IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 17, 2019, accepted February 10, 2019, date of publication February 22, 2019, date of current version April 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2901197

A Weekly Load Data Mining Approach Based on Hidden Markov Model

SHIXIANG LU[1], GUOYING LIN[1], HANLIN LIU[2], CHENGJIN YE[2], HUAKUN QUE[1], AND YI DING[2]

[1]Metrology Center, Guangdong Power Grid Corporation, Guangzhou 510080, China
[2]College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China

Corresponding author: Chengjin Ye (yechenjing@zju.edu.cn)

This work was supported in part by the Science and Technology Project of the China Southern Power Grid Corporation under Grant GDKJXM20161523, in part by the National Natural Science Foundation of China under Grant 51807173, and in part by the China Postdoctoral Science Foundation under Grant 2018M640558.

ABSTRACT With the development of advanced metering infrastructure, massive smart meter readings are generated and stored in smart grids, which makes it possible for detecting of tremendous social value embedded in load data. The majority of the existing load data mining works are performed on the daily time scale without adequate consideration of load information between the days. To better describe the power consumption characteristics of users, a data mining approach based on the weekly load curves is proposed in this paper. First, the piecewise aggregate approximation technique is utilized to reduce the dimensions of the raw weekly load data. Then, a Davies–Bouldin index-based adaptive k-means algorithm is proposed to cluster the studied users into several groups. Finally, a hidden Markov model describing the probabilistic transitions of different load levels is established for each cluster to extract the representative dynamic weekly load features. A feasible tool based on dynamic characteristics of load patterns is invented to evaluate the short-term load forecasting methods, which realizes the pre-check for the forecasting results without future real measurements in the forecasting horizon. Case studies on a real dataset demonstrate that the proposed method is capable of extracting weekly load characteristics of users.

INDEX TERMS Weekly load profiles, dimension reduction, clustering, hidden Markov model evaluation.

I. INTRODUCTION

A. BACKGROUND

With the development of smart grids, smart meters, the basic terminal equipment of Advanced Metering Infrastructure (AMI), has gained increasing popularity worldwide. For instance, in the US, the quantity of smart meters installed has reached 70 million by the end of 2016 [1]; while in China, more than 500 million smart meters will be installed during the 13th Five-Year Plan period (2016-2020) [1]. Consequently, massive load data is generated and stored. With a temporal measurement of 15 minutes, the annual amount of smart meter readings for China reaches 117TB.

Except for the traditional electricity billing, hidden value of the massive smart meter readings is detected by a series of data mining approaches. Typical load data mining procedure includes steps of data cleaning, compression, clustering, forecasting and so on [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Qilian Liang.

B. LITERATURE REVIEW

Among the mentioned load data mining models, load data clustering is considered as the fundamental step for further application of smart meter readings. Classification provides a straightforward recognition about loads, which should assign the loads in the same class sharing similar patterns, while loads in different classes illustrate significant differences. Based on the classification, differentiated mechanisms are designed. Specifically, different Real-time Pricing (RTP), Time-of-use (ToU), or Critical Peak Pricing (CPP) levels are carried out to promote demand response (DR) for different user classes [2]. Researches have been conducted for load data classification or customer segmentation. In general, loads can be clustered through unsupervised and supervised ways. Unsupervised methods such as k-means, hierarchical clustering, and Self- Organizing Map (SOM) classify loads based on the Euclidean distance, density, or other data features [3]. While supervised methods identify unknown loads by learning the existing load classification statistical rules. Typical supervised methods include the support vector

machines (SVM) [4], the decision tree (DT) [5], the logistic regression (LR) [6], [7] and the naive Bayes (NB) classifier [8]. However, with the increasing volume and dimension of load data, traditional clustering methods are difficult to be implemented in a reasonable time. Among these clustering algorithms, the k-means algorithm is more versatile, but the biggest disadvantage of the traditional k-means algorithm is that the cluster number is necessary to be given in advance.

Another important mining procedure for load data is the forecasting. Forecasting results for different time scales have different applications. The long-term forecast mainly provides suggestions for network planning. The medium-term forecast (more than one month) mainly provides a reference for futures trading, reservoir scheduling, overhaul, and fuel plan. The short-term forecast is the basis of the spot transaction. And the very short-term forecast is mainly for real-time scheduling, real-time price forecasting. Generally speaking, load forecasting techniques are mainly divided into statistical models and artificial intelligence models [9]. Statistical models include regression analysis, exponential smoothing and random Time Series. Artificial intelligence models mainly include support vector machine (SVM), artificial neural network (ANN), gray system, and wavelet analysis [10]. However, at present, there are currently no effective evaluation methods to pre-check the forecasting results, thus the planners generally determine the specific utilized forecasting methods subjectively.

Furthermore, the dimension of the original load data needs to be reduced. The dimensionality reduction techniques can be divided into supervised and unsupervised methods. Supervised methods mainly include linear discriminant analysis (LDA) [11] and neural network (NN) [12]. Unsupervised methods include principal component analysis (PCA) [12], independent component analysis (ICA) [13], single value decomposition (SVD) [14], kernel principal component analysis (KPCA) [15] and Fourier analysis (FA) [16]. Besides, for time series data, piecewise aggregate approximation (PAA) can reduce the dimensionality of the original time series while maintaining the original shape.

Besides, the hidden Markov model (HMM) is often used to model the dynamic behavior. In reference [17] the application of HMM in dynamic detection of transmission line outages is introduced; Reference [18] reported a framework of stochastic power management using the HMM; and in reference [19], a transient identification method based on a stochastic approach with the HMM has been suggested and evaluated experimentally. It is worth noting that the HMM represents the dynamic characteristics of the system to be studied, since the user's load curve also has dynamic characteristics too, the HMM provides a good idea for the dynamic behavior modeling of the load.

C. INNOVATIONS

To the best of the authors' knowledge, nowadays, most of the existing load data mining researches are performed on the time scale of a single day, and the selection of load

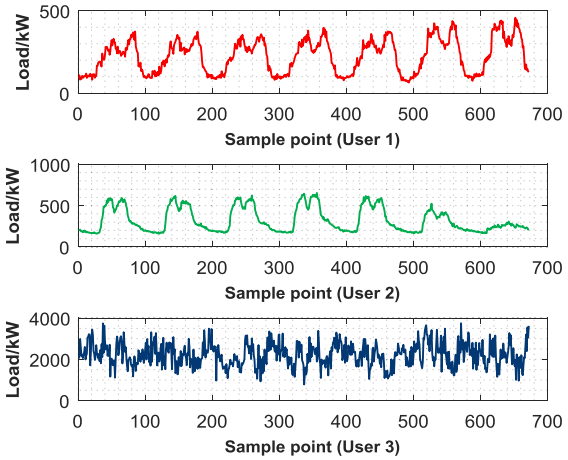

FIGURE 1. Weekly load curve for three types of power users.

characteristic indicators are mainly based on typical daily indices [20], such as the daily maximum load, daily minimum load, daily average load, daily load rate, and so on. To some extent, the utilization of daily approaches has certain rationality. This is because from the perspective of the cyclicality of social production, the behaviors of power consumption have significant similarities between the days.

Nevertheless, it is worth noting that some non-ignorable power consumption characteristics on relatively long time scales cannot be extracted based on the typical daily load curves. **FIGURE 1** shows weekly load curves of three different power users, and the sampling interval in the figure is 15 minutes. For the first and second users, the daily load profiles are basically the same, especially after normalization. If only daily profiles are utilized as the clustering criteria, the first and second users obviously would be classified into one group. However, as can be seen, some certain differences do exist in these two users between days. For the first load, it presents a rising trend, while the second one has an obvious decreasing. If the weekly load profiles are studied, the different trends between days are likely to be preserved, thus more accurate and reasonable clustering results can be obtained. What's more, the third user presents an observable randomicity between days, then it is not feasible to extract and utilize one typical daily load profile for load data mining. So, load data with longer time scale should be used as the basis of load clustering to characterize power consumption features between days for the users' behaviors.

Furthermore, the three loads shown in **FIGURE 1** correspond to three different power usage patterns. From a day's perspective, for the first user, the representative profile can be summarized as a two-spike shape, in which the second spike overtops the first one. For the second user, the latter spikes are generally lower than the former ones in days. Extend to a week, for the first user, the weekends are close to the working days in terms of load level, while for the second one, the weekend loads decrease significantly. In general, the essence of electricity consumption features can be

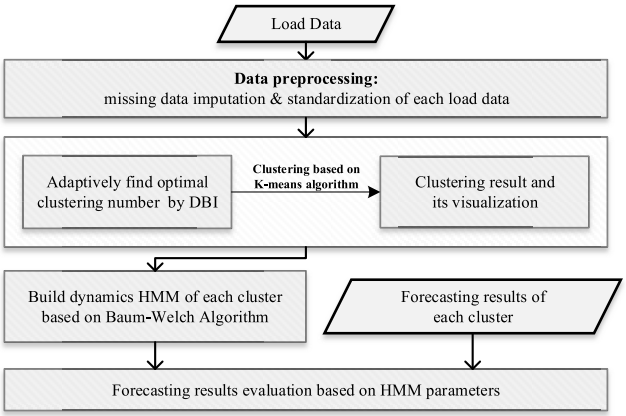

FIGURE 2. Flowchart of the proposed method.

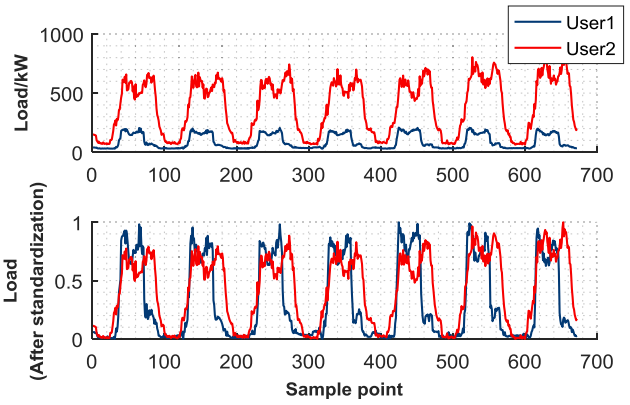

FIGURE 3. Standardization of the weekly load curve for two power users.

considered as the dynamic transition laws between different load levels.

To solve the above problems and better describe the power consumption characteristics of users, a data mining approach based on the weekly load curves is proposed in this paper. The main contributions of the proposed method include:

a) An adaptive k-means method based on PAA technique is proposed to perform clustering. Specifically, a piecewise aggregate approximation (PAA) technology is utilized to transform the dense fluctuating load data into piecewise paned data. Then, A Davies-Bouldin index (DBI) based adaptive k-means algorithm is utilized to cluster the studied users into several groups. The optimal clustering number can be set automatically without prior knowledge.

b) A Hidden Markov Model (HMM) is established to describe the probabilistic transitions of different load levels in the aggregated load curve of each cluster. The proposed model is capable of characterizing the representative dynamic weekly demand features. Moreover, the proposed model also provides a feasible tool to evaluate the performance of the short-term load forecasting. It realizes the pre-check for the forecasting results without the real load data of the forecast day.

D. PAPER ORGANIZATION

The flowchart of the proposed method in this paper is briefly described in **FIGURE 2.**

Specifically, in section II, basic methodologies including data normalization and piecewise aggregate approximation are introduced. Then, an adaptive k-means clustering method based on Davies-Bouldin index (DBI) is proposed, which can find the optimal classification number for the studied dataset. In Section III, the Baum-Welch algorithm is utilized to establish an HMM for each type of load pattern, which represents the dynamic characteristics of it. In Section IV, A feasible tool based on dynamic characteristics of load patterns is invented to evaluate the short-term load forecasting methods. Section V validates the feasibility of the proposed model based on real data sets in Guangdong, China. Finally, the conclusions are presented in Section VI.

II. BASIC METHODOLOGY

A. DATA NORMALIZATION

Since the range of values of raw data varies widely, in some machine learning algorithms, objective functions will not work properly without normalization [21]. Moreover, since we focus on the dynamic characteristic of the loads, which are reflected by the relative power levels of different time periods, thus the original load data needs to be normalized firstly.

Due to the complexity and diversity of the user's power usage behavior, the original load data may not totally or approximately satisfy the Gaussian distribution, the results of commonly used z-scores normalization will be poor [22]. Therefore, the unity-based normalization (also known as min-max normalization) is utilized in this paper [23]. This method ensures that all data is linearly mapped into the interval [0], [1]. The formula of unity-based normalization is as follows.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

where x represents the original load data sequence, $\min(x)$ and $\max(x)$ represent the minimum and maximum values of the load data sequence, respectively. And x' represents the normalized data sequence.

FIGURE 3 shows the data sequence before and after unity-based normalization. It can be easily found that the profiles of the two load curves are quite similar after normalization. Then, these loads with disparate amplitudes are likely to be classified into the same cluster.

B. PIECEWISE AGGREGATE APPROXIMATION

In this paper, the feature used for clustering is the weekly load curve. However, for a raw uncompressed weekly load curve, assuming that the sampling interval of the measuring device is 15 minutes, the data volume of a weekly load curve will reach 672 dimensions.

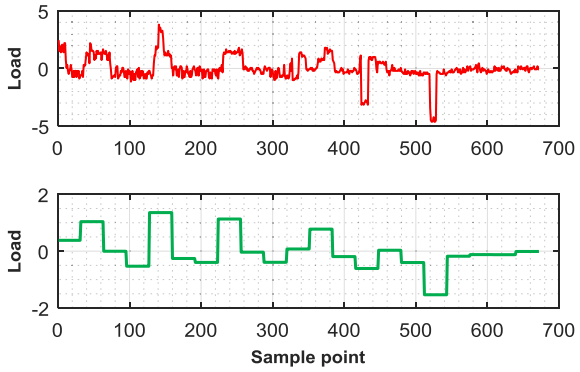

FIGURE 4. PAA result for weekly load curve.

However, clustering uncompressed weekly load data is time consuming, and the high-dimensional data may lead to over-fitting of the clustering algorithm. Since we mainly focus on weekly load profile, so only some typical values or segments are essential to represent it. Thus, the PAA is adopted in this paper to reduce the original dimensions.

Here, we divide a day into three equal time periods, this is in line with social activities of power users: during 0:00-8:00, most of the loads are off; during 8:00-16:00, office and industry facilities are at work; and during 16:00-24:00, most of the offices and industries are off while the residents and malls are on.

The PAA uses the mean of each time period to approximate the whole weekly load data [24]. From a general perspective, the PAA divides the time series of length n into M segments, each segment is of the same length k. The principle of PAA can be expressed as follows.

$$\bar{x}_i = \frac{M}{n} \sum_{j=n/M(i-1)+1}^{(n/M)i} x_j \tag{2}$$

FIGURE 4 shows the converting of a 672-points weekly load curve into a PAA sequence with a length of 21.

As can be seen from **FIGURE 4**, the dimension of the original data is reduced by PAA, and the obtained sequence preserves the original profile information. It means that each daily load curve is represented by 3 representative data values, which cover the average load of 0:00-8:00, 8:00-16:00, and 16:00-24:00 respectively.

C. ADAPTIVE K-MEANS ALGORITHM BASED ON WEEKLY LOAD CURVE

The traditional k-means is not suitable to cluster load profiles directly, because the number of the profile clusters cannot be subjectively determined without any prior knowledge of the dataset. To solve this difficulty, an adaptive k-means that can automatically set k values according to the input dataset is proposed.

First of all, the convergence of the k-means with normalized load data is discussed below. Denote the load data after unity-based normalization and PAA as $\boldsymbol{x} = (x_1, x_2, \cdots, x_N)$, where N is the length of the data. Since $x_i \in [0, 1]$, $i = 1, 2, \cdots N$, the domain of the user load data is an N-dimensional cube with a side length of 1, which obviously satisfies the properties of the convex set [25]:

$$\forall \boldsymbol{x}, \quad \boldsymbol{y} \in S, \ t \in [0, 1] \Rightarrow (1 - t)\boldsymbol{x} + t\boldsymbol{y} \in S \tag{3}$$

Therefore, according to the convex optimization [26], the k-means algorithm here can converge at a fast speed.

The main idea of the proposed k-means is a distance-based iterative process, which is shown as follows [27]:

Algorithm 1 Adaptive k-means Algorithm

Step 1: Randomly assign k seeds for the center vectors $c_1, c_2, \cdots, c_k$;

Step 2: Assign the seeds to the nearest center using a distance measure. The distance measure for load data sequence $D_{PAA}(\bar{X}, \bar{Y})$ is given in **(4).** Constructing non-overlapping clusters of the given dataset based on distance.

$$D_{PAA}(\bar{X}, \bar{Y}) \equiv \sqrt{\frac{n}{M}} \sqrt{\sum_{i=1}^{M} |\bar{x}_i - \bar{y}_i|} \tag{4}$$

Where $\bar{X} = (\bar{x}_1, \bar{x}_2, \cdots \bar{x}_i, \cdots \bar{x}_M)$ and $\bar{Y} = (\bar{y}_1, \bar{y}_2, \cdots \bar{y}_i, \cdots \bar{y}_M)$ are two PAA sequences.

Step 3: Update the centers through the distance measurement method. The chosen center should be the one with the minimum summated distance to all the other samples in the same cluster, which can be expressed as follows.

$$S\left(\bar{X}\right) = \sum_{\bar{Y} \in \Omega_X, \bar{Y} \neq \bar{X}} D_{PAA}\left(\bar{X}, \bar{Y}\right) \tag{5}$$

where, $\Omega_{\bar{X}}$ is the cluster that $\bar{X}$ belongs to. And $\bar{Y}$ is the seed in $\Omega_{\boldsymbol{x}}$ which is different from $\bar{X}$. The $\bar{X}$ with the minimum $S\left(\bar{X}\right)$ should be treated as the center of cluster $\Omega_{\bar{X}}$.

Step 4: Repeat step 2 and 3 until the algorithm converges (the centers are no longer changes).

Then, in order to automatically select the optimal clustering number k, firstly, the sample profiles are mixed into one set; then, a quantitative index is introduced to search the optimal clustering of the mixed profiles.

The key of the proposed adaptive procedure is the clustering evaluation. There are various relevant criteria such as the clustering dispersion indicator (CDI) [28], [29], the scatter index (SI) [3], the Davies-Bouldin index (DBI) [3], [30], [31], and the mean index adequacy (MIA) [28], [32].

Among these indicators, DBI uses quantities and features inherent to the dataset, which is suitable for k-means clustering evaluation. It is defined as follows:

$$I_{DBI} = \frac{1}{K} \sum_{i=1}^{k} \max_{j \neq i} \left(\frac{\overline{C_i} + \overline{C_j}}{D_{i,j}} \right) \tag{6}$$

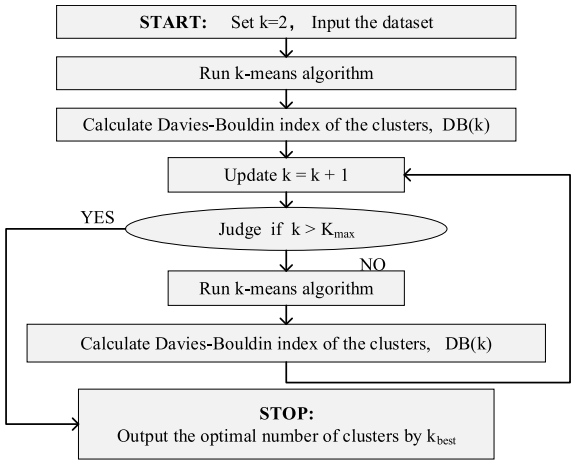

FIGURE 5. Flowchart of the adaptive clustering based on DB index.

where, $\overline{C}_i$ and $\overline{C}_j$ represent the average distance of the seeds in cluster i and j to the center of the corresponding cluster, respectively. $D_{i,j}$ represents the Euclidean distance between the center of the cluster i to cluster j. As can be seen, the smaller the I_{DBI} is, the better the clustering performs.

By comparing indices of different cluster trials, we can obtain the best cluster number k_{best}, which is indicated by the minimum I_{DBI}.

FIGURE 5 shows the procedure of the adaptive k-means. The profiles are continuously subdivided until the DBI of the clusters no longer descends. In this way, the best k is automatically set based on the input profiles without any prior knowledge. To avoid generation of excessive clusters, a threshold is utilized to limit the number of clusters, denoted as k_{max}.

D. THE HIDDEN MARKOV MODELLING OF LOAD DYNAMIC BEHAVIOURS

Since the behavior of a single load is of significant randomness and volatility, its dynamic behavior is difficult to be accurately described. So, the modeling of dynamic characteristics of a single load is of little significance.

On the other hand, the load patterns are of relatively regular power consumption habits. Thus, we aggregate the load in the same cluster to obtain the aggregated load curves which represents the corresponding typical load patterns.

Then, the hidden Markov model (HMM) is utilized to analyze the load dynamic behaviours based on each aggregated load curve.

1) DISCRETIZATION OF AGGREGATED LOAD CURVE

The first step of HMM is to convert the aggregated load curves into sequences. The key of the discretization is the determination of the "breakpoints".

Since the unity-based normalization is utilized, the original load data has been linearly mapped into [0,1]. For discretization, we can just divide the interval into n parts in proportion, and n represents the level of discretization. For example, when $n = 10$, the corresponding discretization rules are as shown in **TABLE 1**.

TABLE 1. A lookup table for discretization rules.

interval	[0,0.1]	[0.1,0,2]	[0.2,0.3]	[0.3,0.4]	[0.4,0.5]
level	1	2	3	4	5
interval	[0.5,0.6]	[0.6,0,7]	[0.7,0.8]	[0.8,0.9]	[0.9,1.0]
level	6	7	8	9	10

Algorithm 2 Mathematical description of two HMM problems

The Evaluation Problem:

Given the HMM $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \pi)$, compute $P(O|\lambda)$, the probability of occurrence of the observation sequence$\boldsymbol{O} = \{O_1, O_2, \cdots, O_T\}$. This problem can be solved by forward algorithm.

The Learning Problem:

Given the observation sequence $\boldsymbol{O} = \{O_1, O_2, \cdots, O_T\}$, find the model $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \pi)$ that best explains the observed data. This problem can be solved by the Baum-Welch algorithm, which will be given below.

In this way, the load curve represented by the PAA sequence can be expressed as a sequence of integers.

2) HIDDEN MARKOV MODEL (HMM)

For weekly load data mining, two issues are mainly addressed in this paper. Firstly, our goal is to build a dynamic model based on a series of observations which is derived from the load curves. That is to say, we want to learn the parameters of the HMM. Secondly, with the known parameters of HMM, we want to calculate the probability of certain observation sequences.

As for HMM, it mainly solves three questions: evaluation problem, decoding problem and learning problem [33]. Obviously, the two questions proposed above correspond to the learning problem and the evaluation question, respectively. The following part gives a mathematical description of the above two problems.

Here, $\boldsymbol{A} = \{a_{ij}\}$ is transition matrix where a_{ij} represents the transition probability from state i to state j. $\boldsymbol{B} = \{b_j(O_t)\}$ is observation emission matrix, where $b_j(O_t)$ represent the probability of observing O_t at state j. $\pi = \{\pi_i\}$ is the initial state probability vector.

For the load curve, compared with the current state, the next observation state is of two options: increase or decrease. Thus, a two-state HMM is utilized in this paper. The state I_1 means that the observation value tends to increase, while the state I_2 means that the observation value tends to decrease. If each state has 10 outputs $O_i = i, i = 1, 2, \cdots, 10$, the **FIGURE 6** shows this HMM topology.

Here, for example, $a_{11} = 0.8$ means that if the current state is I_1, then the probability that the next state holds on is $a_{11} = 0.8$, $a_{12} = 0.2$ means the probability that the next state shifts to state I_2 is $a_{12} = 0.2$. While $b_{1i} = 0.3$ means that if the current state is I_1, then the probability of getting observation value $O_i = i$ is 0.3.

Algorithm 3 The Baum-Welch Algorithm

Expectation:

Calculate $\boldsymbol{Q(\lambda,\lambda')}$, where $\boldsymbol{\lambda'}$ is the current estimate of the HMM parameters. In order to do that, we need to obtain the probability distribution of the hidden variables, and then use it to obtain the expectation of the log-likelihood of the joint probability of the observable sequence and the hidden sequence.

$$\begin{aligned} Q(\lambda, \lambda') = & \sum_{I} \log \pi_{i_1} P(O, I|\lambda') \\ & + \sum_{I} (\sum_{t=1}^{T-1} \log a_{i_{t-1} i_t}) P(O, I|\lambda') \\ & + \sum_{I} (\log b_{i_t}(O_t)) P(O, I|\lambda') \end{aligned} \tag{7}$$

The forward-backward algorithm [34] is used to find the probability distribution of hidden variables in the Expectation step, which uses dynamic programming to reduce the amount of computation greatly.

Maximization:

Find the new model parameters that maximize $Q(\lambda, \lambda')$. Exit if the convergence condition is reached, otherwise return to step 1. The model parameters are calculated as following:

$$\begin{aligned} a_{ij} &= \frac{\sum_{t}^{T-1} P(O, I = i, I = j|\lambda')}{\sum_{t=1}^{T-1} P(O, I|\lambda')} \\ b_j(k) &= \frac{\sum_{t=1}^{T} P(O, I = j|\lambda') \delta(o_t = v_k)}{\sum_{t=1}^{T} P(O, I = j|\lambda')} \end{aligned} \tag{8}$$

where $\boldsymbol{\delta(o_t = v_k)}$ is a Kronecker delta function [35], when $\boldsymbol{o_t = v_k}$, we have $\boldsymbol{\delta(o_t = v_k)} = 1$, otherwise $\boldsymbol{\delta(o_t = v_k)} = 0$.

The following section III shows the detailed Baum-Welch training algorithm and the following section IV shows the detailed procedure of short-term forecasting evaluating.

III. THE BAUM–WELCH ALGORITHM

Briefly, the Baum-Welch algorithm is utilized for parameters estimating of the established HMM, which consists of two steps: Algorithm 3 shows the mathematical expression of these two steps.

More details of the Baum–Welch algorithm can be found in [36]. The historical load data sequences $\boldsymbol{O} = \{O_1, O_2, \cdots, O_T\}$ are input into the Baum–Welch algorithm to train the HMM. The obtained parameter set $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \pi)$ of the HMM model can be utilized to represent the dynamic feature of the load cluster.

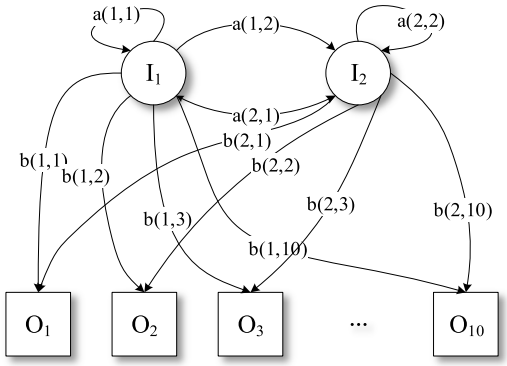

FIGURE 6. HMM structure.

IV. EVALUATION OF FORECASTING METHODS BASED ON HMM

For the evaluation of the forecasting model, the usual approach is to compare the forecasting result with the actual data of the load, and then calculate the MAPE or other indices [37].

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{X_t - F_t}{X_t} \right| \tag{9}$$

where X_t and F_t is the actual and forecasted values, respectively.

As can be seen from the above definition, the actual load data must be known for MAPE evaluation, which is utilized as the benchmark.

In this paper, an HMM based method is proposed to solve this problem. It should be mentioned that it is not appropriate to directly use HMM for load forecasting. This is because the HMM assumes that the output of the model at the next moment is only related to the current time. It is called the non-aftereffect property of the Markov chain [38]. But for the load forecasting, the load information in the previous day or even the previous week should be considered.

However, combined with the forward algorithm, the trained HMM can be used to evaluate the occurrence probability of other sequences. For the forecasting sequences that are more consistent with the load dynamics, the probability values given by the forward algorithms will be larger.

The details of the forward algorithm introduced in algorithm 4.

For a short-term load forecasting problem, various forecasting methods and models are available. By converting the forecasting results into discrete observation sequences, we can determine which sequence has a higher probability. By comparing the probability of each sequence, the different forecasting methods can be evaluated in terms of the dynamic behaviors of the load data.

V. CASE STUDY

A. DATA DESCRIPTION

The load dataset utilized for the case studies in this paper is provided by the Metrology Center of Guangdong Power

Algorithm 4 The forward algorithm

Given the HMM parameter $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \pi)$, the basic idea of the forward algorithm is given below [34]:If the hidden state is q_i at time t, the probability of occurrence of the observation sequence $\boldsymbol{O} = \{O_1, O_2, \cdots, O_t\}$ is defined as the forward probability. Noted as **(10)**.

$$\alpha_t(i) = P(O_1, O_2, \cdots, O_t, I_t = q_i|\lambda) \tag{10}$$

Step 1: Calculate the forward probability of each hidden state at time 1:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad i = 1, 2, \cdots N \tag{11}$$

Step 2: Recursively obtains the forward probability of time $2, 3, \cdots$ T:

$$\alpha_{t+1}(i) = \left[\sum_{j=1}^{N} \alpha_t(j) a_{ji}\right] b_i(O_{t+1}), \quad i = 1, 2, \ldots N \tag{12}$$

Step 3: The final result is shown in **(13).**

$$P(\boldsymbol{O}|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{13}$$

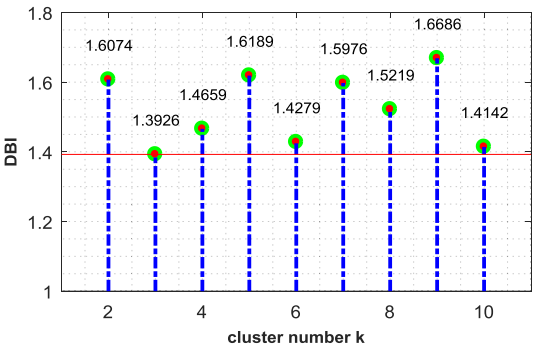

FIGURE 7. DBI for different cluster number k.

Grid Corporation. The dataset contains the smart meter measurements of 200 large users in Foshan, Guangdong province of China in 2016 (366 days), and the sampling interval is 15 minutes. The whole dataset totally has about 7.03 million $(200 \times 35, 136)$ daily load profiles.

After eliminating the missing values or continuous zeroes, the load data of 165 large users for 250 consecutive days are selected as the studied dataset, and there are overall 5,892 weekly profiles.

B. ANALYSIS OF CLUSTERING RESULTS

A DBI-based adaptive k-means is utilized to cluster the mixed load profiles. The DBI results of different clustering schemes are shown in **FIGURE 7**.

As can be seen from **FIGURE 7**, the best clustering number $k_{best} = 3$, which achieves the minimum DBI of 1.3926.

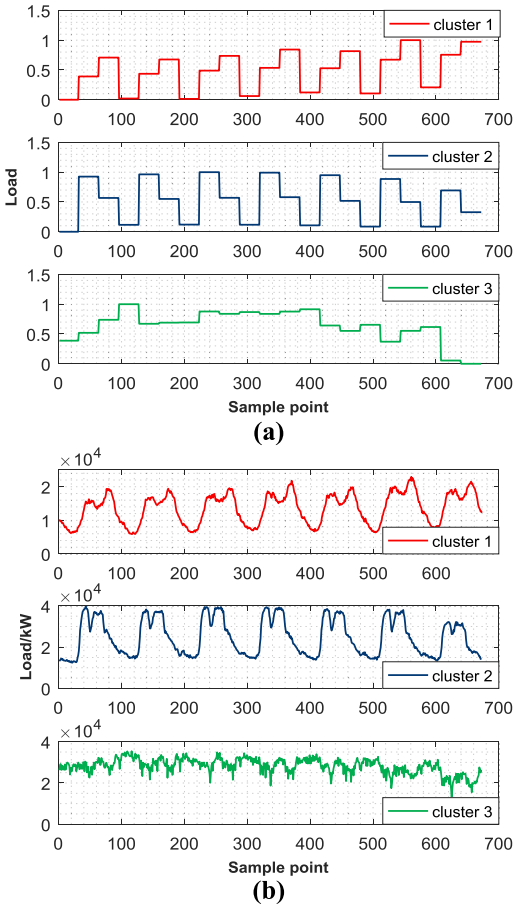

FIGURE 8. (a) 3 cluster centroids represented in PAA; (b) corresponding aggregated load of each cluster.

Thus, the collected 165 users are clustered into three clusters. **FIGURE 8 (a)** illustrates the three centroids or representative load profiles for the three typical user patterns.

After clustering, each load has a label indicating its cluster. By aggregating the load in the same cluster, three aggregate load curves can be obtained as **FIGURE 8 (b)**.

From **FIGURE 8**, we can find that the first and the second cluster both have obvious periodicity but also present some differences. Firstly, the load of the first cluster increases gradually during a whole week, especially on weekends, while the load of the second cluster holds on during the working day, and decrease on weekends. Secondly, the daily peak of the first cluster appears in 16:00-24:00, while the daily peak of the second cluster appears in 8:00-16:00. The cluster 3 is of significant randomness. The loads in this cluster do not have the characteristic of periodicity.

Furthermore, to verify the clustering results, the t-SNE [39] is utilized to visualize the analyzed loads.

From **FIGURE 9**, the loads in cluster 1 or cluster 2 are close to other loads in the same cluster. The few confused and intersecting seeds indicate that the loads in cluster 1 and have regular load patterns and can be classified with good performance. However, loads in cluster 3 are dispersed which means the loads in cluster 3 is difficult to forecast.

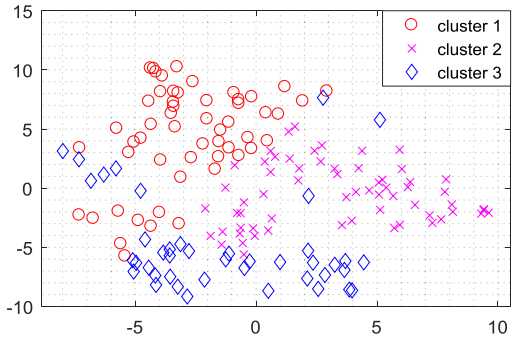

FIGURE 9. Visualization of the clustering results.

C. ANALYSIS OF THE DYNAMIC MODELLING

In order to model the load dynamic behaviors using HMM, the three aggregated load curves are discretized into sequences with ten levels. In detail, after processing these load sequences with unity-based normalization, the specific discretizing principle has been discussed in **TABLE 1**.

After training HMM with the Baum-Welch algorithm, the transition matrix $\boldsymbol{A} = \{a_{ij}\}$ and observation emission matrix $\boldsymbol{B} = \{b_{ij}\}$ can be obtained.

For example, the transition matrix $\boldsymbol{A}_1$ and observation emission matrix $\boldsymbol{B}_1$ of cluster 1 are given as following.

$$\boldsymbol{A}_1 = \begin{pmatrix} 0.9518 & 0.0482 \\ 0.0337 & 0.9663 \end{pmatrix}, \quad \boldsymbol{B}_1' = \begin{pmatrix} 0.3213 & 0 \\ 0.2812 & 0 \\ 0.1272 & 0 \\ 0.1399 & 0 \\ 0.1365 & 0.0462 \\ 0 & 0.1876 \\ 0 & 0.1661 \\ 0 & 0.2036 \\ 0 & 0.2036 \\ 0 & 0.1929 \end{pmatrix} \tag{14}$$

For the transition matrix, $a_{11} = 0.9518$ is significantly larger than $a_{12} = 0.0482$, which means that if the load has an increasing tendency at the current time, the probability that it will still increase at the next moment is much greater than the possibility of decreasing. In general, this kind of dynamic characteristics can be named as inertia load.

In addition, observing the emission matrix $\boldsymbol{B} = \{b_{ij}\}$, we can conclude as follows, when the load level is low, the probability of the load increasing at the next moment is large; while when the load level is high, the load is likely to decrease at the next moment.

Then an HMM for a single user is trained. The **FIGURE 10** shows the weekly load curve of a single power user. Here, this single load curve is also converted into an observation sequence with ten discretization levels too.

After training the observation sequence using the Baum-Welch algorithm, we find that the algorithm does not converge within the upper limit of the iteration (500 times). The transition matrix $\boldsymbol{A} = \{a_{ij}\}$ and observation emission matrix

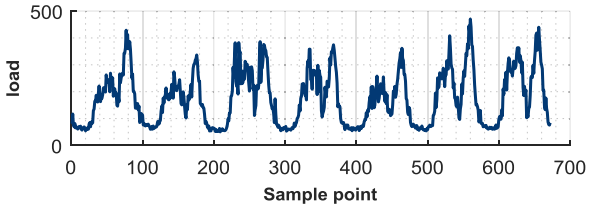

FIGURE 10. Weekly load curve of a single load.

$\boldsymbol{B} = \{b_{ij}\}$ are obtained as follows.

$$\boldsymbol{A} = \begin{pmatrix} 0.0822 & 0.9178 \\ 0.3374 & 0.6626 \end{pmatrix}, \quad \boldsymbol{B}' = \begin{pmatrix} 0.0167 & 0.0264 \\ 0.3055 & 0.2988 \\ 0.1190 & 0.1191 \\ 0.0418 & 0.0416 \\ 0.0450 & 0.0445 \\ 0.0528 & 0.0539 \\ 0.0583 & 0.0559 \\ 0.0812 & 0.0800 \\ 0.0875 & 0.0900 \\ 0.1923 & 0.1898 \end{pmatrix} \tag{15}$$

As can be seen, the load curve in **FIGURE 10** fluctuates greatly. Furthermore, the emission matrix of the aggregated load in (14) is more regular that of the single load in (15). Specifically, in the emission matrix $\boldsymbol{B}_1$ in (14), the structure of the matrix presents a certain symmetry, because it contains several continuous zero elements; while the elements of the emission matrix in (15) are of disorder, this shows the fact that observation result in (14) is limited in only a few values, while the observed result in (15) is more uncertain for a particular hidden state. To conclude, the HMM with emission matrix in (14) has strong regularity, while the emission matrix in (15) is of more possible transitions and presents a high degree of uncertainty.

Therefore, it is reasonable to model the dynamic behavior of weekly load curves based on aggregated or clustering results, rather than a single load curve.

D. EVALUATION OF DIFFERENT FORECASTING METHODS

In this section, the Multiple Linear Regression (MLR) and Neural Network (NN) Model are utilized to illustrate the proposed HMM-based forecasting evaluation.

1) THE UTILIZED MLR AND NN METHODS

In order to construct the training dataset, the input feature parameters and output parameter are described below.

The input parameter is $x = (x_1, x_2 \cdots x_6)$, where x_1 represents the load of the same time on the previous day, x_2 represents load of the same time and same day in the previous week, x_3 represents the previous day's average load, x_4 represents the hour of the day, x_5 represents the day of the week, and x_6 indicates whether the day is a holiday or not. And the output parameter y represents the current load value.

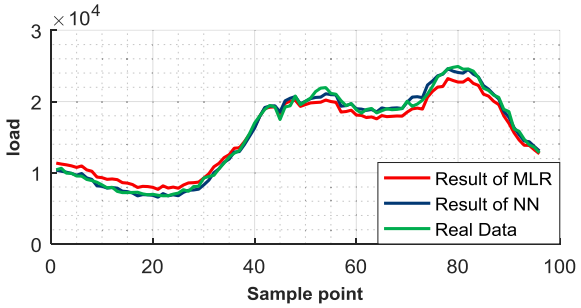

FIGURE 11. Result for cluster 1. Forecasting results and real value.

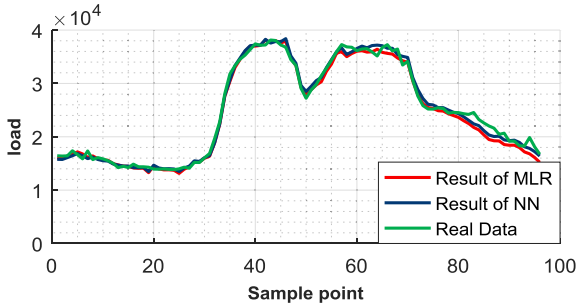

FIGURE 12. Result for cluster 2. Forecasting results and real value.

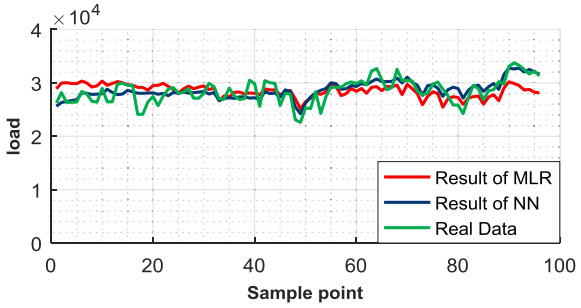

FIGURE 13. Result for cluster 3. Forecasting results and real value.

For MLR, we aim to find the regression coefficients $\beta = (\beta_0, \beta_1 \cdots \beta_6)$ making the input and output satisfy the following equation.

$$y_{MLR} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_6 x_6 \tag{16}$$

The detailed information about MLR can be found in [40].

For NN, a simple backpropagation network is utilized which has an input layer, one hidden layer with 10 nodes and an output layer. The sigmoid function is utilized as the activation function [41]. More details about NN can be found in [42].

2) EVALUATION RESULT

In this section, MLR and NN are utilized to perform load forecasting on the three typical aggregated load patterns which is obtained above, the forecasting results are shown in **FIGURE 11**, **FIGURE 12** and **FIGURE 13**, respectively.

Then, based on the trained HMM, the probabilities of the NN and MLR sequences can be calculated. In order to verify the evaluation effectiveness, the mean absolute percentage error (MAPE) of each forecasting result is assessed.

TABLE 2. HMM probability and MAPE for NN and MLR sequence.

	Forecast Method	HMM Probability	MAPE
Cluster 1	MLR	3.3011×10^{-39}	4.1794%
	NN	4.2424×10^{-35}	2.6722%
	Forecast Method	**HMM Probability**	**MAPE**
Cluster 2	MLR	4.1468×10^{-34}	2.6056%
	NN	1.7888×10^{-36}	3.1055%
	Forecast Method	**HMM Probability**	**MAPE**
Cluster 3	MLR	1.6341×10^{-53}	6.4223%
	NN	1.0997×10^{-40}	4.3402%

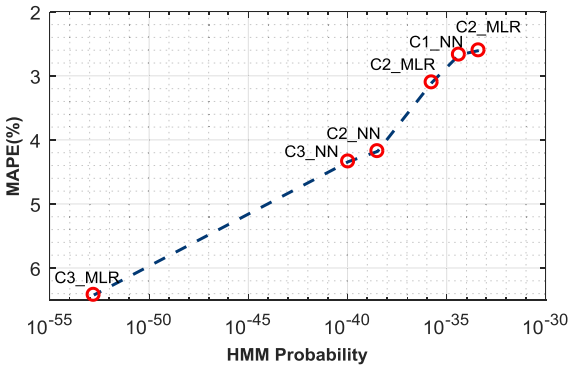

FIGURE 14. HMM Probability and its corresponding MAPE.

The HMM probabilities and MAPE results are illustrated in **TABLE 2**.

For cluster 1 and 3, the NN forecasting results are more accurate, while for Cluster 2, the MLR presents better accuracy. Therefore, for further application, combined forecasting methods should be utilized to improve the accuracy of forecasting but not single MLR or NN.

Further, the data in **TABLE 2** are illustrated in **FIGURE 14**, in which the X-axis is in logarithmic form. It can be found that the points in the coordinate system exhibit an approximate linear relationship.

This linear relationship indicates that there is a close relationship between HMM probability and MAPE. At the same time, it can be proved that the HMM can extract typical load dynamic behavior characteristics from the historical data of the load, and the typical behavior has better consistency in the range of the daily time scale.

VI. CONCLUSIONS

In this paper, a data mining approach based on the weekly load curves is proposed. Firstly, PAA technique is utilized to transform the dense fluctuating load data into piecewise

paned data. Then, a Davies-Bouldin index (DBI) based adaptive k-means algorithm is utilized to cluster the studied users into several groups, where the optimal clustering number can be set automatically without prior knowledge. Finally, a hidden Markov model (HMM) is established to describe the probabilistic transitions of different load levels in the aggregated load curve of each cluster. The proposed model can characterize the representative dynamic weekly demand features. Moreover, it also provides a feasible tool to evaluate the performance of the short-term load forecasting, which realizes the pre-check for the forecasting results without future real measurements in the forecasting horizon.

REFERENCES

[1] A. Cooper, "Electric company smart meter deployments: foundation for a smart grid," Inst. Electr. Innov., Washington, DC, USA, Tech. Rep., 2016.
[2] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, and Y. Zhao, "Load profiling and its application to demand response: A review," *Tsinghua Sci. Technol.*, vol. 20, no. 2, pp. 117–129, Apr. 2015.
[3] G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 933–940, May 2006.
[4] Q.-H. Zhao, M.-H. Ha, G.-B. Peng, and X.-K. Zhang, "Support vector machine based on half-suppressed fuzzy c-means clustering," in *Proc. Int. Conf. Mach. Learn. Cybern.*, vol. 2, 2009, pp. 1236–1240.
[5] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proc. 8th ACM SIGMOD Workshop Res. Issues Data Mining Knowl. Discovery*, 2003, pp. 2–11.
[6] R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, "Development of low voltage network templates—Part I: Substation clustering and classification," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3036–3044, Nov. 2015.
[7] R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, "Development of low voltage network templates—Part II: Peak load estimation by clusterwise regression," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3045–3052, Nov. 2015.
[8] S. Varuna and P. Natesan, "An integration of k-means clustering and naïve Bayes classifier for Intrusion Detection," in *Proc. 3rd Int. Conf. Signal Process., Commun. Netw. (ICSCN)*, Mar. 2015, pp. 1–5.
[9] H. K. Temraz, M. M. A. Salama, and A. Y. Chikhani, "Review of electric load forecasting methods," in *Proc. Can. Conf. Elect. Comput. Eng., Eng. Innov., Voyage Discovery, Conf. (CCECE)*, vol. 1, 1997, pp. 289–292.
[10] K. Zor, O. Timur, and A. Teke, "A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting," in *Proc. 6th Int. Youth Conf. Energy (IYCE)*, 2017, pp. 1–7.
[11] J. Ye, "Least squares linear discriminant analysis," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 1087–1093.
[12] P. Baldi and K. Hornik, "Neural networks and principal component analysis: Learning from examples without local minima," *Neural Netw.*, vol. 2, no. 1, pp. 53–58, 1989.
[13] H. Liao and D. Niebur, "Load profile estimation in electric transmission networks using independent component analysis," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 707–715, May 2003.
[14] N. Abu-Shikhah and F. Elkarmi, "Medium-term electric load forecasting using singular value decomposition," *Energy*, vol. 36, no. 7, pp. 4259–4271, 2011.
[15] C. Y. Jin, "A pattern recognition method for electric nose based on kernel-principal component analysis (PCA) and support vector machine (SVM)," *J. Beijing Univ. Chem. Technol.*, vol. 39, no. 2, pp. 106–109, Feb. 2012.
[16] Z.-C. Yang, "Electric load movement evaluation and forecasting based on the fourier-series model extend in the least-squares sense," *J. Control, Automat. Elect. Syst.*, vol. 26, no. 4, pp. 430–440, 2015.
[17] Q. Huang, L. Shao, and L. Na, "Dynamic detection of transmission line outages using hidden Markov models," in *Proc. Amer. Control Conf.*, 2015, pp. 5050–5055.
[18] T. Ying and Q. Qiu, "A framework of stochastic power management using hidden Markov model," in *Proc. Design, Automat. Test Eur.*, 2008, pp. 92–97.
[19] K.-C. Kwon, J.-H. Kim, and P.-H. Seong, "Hidden Markov model-based real-time transient identifications in nuclear power plants," *Int. J. Intell. Syst.*, vol. 17, no. 8, pp. 791–811, 2002.
[20] X.-J. Wang, L. Chen, and W.-Q. Tao, "Research on load classification based on user's typical daily load curve," in *Proc. IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, 2017, pp. 1–4.
[21] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *J. Mach. Learn. Res.*, vol. 11, no. 1, pp. 2837–2854, 2010.
[22] C. Cheadle, M. P. Vawter, W. J. Freed, and K. G. Becker, "Analysis of microarray data using Z score transformation," *J. Mol. Diag. Jmd*, vol. 5, no. 2, pp. 73–81, 2003.
[23] V. Gajera, Shubham, R. Gupta, and P. K. Jana, "An effective multi-objective task scheduling algorithm using min-max normalization in cloud computing," in *Proc. 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol.*, 2017, pp. 812–816. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7912111/
[24] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowl. Inf. Syst.*, vol. 3, no. 3, pp. 263–286, 2001.
[25] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, vol. 87. Springer Science & Business Media, 2013. [Online]. Available: https://link.springer.com/book/10.1007%2F978-1-4419-8853-9
[26] A. Ben-Tal and A. Nemirovski, "Robust convex optimization," *Math. Oper. Res.*, vol. 23, no. 4, pp. 769–805, 1998.
[27] G. Shi, B. Gao, and L. Zhang, "The optimized K-means algorithms for improving randomly-initialed midpoints," in *Proc. 2nd Int. Conf. Meas., Inf. Control*, vol. 2, Aug. 2013, pp. 1212–1216.
[28] E. Carpaneto, G. Chicco, R. Napoli, and M. Scutariu, "Electricity customer classification using frequency–domain load pattern data," *Int. J. Electr. Power Energy Syst.*, vol. 28, no. 1, pp. 13–20, 2006.
[29] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, "Load pattern-based classification of electricity customers," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 1232–1239, May 2004.
[30] G. J. Tsekouras, N. D. Hatziargyriou, and E. N. Dialynas, "Two-stage pattern recognition of load curves for classification of electricity customers," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1120–1128, Aug. 2007.
[31] G. Chicco and J. S. Akilimali, "Renyi entropy-based classification of daily electrical load patterns," *IET Gener. Transmiss. Distrib.*, vol. 4, no. 6, pp. 736–745, Jun. 2010.
[32] S. Ramos, Z. Vale, J. Santana, and J. Duarte, "Data mining contributions to characterize MV consumers and to improve the suppliers-consumers settlements," in *Proc. IEEE Power Eng. Soc. General Meeting*, Jun. 2007, pp. 1–8.
[33] M. R. Hassan and B. Nath, "Stock market forecasting using hidden Markov model: A new approach," in *Proc. 5th Int. Conf. Intell. Syst. Design Appl. (ISDA)*, Sep. 2005, pp. 192–196.
[34] S.-Z. Yu and H. Kobayashi, "An efficient forward-backward algorithm for an explicit-duration hidden Markov model," *IEEE Signal Process. Lett.*, vol. 10, no. 1, pp. 11–14, Jan. 2003.
[35] R. A. Farrell, E. P. Gray, and R. W. Hart, "Demonstration of a Jacobi polynomial representation of a kronecker delta function," Johns Hopkins Univ., Laurel, MD, USA, Tech. Rep., 1985.
[36] L. R. Welch, "Hidden Markov models and the Baum-Welch algorithm," *IEEE Inf. Theory Soc. Newslett.*, vol. 53, no. 4, pp. 10–13, Dec. 2003.
[37] P. Ray, S. Sen, and A. K. Barisal, "Hybrid methodology for short-term load forecasting," in *Proc. IEEE Int. Conf. Power Electron., Drives Energy Syst. (PEDES)*, Dec. 2015, pp. 1–6.
[38] A. M. Zakharin, "Some Markov models of systems with partial aftereffect," *Cybernetics*, vol. 16, no. 2, pp. 175–188, 1980.
[39] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[40] D. F. Andrews, "A robust method for multiple linear regression," *Technometrics*, vol. 16, no. 4, pp. 523–531, 1974.
[41] Y. Ito, "Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory," *Neural Netw.*, vol. 4, no. 3, pp. 385–394, 1991.
[42] T. C. A. Goh, "Back-propagation neural networks for modeling complex systems," *Artif. Intell. Eng.*, vol. 9, no. 3, pp. 143–151, 1995.

SHIXIANG LU was born in Xuzhou, Jiangsu, China, in 1985. He received the B.S. degree in electrical engineering and automation from the China University of Mining and Technology, Xuzhou, Jiangsu, China, in 2009, and the M.S. and Ph.D. degrees in pattern recognition and intelligent system from the University of Chinese Academy of Sciences, Beijing, China, in 2014.

Since 2014, he has been a Senior Engineer with Guangdong Power Grid Corporation, Guangzhou, Guangdong, China. His research interests include the data mining, and data application in power systems.

GUOYING LIN was born in Meizhou, Guangdong, China, in 1982. He received the B.S. degree in electrical engineering and automation from Zhejiang University, Hangzhou, Zhejiang, China, in 2004, and the M.S. degree in electrical engineering and automation from Shanghai Jiao Tong University, Shanghai, China, in 2007.

Since 2007, he has been a Senior Engineer with the Guangdong Power Grid Corporation, Guangzhou, Guangdong, China. His research interests include the data mining, and data application in power systems.

HANLIN LIU was born in Dalian, Liaoning, China, in 1993. He received the B.S. degree from Zhejiang University, Hangzhou, Zhejiang, China, in 2016, where he is currently pursuing the M.S degree in electrical engineering under the supervision of Prof. Y. Ding.

He has authored or co-authored several articles in load forecasting and load clustering. His current research interests include the application of big data in power systems, and load forecasting and cluster analysis.

CHENGJIN YE was born in Linan, Zhejiang, China, in 1987. He received the B.E. and Ph.D. degrees in electrical engineering from Zhejiang University, China, in 2010 and 2015, respectively. He served as a Distribution System Engineer for the Economics Institute of State Grid Zhejiang Electric Power Co., Ltd., from 2015 to 2017.

Since 2017, he has been with the Smart Grid Operation and Optimization Laboratory, where he has also been a Postdoctoral Researcher with the College of Electrical Engineering, Zhejiang University. His research interests include data-driven power system planning and operation, short-circuit current limitation, and grid resilience enhancement considering extreme natural disasters.

HUAKUN QUE was born in Longyan, Fujian, China, in 1986. He received the B.S. and M.S. degrees in communication engineering from North China Electric Power University, Beijing, China, in 2008 and 2011, respectively.

Since 2011, he has been a Senior Engineer with Guangdong Power Grid Corporation, Guangzhou, Guangdong, China. His research interests include the data mining, and data application in power systems.

YI DING received the B.Eng. degree in electrical engineering from Shanghai Jiao Tong University, China, and the Ph.D. degree in electrical engineering from Nanyang Technological University (NTU), Singapore. He was an Associate Professor with the Department of Electrical Engineering, Technical University of Denmark, Denmark. He also held research and teaching positions in the University of Alberta, Canada, and NTU. He is currently a Professor with the College of Electrical Engineering, Zhejiang University, China.

His research interests include power system planning and reliability evaluation, smart grid, and complex system risk assessment. He was a Consultant as Energy Economist for Asian Development Bank, in 2010. He is a member of IEC working groups for micro-grid standards. He is an Editorial Member of International Journals of *Electric Power System Research* and the *Journal of Modern Power Systems and Clean Energy*. He is also a Guest Editor for the special section of the IEEE TRANSACTIONS ON POWER SYSTEMS.