas is obtained mean value. The algorithm results in this paper are compared with the actual results as shown in Table 2.

| Transformer | Number of users to be identified | Identify the exact number |
|-------------|---------------------------------|--------------------------|
| T1          | 30                              | 28                       |
| T2          | 30                              | 33                       |
| T3          | 30                              | 29                       |

It can be seen from Table 2 that the algorithm in this paper obtains that the household change results of 30 users to be identified in each of station area 1, station area 2 and station area 3 are consistent with the actual records, the accuracy rate is very high, and the engineering applicability is good.

5. Conclusion
In this paper, the current low-voltage station area has the chaotic relationship between households and changes, the manual identification consumes a long time period, the identification accuracy is low, and the input system cost is high. This paper proposes an identification method of household change relations in low-voltage stations based on spectral clustering algorithm of dual-scale similarity. First, according to the different similarities of the voltage curves of users in different stations, the Euclidean distance and cosine distance are considered to establish a two-scale spectral clustering similarity matrix from the distance and shape of the voltage curves; then based on the multi-factor similarity matrix, the voltage curves under the same station are analyzed. Finally, the proposed method is verified from two aspects of simulation and actual engineering examples. The results show that the proposed method has high recognition accuracy, low input cost and good application effect.

Acknowledgments
Supported by National Key Research and Development Program 2018YFB2100100

References
[1] Zhou Feng, Zhou Hui, Diao Yinglong. Development Ideas on Key Technologies of Ubiquitous Power Internet of Things Intelligent Sensing[J]. Proceedings of the Chinese Society of Electrical Engineering, 2020, 40(01): 70-82+375.
[2] Si Yufei, Tan Yanghong, Wang Yan, Kang Wenni, Liu Shan. Cloud-side collaborative structure model for power Internet of Things[J]. Proceedings of the Chinese Society of Electrical Engineering, 2020, 40(24): 7973-7979+8234.
[3] Lu Jun, Luan Wengpeng, Liu Riliang, Wang Peng, Lin Jiaying. Distribution Internet of Things Architecture Based on Comprehensive Perception and Software Definition[J]. Power System Technology, 2018, 42(10): 3108-3115.
[4] Peng Chuning, Luo Ranran, Wang Xiaodong. Research on the technology of supporting ubiquitous power Internet of things for a new generation of smart energy meters[J]. Electrical Measurement and Instrumentation, 2019, 56(15): 137-142.

[5] Tang Zeyang, Zhou Kunpeng, Cao Kan, Wan Li, Xin Junhui, Rao Yize. Verification method of line-to-line relationship based on distribution network operating data[J]. High Voltage Technology, 2018, 44(04): 1059-1068.

[6] Chen Zhaoan, Huang Chun, Zhang Zidan, Jiang Yaqun. A low-voltage distribution network station topology identification method based on T-type gray correlation and KNN algorithm[J]. Power System Protection and Control, 2021, 49(01): 163 -169.

[7] Yu He, Xia Shuibin, Wei Wei, Sun Bingyu, Dong Zhongzhong, Yan Lieqi. The application research of improved grey relational analysis in the intelligent recognition of station area[J]. Electronic Measurement Technology, 2019, 42(08): 32-35.

[8] Tang Donglai, Ni Pingbo, Zhang Jie, Liu Youbo, Li Yu, Zhang Qiang. Recognition method of household change relationship based on discrete Freche distance[J]. Automation of Electric Power Systems, 2021, 45(06): 223-230.

[9] Lian Zikuan. Identification of Topological Relations of Low Voltage Station Based on Data Drive[D]. Zhejiang University, 2020.

[10] Yang Zhichun, Shen Yu, Yang Fan, Le Jian, Su Lei, Lei Yang. Topology identification method of low-voltage distribution network based on data association analysis[J]. Electrical Measurement and Instrumentation, 2020, 57(18): 5-11 +35.

[11] X. Lai, M. Cao, S. Liu and C. Sun, "Low-voltage distribution network topology identification method based on characteristic current," 2021 6th Asia Conference on Power and Electrical Engineering (ACPEE), 2021, pp. 1233-1238, doi: 10.1109/ACPEE51499.2021.9437092.