IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 1, 2019, accepted October 17, 2019, date of publication October 25, 2019, date of current version November 19, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2949726

Short-Term Load Forecasting Based on Deep Learning for End-User Transformer Subject to Volatile Electric Heating Loads

QIFANG CHEN[1], (Member, IEEE), MINGCHAO XIA[1], (Senior Member, IEEE), TENG LU[1], XICHEN JIANG[2], (Member, IEEE), WENXIA LIU[3], (Member, IEEE), AND QINFEI SUN[4]

[1]School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China
[2]Department of Engineering and Design, Western Washington University, Bellingham, WA 98225, USA
[3]State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing 102206, China
[4]State Grid Beijing Electric Power Company, Beijing 100031, China

Corresponding author: Mingchao Xia (mchxia@bjtu.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 51677003, in part by the Project funded by China Postdoctoral Science Foundation under Grant 2018M631326, in part by the State Grid Corporation Science and Technology Project under Grant 52020118000M, and in part by the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources under Grant LAPS18018.

ABSTRACT Short-Term Load Forecasting (STLF) for End-User Transformer Level (EUTL) is challenging due to the high penetration of Electric Heating Loads (EHLs), which exhibit significant uncertainty, nonlinearity, and variability. In this paper, a STLF model is proposed based on the Stacked Auto-Encoder Extreme Learning Machine (SAE-ELM) deep learning framework, which can be used to extract hidden features from the time series load data. In order to improve the capability of extracting deep and diverse features from the data and generate a useful knowledge representation structure, a novel specialized feature indices set is proposed to construct the training sample set. The sliding trend, fluctuation rate, grade of change, and smoothness of the time series are considered and quantified as elements of the training sample set. Then, deep nonlinear features are extracted by using the SAE-ELM with no iterative parameter tuning needed. To illustrate the validity of the proposed model, five numerical cases are conducted. Comparison of results shows that the proposed model improves the capability and sensitivity of dealing with load volatility and forecasting accuracy.

INDEX TERMS Short-term load forecasting, feature representation, deep learning, stacked auto-encoder, extreme learning machine.

I. INTRODUCTION

Short-Term Load Forecasting (STLF) is essential for maintaining the balance between supply and demand in a power system [1], [2]. The rapid development of the smart grid in addition to the deregulation of electricity markets resulted in many smart grid applications becoming increasingly customer-oriented (i.e., demand response [3], [4]) [5]. Load forecasting at End-User Transformer-Level (EUTL) is an important research area that supports many smart applications and is essential for service providers that manage loads and renewable resources. However, due to the uncertainties associated with renewable generation [6] and new electric loads (e.g., air-conditioning [7], electric heating load (EHL) [8], electric vehicle charging stations [9]), the load profile at the transformer is now increasingly non-stationary, making STLF at EUTL challenging.

In recent decades, a number of researchers have investigated STLF. In [10], a distributed STLF method for a bulk power system covering large geographical areas was presented. In order to improve the accuracy of STLF, the bulk power system was partitioned into several subnetworks and their loads were forecasted using weather information. In [11], a STLF model based on support vector regression and two-step hybrid parameters optimization was carried out for distribution system load forecasting. A STLF model considering the hierarchical structure of distribution system was presented in terms of load features in [12]. In order to

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

improve the forecasting accuracy, a hybrid STLF model was proposed for a bulk power system in [13]. In [14], a two-stage hybrid STLF method was proposed. The first stage used a time-series load forecasting model where the forecasted error of the first stage was corrected in the second stage by performing a deep analysis on the impact of relative factors on the model error. Although these aforementioned STLF methods in [10]–[14] may have good STLF performance at the system-level, it cannot be applied to EUTL load forecasting directly. This is because while the load capacity at EUTL is much smaller, the load variability is much larger compared against the load capacity at the system-level. Furthermore, no additional weather information is available at EUTL due to the prohibitive deployment cost and advanced technological requirement. Therefore, the accuracy of the STLF model at EUTL is mainly dependent on the how the historical load data is processed. The raw load data at EUTL always exhibits significant irregularities such as uncertainty, nonlinearity, and large volatility. An effective deep feature extraction method is critical for forming knowledge representation structure to make full use of the deep hidden information from the historical data.

In general, deep feature extraction from historical data can be conducted in two ways: manual data preprocessing by rules and through a deep learning neural network. The seasonality of the load was regarded as an important feature in [2], [15]. In [2], ten different features of the historical load time series were used as latent regressors that were suspected to influence the current hourly load (i.e., loads from the previous week, hours of the day, etc.) in order to develop appropriate regression models. In [15], the representation of seasonality was non-parametric and the daily and weekly cycles were not expressed as mathematical functions but represented as the mean of previously observed load patterns. The feature of seasonality was not an integral part of the forecasting algorithm. In [16], a two-stage linear and nonlinear combined hybrid forecasting model for mixed-use complex STLF was proposed. The labelling process considers the type of day: workday or holiday. Load type is used to decompose the aggregated load into several sub-clusters. These aforementioned custom data preprocessing methods were able to capture some shallow features, but proved unsuitable for mining deep correlation features. In addition, the impacts of some important external factors such as electricity price were not considered. In [17], [18], the wavelet decomposition (WD) based STLF method was used for extracting deep features in the frequency domain. The method reduces the non-stationary characteristics of the load profile by decomposing the original signal into many components with different frequencies. In [19] , the empirical mode decomposition (EMD) method was used to decompose energy price time series data for developing a forecasting model. The data was decomposed into intrinsic mode functions (IMFs) in the time domain with the same length as the original signal, which proved useful for analyzing the features directly. Both WD and EMD were useful in feature extraction from raw time series data, but the relationships and associations between other external factors were not included in the decomposed components.

Deep learning has been used extensively in extracting nonlinear features from data [20]. Frequently used deep learning algorithms for STLF models include deep belief network (DBN) [21], deep recurrent neural network (DRNN) [22], and deep convolutional neural network (DCNN) [23]. In [3], a long short-term memory (LSTM) RNN-based framework was presented to tackle the forecasting problem of a single residential user load. In [24], the DRNN and DCNN were carried out for short-term load forecasting and medium-term load forecasting. In [25], DRNN and DCNN were presented for day-ahead multi-step load forecasting in commercial buildings. In [26], a deep belief network was built for an hourly load forecasting model of a bulk power system. All these deep learning algorithms exhibit excellent performance in STLF.

Stacked auto-encoder (SAE) is a type of feed forward neural network that is composed of an encoder and decoder [27]. It is trained to reconstruct its own input in the output layer in an unsupervised manner. This can be viewed as an advanced feature extractor capable of preserving the hidden abstractions and has an invariant structure in its inputs [28]. A new kind of SAE combined with extreme learning machine (SAE-ELM) was proposed to improve the computational efficiency of deep learning by the authors of [29]. Because of its fast training speed and high generalizability, S-ELM has found applications in many areas, such as representational learning [30], [31], data partitioning [32], pattern recognition [33], and soft sensor modeling [34]. In this paper, a deep learning framework based on SAE-ELM is introduced to construct the STLF model. The contributions of the proposed work are as follows.

1) A novel specialized feature extracting model is proposed to construct a knowledge representation structure for the training sample set with a variety of deep features.
2) The EMD method is utilized to decompose the raw time series data in time domain. Instead of forecasting each individual IMF independently, the EMD method is used as a low pass filter to eliminate the non-stationary IMF.
3) The SAE-ELM based deep learning framework is introduced to construct the STLF model, which is used to predict irregularities at EUTL. By using SAE-ELM, the deep nonlinear features are learned automatically with no iterative parameter turning or fine-turning needed, which renders the training process very fast.

The remainder of this article is organized as follows. The irregularities at EUTL and the specialized feature extraction model are presented in Section II. Section III presents the SAE-ELM model that is utilized in the STLF method. Case studies are conducted and the results of which are discussed in Section IV. Finally, concluding remarks are made in Section V.

II. IRREGULARITY ANALYSIS AND SPECIALIZED FEATURE EXTRACTION

A. IRREGULARITY IN END-USER TRANSFORMER CONTAINING HIGH PENETRATION OF ELECTRIC HEATING LOAD

The power consumption of an EHL is typically much higher than conventional household appliances. These loads have closed-loop controllers that turn the EHL on and off in order to maintain the temperature within setpoints. Thus, the load profile of a household with EHL exhibit large random fluctuations throughout the day. An end-user transformer serves several households, so its load profile is an aggregate of the profiles of the individual households. Due to the small capacity and the high penetration of EHLs, randomness and volatility are prevalent in the load profile of an end-user transformer.

The load profile of a typical day of end-user transformer with high penetration of EHLs is shown in Fig. 1. The noticeable features include high peak-valley difference, sudden large changes, and random fluctuations. There are also other hidden features such as sliding trend, variation with different frequency, and impact from other external factors such as electricity price. All these features need to be extracted from the raw time series data and quantified into a knowledge representation that is easily learned. Then, it is utilized as the input training sample set for the deep learning process in order to ensure that the SAE-ELM model is well trained with enough knowledge.

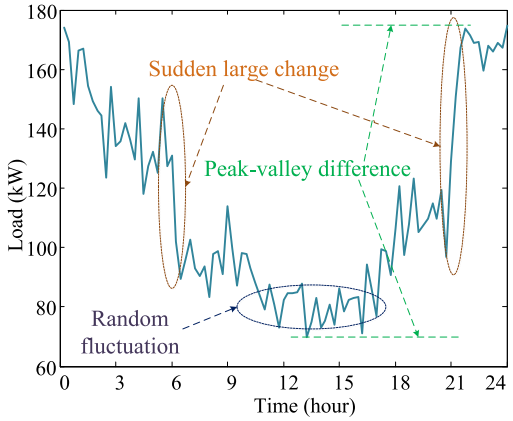


FIGURE 1. Representative daily load profile at EUTL.

B. THE NOVEL SPECIALIZED FEATURE EXTRACTING INDICES

A set of novel specialized feature extracting indices are presented to extract and quantify features from raw data.

1) THE SLIDING TREND (SlT)

Let $\mathbf{P}$ indicate a vector of active power consumed by the load as shown in (1), and t indicate the current time interval. The SlT represents the average sliding trend during a specific time window. It considers both the current state and the m previous states, with m being the window length. Therefore, it is defined by the sliding average model shown in (2).

$$\mathbf{P} = [p_l(1), \cdots, p_l(t), \cdots, p_l(r)] \tag{1}$$

$$SlT(t) = \frac{1}{m}\sum_{i=0}^{m-1} p_l(t-i) \tag{2}$$

where $SlT(t)$ is the sliding trend index; m is the length of sliding trend window; r is the number of time intervals in a day; $p_l(t)$ is the active power of load.

2) THE FLUCTUATING RATE (FlR)

The FlR is defined to capture the fluctuation direction and the relationship between the fluctuation and SlT.

$$FlR(t) = \frac{(p_l(t) - p_l(t-1))}{SlT(t)} \tag{3}$$

3) THE GRADE OF CHANGE (GoC)

The volatility of the load is divided into 5 grades with the GoC quantifying the grade of change.

$$GoC(t) = \begin{cases} 2, & FlR(t) > g_1^+, \text{ grade 5} \\ 1, & \bar{g}_0^+ < FlR(t) \le g_1^+, \text{ grade 4} \\ 0, & g_0^- \le FlR(t) \le \bar{g}_0^+, \text{ grade 3} \\ -1, & g_1^- \le FlR(t) < g_0^-, \text{ grade 2} \\ -2, & FlR(t) < g_1^-, \text{ grade 1} \end{cases} \tag{4}$$

where $GoC(t)$ is the GoC index at time t; g_1^- is the threshold separating grades 1 and 2; g_0^- is the threshold separating grades 2 and 3; $\bar{g}_0^+$ is the threshold separating grades 3 and 4; g_1^+ is the threshold separating grades 4 and 5. Grade 3 indicates that the volatility is small. Grades 2 and 4 denote larger volatility, and grades 1 and 5 represent the highest volatility.

4) THE SMOOTHNESS OF SERIES (SoS)

The EMD method performs auto-adaptive nonlinear analysis on non-stationary time series. The original time series can be decomposed into several IMFs and one residue. The main steps of the EMD method are shown in Table 1.

The raw load time series data can be decomposed into individual frequency components as shown in (6).

$$\mathbf{P} = \sum_{i=1}^{N_{imf}} \mathbf{s}_{imf}^i + \mathbf{s}_{re} \tag{6}$$

where N_{imf} is the number of IMFs.

Let $\mathbf{s}_{imf}^1$ be the highest frequency component, which is also the most non-stationary component. The other components $\mathbf{s}_{imf}^i$ are of decreasing frequencies with increasing i. Because $\mathbf{s}_{imf}^1$ is highly unpredictable, no useful features can be extracted from it.

The SoS index is defined as the stationary series which is extracted from the raw series by neglecting the non-stationary and highly unpredictable IMF. Therefore, the SoS index is

TABLE 1. Steps of the EMD method.

Main steps of EMD
Step 1: Get the original load time series $\mathbf{P}$ ready.
Step 2: Determine the local extreme points, $\mathbf{p}_{\max}$ and $\mathbf{p}_{\min}$, of time series $\mathbf{P}$ and calculate the upper and lower envelopes by interpolation, $\mathbf{p}_{\text{up}}$ and $\mathbf{p}_{\text{low}}$.
Step 3: Compute the mean series $\mathbf{m}$ of the upper and lower envelop using (5). $$\mathbf{m} = \frac{\left(\mathbf{p}_{\text{up}} - \mathbf{p}_{\text{low}}\right)}{2} \quad (5)$$
Step 4: Calculate new IMF series $\mathbf{s}_{\text{imf}} = \mathbf{P} - \mathbf{m}$ and the residue series $\mathbf{s}_{\text{re}} = \mathbf{P} - \mathbf{s}_{\text{imf}}$.
Step 5: Replace the original time series $\mathbf{P}$ with $\mathbf{s}_{\text{re}}$ and repeat steps 2 to 4, until the termination criterion is met.

defined as the sum of all IMFs except the highest frequency component, $\mathbf{s}_{imf}^{1}$, as shown in (7).

$$SoS = \sum_{i=2}^{N_{imf}} \mathbf{s}_{imf}^{i} + \mathbf{s}_{re} \quad (7)$$

Instead of only using the raw load time series as training sample set, all the proposed feature indices along with the raw load time series are used to construct the new knowledge representation set, which is shown in (8). Then, the input training sample set for SAE-ELM is generated from the knowledge representation set.

$$\mathbf{F} = \begin{Bmatrix} p_l(1) & \cdots & p_l(t) & \cdots & p_l(r) \\ SlT(1) & \cdots & SlT(t) & \cdots & SlT(r) \\ FlR(1) & \cdots & FlR(t) & \cdots & FlR(r) \\ GoC(1) & \cdots & GoC(t) & \cdots & GoC(r) \\ SoS(1) & \cdots & SoS(t) & \cdots & SoS(r) \end{Bmatrix} \quad (8)$$

III. STFL MODEL BASED ON SAE-ELM

A. THE STACKED AUTO-ENCODER BASED ON EXTREME LEARNING MACHINE

An auto-encoder (AE) is a neural network that is trained to approximately reconstruct its input in the output layer. It is a feed forward neural network consisting of an encoder and a decoder, a typical architecture of which is shown in Fig. 2.

Let $\mathbf{h} \in \mathbb{R}^{N \times L}$ be the matrix of hidden layer, the input vector $\mathbf{x}$ is mapped to the hidden layer by (9) and reconstructed by (10).

$$\mathbf{h} = g(\mathbf{a}\mathbf{x} + \mathbf{b}) \quad (9)$$

$$\widehat{\mathbf{x}} = \mathbf{h}\boldsymbol{\beta} \quad (10)$$

where $\mathbf{x} \in \mathbb{R}^{N}$ is the input vector. $g(\mathbf{a}, \mathbf{b}, \mathbf{x})$ is the activation function of the hidden layer; in this paper, we use a sigmoid function. $\mathbf{a}$ is the weight matrix mapping the input layer to the hidden layer; $\mathbf{b}$ is the bias vector of hidden layer; $\boldsymbol{\beta}$ is the output weight vector.

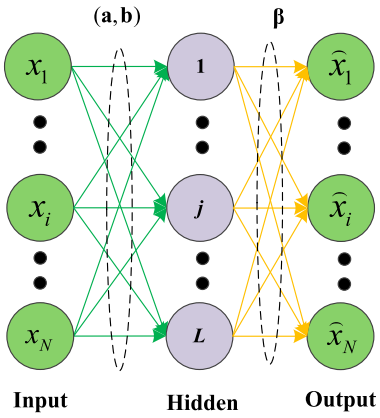


FIGURE 2. The typical architecture of an auto-encoder.

Because the AE model prioritizes the aspects of the input that should be copied, it often learns useful properties of the data. In order to learn sufficient representation for achieving good generalization performance, SAE is often used for mining deep features from data. However, as the number of layers increase, the complexity and computation of the conventional SAE trained with BP algorithm also increases. In this paper, the deep learning framework based on SAE-ELM model is introduced as the core of STLF model, which improves the training efficiency significantly.

The STLF model is composed of an unsupervised deep feature extraction (UDFE) module and a forecasting module (FM), as shown in Fig. 3. The total number of layers of the deep learning framework is $k + 2$, which includes the input and output layers. The UDFE is composed of all layers between the input layer and the $k - 1^{\text{th}}$ hidden layer, wherein each hidden layer includes an AE-ELM. The FM is comprised of the $k - 1^{\text{th}}$ hidden layer, the k^{th} hidden layer and the output layer.

We take the AE-ELM of q^{th} hidden layer to illustrate the process of training AE with ELM. As shown in Fig. 4, the training process of AE-ELM is composed of four main steps.

Step 1

Use the output $\mathbf{H}_{(q-1)}$ of the $q - 1^{\text{th}}$ layer as the input of the q^{th} hidden layer.

Step 2

Map $\mathbf{H}_{(q-1)}$ to the q^{th} hidden layer space with random input weight matrix $\mathbf{a}_{(q)}$, bias vector $\mathbf{b}_{(q)}$, and the activation function $g(\cdot)$ in (11). Then, the output $\widehat{\mathbf{H}}_{(q-1)}$ of the q^{th} hidden layer is given by (12).

$$\mathbf{H}_{(q)}^{\text{AE}} = g\left(\mathbf{a}_{(q)}\mathbf{H}_{(q-1)} + \mathbf{b}_{(q)}\right) \quad (11)$$

$$\widehat{\mathbf{H}}_{(q-1)} = \mathbf{H}_{(q)}^{\text{AE}}\boldsymbol{\beta}_{(q)} \quad (12)$$

where $\boldsymbol{\beta}_{(q)}$ is the output weight of AE in the q^{th} hidden layer.

Step 3

In order to promote generalization, the output weight $\boldsymbol{\beta}_{(q)}$ is obtained by solving the optimization problem

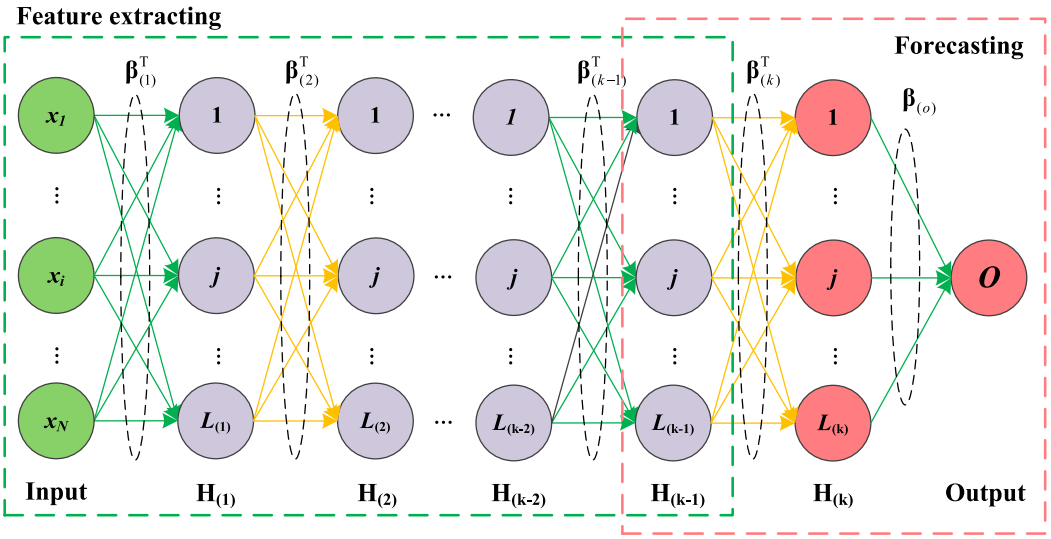


FIGURE 3. Deep learning framework for the STLF model.

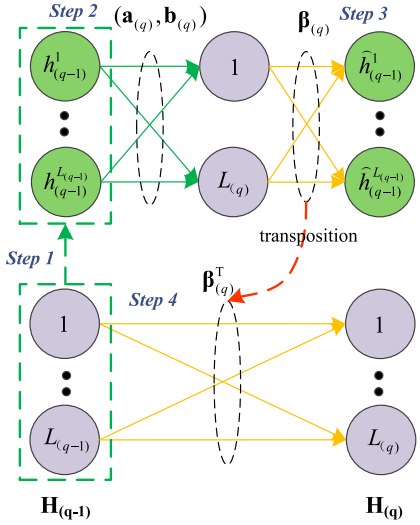


FIGURE 4. The procedure of AE-ELM training.

of (13).

$$\min \ \tfrac{1}{2} \left\|\boldsymbol{\beta}_{(q)}\right\|^2 + \frac{C}{2} \left\|\widehat{\mathbf{H}}_{(q-1)} - \mathbf{H}^{\mathrm{AE}}_{(q)} \boldsymbol{\beta}_{(q)}\right\|^2 \tag{13}$$

where the first item is the regularization term with overfitting. C is the penalty coefficient. The objective function minimizes both the output weight and output error simultaneously.

Next, we take the gradient of (13) with respect to $\boldsymbol{\beta}_{(q)}$ and equate it to zero as shown in (14) and solve for the output weight $\boldsymbol{\beta}_{(q)}$ as in (15) if $\mathbf{H}^{\mathrm{AE}}_{(q)}$ is full column rank; Otherwise, if $\mathbf{H}^{\mathrm{AE}}_{(q)}$ is full row rank, the output weight $\boldsymbol{\beta}_{(q)}$ can be obtained by (16).

$$\boldsymbol{\beta}_{(q)} + C \left(\mathbf{H}^{\mathrm{AE}}_{(q)}\right)^{\mathrm{T}} \left(\widehat{\mathbf{H}}_{(q-1)} - \mathbf{H}^{\mathrm{AE}}_{(q)} \boldsymbol{\beta}_{(q)}\right) = 0 \tag{14}$$

$$\boldsymbol{\beta}_{(q)} = \left(\left(\mathbf{H}^{\mathrm{AE}}_{(q)}\right)^{\mathrm{T}} \mathbf{H}^{\mathrm{AE}}_{(q)} + \frac{\mathbf{I}}{C}\right)^{-1} \left(\mathbf{H}^{\mathrm{AE}}_{(q)}\right)^{\mathrm{T}} \widehat{\mathbf{H}}_{(q-1)} \tag{15}$$

$$\boldsymbol{\beta}_{(q)} = \left(\mathbf{H}^{\mathrm{AE}}_{(q)}\right)^{\mathrm{T}} \left(\mathbf{H}^{\mathrm{AE}}_{(q)} \left(\mathbf{H}^{\mathrm{AE}}_{(q)}\right)^{\mathrm{T}} + \frac{\mathbf{I}}{C}\right)^{-1} \widehat{\mathbf{H}}_{(q-1)} \tag{16}$$

where $\mathbf{I}$ is the identity matrix.

Step 4

After the training of AE is completed, $\boldsymbol{\beta}^{\mathrm{T}}_{(q)}$ is used as the mapping weight between the $q-1^{\mathrm{th}}$ hidden layer and the q^{th} hidden layer. Then the output of the q^{th} hidden layer is obtained by (17).

$$\mathbf{H}_{(q)} = g\left(\mathbf{H}_{(q-1)} \boldsymbol{\beta}^{\mathrm{T}}_{(q)}\right) \tag{17}$$

In regards to FM, it is a typical supervised ELM regression. The output $\mathbf{H}_{(k-1)}$ of the last hidden layer in UDFE is taken as the input. Then, the output weight $\boldsymbol{\beta}_{(o)}$ can be calculated with (18).

$$\begin{cases} \boldsymbol{\beta}_{(o)} = \left(\mathbf{H}^{\mathrm{T}}_{(k)} \mathbf{H}_{(k)} + \frac{\mathbf{I}}{C}\right)^{-1} \mathbf{H}^{\mathrm{T}}_{(k)} \mathbf{y}(t+1), \text{ full column rank} \\ \boldsymbol{\beta}_{(o)} = \mathbf{H}^{\mathrm{T}}_{(k)} \left(\mathbf{H}_{(k)} \mathbf{H}^{\mathrm{T}}_{(k)} + \frac{\mathbf{I}}{C}\right)^{-1} \mathbf{y}(t+1), \text{ full row rank} \end{cases} \tag{18}$$

where $\mathbf{y}(t+1)$ is the target output vector of the training process.

All of the connecting weights between two adjacent layers shown in Fig. 3 are determined after the independent sequential training process. Once the weights are determined, no fine-tuning is required. Finally, the forecasted output $y*(t+1)$ of STLF model is obtained by (19).

$$y*(t+1) = g\left(\mathbf{H}_{(k-1)} \boldsymbol{\beta}^{\mathrm{T}}_{(k)}\right) \boldsymbol{\beta}_{(o)} \tag{19}$$

B. THE FORMATION OF TRAINING SAMPLE SET

The training sample set is one of the most important components of the STLF model. It determines what can be extracted and learned by the deep learning process. In order to make the training sample set informative, the aforementioned knowledge representation set based on feature indices are utilized to construct the attributes series of each time interval, as shown in Fig. 5.

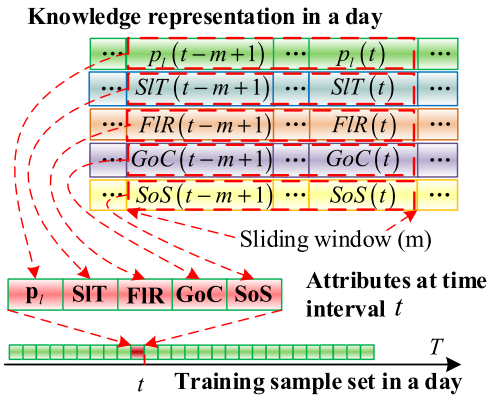


FIGURE 5. AE-ELM training sample set.

Denote t to be the current time interval and $(t+1)$ to be the prediction time interval. A sliding window with length m is used to update features. Then, m samples from time interval $(t-m+1)$ to t of each index in knowledge representation structure are used to construct the attribute series at time interval t sequentially. Finally, the sample series in a day is formed by the attribute series of all time intervals.

C. THE PROCESS OF STLF METHOD

The entire STLF method is shown in Fig. 6, which consists of the three main steps of input attribute set selection, determination of network parameters, and final forecasting.

1) INPUT ATTRIBUTE SET SELECTION

Fig. 5 shows the training sample set for the duration of one day. The day is split into 96 intervals with an attribute series associated with each time interval. In order to forecast the load of the $(t+1)^{\text{th}}$ time interval, all the attribute series at the t^{th} time interval in all previous days are input into SAE-ELM in parallel, which is shown as step 1 in Fig. 6 and described by (20).

$$\mathbf{A}=\left\{\begin{array}{ccc} y^1(t+1), & \left[\mathbf{p}_{l,t}^1, \mathbf{SlT}_t^1, \mathbf{FlR}_t^1, \mathbf{GoC}_t^1, \mathbf{SoS}_t^1\right] \\ \vdots & \vdots & \vdots \\ y^d(t+1), & \left[\mathbf{p}_{l,t}^d, \mathbf{SlT}_t^d, \mathbf{FlR}_t^d, \mathbf{GoC}_t^d, \mathbf{SoS}_t^d\right] \end{array}\right\} \tag{20}$$

where $y^d(t+1)$ is the target output of the $(t+1)^{\text{th}}$ time interval in the d^{th} day. The term in square bracket is the input attribute series at the t^{th} time interval.

2) DETERMINATION OF NETWORK PARAMETERS

First, the number of nodes in the input layer is determined by the dimension of the input attribute series. Then, the attribute set in (20) is taken as input to train deep learning neural network based on SAE-ELM to determine connecting weights in UDFE and FM process according to (9) - (19).

3) FINAL FORECASTING

The computed network parameters are used to construct the STLF model. Then, the attribute set at the current time interval t in the prediction day is used as input to obtain the forecasted load $\widehat{y}(t+1)$.

IV. NUMERICAL STUDY

A. DATA PREPERATION

The heating season in China typically lasts from Nov. 11 to Mar. 15 of the following year. During this period, the heating electricity price in the evening is much lower than that of non-heating season, which greatly increases the electric heating load. In this paper, the data of one residential transformer in Beijing from Jan. 1, 2016 to Mar. 15, 2016 is used as historical load series, as shown in Fig. 7.

Data from Jan. 1, 2016 to Mar. 8, 2016 is used as the initial time series to construct the knowledge representation set and the remaining data from Mar. 9, 2016 to Mar. 15, 2016 is used as the original testing data. The historical data is recorded every 15-minute and there are 96 time intervals in a day. All the historical data is first extracted by the proposed specialized feature indices and formed as a feature set. Then, the feature set is constructed as a training sample set. Before input to the training model, the elements in training sample set need to be normalized to $[-1, 1]$ within their own categories by (21). Then, it is formed as training sample set day by day according to Fig. 5.

$$y=\left(y^{\max}-y^{\min}\right)\frac{x-x^{\min}}{x^{\max}-x^{\min}}+y^{\min} \tag{21}$$

where $y^{\max}$ is the upper bound of normalization range, set to 1; $y^{\min}$ is the lower bound of normalization range, set to -1. x is the data to be normalized; $x^{\max}$ is the maximum value of x and $x^{\min}$ is minimum value of x.

B. CASES DESIGN AND SETTINGS

For validation, the proposed STLF method is compared against other methods such as the moving average method (MAM), backpropagation neural network (BPNN), and long short term memory (LSTM) method.

(1) MAM is a type of empirical forecasting method, which uses the average value of a rolling time window as forecasted value $y_{mam}(t+1)$. In this paper, the rolling time window of length 3 is used as benchmark.

$$y_{mam}(t+1)=\frac{1}{3}\sum_{i=t-2}^{t}x(i) \tag{22}$$

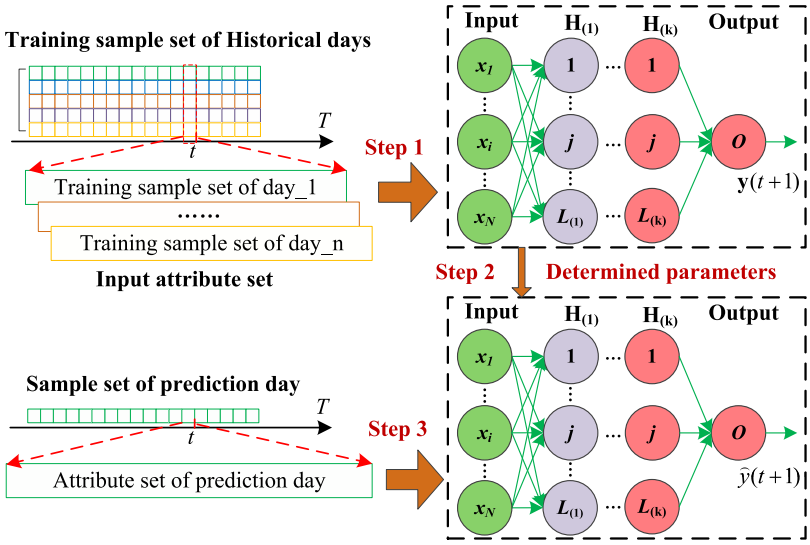


FIGURE 6. The complete STLF method.

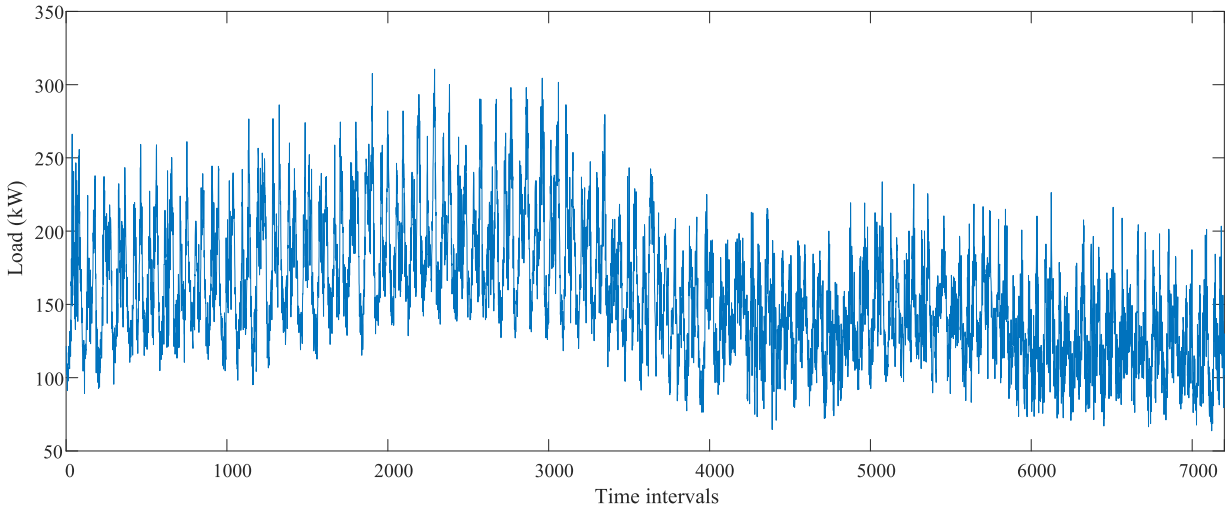


FIGURE 7. The load time series from Jan. 1st 2016 to Mar. 15th 2016.

(2) BPNN is used as the core of STLF model. Historical load time series is formed as training input. The attributes of the training input are the loads of the five previous time intervals.
(3) STLF model based on LSTM is also selected as one of the benchmarks. The time series data from Jan. 1, 2016 to Mar. 8, 2016 is used as training input.
(4) The proposed STLF model is based on SAE-ELM. The length of sliding window of input training sample set is 3. The four thresholds of GoC are −0.2, −0.08, 0.08, and 0.2, respectively.

Some key parameters of the aforementioned methods are listed in Table 2. For BPNN, the number of hidden layers is set to 1 and the number of hidden nodes is 10. The number of hidden layers of LSTM is set to 2 and the number of hidden nodes of each layer is 200. The number of hidden layers of SAE-ELM is set to 2 and the number hidden nodes in each layer is 150.

TABLE 2. Key parameters.

Method	Hidden layers	Hidden nodes	Epochs
MAM	/	/	/
BPNN	1	10	250
LSTM	2	200	250
SAE-ELM	2	150	/

C. RESULTS

The results of the case studies are presented in Table 3 and in Fig. 8. In Table 3, the mean absolute percentage error (MAPE), the maximum absolute percentage error (MaxAPE), the root mean square error (RMSE), and the maximum

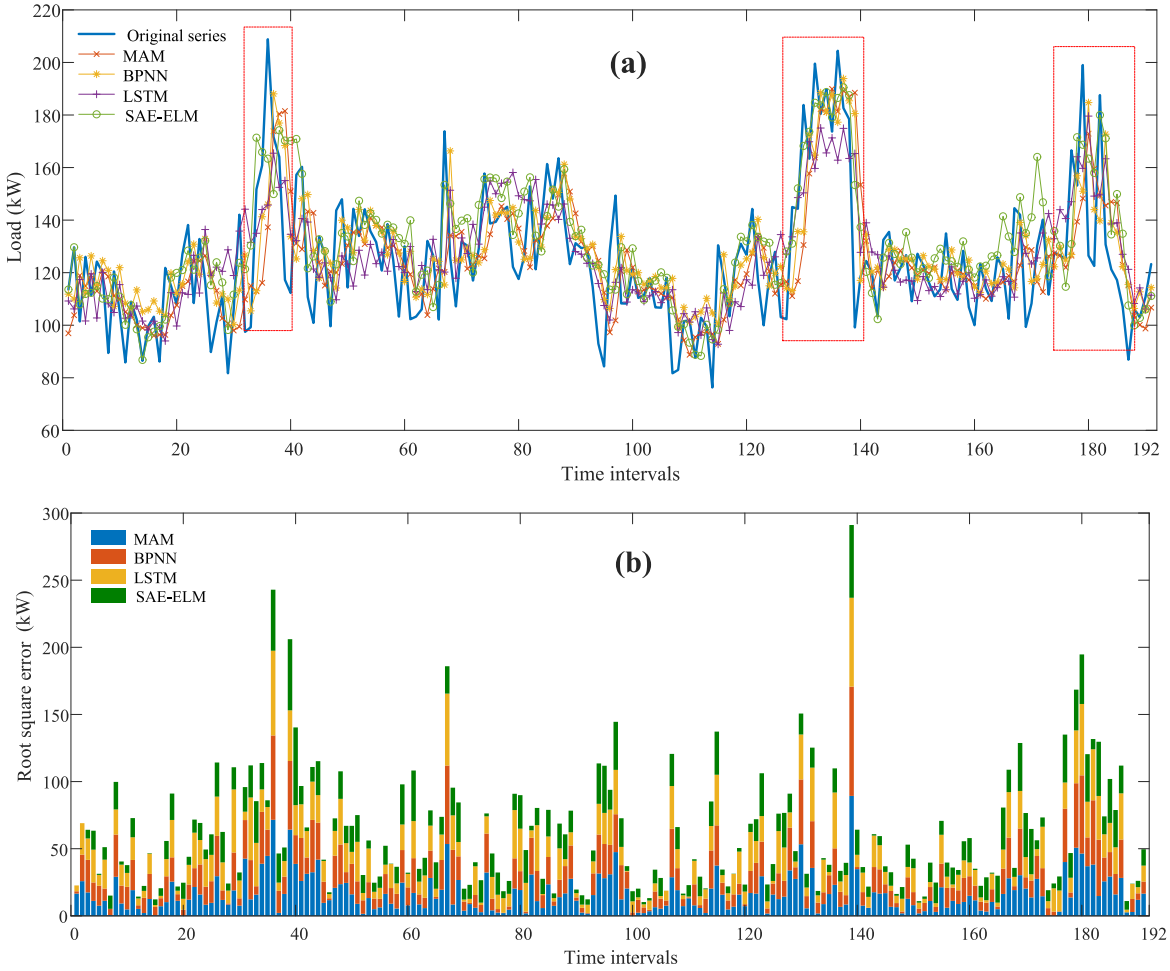


FIGURE 8. Comparison between the proposed SAE-ELM trained by feature set and the SAE-ELM trained by time series. (a) the comparison results among the four methods and the original series; (b) the comparison of root square error of four methods.

TABLE 3. Overall forecasting errors of 7 days.

Method	MAM	BPNN	LSTM	SAE-ELM
MAPE (%)	13.71	13.72	13.15	11.83
Max-APE (%)	90.1	82.08	78.94	68.1
RMSE (kW)	21.3	20.12	19.71	17.32
Max-RSE (kW)	89.35	81.39	66.19	55.25
CI (s)	/	6.26	167.65	2.73

root square error (Max-RSE) are used in comparing the performance of the four methods.

For clarity, only the results of the first two days (Mar. 9, 2016 and Mar. 10, 2016) are shown in Fig. 8. It can be concluded that the load time series fluctuates significantly. The minimum load is about 80 kW and the maximum load is about 210 kW. The peak-valley difference is about 62%.

Among the four methods, the MAPE of the proposed STLF method based on SAE-ELM is the lowest, followed by LSTM, MAM, and BPNN. The MAPE of the proposed STLF method based on SAE-ELM has decreased by 10.04%, 13.78%, 13.71% when compared against the LSTM method, BPNN method, and MAM method, respectively. Because the EHLs are affected by many external factors, the load active power sometimes changes abruptly. Due to the high variability of the load, this may result in high forecasting error. Therefore, the Max-APEs and Max-RSE of all the four methods are high. The proposed STLF model based on SAE-ELM has the lowest Max-APE at 68.1% and the lowest Max-RSE at 55.25 kW. Comparing the RMSE index, the proposed STLF model based on SAE-ELM has the best performance, which is 11.97% better than LSTM, 13.77% better than BPNN, and 18.54% better than MAM.

Comparing the computational index (CI) in Table 3, it can be seen that the proposed STLF model based on SAE-ELM has the highest computational efficiency. It is about 1.3 times faster than BPNN and 60.41 times faster than LSTM. This is because no iterative parameter tuning and fine-tuning are needed in the training process for the STLF model. Therefore, the proposed STLF model based on SAE-ELM is better than the other three methods across the five indices.

As shown in Fig. 8, the proposed STLF model based on SAE-ELM is capable of following the variation during most

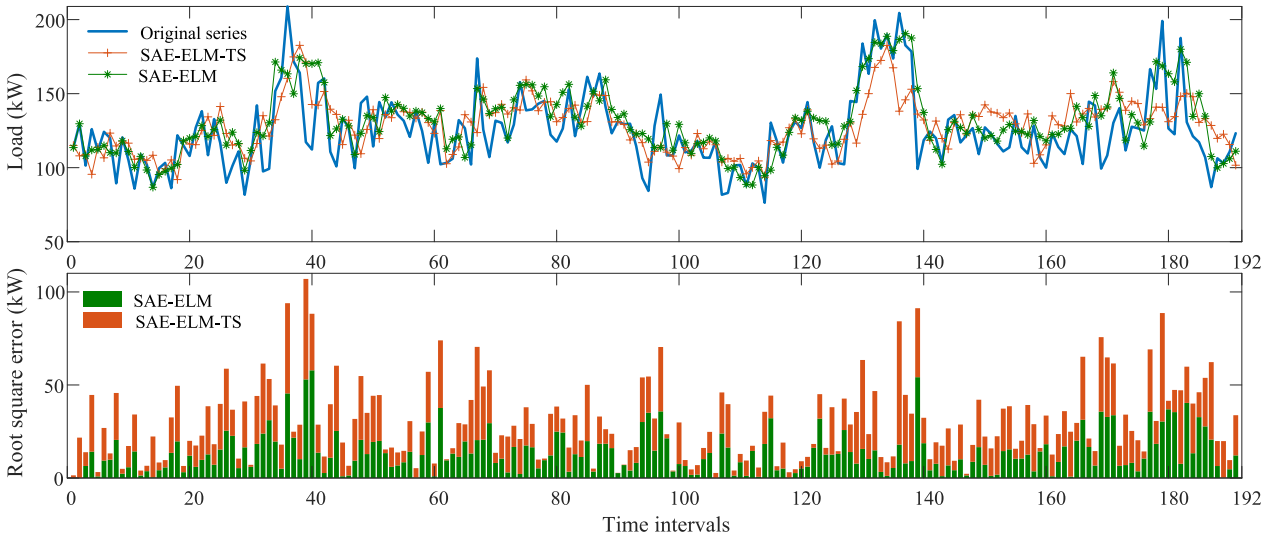


FIGURE 9. The comparison results between the proposed SAE-ELM trained by feature set and the SAE-ELM trained by time series.

time intervals. Therefore, the root square error is lowest in most time intervals. Due to the impacts of external factors, there are three large volatile periods in Fig. 8 (a). Fortunately, the proposed STLF model based on SAE-ELM is capable of capturing most of the sudden change features effectively while the other three methods may have a one to two time interval delay. This is because the fluctuation and the grade of change are extracted and represented as part of the training sample set by the proposed specialized feature indices set. In summary, it can be concluded from the four comparison cases that:

(1) The proposed STLF model based on SAE-ELM is better in dealing with small scale load forecasting with large volatility.
(2) Due to the better deep feature extracting and learning ability, the STLF models based on deep learning framework (SAE-ELM and LSTM) generally performs better than other conventional methods.
(3) The specialized feature extracting indices can help capture and represent deep diverse information and therefore contribute to the learning process and improve the model accuracy.

D. ANALYSIS OF THE EFFECT OF SPECIALIZED FEATURE SET

In order to further analyze the advantages of the proposed specialized feature set, one more case study is conducted. In this comparison, the STLF model also based on the same SAE-ELM framework but trained by historical time series is compared to the proposed STLF model.

The comparison results are shown in Fig. 9. The performance of the STLF method based on SAE-ELM-TS in dealing with volatility is worse compared to the proposed STLF method. This is because the STLF method based on SAE-ELM-TS is only trained by time series data instead of by the specialized feature indices. This is especially true for predicting sudden large changes, which is much more difficult. Comparing with the proposed STLF model, it is usually with one time interval delay. The root square error of SAE-ELM-TS is also larger than that of the proposed STLF model in most time intervals. The RMSE of SAE-ELM-TS is 23.72 which is much larger than that of the proposed STLF model.

It can be concluded that (i) the proposed specialized feature extracting indices can capture deep diverse features from historical time series; (ii) the training sample set constructed from feature indices can help improve the effect of deep learning process and improve the capability and sensitivity in dealing with volatility. Therefore, the forecasting accuracy is improved.

V. CONCLUSION

In this paper, a STLF model at EUTL with high penetration of EHLs is proposed based on a novel specialized feature indices set and SAE-ELM deep learning framework. Due to the large power excursions and significant randomness of EHLs, the raw load time series at EUTL exhibit a variety of irregularities, such as uncertainty, nonlinearity, and large volatility. Therefore, a novel specialized feature indices set is proposed to analyze and extract deep diverse features of raw load time series to construct the training sample set. The novel specialized feature indices set is proven to be helpful in improving the deep learning process and the capability and sensitivity in dealing with volatility. By using SAE-ELM, the deep nonlinear features are learned automatically with no iterative parameter turning and fine-turning needed, which makes the learning process much faster than the LSTM, BPNN, and MAM methods.

Although deep learning is a promising framework for demand forecasting, it lacks strong capability in extracting

deep features from highly volatile time series. Therefore, more effective methods need to be further studied.

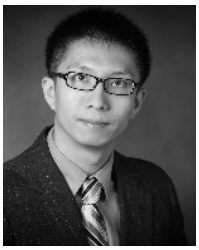
REFERENCES

[1] B. Li, J. Zhang, Y. He, and Y. Wang, "Short-term load-forecasting method based on wavelet decomposition with second-order gray neural network model combined with ADF test," *IEEE Access*, vol. 5, pp. 16324–16331, 2017.
[2] A. Tarsitano and I. L. Amerise, "Short-term load forecasting using a two-stage sarimax model," *Energy*, vol. 133, pp. 108–114, Aug. 2017.
[3] Q. Chen, F. Wang, B.-M. Hodge, J. Zhang, Z. Li, M. Shafie-Khah, and J. P. S. Catalão, "Dynamic price vector formation model-based automatic demand response strategy for PV-assisted EV charging stations," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2903–2915, Nov. 2017.
[4] F. Wang, K. Li, C. Liu, Z. Mi, M. Shafie-Khah, and J. P. S. Catalão, "Synchronous pattern matching principle-based residential demand response baseline estimation: Mechanism analysis and approach description," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6972–6985, Nov. 2018.
[5] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.
[6] Q. Chen, M. Xia, Y. Zhou, H. Cai, J. Wu, and H. Zhang, "Optimal planning for partially self-sufficient microgrid with limited annual electricity exchange with distribution grid," *IEEE Access*, vol. 7, pp. 123505–123520, 2019, doi: 10.1109/ACCESS.2019.2936762.
[7] B. Yildiz, J. I. Bilbao, and A. B. Sproul, "A review and analysis of regression and machine learning models on commercial building electricity load forecasting," *Renew. Sustain. Energy Rev.*, vol. 73, pp. 1104–1122, Jun. 2017.
[8] J. J. Shah, M. C. Nielsen, T. S. Shaffer, and R. L. Fittro, "Cost-optimal consumption-aware electric water heating via thermal storage under time-of-use pricing," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 592–599, Mar. 2015.
[9] Q. Chen, N. Liu, C. Hu, L. Wang, and J. Zhang, "Autonomous energy management strategy for solid-state transformer to integrate PV-assisted EV charging station participating in ancillary service," *IEEE Trans. Ind. Informat.*, vol. 13, no. 1, pp. 258–269, Feb. 2017.
[10] D. Liu, L. Zeng, C. Li, K. Ma, Y. Chen, and Y. Cao, "A distributed short-term load forecasting method based on local weather information," *IEEE Syst. J.*, vol. 12, no. 1, pp. 208–215, Mar. 2018.
[11] H. Jiang, Y. Zhang, E. Muljadi, J. J. Zhang, and D. W. Gao, "A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3341–3350, Jul. 2018.
[12] X. Sun, P. B. Luh, K. W. Cheung, W. Guan, L. D. Michel, S. S. Venkata, and M. T. Miller, "An efficient approach to short-term load forecasting at the distribution level," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2526–2537, Jul. 2016.
[13] J. Zhang, Y.-M. Wei, D. Li, Z. Tan, and J. Zhou, "Short term electricity load forecasting using a hybrid model," *Energy*, vol. 158, pp. 774–781, Sep. 2018.
[14] Y. Wang, Q. Xia, and C. Kang, "Secondary forecasting based on deviation analysis for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 500–507, May 2011.
[15] B. A. HØverstad, A. Tidemann, H. Langseth, and P. Öztürk, "Short-term load forecasting with seasonal decomposition using evolution for parameter tuning," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1904–1913, Jul. 2015.
[16] K. Park, S. Yoon, and E. Hwang, "Hybrid load forecasting for mixed-use complex based on the characteristic load decomposition by pilot signals," *IEEE Access*, vol. 7, pp. 12297–12306, 2019.
[17] C. Guan, P. B. Luh, L. D. Michel, Y. Wang, and P. B. Friedland, "Very short-term load forecasting: Wavelet neural networks with data pre-filtering," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 30–41, Feb. 2013.
[18] S. Li, P. Wang, and L. Goel, "A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 1788–1798, May 2016.
[19] S. Lahmiri, "Comparing variational and empirical mode decomposition in forecasting day-ahead energy prices," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1907–1910, Sep. 2017.
[20] M. Khodayar, O. Kaynak, and M. E. Khodayar, "Rough deep neural architecture for short-term wind speed forecasting," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2770–2779, Jul. 2017.
[21] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
[22] A. Rahman, V. Srikumar, and A. D. Smith, "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks," *Appl. Energy*, vol. 212, pp. 372–385, Feb. 2018.
[23] H. Wang, H. Yi, J. Peng, G. Wang, Y. Liu, H. Jiang, and W. Liu, "Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network," *Energy Convers. Manage.*, vol. 153, pp. 409–422, Dec. 2017.
[24] L. Han, Y. Peng, Y. Li, B. Yong, Q. Zhou, and L. Shu, "Enhanced deep networks for short-term and medium-term load forecasting," *IEEE Access*, vol. 7, pp. 4045–4055, 2018.
[25] M. Cai, M. Pipattanasomporn, and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques," *Appl. Energy*, vol. 236, pp. 1078–1088, Feb. 2019.
[26] T. Ouyang, Y. He, H. Li, Z. Sun, and S. Baek, "Modeling and forecasting short-term power load with copula model and deep belief network," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 3, no. 2, pp. 127–136, Apr. 2019.
[27] Z. Zhang, S. Li, Y. Xiao, and Y. Yang, "Intelligent simultaneous fault diagnosis for solid oxide fuel cell system based on deep learning," *Appl Energy*, vols. 233–234, pp. 930–942, Jan. 2019.
[28] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.
[29] H. Zhou, G. B. Huang, Z. Lin, H. Wang, and Y. C. Soh, "Stacked extreme learning machines," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 2013–2025, Sep. 2015.
[30] E. Cambria *et al.*, "Extreme Learning Machines [Trends & Controversies]," *IEEE Intell. Syst.*, vol. 28, no. 6, pp. 30–59, Nov./Dec. 2013.
[31] C. M. Wong, C. M. Vong, P. K. Wong, and J. Cao, "Kernel-based multilayer extreme learning machines for representation learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 757–762, Mar. 2018.
[32] Y. Yang, Q. M. J. Wu, Y. Wang, K. M. Zeeshan, X. Lin, and X. Yuan, "Data partition learning with multiple extreme learning machines," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1463–1475, Aug. 2015.
[33] J. Tang, C. Deng, and G.-B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 809–821, Apr. 2015.
[34] L. Yao and Z. Ge, "Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1490–1498, Feb. 2018.

QIFANG CHEN (S'13–M'17) received the B.S. and M.S. degrees in communication engineering and electric engineering respectively, from Xiangtan University, Hunan, in 2010 and 2013, respectively, and the Ph.D. degree in electrical engineering from North China Electric Power University, Beijing, China, in 2017. He is currently a Postdoctoral Researcher with the School of Electrical Engineering, Beijing Jiaotong University, Beijing, China. His research interests include integrated energy systems, demand response, and flexible load modeling and control.

MINGCHAO XIA (M'03–SM'17) received the B.S. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1998 and 2003, respectively. He is currently a Professor with the School of Electrical Engineering, Beijing Jiaotong University. His research interests include the energy internet, smart power distribution system control and optimization, power electronics in power distribution, and flexible load control.

TENG LU received the B.S. degree in electrical engineering from the China University of Mining and Technology, Beijing, in 2018. She is currently pursuing the M.S. degree with the School of Electrical Engineering, Beijing Jiaotong University. Her research interest includes in modeling and control of flexible loads in power grid.



XICHEN JIANG (M'16) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Illinois, Urbana-Champaign, in 2010, 2012, and 2016, respectively. He is currently an Assistant Professor with the Department of Engineering and Design, Western Washington University. His research interests include controls theory, power system reliability, and cyber-physical systems. In particular, his research applies statistical and set-theoretic methods to detect and mitigate attacks or faults on the power system networks.

WENXIA LIU (M'11) received the B.S. degree in radio engineering from the Nanjing University of Science and Technology, in 1990, the M.S. degree in power system and automation from Northeast Electric Power University, in 1995, and the Ph.D. degree in power system and automation from North China Electric Power University (NCEPU), in 2009. She is currently a Professor with the School of Electrical and Electronic Engineering, NCEPU. Her research interests include in intelligent planning of power systems, power system risk assessment, power communication systems, and reliability and planning of cyber-physical systems (CPS).

QINFEI SUN received the B.S., M.S., and Ph.D. degrees in electrical engineering from the College of Information and Electrical Engineering, China Agricultural University, in 2010, 2012, and 2016, respectively. He was a Visiting Ph.D. Student with the Microgrids Research Program, Department of Energy Technology, Aalborg University, Aalborg, Denmark, from 2014 to 2015. He is currently an Engineer with the State Grid Beijing Electric Power Research Institute. His research interests include power electronics for distributed generation, operation, control, and power quality of microgrid, demand-side response, and especially for electric heat and cooling supply of buildings.