Title: Topology Identification of Low-Voltage Distribution Network Based on Deep Convolutional Time-Series Clustering

Authors: Qingzhong Ni and Hui Jiang

Abstract: Accurate topology relationships of low-voltage distribution networks are important for distribution network management. However, the topological information in Geographic Information System (GIS) systems for low-voltage distribution networks is prone to errors such as omissions and false alarms, which can have a heavy impact on the effective management of the networks. In this study, a novel method for the identification of topology relationships, including the user-transformer relationship and the user-phase relationship, is proposed, which is based on Deep Convolutional Time-Series Clustering (DCTC) analysis. The proposed DCTC method fuses convolutional autoencoder and clustering layers to perform voltage feature representation and clustering in a low-dimensional feature space simultaneously. By jointly optimizing the clustering process via minimizing the sum of the reconstruction loss and clustering loss, the proposed method effectively identifies the network topology relationships. Analysis of examples shows that the proposed method is correct and effective.

Keywords: deep clustering; topology relationship; convolutional autoencoder

1. Introduction

As the end link of the power system, the low-voltage distribution network plays a crucial role in providing electricity to consumers. With the increasing integration of distributed energy resources into the power system, obtaining an accurate real-time topology structure of the distribution network becomes imperative [1–3]. The precise mapping of the distribution network’s topology, which includes the relationship between users and transformers, as well as the phase relationship of users, is essential for the effective management of low-voltage distribution networks. This mapping enables precise calculation of line loss and facilitates the resolution of three-phase imbalances [4,5]. However, inaccurate and untimely updates of the records regarding the topology of low-voltage distribution networks on the distribution system have occurred due to frequent municipal renovations over the years. Consequently, the operation and maintenance management of the distribution network faces significant challenges [6]. Therefore, it is necessary to identify the topology relationship of low-voltage distribution networks.

Existing methods for identifying the topology relationship of distribution networks primarily involve manual inspection, injection signals, and data analysis. The manual inspection method relies on personnel to conduct on-site line checks and manually record the topology relationship of a distribution network. However, this approach is inefficient and costly. The injection signal method determines the connection relationship of the topology by injecting a pulse current or carrier signal based on the reception results of the device [7,8]. Nevertheless, this method necessitates a substantial amount of terminal equipment, entails high investment costs, and poses operational and maintenance challenges. The data analysis method entails analyzing the spatial and temporal characteristics of electrical quantities to identify the topology relationship of low-voltage distribution networks automatically. However, this method requires a large amount of real-time electrical data, and data quality must be of high standards.
The increasing adoption of Advanced Metering Infrastructure (AMI) has been driven by the advancements in smart technology within the distribution system. This system enables real-time measurement of critical electrical information, including electric power, voltage, and current. Consequently, it establishes a comprehensive and up-to-date database that greatly facilitates the identification of topology relationships [9]. Therefore, the data analysis method has become an important research direction in topology relationship identification. The voltage characteristics exhibit a correlation with the topology of the low-voltage distribution network, implying that users connected to the same transformer and phase exhibit similar voltage profiles.

To examine the relationship between distribution network topology and voltage similarity, the primary methods employed for data analysis are clustering and correlation analysis. Clustering methods, which are unsupervised learning algorithms, are suitable for identifying the distribution network topology relationship as they do not depend on accurate data labels. Typically, conventional clustering methods for topological relationship identification involve a two-stage process that applies clustering after feature extraction. In reference [10], a phase relationship identification method based on Principal Component Analysis (PCA) dimension reduction and K-means clustering was proposed. Reference [11] proposed a phase relationship identification method that utilizes Non-negative Matrix Factorization (NMF) and label propagation. Furthermore, references [12,13] suggested a method for identifying distribution network topology relationships based on the t-SNE algorithm and unsupervised clustering; references [14,15] introduced a clustering method to identify distribution network topology relationships by constructing voltage features.

However, these methods face challenges regarding initial cluster center selection and interference caused by anomalous data. Additionally, traditional data dimension reduction is independent of the clustering task, which means that low-dimensional voltage features may not necessarily be suitable for the clustering task [16].

The correlation analysis method determines the connection relationship between the user and transformer or the user and phase by analyzing voltage correlation coefficients. Reference [17] proposed a Kalman filter and a Pearson correlation coefficient method for identifying the topology relationship of low-voltage distribution networks. Reference [18] proposed calibrating the topology of low-voltage distribution networks based on the Pearson correlation coefficient and the K-Nearest Neighbor (KNN) algorithm. Nevertheless, the correlation analysis method relies on thresholding the correlation coefficient, so the threshold needs to be adjusted when applied in different distribution networks. Additionally, the recognition accuracy is not high when voltage data are close [19].

In recent years, deep clustering has garnered considerable attention [20,21]. This algorithm has found widespread application in various fields [22,23]. By employing a deep autoencoder, the deep clustering algorithm extracts nonlinear features from high-dimensional data, thereby enabling one-stage clustering through joint optimization of the feature extraction and clustering processes. The pioneering work of reference [24] introduced the application of deep learning in time series clustering, yielding favorable clustering outcomes. Furthermore, Reference [25] presented an improved deep convolutional embedded clustering algorithm for enhancing the quality of image data clustering. Their approach utilizes a Convolutional AutoEncoder (CAE) to effectively learn meaningful features while preserving the local structure of the input images.

The voltage time series in the distribution network exhibits characteristics of large data volume and high dimensionality. Therefore, this study presents a novel framework called Deep Convolutional Time-Series Clustering (DCTC) for identifying the topology relationship of low-voltage distribution networks. The DCTC framework utilizes a deep neural network that incorporates a convolutional autoencoder and a joint optimization clustering model. The contributions and innovations of this study are summarized as follows:

(1) This study presents an end-to-end unsupervised clustering framework aimed at identifying the topology relationship in low-voltage distribution networks. This end-to-end framework can automatically extract low-dimensional features from voltage
data without the need for manual feature design. The algorithm integrates a two-stage process, where clustering is applied after voltage feature extraction, into a single model. By inputting a user voltage dataset, the user-phase and user-transformer connection relationship can be directly obtained.

2) The DCTC framework adopts a convolutional autoencoder to reduce the dimensionality of the input voltage data. Unlike linear dimensionality reduction methods such as PCA and NMF, the convolutional autoencoder performs nonlinear mapping to decrease the dimensionality of voltage time series. This nonlinear mapping effectively extracts key features from voltage time series, enhancing the accuracy of distribution network topology relationship identification.

3) The DCTC framework proposed in this study adopts a joint optimization of the convolutional autoencoder and clustering process by minimizing the sum of the reconstruction loss and clustering loss, which enables the feature extraction and clustering to enhance mutually. Consequently, it achieves high accuracy in identifying the distribution network topology relationship.

2. Problem Formulation

The low-voltage distribution network comprises various components such as lines, cables, distribution transformers, compensators, and other auxiliary facilities, enabling the efficient distribution of electricity to urban users. In China, the deployment of smart meters has extensively covered the power supply areas in most cities, thereby providing valuable data support for identifying the relationship between single-phase users and transformers, as well as the phase relationship, through power big data analysis. These smart meters collect data, including users’ electric energy, voltage amplitude, and active power, every 15 min, while the concentrator gathers power data from transformers. Power data are transmitted to the intelligent distribution terminal via High Power Line Communication (HPLC). At the intelligent distribution terminal, collected data are analyzed, computed, and subsequently transmitted to the main distribution station.

The low-voltage distribution network topology relationship mainly refers to the subordination relationship between users and transformers and the phase information belonging to users. The challenges encountered in the topology relationship of the low-voltage distribution network can be summarized as follows:

1) The topology relationship information of users is not recorded.
2) The topology relationship information of users is incorrectly registered to adjacent transformers and phase lines.

The 380 V low-voltage distribution network primarily operates in a radial configuration. In this structure, the direction of power flow is clearly defined so that the subordination between the user and transformer and the relationship of the phase sequence can be clarified [26]. Figure 1 depicts the daily voltage profiles of the three users connected to Transformer 1 (namely, Supply Area 1) and three users connected to Transformer 2 (namely, Supply Area 2), while Figure 2 shows the daily voltage profiles of the six users for a transformer with different phases.

Due to loading uncertainties, the voltage profiles of users connected to different distribution transformers exhibit certain differences. Conversely, the voltage profiles of users connected to the same distribution transformer tend to exhibit higher similarity, as demonstrated in Figure 1. Additionally, customers sharing the same phase at a given transformer display a notable resemblance in their voltage profiles, as depicted in Figure 2. By constructing a voltage time series and analyzing voltage similarity, it becomes possible to determine the transformer affiliation and phase allocation of the users.
In this study, we propose the DCTC framework to perform voltage time series clustering, as shown in Figure 3. DCTC consists of an autoencoder and clustering layer. DCTC extracts the low-dimensional feature representation in the voltage time series through an autoencoder, and the clustering layer can assign clusters to the low-dimensional feature representation of the voltage time series. Lastly, the reconstruction loss and the clustering loss are jointly optimized to achieve end-to-end clustering. The reconstruction loss is measured by Mean Squared Error (MSE), and the clustering loss is defined as Kullback–Leibler Divergence (KL Divergence) [27].

Figure 3. The structure of DCTC.

3.1. Autoencoder Structure

The autoencoder is mainly composed of an encoder and a decoder. The autoencoder obtains the low-dimensional feature representation of the high-dimensional input data through the encoder, and the low-dimensional feature representation is reconstructed by the decoder to be close to the input data. The convolutional autoencoder replaces the fully
connected layer with a convolutional layer to improve the performance of data feature extraction. Because the fully connected layer connects each node in the neural network to the nodes in the previous layer, there are a large number of parameters that need to be trained. Through weight sharing and sparse connection, the convolutional layer is able to reduce training parameters, reducing training time and improving the effectiveness of feature extraction [28].

Figure 3 shows the structure of the convolutional autoencoder. The encoder uses the superposition of two 1D convolutional layers and a dense layer to reduce the dimension of the voltage time series and extract the hierarchical voltage features. Different from the approach of adding a pooling layer after the convolutional layer to reduce the dimensionality of the data, DCTC increases the strides parameter of the convolutional layer in a way that converts voltage data into a more compact representation and improves the data conversion ability [25]. As shown in Figure 3, for the convolutional layer, the strides are set to 2.

Let \( \{u_i \in U\}_{i=1}^n \) be \( n \) voltage time series. At first, the voltage time series is input into the encoder. Through the computation of two 1D convolution layers and a dense layer, a low-dimensional feature representation of the voltage time series \( \{z_i \in Z\}_{i=1}^n \) can be obtained. Then the feature representation is input into the decoder for reconstructing the input data. Let \( f_W \) and \( g_W \) be the mappings of encoder and decoder, respectively, where \( W \) and \( W' \) are the corresponding parameter sets. As shown in Equation (1), Mean Square Error (MSE) is used as the reconstruction loss \( L_{rec} \). The parameters of the 1D convolutional autoencoder are continuously updated by minimizing the loss function. By training appropriate parameters, the mapping \( f_W \) produces the key features of the input voltage time series.

\[
L_{rec} = \frac{1}{n} \sum_{i=1}^{n} \| u_i - g_W(f_W(u_i)) \|^2 
\]  

(1)

### 3.2. Clustering Layer

The low dimensional voltage feature representation \( \{z_i \in Z\}_{i=1}^n \) extracted by the autoencoder is used as input to the clustering layer. Firstly, K-means clustering is used to initialize cluster centers \( w_k \) \((k = 1, 2, 3, \cdots, K)\), where \( K \) is the number of clusters. After that, we calculate the similarity of \( z_i \) and \( w_k \) and map \( z_i \) to the soft assignments \( q_{ik} \) by Student’s t-distribution [29], as shown in Equation (2).

\[
q_{ik} = \frac{(1 + \frac{\text{siml}(z_i, w_k)}{\alpha})^{-\frac{\alpha+1}{2}}}{\sum_{k=1}^{K} (1 + \frac{\text{siml}(z_i, w_k)}{\alpha})^{-\frac{\alpha+1}{2}}}
\]

(2)

where \( \alpha \) are the degrees of freedom and \( \text{siml}(z_i, w_k) \) denotes the similarity measure of \( z_i \) and \( w_k \). Here we set \( \alpha = 1 \).

As the low-dimensional feature of the voltage time series of the same phase with the same transformer tends to have similar overall profiles, they can be far apart in terms of geometric distance. Therefore, this study adopts the correlation-based distance [30] instead of the Euclidean distance to measure the similarity of \( z_i \) and \( w_k \), as shown in Equations (3) and (4).

\[
\text{corr}(z_i, w_k) = \frac{(z_i - \overline{z})(w_k - \overline{w})}{\|z_i - \overline{z}\| \|w_k - \overline{w}\|} 
\]

(3)

\[
CD(z_i, w_k) = \sqrt{1 - (\text{corr}(z_i, w_k))^2} 
\]

(4)

We set \( \overline{z} = (\overline{z_1}, \overline{z_2}, \cdots, \overline{z_l}) \) and \( \overline{w} = (\overline{w_1}, \overline{w_2}, \cdots, \overline{w_l}) \) where \( \overline{z} \) and \( \overline{w} \) denote the average value of \( z_i \) and \( w_k \), respectively.
In order to normalize the loss contribution of each centroid and put more emphasis on data points assigned with high confidence [20], the auxiliary target distribution $p_{ik}$ is defined by Equation (5):

$$p_{ik} = \frac{q_{ik}^2 / f_k}{\sum_{k=1}^{K} q_{ik}^2 / f_k} \quad (5)$$

where $f_k = \sum_{i=1}^{n} q_{ik}$ denotes the probability of all low-dimensional feature vectors being assigned cluster centers.

3.3. Clustering Loss and Joint Optimization

The structure of the decoder is retained after the feature extraction of the 1D convolutional autoencoder [25], which can avoid distortion of the low-dimensional feature space [21]. To ensure the coherence between the soft assignments $q_{ik}$ and the auxiliary target distribution $p_{ik}$, this study defines KL Divergence as the clustering loss as shown in Equation (6):

$$L_{clu} = \sum_{i=1}^{n} \sum_{j=1}^{k} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (6)$$

As shown in Equation (7), DCTC jointly optimizes the reconstruction loss of the convolutional autoencoder and the clustering loss to improve the accuracy of the clustering assignment.

$$L_{DCTC} = L_{rec} + \gamma L_{clu} \quad (7)$$

where $\gamma$ is a coefficient that controls the degree of distorting embedded space.

Firstly, the parameters of a convolutional autoencoder are pre-trained to generate valid feature representations of the input voltage time series, and the cluster centers are initialized. The soft assignments and auxiliary target distribution are computed, and the joint optimization of the clustering layer and convolutional autoencoder is performed by minimizing Equation (7). The autoencoder weights and cluster centers are updated using the Adaptive Moment Estimate (Adam) and backpropagation. Furthermore, the auxiliary target distribution is updated during every Adam update using Equation (5).

4. Low-Voltage Distribution Network Topology Relationship Identification

In this study, we adopt the DCTC clustering algorithm for user-transformer relationship identification and phase identification.

4.1. Data Normalization

Currently, the intelligent distribution transformer terminal collects voltage data of transformers and users every 15 min and uses these data to complete the identification of the transformer and phase to which the low-voltage users belong. The voltage is uploaded to the meter data management system through the communication network. During the collection and download process, problems such as meter faults and changes in the transmission network environment can occur, which can lead to some missing voltage–time data points. This study uses linear interpolation for filling voltage data. After filling in the missing values, the voltage matrix $\mathbf{U}$ fills in the missing values as follows.

$$\mathbf{U} = \begin{bmatrix} \pi_{1,1} & \pi_{1,2} & \cdots & \pi_{1,d} \\ \pi_{2,1} & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \pi_{n,1} & \cdots & \cdots & \pi_{n,d} \end{bmatrix} \quad (8)$$

where $n$ represents the number of voltage data samples, $d$ represents the number of voltage data nodes, and data are collected once every 15 min, the number of nodes in a day $d = 96$. To improve the accuracy and training speed of the model, it is necessary to normalize the
voltage data. The formula used in this study is maximum-minimum normalization, as shown in Equations (9) and (10):

\[ u_{ij} = \frac{\bar{u}_{ij} - \min(\bar{u}_j)}{\max(\bar{u}_j) - \min(\bar{u}_j)} \quad i = 1, 2, 3, \ldots, n \quad j = 1, 2, 3, \ldots, d \] (10)

where \( \bar{u}_j \) denotes voltage data from all users at time \( j \), \( \max(\bar{u}_j) \) and \( \min(\bar{u}_j) \) denote the maximum and minimum value of all users at time \( j \).

4.2. Evaluation Criteria

We use cluster Accuracy (ACC), Normalized Mutual Information (NMI), and Adjusted Rand Index (ARI) as our measure of clustering performance. These three evaluation indicators are defined as follows.

4.2.1. Accuracy (ACC)

Accuracy is used to compare the clustered labels with the true labels of the data. This provides a visual indication of the proportion of samples correctly assigned out of the total sample. Its definition is as follows:

\[ ACC = \max_m \sum_{i=1}^{N} \frac{1}{N} \{ r_i = map(s_i) \} \] (11)

where \( s_i \) is the cluster assignment label, \( r_i \) is the true label, and \( N \) is the number of users, \( map(s_i) \) as a mapping function containing all possible one-to-one mappings between the clustered label and the true label.

4.2.2. Normalized Mutual Information (NMI)

Normalized Mutual Information is used to measure the amount of information contained in a cluster result about another cluster result. Its range of values is \([0,1]\), and it is defined as follows:

\[ NMI = \frac{2MI(r,s)}{H(r) + H(s)} \] (12)

where \( r \) and \( s \) are the true label and the cluster assignment label, respectively, \( H() \) represents the information entropy [31] and \( MI() \) represents the mutual information [32] of the cluster assignment label and the true label.

4.2.3. Adjusted Rand Index (ARI)

Adjusted Rand index is an improvement of RI (rand index) [33]. RI indicates the percentage of all users whose clustered labels are exactly the same as the real labels. Adjusted Rand Index is defined as follows.

\[ ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]} \] (13)

where \( E[RI] \) denotes the mathematical expectation of RI and \( \max(RI) \) denotes the maximum value of RI.

Accuracy and Normalized Mutual Information take values in the range \([0,1]\). Larger values indicate a better clustering performance. Adjusted Rand Index takes values within \([-1,1]\), 1 indicates the best clustering performance, and -1 indicates the worst clustering performance.

4.3. Topology Relationship Identification Algorithm

The user-transformer relationship identification based on DCTC is summarized in Algorithm 1, and the phase identification follows a similar approach. The topology identifi-
cation of the low-voltage distribution network identification process is as follows: First, we set the number of clusters to the number of transformers and then perform Algorithm 1 to obtain the cluster labels and analysis the user-transformer relationship. For the cluster containing the voltage of Transformer 1, the users are classified as Transformer 1, etc. Next, we select voltage data for phase identification, set the number of clusters to three, and perform Algorithm 1. For the class containing phase A voltage, the users’ phase is determined to be phase A, etc. Finally, the complete user-transformer relationship and phase relationship of each user is obtained.

**Algorithm 1 User-transformer relationship identification Based on DCTC**

**Input:** Voltage dataset: \( \mathbf{U} \); Number of transformers: \( K \); Maximum iterations: \( MaxIter \); Stopping threshold: \( \delta \);

**Output:** Cluster labels: \( s \)

1. Obtain the normalized voltage matrix \( \mathbf{U} \) according to Section 4.1.
2. Pretrain the parameters of convolutional autoencoder in Equation (1).
3. Initialize cluster centers.
4. for \( \text{iter} \in \{1, 2, 3, \ldots, MaxIter\} \) do:
   5. Compute the low dimensional voltage feature \( z_i \).
   6. Compute \( q_{ik} \) and \( p_{ik} \) using Equations (2) and (5).
   7. Optimize Equation (7) by using Adam and backpropagation.
   8. if the cluster assignment change between successive iterations < \( \delta \) then
   9. Stop training.
11. end if
11. end for

5. Results

This study takes the low-voltage distribution network of a province in the east of China in 2018 as an example to identify the topology relationship. This distribution network contains four three-phase transformers and 312 single-phase voltage meters at user sides. Three single-phase voltage meters are fixed at each transformer’s low voltage side, so the voltage dataset size is 324. Voltage data were collected at a frequency of 15 min for a total of 96 data samples in one day. The user’s actual phase information of four transformers is shown in Table 1. Furthermore, the network of the feeder connected to Transformer 4 is shown in Figure 4.

<table>
<thead>
<tr>
<th>Transformer</th>
<th>Serial Number</th>
<th>Number of Users in Phase A</th>
<th>Number of Users in Phase B</th>
<th>Number of Users in Phase C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>59</td>
<td>38</td>
<td>57</td>
<td>57</td>
<td>154</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>45</td>
<td>38</td>
<td>38</td>
<td>118</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>7</td>
<td>16</td>
</tr>
</tbody>
</table>

**Figure 4.** The network of the feeder connected to Transformer 4.
Table 1. The user’s actual phase information of 4 transformers.

<table>
<thead>
<tr>
<th>Transformer Serial Number</th>
<th>Number of Users in Phase A</th>
<th>Number of Users in Phase B</th>
<th>Number of Users in Phase C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>59</td>
<td>38</td>
<td>57</td>
<td>154</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>45</td>
<td>38</td>
<td>118</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>16</td>
</tr>
</tbody>
</table>

5.1. Parameter Setting

The parameter settings of the DCTC model are shown in Table 2. The encoder and decoder are mirrored to generate a 24 dimensional voltage feature. Firstly, the autoencoder networks are pre-trained 100 times with a batch size of 32. The coefficient of the clustering loss \( \gamma \) is set to 0.1 by experiments. The maximum number of iterations is set to 200. The training is stopped when the clustering result meets the discriminative criterion, i.e., less than 0.1% of the cluster assignments are unchanged for five consecutive iterations.

Table 2. DCTC parameters and structure.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Layer</th>
<th>Parameter</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encoder</td>
<td>Input</td>
<td>—</td>
<td>(None, 96, 1)</td>
</tr>
<tr>
<td></td>
<td>Conv1D</td>
<td>Filters = 50, Kernel size = 10, strides = 2</td>
<td>(None, 48, 50)</td>
</tr>
<tr>
<td></td>
<td>Conv1D</td>
<td>Filters = 50, Kernel size = 10, strides = 2</td>
<td>(None, 24, 50)</td>
</tr>
<tr>
<td></td>
<td>Dense</td>
<td>Units = 1</td>
<td>(None, 24, 1)</td>
</tr>
<tr>
<td></td>
<td>Dense</td>
<td>Units = 50</td>
<td>(None, 24, 50)</td>
</tr>
<tr>
<td>Decoder</td>
<td>Conv1DTranspose</td>
<td>Filters = 50, Kernel size = 10, strides = 2</td>
<td>(None, 48, 50)</td>
</tr>
<tr>
<td></td>
<td>Conv1DTranspose</td>
<td>Filters = 1, Kernel size = 10, strides = 2</td>
<td>(None, 96, 1)</td>
</tr>
</tbody>
</table>

5.2. User-Transformer Relationship Identification

To evaluate the effectiveness of the proposed method in identifying user-transformer relationships, it was compared with the correlation analysis method [17]. We compared the accuracy of the correlation analysis method with the method proposed in this study.

User-transformer relationship identification has low data volume requirements, and it is possible to select the voltage measurements in half a day for analysis. For the purposes of this study, the 48 point voltage measurement data of users from 00:00 to 12:00 on the same day are selected for the user-transformer relationship identification. As shown in Table 3, the overall user-transformer relationship identification accuracy of the correlation analysis method is 93.8%. For Transformer 1, seventeen users were incorrectly assigned to Transformer 2, while two users were incorrectly assigned to Transformer 4, and one user was assigned to Transformer 3. The identification method proposed in this study achieves an accuracy of 100% for the user-transformer relationship identification in this example. Based on the above results, it can be inferred that the method proposed in this study has significant advantages over the traditional correlation analysis method in identifying the user-transformer relationship.

Table 3. The accuracy (%) of user-transformer identification under different methods.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Transformer 1</th>
<th>Transformer 2</th>
<th>Transformer 3</th>
<th>Transformer 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation Analysis</td>
<td>87.3%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>93.8%</td>
</tr>
<tr>
<td>DCTC method</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
5.3. Phase Relationship Identification

To demonstrate the efficacy of the method presented in this study for phase identification, we compared DCTC against several existing algorithms, namely the PCA combined K-means clustering method proposed in [10], the autoencoder (AE) combined K-Means clustering algorithm proposed, and DTC proposed in [24]. We use the clustering evaluation criteria defined in Section 4.2 to assess the performance of phase relationship identification.

In the AE combined K-Means algorithm, the parameters of the autoencoder are the same as that of the DCTC proposed in this study. By applying PCA and AE for dimensionality reduction, the input voltage dataset was reduced from its original 96 dimensions down to 24 dimensions. Euclidean distance is used as a similarity measure of the K-means clustering algorithm, and Euclidean distance and correlation distance are used as the similarity measure of the DTC and DCTC.

Two feeders connected to Transformers 1 and 2, respectively, with a larger number of users, are chosen for phase identification. Tables 4 and 5 show the performance of these four methods. The results are analyzed from the following three perspectives:

(1) Compared with the PCA combined K-means clustering method, the ACC values, the NMI values, and the ARI values of AE combined K-means clustering method was improved, respectively, by 1.3%, 3.7%, and 4.9% for Transformer 1 and 18.2%, 36.5%, 37.6% for Transformer 2. It can be seen that the clustering effect of AE is better than that of PCA dimension reduction under the K-means clustering method. The reason for this is that the convolutional autoencoder can learn the nonlinear relationship of voltage data and has a stronger feature extraction ability.

(2) Compared to the DTC algorithm based on Euclidean Distance, the ACC values, the NMI values, and the ARI values of the DCTC algorithm were improved, respectively, by 3.9%, 8.4%, 11% for Transformer 1, and 7.5%, 24.9%, 20% for Transformer 2. Compared to the DTC algorithm based on Correlation-Based Distance, the ACC values, the NMI values, and the ARI values of the DTC algorithm were improved, respectively, by 2.6%, 8.2%, 6.7% for Transformer 1 and 0.8%, 3.5%, 2.3% for Transformer 2. It can be seen that the DCTC algorithm has a better clustering effect. This is due to the fact that long-term memory neural networks (LSTMs) increase the computational cost and lead to slow training speed, and LSTMs may not be appropriate for an application on voltage data.

(3) In comparison to the DCTC algorithm based on Euclidean Distance, the ACC values, the NMI values, and the ARI values of the DCTC algorithm based on Correlation-Based Distance can reach 100%, 100%, and 100%, respectively, for Transformer 1 and Transformer 2. This is because the Correlation-Based Distance can take into account the overall profile of low-dimensional features, and the similarity of low-dimensional features can be measured more effectively.

Table 4. Clustering evaluation criteria of user-phase identification for Transformer 1.

<table>
<thead>
<tr>
<th>Method</th>
<th>Euclidean Distance</th>
<th>Correlation-Based Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACC</td>
<td>NMI</td>
</tr>
<tr>
<td>PCA + K-means</td>
<td>0.968</td>
<td>0.885</td>
</tr>
<tr>
<td>AE + K-means</td>
<td>0.981</td>
<td>0.922</td>
</tr>
<tr>
<td>DTC</td>
<td>0.923</td>
<td>0.780</td>
</tr>
<tr>
<td>DCTC</td>
<td>0.962</td>
<td>0.864</td>
</tr>
</tbody>
</table>

In conclusion, the DCTC algorithm proposed in this study for identifying the topology relationship of low-voltage distribution networks demonstrates superior performance compared to other dimensionality reduction and clustering combination methods.
Table 5. Clustering evaluation criteria of user-phase identification for Transformer 2.

<table>
<thead>
<tr>
<th>Method</th>
<th>Euclidean Distance</th>
<th>Correlation-Based Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACC</td>
<td>NMI</td>
</tr>
<tr>
<td>PCA + K-means</td>
<td>0.818</td>
<td>0.635</td>
</tr>
<tr>
<td>AE + K-means</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>DTC</td>
<td>0.785</td>
<td>0.452</td>
</tr>
<tr>
<td>DCTC</td>
<td>0.86</td>
<td>0.701</td>
</tr>
</tbody>
</table>

6. Conclusions

In this study, a method for identifying the topology relationship of low-voltage distribution networks based on the DCTC algorithm is proposed. The DCTC algorithm uses a convolutional autoencoder to extract low dimensional voltage features from high dimensional meter voltage data and joint train reconstruction loss and clustering loss to improve the clustering effect. In comparative studies, the DCTC algorithm exhibits superior recognition accuracy compared to alternative approaches for identifying the topology relationship of the distribution network.

However, it is important to acknowledge that the proposed method in this study has certain limitations and shortcomings. This method is limited to identifying single-phase smart meters only, and it is not suitable for the identification of three-phase loads, which is the focus of future research. Furthermore, our future work aims to extend the application of the DCTC algorithm to other time series datasets in order to address challenges encountered across various domains.

Author Contributions: Conceptualization, H.J.; methodology, H.J.; software, Q.N.; validation, Q.N.; formal analysis, Q.N.; investigation, Q.N.; resources, Q.N.; data curation, Q.N.; writing—original draft preparation, Q.N.; writing—review and editing, Q.N.; visualization, Q.N.; supervision, Q.N.; project administration, H.J.; funding acquisition, H.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Shenzhen Science and Technology Innovation Foundation under Grant JCYJ20190808165201648.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. These data are not publicly available because the user load data in this article is suspected to be commercially confidential.

Conflicts of Interest: The authors declare no conflict of interest.

References
2. Liu, G.; Ferrari, M.F.; Ollis, T.B.; Sundararajan, A.; Olama, M.; Chen, Y. Distributed Energy Management for Networked Microgrids with Hardware-in-the-Loop Validation. *Energies* 2023, 16, 3014. [CrossRef]


