

# High-resolution Load Profile Clustering Approach Based on Dynamic Largest Triangle Three Buckets and Multiscale Dynamic Warping Path Under Limited Warping Path Length

Mi Wen, Yue Ma, Weina Zhang, Yingjie Tian, and Yanfei Wang

**Abstract**—With the popularity of smart meters and the growing availability of high-resolution load data, the research on the dynamics of electricity consumption at finely resolved time-scales has become increasingly popular. Many existing algorithms underperform when clustering load profiles contain a large number of feature points. In addition, it is difficult to accurately describe the similarity of profile shapes when load sequences have large fluctuations, leading to inaccurate clustering results. To this end, this paper proposes a high-resolution load profile clustering approach based on dynamic largest triangle three buckets (LTTBs) and multiscale dynamic time warping under limited warping path length (LDTW). Dynamic LTTB is a novel dimensionality reduction algorithm based on LTTB. New sequences are constructed by dynamically dividing the intervals of significant feature points. The extraction of fluctuation characteristics is optimized. New curves with more concentrated features will be applied to the subsequent clustering. The proposed multiscale LDTW is used to generate a similarity matrix for spectral clustering, providing a more comprehensive and flexible matching method to characterize the similarity of load profiles. Thus, the clustering effect of a high-resolution load profile is improved. The proposed approach has been applied to multiple datasets. Experiment results demonstrate that the proposed approach significantly improves the Davies-Bouldin indicator (DBI) and validity index (VI). Therefore, better similarity and accuracy can be achieved using high-resolution load profile clustering.

**Index Terms**—Load profile clustering, largest triangle three buckets (LTTB), dynamic time warping (DTW), spectral clustering.

Manuscript received: July 1, 2022; revised: September 20, 2022; accepted: October 22, 2022. Date of CrossCheck: October 22, 2022. Date of online publication: November 28, 2022.

This work was supported by the Joint Fund of National Natural Science Foundation of China (No. U1936213), National Natural Science Foundation of China (No. 61872230), Program of Shanghai Academic Research Leader (No. 21XD1421500), and Shanghai Science and Technology Commission Project (No. 20020500600).

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

M. Wen, Y. Ma (corresponding author), W. Zhang, and Y. Wang are with the College of Computer Science and Technology, Shanghai University of Electric Power, Shanghai, China (e-mail: miwen@shiep.edu.cn; oppositem@mail.shiep.edu.cn; mszhangwn@shiep.edu.cn; shuangerwawai@mail.shiep.edu.cn).

Y. Tian is with the Shanghai Electric Power Research Institute, Shanghai, China (e-mail: 13901712348@163.com).

DOI: 10.35833/MPCE.2022.000386

## I. INTRODUCTION

WITH the development of smart grids and advanced metering infrastructure, a vast amount of fine-grained electricity consumption data has been generated [1]. Mining the electricity consumption behavior of consumers and analyzing the potential connections among power consumption data allow the distribution of loads in regions to be clearly ascertained and improve the deployment efficiency of the power grid [2]. Merging consumers with similar behaviors is an efficient way to explore the typical electricity consumption behavior of different types of consumers [3]. The high-resolution load profiles collected by smart meters contain more information but present a high degree of temporal variability [4], which leads to indeterminate peaks and valleys in the power consumption curves. In recent years, effectively clustering high-resolution load profiles and extracting typical power consumption behaviors have become significant challenges and an extremely popular research area [5].

Numerous methods, including  $K$ -medoids [6],  $K$ -means [7], fuzzy C-means [8], and spectral clustering [9], [10], have been used for load profile clustering. However, the popular advanced monitoring systems in power distribution networks provide more details related to the power consumption behavior and complicate the load profile analysis. Many small fluctuations will be recorded in a load curve sampled at a high frequency. These data floating under a stable average value make it difficult to describe the user's power consumption behavior and will increase the calculation cost. Most cluster analyses use the Euclidean distance as the basis for evaluating the similarity, ignoring consumption behaviors that might be offset on the timeline. Few studies have utilized the shapes of the load profiles as significant properties for clustering and classification.

To retain the major information related to electricity usage activities and improve computational efficiency, some studies have utilized different dimensionality reduction technologies to reduce the dimensions of the load data such as piecewise aggregate approximation (PAA) [9] and discrete wavelet transform (DWT) [11]. These methods transform the original data to extract more compact features. For a curve that has a large peak-valley difference over a short time period,

the result after dimensionality reduction might be distorted. The largest triangle three buckets (LTTB) algorithm [12] was proposed by Steinarrsson for the first time as an effective downsampling dimensionality reduction algorithm. It maximizes the shape similarity to the original data and guarantees that the samples contained in the output are present in the original input. In [13], density-based spatial clustering of applications with noise (DBSCAN) was used for load profile clustering after using LTTB algorithm to reduce the number of dimensions. The LTTB algorithm can represent the original sequence in most cases. Nevertheless, not all of the buckets can be visually represented with only one point in a static equal division. Further, the largest triangle dynamic (LTD) algorithm was proposed to extract feature data points from rapidly changing irregular data. It can adjust the bucket length to adapt to fluctuations in the curve, but the unlimited assignment of regions is prone to distort the time axis. In addition, this algorithm has never been applied to load profile analysis.

Dynamic time warping (DTW) can realign the matching relationship between data points to alleviate the problem of time drift [14]. Therefore, DTW has been used in cluster analyses of load profiles [15], [16] to provide a more accurate evaluation of the similarity based on the shape of a curve. Reference [10] used improved fast DTW to calculate coarse-grained data to form a similarity matrix of spectral clustering. Although the calculation efficiency was improved, the accuracy was decreased. Traditional DTW may be prone to causing pathological alignment. A data point may be sufficiently close to match a large subsection of the other time series and affect the assessment of the similarity. Existing studies have effectively improved DTW and suppressed pathological alignments. The LimitDTW algorithm was proposed in [17] to restrict the search area to the regions close to the diagonal area of the distance matrix. However, limiting the number of connection points may cause incorrect alignment owing to its rigidity. Reference [18] proposed a fast derivative dynamic time warping (FDDTW) algorithm that works on the derivative of the raw data to speed up the calculation of the elastic dissimilarity and precisely reveal the load shape features, but the similarity of the curves can be evaluated by indicators other than the trend. Reference [19] proposed a method to limit the step size of a generated path to provide a flexible matching relationship. However, the maximum allowable step size cannot be consistently determined using this method.

To address these problems, a high-resolution load profile clustering approach based on dynamic LTTB and a multiscale dynamic time warping under limited warping path length (LDTW) is proposed in this paper. The proposed approach reduces the dimensions of high-resolution load profiles using the dynamic LTTB and obtains a new representative sequence composed of the original data. It can dynamically select the region where the appropriate feature points are located according to the degree of fluctuation in the curve. The use of raw data to form a new sequence ensures that the characteristics will not be flattened because of vio-

lent fluctuations over a short time period. Subsequently, the multiscale LDTW is used as a metric to describe the similarity of sequences with regards to shape and value. The method of limiting the overall step size is used to determine the matching relationship of the data points in multiscale LDTW, suppressing pathological alignments without limiting the matching relationship of each point. The similarity matrix of spectral clustering is created accordingly. Finally, spectral clustering improves the similarity of the load profiles in the same cluster and increases the accuracy of load profiling.

The main contributions of this paper are as follows.

1) An improved LTTB algorithm is proposed to address the high volatility of high-resolution load profiles. A dynamic method for determining a region of representative points in the load profiles is provided, and the new sequence is composed of the original data to reduce the possibility of distortion. It can overcome the fixed bucket limitation and maintain a balance between the retention of key information and the computational cost.

2) Multiscale LDTW is proposed to construct a similarity matrix for spectral clustering. The determination of the matching relationship between data points by limiting the overall step size is applied to load profile clustering for the first time. This flexible method suppresses pathological alignment. The discrimination ability of the shape similarity is enhanced, and the accuracy of high-dimensional load profile clustering is improved.

3) An approach to high-resolution load profile clustering is proposed, which improves the effect of clustering and provides a more accurate description of the power consumption behavior.

## II. PROPOSED APPROACH

The main process of load profile clustering includes five steps: input raw data to generate load profiles, preprocess the raw load profiles, reduce the dimensionality of the load profiles using dynamic LTTB, cluster load profile using multiscale LDTW, and extract typical load profiles.

### A. Data Preprocessing

We regard the load data that fall outside three standard deviations of the mean as outlier data and fill in missing data using the  $k$ -nearest neighbor (KNN) algorithm. We normalize the data using the min-max method and convert data at different levels into a unified measure. The processed data are arranged to form the load profiles for clustering.

Assuming that the original data  $X = \{x_1, x_2, \dots, x_n\}$  have been screened for outliers and that the missing values are filled in, we convert the original data into a new sequence  $X' = \{x_{1'}, x_{2'}, \dots, x_{n'}\}$ . The  $i^{\text{th}}$  datum  $x_{i'}$  is calculated using (1).

$$x_{i'} = \frac{x_i - x_{i,\min}}{x_{i,\max} - x_{i,\min}} \quad i = 1, 2, \dots, n \quad (1)$$

where  $x_{i'}$  is the  $i^{\text{th}}$  record after normalization using the extreme values; and  $x_{i,\min}$  and  $x_{i,\max}$  are the minimum and maximum records, respectively.

### B. Dynamic LTTB

The proposed dynamic LTTB has a dynamic bucket size that is different from LTTB with equal-sized buckets. Inspired by the LTD algorithm, dynamic LTTB reassigns each bucket to capture representative feature points. If the fluctuation in a bucket is considerable, it is likely that the data points are better represented as two buckets. Two adjacent buckets can be merged if the fluctuation is low. The steps of the proposed method are as follows.

*Step 1:* separate all the data points into roughly equal-sized buckets, with the first and last data points from the original data serving as the first and last buckets, respectively.

*Step 2:* evaluate the fluctuation in the data in each bucket by the variance and adjust the sizes of the buckets accordingly. Select the highest-ranking bucket and divide it into two equal-sized buckets so that an extra bucket can be created. We maintain the total number of iterative buckets by locating and combining an adjacent bucket pair with the lowest total variance sum. Meanwhile, we focus on limiting the size of a single bucket. Taking division and merging as one operation, we obtain the final bucket distribution through iteration. The number of iterations for adjusting the bucket size is calculated using (2).

$$N = \frac{n}{T \times 10} \quad (2)$$

where  $N$  is the number of iterations;  $n$  is the original count of data points; and  $T$  is the downsampling threshold.

To guarantee that a bucket does not cause an excessive time shift after redivision, which will have a negative effect on the description of the load curve characteristics, we need to limit the size of a single bucket. Let the maximum length of the barrel be  $m$  times the original length and the minimum length be  $n$  times the original length. The range of the limit length can be adjusted according to different datasets. We select different values of  $m$  and  $n$  for multiple power load datasets having different lengths for the experiments. During the experiments, we need to select different values of  $m$  and  $n$  to limit the splitting of the buckets, and we compare the effects of dimensionality reduction for different values of  $m$  and  $n$  to determine the appropriate values.

*Step 3:* go through all the buckets and choose the representative points from each. The first and last buckets have only one point that is selected by default. The dynamic LTTB moves from left to right, working with three buckets at a time. The first point in the left corner of the triangle is always fixed as the previously selected point. We select the midpoint of the third bucket as a temporary point. Thus, the dynamic LTTB has two fixed points. The remaining point to be determined is the intermediate data point. We use the effective area (EA) of the data point of the current bucket to determine the most representative point. The EA of a point is the area of a triangle formed by its two adjacent points. The construction of the largest triangle across the buckets is depicted in Fig. 1 using the previously selected points  $A$  and  $B$  and a temporary point  $C$ . Each circle represents a data point, and the blue circles are the selected representative data points.

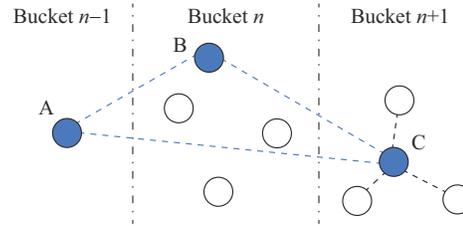


Fig. 1. Largest triangle formed by three adjacent buckets.

By adjusting the buckets of the original data and restricting bucket segmentation dynamically, representative data points can be extracted more effectively, and the interval for selecting the feature data can be selectively controlled when the curves have different fluctuations.

### C. Spectral Clustering Based on Multiscale LDTW

The description of the similarity between curves and the choice of clustering algorithm are the important components of the power load profile clustering analysis. The DTW algorithm is frequently used to compare similarities across time series and has a positive impact on curve comparisons. On this basis, the proposed multiscale LDTW offers a more adaptable and multiangle data point matching relationship. Spectral clustering has been widely used because of its excellent clustering effects. This paper proposes spectral clustering based on multiscale LDTW to improve clustering outcomes. It employs multiscale LDTW as a similarity measure to produce a similarity matrix for spectral clustering.

#### 1) Multiscale LDTW

The proposed multiscale LDTW inherits the concept of DTW with a limited warping path length. In the matching process, the soft restriction method of limiting the number of steps in the entire series matching process is used to generate an evaluation index for clustering. On this basis, the original distance measurement method is improved. Moreover, a method for determining the step length for the restriction of series with different characteristics is proposed to balance the topological and direct alignments of the series.

The main idea of multiscale LDTW is to determine the step size of a path to provide more flexible restriction for data point matching and find the optimal matching path. We can observe the process in reverse. Let the total step size of a path be  $S$ , the current step size in the matching process be  $s$ , and the corresponding distance be  $l$ . The last point of the two series must match, and  $S=s$  at this time.  $S=s-1$  must originate from the left, lower, or lower left side according to DTW rules. The previous step cannot be used if the step size for these three positions does not satisfy  $S=s-1$ . An additional dimension  $s$  is added to determine whether the possible step size of the previous data point conforms to the total step size  $S$ . When filling the distance matrix  $Dist$ , each position  $Dist[i,j,s]$  needs to include the path length corresponding to all possible path lengths, and the minimum is selected in this step.

Therefore, we can improve the calculation process of DTW and obtain:

$$Dist[i,j,s]=\begin{cases} 0 & i=0,j=0 \\ \infty & i=0,j>0 \text{ or } i>0,j=0 \\ dist(p_i,q_j)+\min\{Dist[i-1,j-1,s-1], \\ Dist[i-1,j,s-1],Dist[i,j-1,s-1]\} & i>0,j>0 \end{cases} \quad (3)$$

where  $i$  and  $j$  represent the positions of the data points in the series  $P$  and  $Q$ , respectively; and  $dist(p_i,q_j)$  is the distance between the data points  $p_i$  and  $q_j$ .

The numerical values and derivatives of the series measure the characteristics of the series from different perspectives. The difference in the numerical values indicates the difference between the series themselves, whereas the derivative more closely reflects whether the changes in the trends of the series are similar. In this study, we use a combination of numerical and derivative differences as the distance measure and provide adjustable weights. An element of the distance matrix is calculated as:

$$dist(x,y)=\alpha d_E(x,y)+\beta d_D(x,y) \quad (4)$$

where  $dist(x,y)$  is the distance metric obtained after weighting;  $d_E(x,y)$  and  $d_D(x,y)$  are the numerical and derivative differences of each corresponding data point of the two series, respectively; and  $\alpha$  and  $\beta$  are the corresponding weights ( $\alpha+\beta=1$ ).

The weights can be adjusted according to the characteristics of different time series. To calculate the derivative of a data point, we use the following approximation:

$$Dev(p_i)=\frac{(p_i-p_{i-1})+(p_{i+1}-p_i)/2}{2} \quad (5)$$

The use of the total step size as the limiting condition can improve flexibility during the matching process, but there is a problem of determining a limited step size for the profiles. In this study, the standard deviation is used to measure the fluctuation in the sequences in the datasets, and the limited step size  $L$  is determined accordingly.

The proposed approach determines whether it is necessary to provide a more relaxed step size limitation by measuring the fluctuation in the data points compared with the points at the same position. Let the original step size be the length of the series. The standard deviation of the points at the same position is calculated and recorded, and the difference between the two paired points is compared with the standard deviation of the position. When the difference is small, the fluctuation is not sufficient to increase the step size. Otherwise, the original step size is increased by 1. Then, we iterate over the data points at each location to obtain the final step size  $L$ .

The calculation process for multiscale LDTW is summarized in Algorithm 1.

## 2) Improved Spectral Clustering

Spectral clustering is becoming increasingly popular because of its excellent performance for time-series clustering. The fundamental concept behind spectral clustering is the conversion of all of the power data into points in space. The edge that links the weights of these points between two spots separated by a long distance is lower.

---

### Algorithm 1: calculation process for multiscale LDTW

---

- Input:** sequences  $A,B$ ; limited step size  $L$ ; distance matrix  $Dist$  formed by each combination of data points between the sequences calculated by (4)
- Output:** distance between sequences  $A$  and  $B$  calculated by multiscale LDTW
1. Let  $N$  be the length of the sequence
  2. Let  $D[p,q,s]=(diff,(p',q'))$  be the matrix recording the cumulative distances computed during the matching process.  $p$  and  $q$  represent the positions of the data points matched by the current series  $A$  and  $B$ , respectively;  $diff$  records the accumulated distance to the current position; and  $(p',q')$  is used to record the matching situation of the previous step
  3. Initialize  $D[1,1,0]=(Dist[1,1,0],(0,0))$
  4. Calculate  $D$  when  $p=1$  or  $q=1$  using (3)
  5. **for**  $p=1$  to  $N$  **do**
  6.   **for**  $q=1$  to  $N$  **do**
  7.     **for**  $s=\min(p,q)$  to  $2N-1$  **do**
  8.       Calculate  $D[p+1,q+1,s][0]$  by (3) and record the coordinate corresponding to the minimum value using  $D[p+1,q+1,s][1]$
  9.     **end for**
  10.   **end for**
  11. **end for**
  12.  $MinStep=N, MaxStep=L$
  13. The smallest  $D[N,N,s][0]$  in the range from  $MinStep$  to  $MaxStep$  is the multiscale LDTW distance
- 

To achieve the goal of clustering, the edge weight between distinct subgraphs after cutting is as low as possible, and the edge weight sum within the subgraph is as high as possible by cutting the graph made of all data points [20]. In this study, a similarity matrix that is in line with the load profile characteristics is generated using multiscale LDTW, replacing the original distance metric. The steps of spectral clustering based on multiscale LDTW are as follows.

*Step 1:* create a graph  $G$  for a given set of samples containing  $n$  load profiles, with each series acting as a vertex on the graph. Calculate the multiscale LDTW distance between each sample.

*Step 2:* use multiscale LDTW as the distance measure  $f_m$  to calculate the distance between each series to form the similarity matrix  $S$ .

$$S(i,j)=f_m(i,j) \quad (6)$$

$$S=\begin{bmatrix} S(1,1) & S(1,2) & \dots & S(1,n) \\ S(2,1) & S(2,2) & \dots & S(2,n) \\ \vdots & \vdots & & \vdots \\ S(n,1) & S(n,2) & \dots & S(n,n) \end{bmatrix}$$

Construct the degree matrix  $D$ . Each diagonal element of  $D$  is the sum of the elements of each row of the associated similarity matrix  $S$ , and all other elements are 0.

$$d(i,j)=\sum_{j=1}^n S(i,j) \quad (7)$$

$$D=\begin{bmatrix} d(1,1) & 0 & \dots & 0 \\ 0 & d(2,2) & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & d(n,n) \end{bmatrix}$$

Calculate the Laplacian matrix  $L$  using  $S$  and  $D$  and normalize the Laplacian matrix to produce superior clustering results.

$$L = D^{-\frac{1}{2}}(D - S)D^{-\frac{1}{2}} \quad (8)$$

Step 3: compute the eigenvectors of  $L$  and arrange the vec-

tors in ascending order of eigenvalues to form the matrix  $H$ . To form a new matrix  $Y_{k \times n}$ , the first  $k$  eigenvectors are chosen. The row vector of matrix  $Y$  is used as the new feature of the original series to obtain the clustering result.

The flowchart of the proposed approach is shown in Fig. 2.

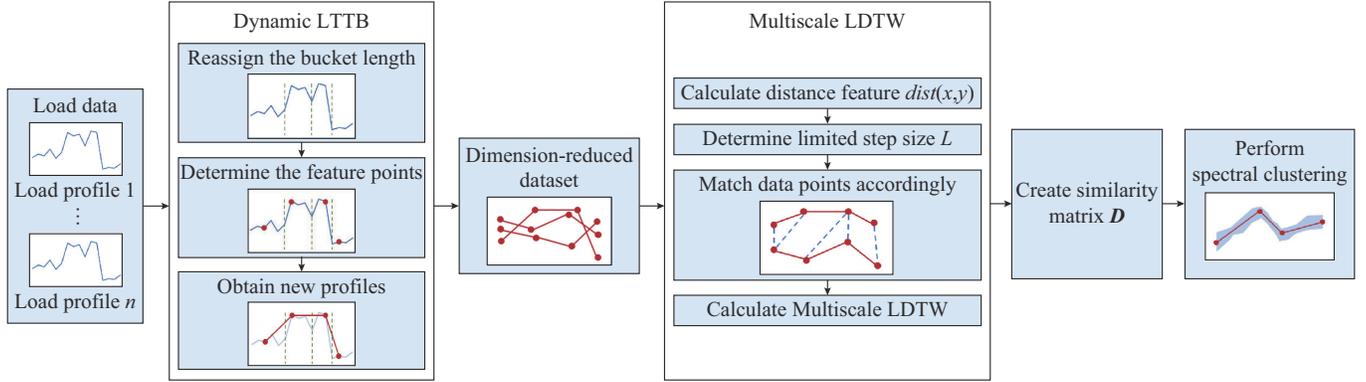


Fig. 2. Flowchart of proposed approach.

### III. PERFORMANCE EVALUATION

In this section, the proposed approach is evaluated with multiple high-resolution load profile datasets. And it is compared with popular approaches in terms of the effects of dimensionality reduction and the similarity measures.

#### A. Descriptions of Datasets

In this study, multiple publicly available load datasets are used to evaluate the proposed approach. We choose the smart building energy dataset CU-BEMS [21] and select three datasets related to power loads in the University of California time series archive [22] to evaluate the performance of the proposed approach. High-resolution data are measured every few seconds to 30 min [23]. The datasets used in this study have a sampling frequency ranging from 1 to 15 min. We evaluate the volatility of the datasets using the coefficient of variation (CV). The CV of the power load data exceeds 0.8. Therefore, each dataset has a high degree of volatility. The basic information of the datasets used in this study is summarized in Table I.

TABLE I  
BASIC INFORMATION OF DATASETS

Dataset	Length	Size of samples	CV
CU-BEMS	1440	101	1.42
Computer	720	241	0.83
Powercons	144	176	1.06
Electric devices	96	1367	1.23

#### B. Performance Measures

This subsection introduces the performance measures used in this study. The effects of dimensionality reduction are evaluated using the average distinguished information (ADI). The Davies-Bouldin indicator (DBI) and validity index (VI)

are used to measure the clustering effect. The DBI is a commonly used indicator to measure the clustering effect. By comparing the similarity based on both distance and correlation, the VI considers whether the changes in the trends between series are similar in a more thorough manner.

1) ADI [9]: ADI assesses the retention of the original sequence features. Representation data with a larger ADI implying that the performance of an algorithm is better for maintaining distinct characteristics.

$$ADI = \frac{\frac{1}{M} \sum_{j=1}^L \sqrt{\sum_{i=1}^M (y_{ij} - \bar{y}_{Nj})^2}}{L_s} \quad (9)$$

where  $y_{ij}$  is the element in a two-dimensional representation dataset  $Y$ ;  $M$  is the total number of series;  $\bar{y}_{Nj}$  is the mean of all the  $j^{\text{th}}$  data of the series; and  $L_s$  is the length of each series.

2) DBI [24]: DBI is commonly used to evaluate the within-cluster scatter and between-cluster separation.

$$\left\{ \begin{array}{l} S_i = \left\{ \frac{1}{T_i} \sum_{j=1}^{T_i} |X_j - A_i|^q \right\}^{\frac{1}{q}} \\ M_{ij} = \left\{ \sum_{k=1}^{N_i} |a_{ki} - a_{kj}|^p \right\}^{\frac{1}{p}} \\ DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} \frac{S_i + S_j}{M_{ij}} \end{array} \right. \quad (10)$$

where  $X_j$  is the  $j^{\text{th}}$  data point in class  $i$ ;  $A_i$  is the center of class  $i$ ;  $T_i$  is the number of data points in class  $i$ ;  $N_i$  is the number of data points in  $A_i$ ;  $a_{ki}$  is the value of the  $k^{\text{th}}$  attribute of the center point of class  $i$ ;  $S_i$  is the mean distance between the samples in the  $i^{\text{th}}$  class and their cluster centroids; and  $M_{ij}$  is the distance between the  $i^{\text{th}}$  class and the  $j^{\text{th}}$  cluster

centroid. A smaller DBI indicates a better clustering effect.

3) VI [25]: VI considers whether the changes in the trends between series are similar by evaluating the similarity on the basis of both the distance and correlation. Furthermore, a change in the value of  $\mu$  can alter the proportion of distance and correlation depending on the situation; normally,  $\mu$  is 0.5.

$$\begin{cases} d(x) = \frac{\bar{d}_i(x)}{\bar{d}_0(x)} \\ r(x) = \frac{\bar{r}_0(x)}{\bar{r}_i(x)} \\ VI = \mu d(x) + (1 - \mu)r(x) \end{cases} \quad (11)$$

where  $\bar{d}_i(x)$  is the average value of the distances from all of the objects of the dataset to the corresponding clustering centers;  $\bar{d}_0(x)$  is the average distance between cluster centers;  $\bar{r}_i(x)$  is the average correlation between the consumption profiles and the corresponding clustering centers; and  $\bar{r}_0(x)$  is the average correlation between clustering centers.

### C. Effects of Dimensionality Reduction

In this subsection, we analyze the effects of dimensionality reduction from three aspects, compare the information retention and clustering quality after performing equally-spaced sampling (ES), PAA, LTTB, and dynamic LTTB, and show the details of data processing through an example. We use different methods to reduce the dimensions of the four datasets with different lengths, and the resulting ADIs are listed in Table II. The ADIs of the dynamic LTTB are 0.16, 0.13, and 0.06 higher than those of the other methods on average. It can be observed that dynamic LTTB performs better than the other methods and can retain more original features.

TABLE II  
COMPARISON OF ADIS FOR DIFFERENT DIMENSIONALITY REDUCTION METHODS

Dataset	New dimension	ADI			
		ES	PAA	LTTB	Dynamic LTTB
CU-BEMS	48	0.72	0.78	0.83	0.89
	24	0.71	0.69	0.74	0.88
Computer	48	0.18	0.22	0.25	0.28
	24	0.13	0.20	0.18	0.23
Powercons	24	0.24	0.30	0.33	0.35
	12	0.20	0.26	0.31	0.32
Electric devices	24	0.19	0.25	0.41	0.48
	12	0.22	0.17	0.39	0.47

Further, we cluster the dimensionless load profiles created by different methods as clustering targets and compare their clustering effects. In Table III, we consider a dimensionality reduction curve with a short length as an example to obtain the DBIs and VIs of the four datasets. Compared with direct clustering of the original data, DBI and VI only increase by 0.13 and 0.05, respectively, when using dynamic LTTB, which are lower than those of the other methods.

TABLE III  
COMPARISON OF DBIS AND VIs OF FOUR DATASETS OBTAINED BY DIFFERENT DIMENSIONALITY REDUCTION METHODS

Dataset	Index	Original data	ES	PAA	LTTB	Dynamic LTTB
CU-BEMS	DBI	0.74	1.92	1.21	1.06	0.78
	VI	0.11	0.34	0.19	0.16	0.13
Computer	DBI	2.52	4.11	2.97	2.89	2.77
	VI	0.29	0.67	0.38	0.30	0.26
Powercons	DBI	1.51	2.57	1.73	1.69	1.63
	VI	0.74	1.14	0.91	0.93	0.87
Electric devices	DBI	3.31	4.82	4.05	3.48	3.41
	VI	0.48	0.91	0.71	0.63	0.56

To demonstrate the effects of different dimensionality reduction methods more intuitively, we select a series in the Powercons dataset as an example to show the series obtained by different dimensionality reduction methods. The original series length decreases from 144 to 12. The results are shown in Fig. 3.

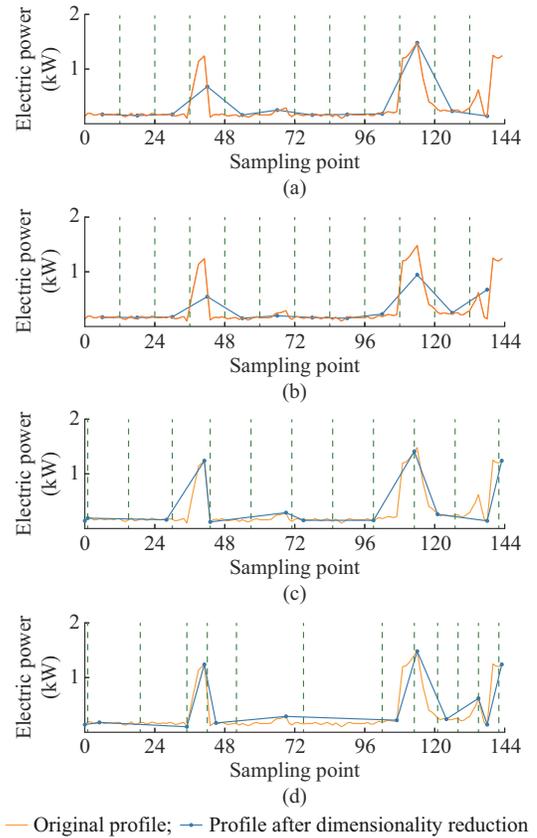


Fig. 3. Example series from Powercons dataset showing effects of different dimensionality reduction methods. (a) ES. (b) PAA. (c) LTTB. (d) Dynamic LTTB.

The sampling time of this dataset is 10 min. The green dotted line represents the division line for each bucket. From this example, we can observe the advantages of dynamic LTTB for dimensionality reduction. The results of ES can be considered as random selection, which is likely to miss key data features. The PAA uses the average value of the data in

each bucket as a representative. If the data points in the bucket significantly fluctuate, the representative point may poorly describe the characteristics of the original series compared with the average. LTTB is more flexible in selecting representative points, but it does not perform well in cases where multiple key data points are located in the same bucket. The dynamic LTTB considers the locations of potential key data points during the process of generating buckets and accordingly creates a flexible division. In the case of load profiles with large fluctuations and uncertain locations, the feature retention of the original sequence can be improved, which is conducive to the subsequent clustering.

#### D. Effects of Similarity Method

We use different similarity methods to cluster different datasets and compare their clustering performances and the matching relationships of data points. We perform clustering using the similarity matrix calculated by DTW, derivative dynamic time warping (DDTW), LimitDTW, and multiscale LDTW and evaluate the clustering quality using DBI and VI. The results are presented in Table IV. The experiment results demonstrate that the proposed multiscale LDTW achieves excellent performance during the clustering process for load profiles with different lengths. DTW is particularly prone to pathological alignment when the curves have large fluctuations; thus, the clustering quality is poor. Other methods provide different solutions to this problem. Therefore, the results are generally better. Among these methods, DBI and VI of multiscale LDTW are 0.5 and 0.09 lower than those of DDTW on average, respectively, and 0.27 and 0.1 lower than those of LimitDTW on average. Multiscale LDTW has better performance.

TABLE IV  
COMPARISON OF CLUSTERING QUALITY WITH DIFFERENT SIMILARITY METHODS

Dataset	Index	DTW	DDTW	LimitDTW	Multiscale LDTW
CU-BEMS	DBI	1.52	0.91	1.03	0.78
	VI	0.26	0.17	0.21	0.13
Computer	DBI	3.43	3.07	2.96	2.77
	VI	0.41	0.28	0.29	0.26
Powercons	DBI	2.18	1.74	1.95	1.63
	VI	1.31	0.91	0.98	0.87
Electric devices	DBI	3.97	4.87	3.73	3.41
	VI	0.77	0.83	0.72	0.56

To demonstrate the improvement in data point matching with multiscale LDTW more clearly, we take the matching paths of two load profiles in the Powercons dataset obtained by DTW and multiscale LDTW. The sampling time of this dataset is 10 min. The calculated suitable step size for this example is 179. The connection relationships between the data points are shown in Fig. 4. The diagram plots the two curves (series A and series B) on one coordinate system to represent the matching relationship of each data point. We can clearly observe an improvement in the situation where one point matches too many data points. Therefore, the prob-

lem of pathological alignment is significantly alleviated. In addition, multiscale LDTW does not impose any restrictions on the specific matching range of a single data point. It only regulates the total length of the path via the computed maximum step size, allowing for more flexible creation of the optimal path during matching. Combining the above two points, it will result in more logical and accurate clustering results as well as typical load profiles.

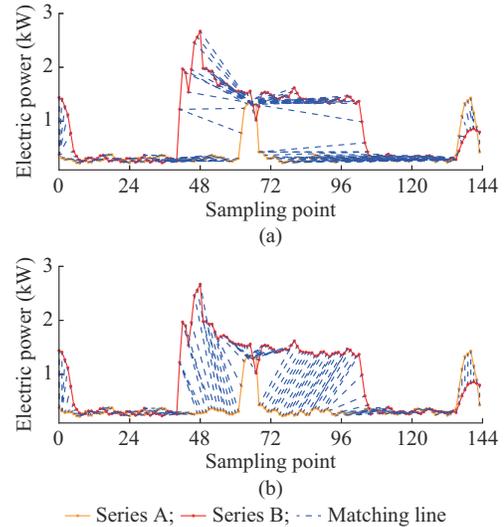


Fig. 4. Connection relationships between data points obtained with DTW and multiscale LDTW. (a) DTW. (b) Multiscale LDTW.

The clustering results for the CU-BEMS dataset in Fig. 5 are obtained using the proposed approach. The load profiles are divided into six clusters, where the gray curve represents the original curve and the red curve represents the typical curve of each cluster.

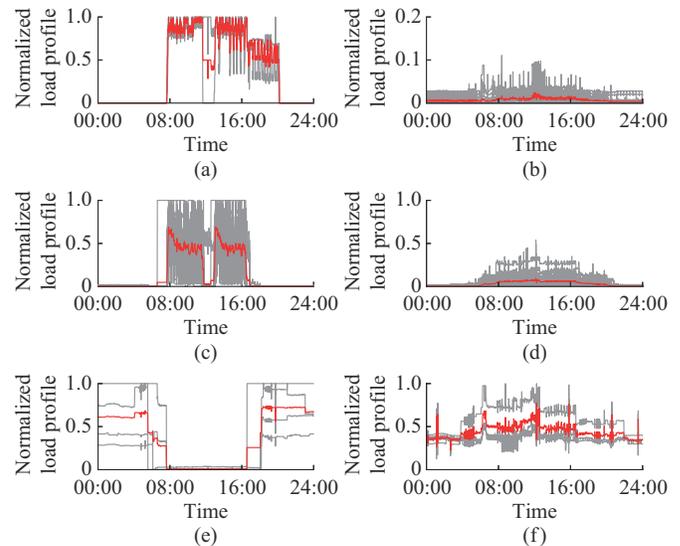


Fig. 5. Clustering results for CU-BEMS dataset with proposed approach. (a) Cluster 1. (b) Cluster 2. (c) Cluster 3. (d) Cluster 4. (e) Cluster 5. (f) Cluster 6.

According to the results, we can obtain the following characteristics of each cluster. There is a continuous power con-

sumption peak from 08:00 to 20:00 in cluster 1, and it decreases slightly at noon. The load of cluster 2 is very low for a long time. The peak periods of power consumption for cluster 3 are 08:00-12:00 and 13:00-16:00. The overall load level of cluster 4 is low, but the load from 06:00 to 20:00 is relatively high. The peak load of cluster 5 occurs from 00:00 to 08:00 and from 16:00 to 24:00. The load of cluster 6 is maintained at a stable level. Compared with previous methods, the proposed approach achieves a more accurate division. Each cluster has distinct characteristics considering the similarity between samples from multiple perspectives.

### E. Computational Efficiency

In this subsection, we compare DTW, DDTW, and LimitDTW with multiscale LDTW. The average time required to generate a similarity matrix using different methods for four different datasets is shown in Fig. 6. The average time of DDTW is longer because the calculation of the derivative takes more time, but the derivatives of the load profiles can provide an assessment of the similarity between the trends. LimitDTW is slightly faster than DTW because it reduces the search scope. Multiscale LDTW is more time-consuming than the other methods. Let the lengths of two sequences be  $M_1$  and  $N_1$ , and let  $L_1$  be the maximum step size of the restriction. Because the situations for different paths need to be considered, the time complexity of multiscale LDTW is  $O(L_1 M_1 N_1)$ , which is higher than that of DTW ( $O(M_1 N_1)$ ). Although the performance with respect to this aspect is unsatisfactory, the structure of multiscale LDTW is suitable for parallel computation. The calculations of different step sizes and the distances between different time series are independent and can be performed simultaneously.

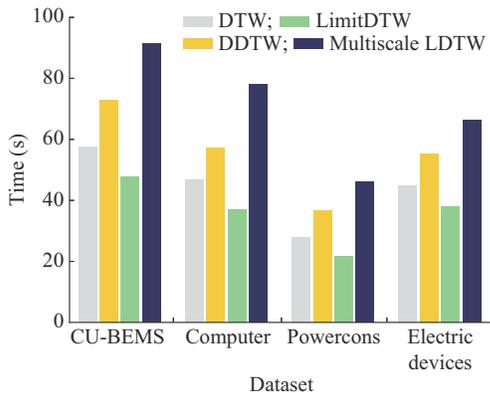


Fig. 6. Average time required to generate a similarity matrix using different methods for four different datasets.

## IV. CONCLUSION

This paper presents a high-resolution load profile clustering approach based on dynamic LTTB and multiscale LDTW that improves the effects of feature extraction and clustering in the data with substantial fluctuations. The proposed dynamic LTTB can scan the possible positions of data points with key features and flexibly adjust the distribution of the buckets in each sequence. The proposed multiscale LDTW suppresses the pathological alignments of the data by limiting the overall step size of the matching between the da-

ta points in sequences. The adjustable numerical and derivative distances are combined for measurement. Compared with other popular methods, the proposed approach exhibits better accuracy for most datasets.

In future work, an efficient approach to DTW that can quickly determine the matching relationships between data points is worth studying.

## REFERENCES

- [1] Y. Liu, G. Wang, W. Guo *et al.*, "Power data mining in smart grid environment," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 3169-3175, Feb. 2021.
- [2] Y. Li, D. Han, and Z. Yan, "Long-term system load forecasting based on data-driven linear clustering method," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 2, pp. 306-316, Mar. 2018.
- [3] M. Wen, R. Xie, K. Lu *et al.*, "Feddetect: a novel privacy-preserving federated learning framework for energy theft detection in smart grid," *IEEE Internet of Things Journal*, vol. 9, no. 8, pp. 6069-6080, Sept. 2021.
- [4] Y. Wang, Q. Chen, C. Kang *et al.*, "Load profiling and its application to demand response: a review," *Tsinghua Science and Technology*, vol. 20, no. 2, pp. 117-129, Apr. 2015.
- [5] C. Si, S. Xu, C. Wan *et al.*, "Electric load clustering in smart grid: methodologies, applications, and future trends," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 2, pp. 237-252, Mar. 2021.
- [6] Z. Wang and H. Wang, "Analyzing seasonal variation in residential load patterns via two-stage clustering and relative entropy: poster," in *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*, Torino, Italy, Jun. 2021, pp. 286-287.
- [7] D. Zhang, X. Zhao, Y. Guo *et al.*, "Cluster analysis of smart meter load based on electricity behavior characteristics," in *Proceedings of International Conference on Genetic and Evolutionary Computing*, Jilin, China, Jan. 2022, pp. 429-437.
- [8] A. Dedić, T. Konjić, M. Čalasan *et al.*, "Fuzzy C-means clustering applied to load profiling of industrial customers," *Electric Power Components and Systems*, vol. 49, no. 11-12, pp. 1-17, Apr. 2022.
- [9] S. Lin, F. Li, E. Tian *et al.*, "Clustering load profiles for demand response applications," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1599-1607, Nov. 2017.
- [10] Z. Bi, Y. Leng, Z. Liu *et al.*, "An improved spectral clustering algorithm using fast dynamic time warping for power load curve analysis," in *Proceedings of International Conference on Mobile Computing, Applications, and Services*, Athens, Greece, Dec. 2020, pp. 143-159.
- [11] J. Menezes and N. Poojary, "Dimensionality reduction and classification of hyperspectral images using DWT and DCCF," in *Proceedings of 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, Muscat, Oman, Mar. 2016, pp. 1-6.
- [12] S. Steinarrson, *Downsampling Time Series for Visual Representation*. Ph.D. dissertation, University of Iceland, Reykjavik, Iceland, 2013.
- [13] J. L. Chen, Y. Huang, W. Qiu *et al.*, "Research on feature extraction method of power grid online data based on big data," *Journal of Physics: Conference Series*, vol. 2030, p. 012064, Sept. 2021.
- [14] H. Ding, G. Trajcevski, P. Scheuerermann *et al.*, "Querying and mining of time series data: experimental comparison of representations and distance measures," *Proceedings of the VLDB Endowment*, vol. 1, no. 2, pp. 1542-1552, Aug. 2008.
- [15] G. L. Ray and P. Pinson, "Online adaptive clustering algorithm for load profiling," *Sustainable Energy, Grids and Networks*, vol. 17, p. 100181, Mar. 2019.
- [16] T. Teeraratkul, D. O'Neill, and S. Lall, "Shape-based approach to household electric load curve clustering and prediction," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5196-5206, Mar. 2017.
- [17] M. Gao, T. Gong, R. Lin *et al.*, "A power load clustering method based on limited DTW algorithm," in *Proceedings of 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Chengdu, China, Mar. 2019, pp. 253-256.
- [18] H. Liang and J. Ma, "Develop load shape dictionary through efficient clustering based on elastic dissimilarity measure," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 442-452, Jan. 2021.
- [19] Z. Zhang, R. Tavenard, A. Bailly *et al.*, "Dynamic time warping under limited warping path length," *Information Sciences*, vol. 393, pp. 91-107, Jul. 2017.

- [20] A. Ghosal, A. Nandy, A. K. Das *et al.*, "A short review on different clustering techniques and their applications," in *Emerging Technology in Modelling and Graphics*. Berlin: Springer, 2019, pp. 69-83.
- [21] M. Pipattanasomporn, G. Chitalia, J. Songsiri *et al.*, "CU-BEMS, smart building electricity consumption and indoor environmental sensor datasets," *Scientific Data*, vol. 7, no. 1, pp. 1-14, Jul. 2020.
- [22] H. A. Dau, A. Bagnall, K. Kamgar *et al.*, "The UCR time series archive," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1293-1305, Nov. 2019.
- [23] R. Granell, C. J. Axon, and D. C. Wallom, "Impacts of raw data temporal resolution using selected clustering methods on residential electricity load profiles," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3217-3224, Dec. 2014.
- [24] A. Rajabi, L. Li, J. Zhang *et al.*, "A review on clustering of residential electricity customers and its applications," in *Proceedings of 2017 20th International Conference on Electrical Machines and Systems (ICEMS)*, Sydney, Australia, Aug. 2017, pp. 1-6.
- [25] Y. Wang, L. Li, and Q. Yang, "Application of clustering technique to electricity customer classification for load forecasting," in *Proceedings of 2015 IEEE International Conference on Information and Automation*, Lijiang, China, Aug. 2015, pp. 1425-1430.

**Mi Wen** received the M.S. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2005, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2008, both in computer science. She is currently a Professor of the College of Computer Science and Technology with Shanghai University of Electric Power, Shanghai, China. She is an Associate Editor of *Peer-to-Peer Networking and Applications* (Springer). She is presently a Director of Shanghai Computer Society. Her research interests include privacy-preserving in wireless networks, big

data, and smart grid.

**Yue Ma** received the B.E. degree from the Department of Electrical Engineering and Automation, Zhejiang Sci-Tech University, Hangzhou, China, in 2020. He is currently pursuing the M.S. degree in electric power information technology in Shanghai University of Electric Power, Shanghai, China. His research interests include load profile clustering and load forecasting.

**Weina Zhang** received the M.S. degree in system analysis and integration from South China University of Technology, Guangzhou, China, in 2008, and the Ph.D. degree in computer science and technology at South China University of Technology, in 2020. Now she is a Lecturer in the College of Computer Science and Technology in Shanghai University of Electric Power, Shanghai, China. Her research interests include recommendation algorithms with interpretability, stability and security, and power load forecasting.

**Yingjie Tian** received the M.S. degree in power system and automation from the Hunan University, Changsha, China, in 1993. He is currently the Deputy Chief Engineer and Senior Engineer of the Electric Power Research Institute of State Grid Shanghai Electric Power Company, Shanghai, China. His research interests include electric power big data and electric power artificial intelligence.

**Yanfei Wang** received the B.E. degree from the College of Computer Science and Technology, Qingdao University of Technology, Qingdao, China, in 2018. She is currently pursuing the M.S. degree in electric power information technology in Shanghai University of Electric Power, Shanghai, China. Her research interests include deep learning, adversarial attacks, and defenses.