IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 27, 2018, accepted September 19, 2018, date of publication October 4, 2018, date of current version October 31, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2873712

Short-Term Non-Residential Load Forecasting Based on Multiple Sequences LSTM Recurrent Neural Network

RUNHAI JIAO, TIANMING ZHANG, YIZHI JIANG, AND HUI HE
School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

Corresponding author: Runhai Jiao (runhaijiao@ncepu.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2018ZD06 and in part by the National Natural Science Foundation of China under Grant 61503137.

ABSTRACT The energy consumption by non-residential consumers in China accounts for a significant proportion of the total energy consumption in the society. Thus, accurate non-residential load forecasting plays an increasingly essential role in the future grid planning and operation. In this paper, a method based on LSTM recurrent neural network is proposed to predict the load of non-residential consumers using multiple correlated sequence information. First, the k-means is employed to analyze the daily load curves of non-residential consumers, classify and mine the consumer's energy consumption behavior patterns. Then, the Spearman correlation coefficient is applied to investigate the time correlation under multiple time series for non-residential consumers. It is found that there exist multiple related time sequences such as adjacent time points, the same time points in adjacent days, and the same day in adjacent weeks among data samples for a specific consumer. Therefore, a non-residential load forecasting framework based on multiple sequence LSTM recurrent neural networks is presented. The proposed framework is tested on a real data set, which contains 48 non-residential consumers' energy consumption data in China, and outperforms the other load forecasting approaches. Experiment results show that this method can effectively utilize the multiple sequence information, and successfully capture the dependencies among these sequences.

INDEX TERMS Short-term non-residential load forecasting, recurrent neural network (RNN), long short-term memory (LSTM), spearman correlation coefficient.

I. INTRODUCTION

Since the electric power is a non-storable product, electrical operators should ensure a precise balance between the electricity production and consumption at any moment [1]. Therefore, it is necessary to formulate a scheduling plan, and accurate short-term load forecasting (STLF) can significantly facilitate the power system operations. Short-term load forecasting mainly refers to predicting the power load in the next few hours, a day or several days. Moreover, the forecasting errors have a considerable influence on subsequent safety check of the power grid. Short-term load forecasting is also of great significance for dynamic state estimation and load dispatching [2], [3].

Smart grid is a hot spot in the current global power industry, while consumers, play an important role in smart grid demand response [4]. Consumers can be divided into residential consumers, business consumers, and industrial consumers according to their energy consumption scale and habits. Non-residential consumers, including business consumers and industrial consumers, occupy a significant portion of energy consumption. They are generally deployed with advanced meter infrastructure (AMI). Such massive deployment has also created opportunities for short-term load forecasting and demand-side management [5]. According to the statistics released by the National Energy Administration of the People's Republic of China in 2017 on the national power industry, non-residential energy consumption accounts for a large proportion of total social energy consumption, which is approximately 86% [6]. Therefore, compared with residential consumers, non-residential consumers are the most crucial demand response resources in smart grid. By mining the patterns of non-residential energy consumption, it is essential to improve the accuracy of short term load forecasting for power system operation. Nevertheless, non-residential consumers' power consumption is highly susceptible to many factors such as weather, temperature, economy, and policies. Regarding short-term load forecasting methodologies, many approaches had been reported in the literature to address this problem. However, very few of them have confronted with individual consumers directly, especially the non-residential consumers. Hence, traditional forecasting methods are difficult to achieve accurate predictions [7]. The issues on load forecasting of short-term non-residential consumers remain open.

In this paper, it aims to address this issue on short-term load forecasting for the individual non-residential consumer. Firstly, the daily load curves analysis of non-residential consumers is conducted to explore the adjacent-time point correlation, the day-related correlation, and week-related correlation. Then, a non-residential load forecasting framework based on multiple sequence recurrent neural networks (RNN) is proposed. Since daily load curves show that individual consumer load is strongly correlated with adjacent time point, day and week, three different time series features are used as input to predict the load of the individual non-residential consumer. Experiment results show that this approach outperforms the other state-of-the-art methods. The main contributions of this paper are as follows:

- K-means is utilized in clustering daily load curves to explore the patterns of non-residential energy consumption.
- The Spearman correlation coefficient is employed to measure the time series correlation of non-residential consumers. The calculation results show that most of the non-residential consumers' load data have strong correlations in adjacent time points, days, and weeks.
- The load forecasting framework, which based on Long Short-Term Memory (LSTM), selects multiple related time sequences as input to improve forecasting accuracy.
- An empirical study on 48 non-residential consumers collected by AMI in a south region of China demonstrates that the proposed method has a better forecasting accuracy than other rival methods.

The rest of this paper is organized as follows. Section II reviews the load forecasting studies. Section III introduces the dataset, cluster analysis and Spearman correlation coefficient statistics on the dataset. Section IV proposes a short-term non-residential load forecasting framework based on LSTM. Section V describes the model evaluation indicators. Section VI presents the comparison algorithms, experimental settings, experimental results, and analysis. Section VII concludes the paper and points out future work.

II. LITERATURE STUDY

Traditional models for short-term load forecasting fall into two categories. The first type is the time-series model most commonly used for processing time-varying load data. Typical time series methods include Box-Jenkins model [8], Autoregressive Moving Average model (ARMA) [9], [10], and Autoregressive Integrated Moving Average model (ARIMA) [11]. These prediction approaches based on time series models are simple implemented and widely used, but they require high stability of the load sequence. However, compared with system load, the uncertainty of energy consumption of an individual non-residential consumer is exceptionally high, and the load sequences of different types of consumers vary greatly. Therefore, the prediction performance of time series model is usually poor when forecasting the energy consumption of an individual non-residential consumer.

To overcome these deficiencies, the second type of load forecasting model based on artificial intelligence is coming into view. Among them, Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest (RF) are representative methods of shallow machine learning algorithms [12]. Most of the state-of-the-art methods in short-term load forecasting are derived from these shallow machine learning algorithms because of their excellent nonlinear function fitting ability.

Since ANN has the advantages of high tolerance to odd samples, strong nonlinear mapping ability, self-adaptation, and self-organization [13], Jetcheva [1] proposed a combined forecasting method based on clustering and neural network. Wan *et al.* [14] introduced a short-term prediction model combining fuzzy logic control and improved BPNN algorithm, which improved the prediction accuracy and robustness. However, the drawback of the BP neural network is that it easily falls into local optima. Many experiments on several UCI benchmark datasets by Huang *et al.* [15], [16] showed that: ELM has high training efficiency and low error rate regarding regression and classification, which overcomes traditional neural network over-fitting. However, the choice of the network structure of the ELM network should be strict. Otherwise, it will lead to under-fitting or over-fitting.

SVM is also one of the most popular machine learning methods in load forecasting. Lin *et al.* [17] proposed a Least Squares Support Vector Machine (LSSVM) model optimized by the particle swarm optimization algorithm. Jiang *et al.* [18] presented an SVM-based method for distribution networks. However, SVM uses quadratic programming, which involves the calculation of higher-order matrices, to solve support vectors. When the number of samples is large, the storage and calculation of the matrix will consume a significant amount of machine memory and computation time [19]. Therefore, SVM-based algorithms are difficult for training large-scale dataset.

RF is a typical ensemble method, and it has a good performance of generalization and robustness. Wu *et al.* [20] introduced a two-stage method of random forest and grey relation projection. Cheng *et al.* [21] used a random forest-based integrated system for load forecasting, and gained better results than other machine learning algorithms. However, RF has problems with over-fitting on odd samples.

To summarize, the common disadvantage of the above algorithms lies in the lack of consideration of the correlation among the time series data. These methods only establish a non-linear mapping relationship between the features and the load and neglect the correlation relationship among consecutive load samples. In fact, as a typical time series, the load samples are not only non-linear with their corresponding features, but there also exist high correlations among load samples. For a given consumer, the change of the load data is a continuous process. The load at the current moment and the one at the previous moment are not mutually independent, because the current load depends not only on the current features but also on the features and loads of the past moments. Therefore, most of the existing methods only establish a non-linear relationship between the features at a single moment and corresponding load. They ignore the strong correlation among the load samples, which leads to limited prediction accuracy.

Penya *et al.* [22] proposed a two-stage forecasting method. They firstly classified load data according to the type of the day in a week and then constructed a forecasting model for each type of day. According to the periodicity of load data, the dataset was divided into three categories: weekday, Saturday, and Sunday. They also tried to add weather features [23] to these three prediction models to obtain higher accuracy. However, because of the high uncertainty of energy consumption, it is hard to characterize the energy consumption patterns for an individual non-residential consumer just by using the type of the day in a week.

With the rapid development of deep learning, some scholars began to pay attention to the deep neural network and used the time correlation information to build load forecasting models based on the recurrent neural network. Kong *et al.* [7] presented a short-term load forecasting framework based on LSTM for all residential consumers. This method used the time series correlation of load data at adjacent moments to predict the load of residential consumers and obtained better prediction accuracy than rival approaches. However, unlike residential consumers, the production and management of non-residential consumers are affected by many factors, such as economic, political, and significant events. The energy consumption patterns of non-residential consumers cannot be simply summarized by weekdays and weekends. Moreover, the energy consumption behavior varies among different type of industries and individuals, so it is difficult to model all non-residential consumers load and energy consumption patterns using a single model. Hence, this paper proposes a short-term load forecasting method based on multiple sequences LSTM recurrent neural network. The proposed method can not only establish the nonlinear relationship between features and load, but also capture the correlation between adjacent-time point correlation, day-related correlation, and week-related correlation to improve the accuracy of load forecasting.

III. DATASET AND CORRELATION ANALYSIS

A. DATASET

The dataset contains 48 non-residential consumers' load data collected by AMI in South China from July 1, 2015 to August 31, 2015 with a sampling interval of 15 minutes. These non-residential consumers belong to the following industries, such as catering industry (restaurant), electronic equipment manufacturing industry (electronic), steel smelting industry (steel), aluminum smelting industry (alum), printing industry (print), coal mining industry (coal), gas production and supply industry (gas), paper industry (paper), and hotel industry (hotel).

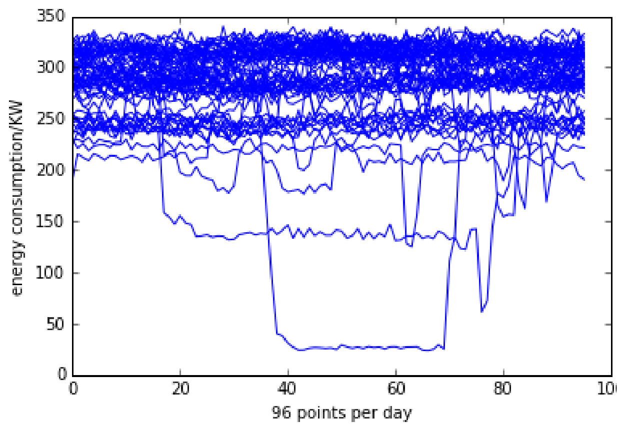


FIGURE 1. Daily load curves of an aluminum smelting industry consumer.

Fig.1 shows the daily load curves of an aluminum smelting industry consumer. Apart from a small amount of noisy data, the consumer's daily load power is stable between 200 kW and 350 kW, and the daily load curves are relatively smooth. Fig. 2 shows the daily load curves of a catering industry consumer. It can be seen that the two dining periods from 11:30 to13:30 and from 18:00 to 20:00 are the peak periods of energy consumption. Moreover, the energy consumption during the dinner time is higher than that during the lunchtime. After 22:00 energy consumption dramatically declines and remains stable between 20 kW and 50 kW. Fig.3 shows the daily load curves of a coal mining industry consumer. The daily load curves of the consumer are relatively messy, and there is no apparent pattern. The total energy consumption is between 0 KW and 14000 KW.

Since each non-residential consumer has different ways of using electricity, their patterns of energy consumption are entirely different. Hence, it is reasonable to train forecasting models individually for each consumer.

B. CLUSTER ANALYSIS

For a given non-residential consumer, its production and management generally have strong periodicity and regularity according to its industry. If there are no significant events or policy changes, we suppose that the load curves for a specific non-residential consumer should have the following properties:

- The load curve fluctuation in a short time should be small (adjacent-time point correlation).

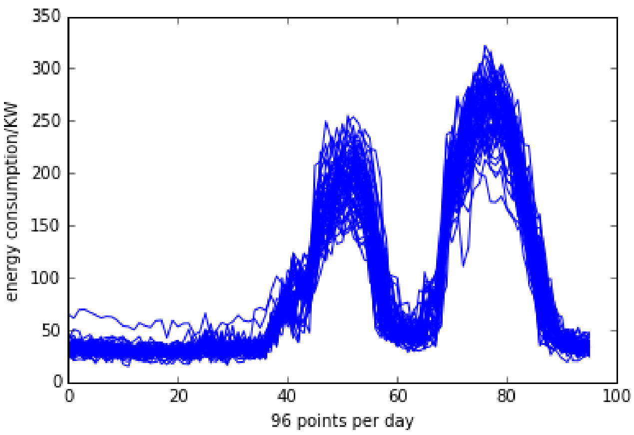


FIGURE 2. Daily load curves of a catering industry consumer.

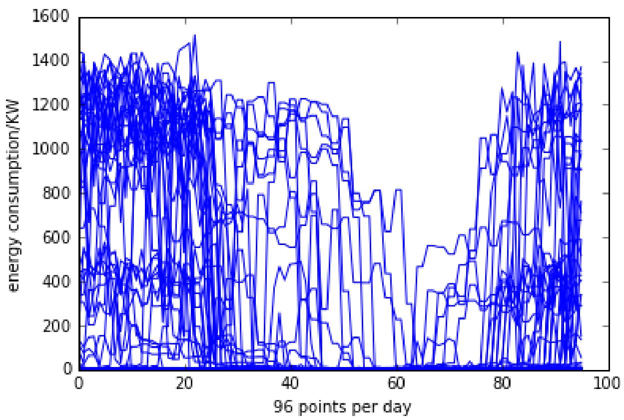


FIGURE 3. Daily load curves of a coal mining industrial consumer.

- The loads at the same time point in different days are not significantly different (day-related correlation).
- The daily load curves with same week types are similar (week-related correlation). The week types include Monday, Tuesday... and Sunday.
- The difference in daily load curve between working days and non-working days depends on the consumer and its industry.

In order to verify our hypothesis, k-means Clustering was utilized to explore the patterns of non-residential energy consumption. Since we assumed that activities of non-residential consumers can be divided into two types (working or non-working), the number of clusters was set to 2 in the experiment.

As shown in Fig.4, the daily load curves of a catering industry consumer are clustered in July. Though the number of clusters is set to 2, it can be seen that the consumer's daily load curves only belong to one cluster. Therefore, there is only one energy consumption pattern that this restaurant opens from noon to night, and the peaks of energy consumption are from12:00 to 15:00 in the noon and from 19:30 to 1:00 in the evening.

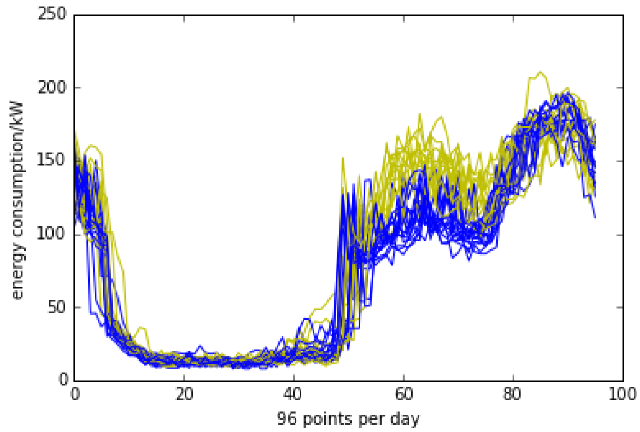


FIGURE 4. Clustering results of daily load curves for a catering industry consumer.

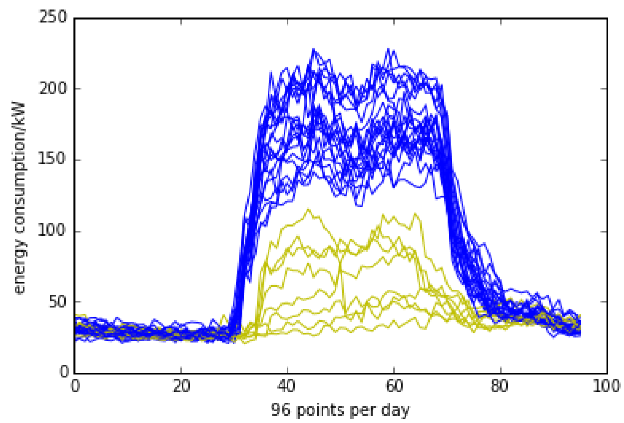


FIGURE 5. Clustering results of daily load curves for an electronic equipment manufacturing industrial consumer.

The daily load curves of electronic equipment manufacturing industrial consumer in July are divided into two clusters, as shown in Fig.5. The blue cluster is significantly higher than the yellow one. Combined with the date of energy consumption shown in Fig.6, it can be seen that the blue clusters are mostly distributed on weekdays, and the yellow clusters are distributed on weekends. Besides, it can be inferred that most of the production is scheduled on weekdays, and energy consumption patterns are different between working days and non-working days.

However, the way of energy consumption among all non-residential consumers differs significantly. Fig.7 shows a hotel industry consumer's clustering results. There is no apparent pattern for this consumer, for the energy consumption of a hotel is affected by many factors, such as seasons, policies, economy, weather, and even the travelers' lifestyles. Combined with the date of energy consumption shown in Fig.8, this consumer has no distinct pattern in time series, and it is difficult to summarize the regularity of energy consumption accurately.

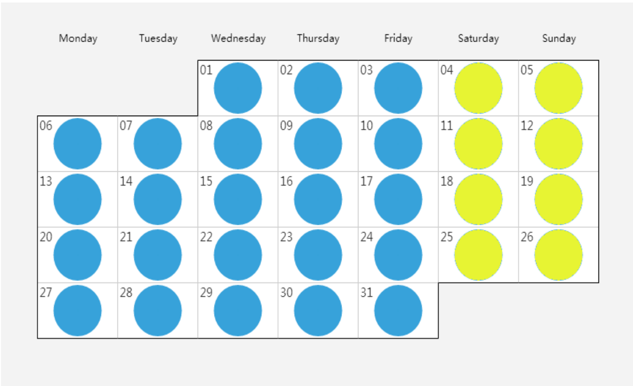


FIGURE 6. **Clustering results on the date for an electronic equipment manufacturing industrial consumer.**

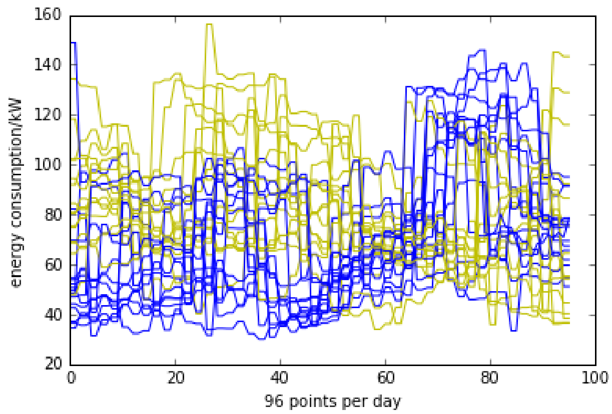


FIGURE 7. **Clustering results of daily load curves for a hotel industrial consumer.**

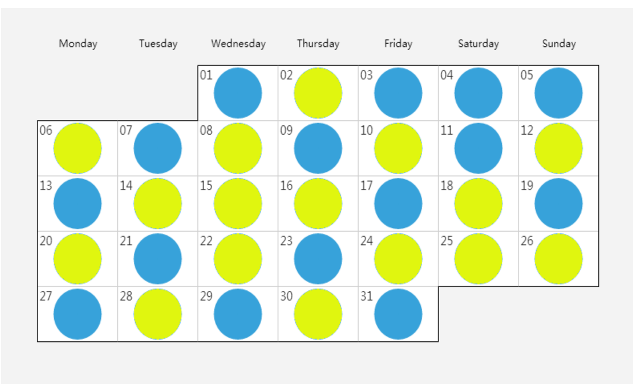


FIGURE 8. **Results of clustering on the date for a hotel industrial consumer.**

C. TIME SERIES CORRELATION

Since the consumer load data is not from a normal distribution, the Spearman correlation coefficient is employed to measure the time series correlation for individual non-residential consumer [25].

Let X and Y be the matrices whose correlation needs to be calculated. All the dimensions are n. It can be expressed as: $(X, Y) = \{(X_1, Y_1), ((X_2, Y_2)), \ldots, (X_n, Y_n)$

Hypothesis testing:

H_0 : X and Y have no correlation.

H_1 : X and Y have correlation.

Let R_i denote the rank of X_i in $(X_1, X_2, \ldots, X_n)$, Q_i denote the rank of Y_i in $(Y_1, Y_2, \ldots, Y_n)$, If X_i is synchronous with Y_i, R_i is also synchronous with Q_i, vice versa. In order to define the consistency between ranks, the Spearman correlation coefficient is defined as follows [25]:

$$r_s = \frac{\sum_{i=1}^{n}[(R_i - \frac{1}{n}\sum_{i=1}^{n} R_i)(Q_i - \frac{1}{n}\sum_{i=1}^{n} Q_i)]}{\sqrt{\sum_{i=1}^{n}(R_i - \frac{1}{n}\sum_{i=1}^{n} R_i)^2}\sqrt{\sum_{i=1}^{n}(Q_i - \frac{1}{n}\sum_{i=1}^{n} Q_i)^2}} \tag{1}$$

$$\sum\nolimits_{i=1}^{n} R_i = \sum\nolimits_{i=1}^{n} Q_i = \frac{n(n+1)}{2} \tag{2}$$

$$\sum\nolimits_{i=1}^{n} R_i^2 = \sum\nolimits_{i=1}^{n} Q_i^2 = \frac{n(n+1)(2n+1)}{6} \tag{3}$$

So, r_s can be simplified as:

$$r_s = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n}(R_i - Q_i)^2 \tag{4}$$

The t-test is used for the correlation test in the parameter statistics. Under the null hypothesis, the t-test statistic can also be defined as:

$$T = r_s\sqrt{\frac{n-2}{1-r_s^2}} \tag{5}$$

This statistic obeys the t distribution of $v = n - 2$ under the null hypothesis. When $T > t_{\alpha,v}$, it suggests that the two variables are correlated with each other. On the contrary, when $T \leq t_{\alpha,v}$, it indicates that the two variables are not correlated.

Since the Spearman correlation coefficient removes some interference factors that affect the overall data, it is more applicable when the data volume is large. Therefore, the Spearman correlation coefficient is employed to calculate the time series correlations (including adjacent-time point correlation, day-related correlation, and week-related correlation) in load forecasting.

For the catering industry consumer, the electronic equipment manufacturing industrial consumer, and the hotel industry consumer mentioned in section B, we obtain the Spearman correlation coefficients of the adjacent-time point correlation, day-related correlation, and week-related correlation according to formula 1 to 5. Then the calculated T value is compared with $t_{0.01,10}$. If $T > t_{0.01,10}$, it accepts the hypothesis H_1, and X and Y are correlated. Otherwise, it accepts the hypothesis H_0, X and Y are not correlated. The calculation results are shown in Table 1.

Table 1 shows that Spearman correlation coefficient of the catering industrial consumer and the electronic equipment manufacturing industry has a high correlation

TABLE 1. Consumer load data time series correlation results.

Consumer	Adjacent-time point correlation coefficient	Day-related correlation coefficient	Week-related correlation coefficient
catering industry	0.63	0.90	0.91
	Accept H_1 hypothesis	Accept H_1 hypothesis	Accept H_1 hypothesis
Electronic equipment manufac-turing industrial	0.68	0.82	0.89
	Accept H_1 hypothesis	Accept H_1 hypothesis	Accept H_1 hypothesis
Hotel industry	0.18	0.20	0.11
	Accept H_0 hypothesis	Accept H_0 hypothesis	Accept H_0 hypothesis

in the adjacent-time point series, day-related series, and week-related series, and they all accept the hypothesis H_1. However, the hotel industry consumer's Spearman correlation coefficient is low, and there is no apparent correlation in the adjacent-time point series, day-related series, and week-related series.

IV. LSTM-BASED NON-RESIDENTIAL LOAD FORECASTING FRAMEWORK

In order to utilize different sequence correlations and capture the long-distance dependencies in the load data, we propose an LSTM-based non-residential load forecasting framework which takes three-time sequences as input.

A. RECURRENT NEURAL NETWORK

RNN is a kind of deep neural network, which is good at processing sequence data. Different from conventional feed-forward neural networks, RNN preserves, learns, and records historical information in sequence data by connecting periodically hidden layers nodes [26]. As shown in Fig.9, the structure of the RNN includes the input layer, hidden layer, and output layer, U, V, W are the weights of input layer to hidden layers, hidden layers to output layers.

Since RNN shares parameters between layers, it dramatically reduces the number of parameters and thereby shortens the training time [27]. However, RNN is quickly involved in gradient vanishing or gradient exploding when processing long sequence data.

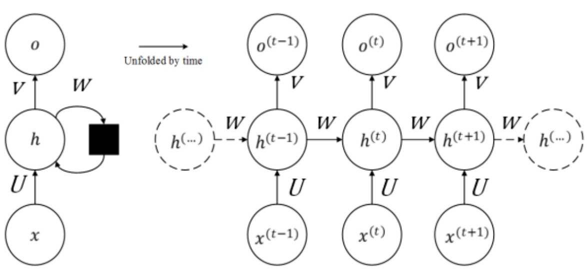


FIGURE 9. The structure of Recurrent Neural Network.

B. THE LSTM MODEL

The structure of an LSTM cell is shown in Fig.10. Each LSTM cell contains three control gates, which are the input gate i_t, output gate o_t, and forget gate f_t. The input data of LSTM at time t is x_t, the output value is h_t, c_t is the memory state, and $\tilde{c}_t$ is the candidate state. The LSTM memory cell is updated as follows [28], [29]:

$$f_t = \sigma(w_f \times [h_{t-1}, x_t] + b_f) \tag{6}$$
$$i_t = \sigma(w_i \times [h_{t-1}, x_t] + b_i) \tag{7}$$
$$\tilde{c}_t = \tanh(w_c \times [h_{t-1}, x_t] + b_c) \tag{8}$$
$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \tag{9}$$
$$o_t = \sigma(w_o \times [h_{t-1}, x_t] + b_o) \tag{10}$$
$$h_t = o_t \times tanh(c_t) \tag{11}$$

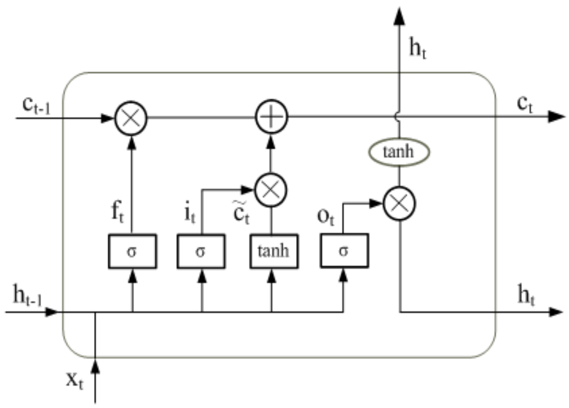


FIGURE 10. The structure of an LSTM cell.

In the above formulas, w_f, w_i, w_c, w_o are the weight matrices of the forget gate, input gate, candidate state, and output gate; b_f, b_i, b_c, b_o are the offsets, respectively. σ indicates that the activation function is Sigmoid function, and *tanh* stands for hyperbolic tangent function. $\times$ means an element-wise multiplication.

Compared with RNN, LSTM adaptively controls how much information flows through these gates. Besides, it is based on the idea of creating paths through time have derivatives that neither vanish nor explode, and these cells accumulate information (such as evidence for a particular feature or category) over a long duration. Therefore, LSTM captures the long-distance dependencies among sequence data, and it is widely used in tasks such as natural language processing and speech recognition.

C. NON-RESIDENTIAL LOAD FORECASTING FRAMEWORK

The LSTM-based short-term load forecasting framework is shown in Fig.11. There exists the time series correlation (including adjacent-time point correlation, day-related correlation, and week-related correlation) in most of the non-residential consumers' load data. Therefore, three LSTM recurrent neural networks are paralleled to capture three kinds of correlation information in time series. Furthermore,

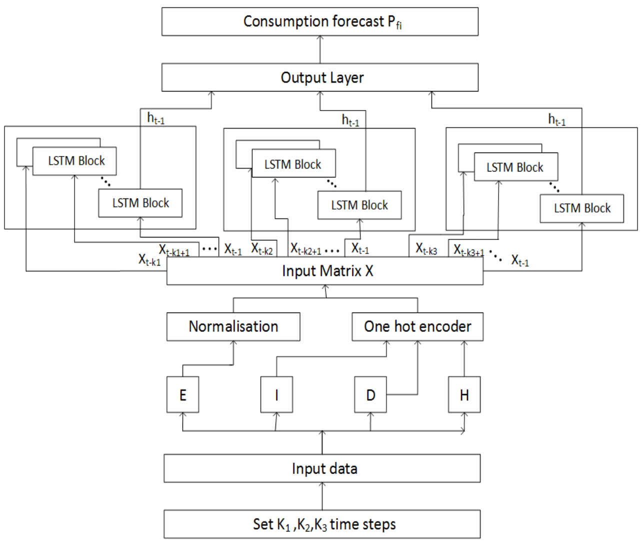


FIGURE 11. LSTM-based non-residential load forecasting framework.

the LSTMs learn the consumers' energy consumption patterns through the three-time series, so that the forecasting accuracy will be improved.

In the framework, the steps are as follows:

1) The inputs of the framework are time series load data. Firstly, they can be formed as three kinds of sequences.

- The time step of the adjacent-time point sequence is set to K_1, and the load sequence consists of consecutive K_1 time points before the predicted time point. $E_1 = \{e_{t-K_1}, e_{t-K_1+1}, \ldots, e_{t-1} \in R^{k_1}$, E_1 is the adjacent-time point feature vector.
- The time step of the day-related time sequence is set to K_2, and the load sequence is composed of the same time points in consecutive K_2 days before the predicted time point. $E_2 = \{e_{t-K_2}, e_{t-K_2+1}, \ldots, e_{t-1} \in R^{k_2}$, E_2 is the day-related feature vector.
- The time step of the week-related time sequence is set to K_3, and the load sequence is formed by the same time point of the same day in contiguous K_3 weeks before the predicted time point. $E_3 = \{e_{t-K_3}, e_{t-K_3+1}, \ldots, e_{t-1} \in R^{k_3}$, E_3 is the week-related feature vector.

Then, time information is incorporated into the inputs.

- The incremental sequence of the time indices for the past K time steps $I \in R^k$, where the range for each element of I is (1 to 96).
- The corresponding day of week indices for the past K time steps D, each of which ranges from 0 to 6.
- The corresponding binary holiday marks H, each of which can either be 0 or 1.

Since LSTM-based recurrent neural networks are sensitive to the data scale, we adopt the Min-Max method to normalize the load sequence data E_1, E_2, E_3, and the normalized data range is transformed into [0, 1], defined as follows:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{12}$$

E_1, E_2, and E_3 are normalized as $\tilde{E}_1$, $\tilde{E}_2$, and $\tilde{E}_3$ and merged as $\tilde{E}$, while the vectors I, D, H are encoded by one hot encoder. The one hot encoder maps an original element from the categorical feature vector with M cardinality into a new vector with M elements, where only the corresponding new element is one while the rest of new elements are zeroes.

Once the four vectors are scaled to $\tilde{E}, \tilde{I}, \tilde{D}$, and $\tilde{H}$, the input for LSTM layer is a matrix of the concatenation of the four, i.e.:

$$X = \{\tilde{E}^T, \tilde{I}^T, \tilde{D}^T, \tilde{H}^T \tag{13}$$

2) Since $\tilde{E}$ consists of three sequence vectors, each of them is fed to corresponding LSTM recurrent neural network. Each row of the input matrix is the scaled features for the corresponding time step, which is fed to corresponding LSTM block. The outputs of three parallel LSTM networks are fed to an output layer, which maps the LSTM outputs into a single value, i.e., the energy consumption forecast P_{ft} of the target time interval.

V. PERFORMANCE EVALUATION

The load forecasting performance of the model is evaluated by different statistical indicators including the Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The above statistical indicators are described briefly as follows:

The MAE represents the average quantity of total absolute error between the estimated and measured values:

$$e_{MAE} = \frac{1}{N}\sum_{t=1}^{N}|P_{ft} - P_{rt}| \tag{14}$$

P_{ft} is the predicted load at time t; P_{rt} is the real load at time t; N is the number of samples.

The RMSE determines the precision by comparing the deviation between the estimated and real data. The RMSE is always a non-negative value and is calculated by:

$$e_{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(P_{ft} - P_{rt})^2} \tag{15}$$

The MAPE is close to the MAE but each gap between estimated and measured data is divided by the measured data in order to consider the relative gap.

$$e_{MAPE} = \frac{1}{N}\sum_{t=1}^{N}|\frac{P_{ft} - P_{rt}}{P_{rt}}| \times 100\% \tag{16}$$

VI. RESULTS AND DISCUSSION

A. EXPERIMENT SETTINGS AND COMPARISON ALGORITHMS

In the experiment, the period spans over 62 days. The data are split into three subsets for different purposes, namely training set (from 01-July-2015 to 13-August-2015), validating set (from 14-August-2015 to 25-August-2015), and testing set (from 26-August-2015 to 31-August-2015). In other words, the data split is 0.7/0.2/0.1. The training set is used to train forecasting models, and the validating set is used to select the best performing models, while the testing set is used for result evaluation at last. In this case, there are six days in total for evaluation. Considering the dataset are with 15-minute intervals, there are 27,648 time points to forecast for all consumers. For the empirical algorithms, both the training and validating sets are combined to derive the statistics for each consumer.

In order to verify the predictive performance of the proposed algorithm, we also selected one empirical algorithm and some machine learning algorithms for comparison.

1) EMPIRICAL ALGORITHM AND SETTINGS

Autoregressive Integrated Moving Average (ARIMA) model is a simple algebraic model describing stationary random time series. It is composed of an autoregressive model (abbreviated as AR model) and a moving average model (abbreviated as MA model). In $ARIMA(p, d, q)$, p is the autoregressive term, q is the moving average term, and d is the degree of differencing (the number of times the data have had past values subtracted). In this experiment, d is set to 2, p and q are determined by the autocorrelation graph and partial correlation graph of the consumer load data.

2) MACHINE LEARNING ALGORITHMS AND THEIR SETTINGS

Extreme Learning Machine (ELM) is a single hidden-layer feedforward neural network. ELM only needs to set the number of hidden nodes in the network. It is not necessary to adjust the input weights of the network and the offset of the hidden nodes during the execution of the algorithm, and the only optimal solutions can be generated. In the experiment, we tried several time steps, including {3, 5, 8, 10}, and use an exhaustive search strategy to obtain the optimal number of hidden layer nodes. The activation function is sigmoid. Due to the nature of random initialization of ELM, there is no concept of an epoch in the original ELM framework [7]. Therefore, the training of ELM repeats 200 times to obtain the best model.

Random Forest (RF) uses the decision tree as a base learner and further incorporates random feature selection in decision trees. In the experiment, we set the number of trees to {100, 200, 500} and set the maximum depth of the trees to 5.

In the **experiment of Support Vector Regression (SVR)** algorithm, we tried some standard kernel functions, such as RBF, sigmoid, linear, and polynomial.

3) THE PROPOSED ALGORITHM AND SETTINGS

In this experiment, we employed the Keras library to build this forecasting model. After parameter tuning, the number of training epochs of the LSTM model was 200. The number of nodes in the LSTM three parallel networks was 64, 40, and 32. The activation function of the LSTM layer is sigmoid. Adam is the optimizer in the training process. In this experiment, the LSTM time steps $K_2 = 3$ and $K_3 = 3$ are set by tuning parameter, and try to set $K_1 = \{3, 5, 8, 10\}$.

To summarize, the experiment settings are listed in Table 2.

TABLE 2. Experiment settings.

Algorithm	Parameters				
LSTM	Time Point-related LSTM	Hidden layers	Hidden nodes	Time steps	Epochs
		1	64	3,5,8,10	200
	Day-related LSTM	Hidden layers	Hidden nodes	Time steps	Epochs
		1	40	3	200
	Week-related LSTM	Hidden layers	Hidden nodes	Time steps	Epochs
		1	32	3	200
ARIMA	d	p		q	
	2	Varied		Varied	
RF	Number of trees		depth		
	100, 200, 500		5		
SVR	Kernel function				
	RBF, Sigmoid, Linear, Polynomial				
ELM	Activation function		Hidden nodes	Time steps	
	Sigmoid		Varied	3,5,8,10	

B. EXPERIMENTAL RESULTS AND ANALYSIS

1) PERFORMANCE FOR INDIVIDUAL NON-RESIDENTIAL CONSUMERS

We performed single-consumer predictions on the dataset of 48 consumers and finally averaged the predicting MAPE by each consumer as the final experimental results.

Generally, the forecasts for each time interval of each consumer are not as accurate as the forecasts for substation loads. This level of accuracy is anticipated because it was reported that the forecasting accuracy tends to drop significantly as the level of aggregation decreases [30], from 1.97% MAPE at the national level and 5.15% at university campus level to 13.8% at the village level. Kong *et al.* [7] also provided a range of MAPE between 44%-136% for residential load forecasting. Therefore, in our case, the accuracies achieved in Table 3 for various methods are reasonable.

From Table 3, it can be seen that the LSTM prediction model has the best predictive performance under each statistical indicators. The multiple sequence LSTM-based framework is generally the best. When the time point-related LSTM time step is set to 5, the optimal predictive performance can be obtained. The forecasting performance of RF is second only to LSTM. This is due to the fact that ensemble learning has good generalization performance. The traditional ARIMA predicting MAPE is 35.87%, slightly better than SVR and ELM. The predictive performance of SVR and ELM belongs to one level, their predicting MAPE range from 70% to 90%. With the increasing time steps, the MAPE of ELM increases. This may be caused that ELM cannot utilize sequence information. When the time step is increasing, ELM takes them as noise and the performance decreases.

TABLE 3. The experimental results of different models.

Algorithm/Scenario	MAE (KW)	MAPE (%)	RMSE (KW)
LSTM-3-3-3 time steps	24.17	22.97	36.83
LSTM-5-3-3 time steps	**23.84**	**22.45**	**36.07**
LSTM-8-3-3 time steps	23.89	22.88	36.12
LSTM-10-3-3 time steps	24.03	23.32	36.46
ARIMA	31.07	35.87	42.01
RF-100 estimators	27.53	26.92	40.32
RF-200 estimators	27.40	26.89	40.23
RF-500 estimators	27.48	26.90	40.28
SVR-RBF kernel	40.90	77.56	49.60
SVR-sigmoid kernel	42.92	82.16	51.63
SVR-linear kernel	35.89	63.85	45.42
SVR-polynomial kernel	92.51	151.50	112.01
ELM-3 time steps	58.78	85.47	74.54
ELM-5 time steps	62.26	91.72	78.64
ELM-8 time steps	64.67	95.79	81.49
ELM-10 time steps	65.37	106.33	82.07

2) PERFORMANCE WITH DIFFERENT FEATURES

As analyzed in section III, there exist multiple related time sequences such as adjacent time points (represented as E_1), the same time points in adjacent days (represented as E_2), and the same day in adjacent weeks (represented as E_3) among data samples for a specific consumer. Therefore, we performed four comparison experiments with different features.

- Feeding E_1 into the LSTM-based framework for prediction, which only utilizes the adjacent-time point correlation information;
- Feeding E_1 and E_2 into the LSTM-based framework for prediction, which employs both the adjacent-time point correlation information and the day-related correlation information;
- Feeding E_1 and E_3 into the LSTM-based framework for prediction, which uses both the adjacent-time point correlation information and the week-related correlation information;
- Feeding E_1, E_2 and E_3 into the LSTM-based framework for prediction, which takes advantage of all the three kinds of sequence information.

The four experiments aimed to verify the contributions of different features to the model. Similarly, we performed single-consumer predictions on the dataset of 48 consumers and finally averaged the predicting MAPE by each consumer in the final experimental results. The LSTM time point-related time step was set to 5, as shown in Fig.12.

From Fig.12, it can be seen that the MAPE decreases with more features. The week-related features contribute more than day-related features in the LSTM-based framework. Moreover, the result of feeding the three related features E_1, E_2 and E_3 into the framework is the best overall consumers.

From the perspective of single-consumer load forecasting, one non-residential consumer is selected from each industry, and the MAPEs are shown in Table 4.

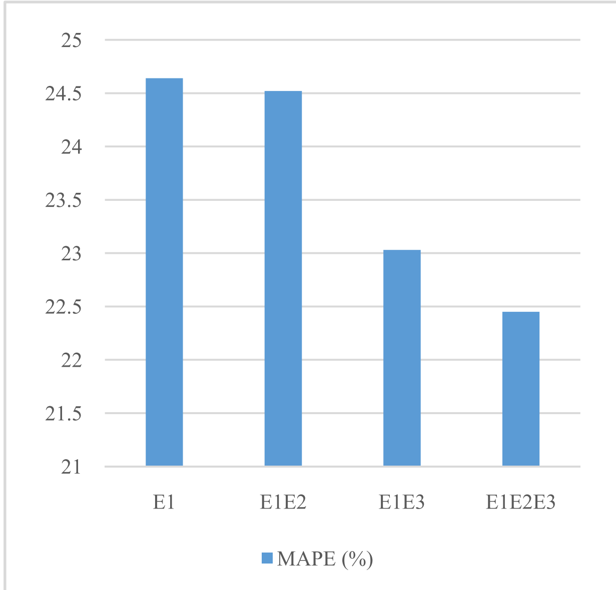


FIGURE 12. Experimental results of different features.

TABLE 4. Mape of different feature matrices for an individual non-residential consumer.

Consumer	E_1(%)	E_1E_2(%)	E_1E_3(%)	$E_1E_2E_3$(%)
alum3	47.63	42.94	47.58	**38.53**
print12	11.76	11.74	11.61	**11.52**
coal4	44.01	**41.23**	42.83	42.23
electro-nic4	30.00	23.38	17.68	**16.93**
gas14	27.17	26.96	26.93	**26.18**
paper5	47.07	46.01	43.02	**40.75**
restaurant1	9.39	8.10	7.92	**7.82**
steel4	57.56	55.82	**53.82**	54.70
hotel14	19.04	18.70	18.69	**18.52**

From Table 4, it can be seen that the contributions of day-related features and week-related features sometimes are high and sometimes are low. We infer that it depends on the each consumer's energy consumption patterns, as analyzed in section III. Furthermore, most of the best results belong to models with three sequence features.

In order to reveal the most effectiveness of the proposed framework, we counted the best forecasting algorithms among all the 48 non-residential consumers. As shown in Fig.13, there are 35 consumers getting best results with $E_1E_2E_3$, 4 consumers acquiring best results with E_1E_2, 7 consumers obtaining best results with E_1E_3, and only 2 consumers gaining best results with E_1. To sum up, in the proposed method, incorporating with three different kinds of sequence information helps improve the forecasting accuracy.

3) DIFFERENT CONSUMER FORECASTING PERFORMANCE COMPARISON

In Fig.13, it can be seen that not all the consumers' results will be the best when feeding feature vectors $E_1E_2E_3$. This depends on the consumers' energy consumption pattern.

If the patterns are irregular, LSTM cannot capture the consumer's time series correlation. Adding more time series features will decrease the forecasting performance.

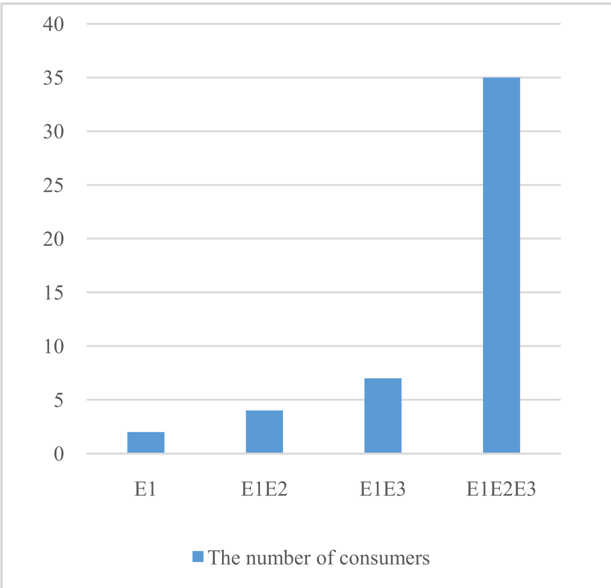


FIGURE 13. Statistics of the best predictive performance algorithm for each consumer.

In section III, we analyzed the load data features of a catering industry consumer, an electronic equipment manufacturing industrial consumer, and a hotel industry consumer, then used the clustering results and Spearman correlation coefficient to determine whether the consumers' energy consumption pattern is regular.

We utilized the proposed framework to forecast the catering industry consumer, the electronic equipment manufacturing consumer, and the hotel industry consumer and performed the four comparison experiments. The forecasting MAPE are shown in Table 5.

TABLE 5. Mape of different consumers.

Consumer	E_1(%)	E_1E_2(%)	E_1E_3(%)	$E_1E_2E_3$(%)
the catering industry	17.72	16.04	16.10	**15.45**
the electronic industry	26.60	20.11	19.87	**19.57**
the hotel industry	**17.00**	18.57	17.66	19.50

In Table 1, Spearman correlation coefficients of the restaurant industry consumer and the electronic equipment manufacturing industry consumer are high in the adjacent-time point correlation, day-related correlation, and week-related correlation, which all accept the hypothesis H_1. Therefore, in Table 5, with the addition of day-related features and week-related features, the prediction results are getting better. When the three feature vectors $E_1E_2E_3$ are fed to the model, the predictive performance is optimal, which predicting MAPEs are 2.27% and 7.03% lower than the one with feature vector E_1. However, when using LSTM to predict the load data for the hotel industry consumer, LSTM may not be able to learn valid information and did not achieve the better result because the consumer load data has no correlated information in time series. Beacause as shown in Table 1, Spearman correlation coefficients of the hotel industry consumer is low in the adjacent-time point correlation, day-related correlation, and week-related correlation, which all accept the hypothesis H_0.

As mentioned before, LSTM can capture long-distance dependencies in sequences, so that it can discover the energy consumption patterns of individual non-residential consumers. Moreover, the Spearman correlation coefficient is a typical indicator for correlation estimation. For a specific consumer, if the Spearman correlation coefficient for this consumer is high and the hypothesis H_1 is accepted, such time series features can be added to the proposed framework to improve the forecasting performance.

VII. CONCLUSION

Non-residential energy consumption accounts for a large proportion of total energy consumption and it is closely related to the stable operation of the grid. Therefore, it is important to address the issues of short-term non-residential load forecasting.

In this paper, k-means clustering is utilized to analyze the consumers' energy consumption patterns, while the Spearman correlation coefficient provides a measurement of correlation for sequence data. Then, a LSTM-based framework is proposed to forecast non-residential consumers' short-term energy consumption, which employs the sequence correlated features as input. For a specific consumer, if the Spearman correlation coefficient for this consumer is high and the hypothesis H_1 is accepted, such time series features can be added to the proposed framework to improve the forecasting performance. Case studies show that the proposed method obtains the best forecasting results in real dataset.

In the future work, we will consider the impact of some external factors in load forecasting, such as the current economic orientation and policy support in the region, and adding prior knowledge to gain better performance.

REFERENCES

[1] J. G. Jetcheva, M. Majidpour, and W.-P. Chen, "Neural network model ensembles for building-level electricity load forecasts," *Energy Buildings*, vol. 84, pp. 214–223, Dec. 2014.

[2] R. Ballini and R. R. Yager, "OWA filters and forecasting models applied to electric power load time series," *Evolving Syst.*, vol. 5, no. 3, pp. 159–173, 2014.

[3] J. Zheng, C. Xu, Z. Zhang, and X. Li, "Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network," in *Proc. 51st Annu. Conf. Inf. Sci. Syst.*, Mar. 2017, pp. 1–6.

[4] S. Rusitschka, K. Eger, and C. Gerdes, "Smart grid data cloud: A model for utilizing cloud computing in the smart grid domain," in *Proc. 1st IEEE Int. Conf. Smart Grid Commun.*, vol. 23, Oct. 2010, pp. 483–488.

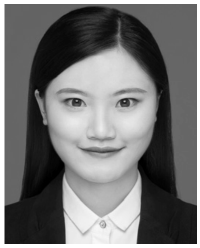
[5] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer, "Short-term load forecasting in a non-residential building contrasting models and attributes," *Energy Buildings*, vol. 92, pp. 322–330, Apr. 2015.

[6] *National Power Statistic*. Accessed: Dec. 18, 2017. [Online]. Available: http://data.stats.gov.cn/easyquery.htm?Cn=C01

[7] 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*, to be published.

[8] P. Vähäkyla, E. Hakonen, and P. Léman, "Short-term forecasting of grid load using Box-Jenkins techniques," *Int. J. Elect. Power Energy Syst.*, vol. 2, no. 1, pp. 29–34, 1980.
[9] H. Li, Y. Zhao, Z. Zhang, and X. Hu, "Short-term load forecasting based on the grid method and the time series fuzzy load forecasting method," in *Proc. Int. Conf. Renew. Power Gener.*, Oct. 2015, pp. 1–6.
[10] M. T. Hagan and S. M. Behr, "The time series approach to short term load forecasting," *IEEE Trans. Power Syst.*, vol. PWRS-2, no. 3, pp. 785–791, Aug. 1987.
[11] G. Juberias, R. Yunta, J. G. Moreno, and C. Mendivil, "A new ARIMA model for hourly load forecasting," in *Proc. IEEE Transmiss. Distrib. Conf.*, vol. 1, Apr. 1999, pp. 314–319.
[12] H. Z. Wang, J. Zhou, and K. Liu, "Summary of research on the short-term load forecasting method of the electric power system," *Elect. Automat.*, vol. 37, no. 1, pp. 1–3, Jan. 2015.
[13] D. C. Park, M. A. El-Sharkawi, R. J. Marks, II, L. E. Atlas, and M. J. Damborg, "Electric load forecasting using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 6, no. 2, pp. 442–449, May 1991.
[14] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.
[15] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, 2006.
[16] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 2, Jul. 2004, pp. 985–990.
[17] W.-M. Lin, C.-S. Tu, R.-F. Yang, and M.-T. Tsai, "Particle swarm optimisation aided least-square support vector machine for load forecast with spikes," *IET Gener. Transmiss. Distrib.*, vol. 10, no. 5, pp. 1145–1153, Apr. 2016.
[18] H. Jiang, Y. Zhang, E. Muljadi, J. J. Zhang, and 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.
[19] X. Wu, J. He, T. Yip, and P. Zhang, "A two-stage random forest method for short-term load forecasting," in *Proc. IEEE Eindhoven PowerTech*, Nov. 2016, pp. 1–6.
[20] X. Wu, J. He, T. Yip, and P. Zhang, "A two-stage random forest method for short-term load forecasting," in *Proc. IEEE Eindhoven PowerTech*, Jun. 2015, pp. 1–6.
[21] Y.-Y. Cheng, P. P. K. Chan, and Z.-W. Qiu, "Random forest based ensemble system for short term load forecasting," in *Proc. Int. Conf. Mach. Learn. Cybern.*, vol. 1, Jul. 2012, pp. 52–56.
[22] Y. K. Penya, C. E. Borges, and I. Fernández, "Short-term load forecasting in non-residential buildings," in *Proc. IEEE Africon*, vol. 83, no. 4, Sep. 2011, pp. 1–6.
[23] Y. K. Penya, C. E. Borges, D. Agote, and I. Fernández, "Short-term load forecasting in air-conditioned non-residential buildings," in *Proc. IEEE Int. Symp. Ind. Electron.*, vol. 83, Jun. 2011, pp. 1359–1364.
[24] J. L. Myers, A. D. Well, and R. F. Lorch, Jr., *Research Design and Statistical Analysi*, 2nd ed. Mahwah, NJ, USA: L.Erlbaum Associates, 2013.
[25] J. L. Myers, A. D. Well, and R. F. Lorch, Jr., *Research Design and Statistical Analysis*, 2nd ed. Mahwah, NJ, USA: Erlbaum Associates, 2003.
[26] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *Proc. 32nd Int. Conf. Mach. Learn. (JMLR)*, vol. 37, Jul. 2015, pp. 2342–2350.
[27] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *Proc. 32nd Int. Conf. Mach. Learn. (JMLR)*, 2015, pp. 2342–2350.
[28] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[29] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," in *Proc. 9th Int. Conf. Artif. Neural Netw. (ICANN)*, vol. 2, Sep. 1999, pp. 850–855.
[30] A. Marinescu, C. Harris, I. Dusparic, S. Clarke, and V. Cahill, "Residential electrical demand forecasting in very small scale: An evaluation of forecasting methods," in *Proc. 2nd Int. Workshop Softw. Eng. Challenges Smart Grid*, vol. 19, May 2013, pp. 25–32.

RUNHAI JIAO received the Ph.D. degree in computer science from Beihang University in 2008. He is currently an Associate Professor at North China Electric Power University. He has also been a Visiting Scholar at the Galvin Center for Electricity Innovation, Illinois Institute of Technology. He has authored over 20 papers in the international journals and meetings. His research interests include data mining, load forecasting, power transmission network planning, and distribution network automation.



TIANMING ZHANG is currently pursuing the master's degree in computer technology at North China Electric Power University. Her current research interests include deep learning and its application in load forecasting.

YIZHI JIANG is currently pursuing the master's degree at North China Electric Power University. Her research interests include machine learning and its applications in power systems, and data mining.

HUI HE received the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications in 2009. She is currently a Lecturer at North China Electric Power University. She has also been a Visiting Scholar at the Montreal Institute for Learning Algorithms, University of Montreal. She has authored over 20 papers in the international journals and conferences. Her research interests include renewable energy forecasting, deep learning, and electricity markets.