Energy 160 (2018) 1186–1200

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

A deep learning model for short-term power load and probability density forecasting

Zhifeng Guo [a, b], Kaile Zhou [a, b, c, *], Xiaoling Zhang [c], Shanlin Yang [a, b]

[a] School of Management, Hefei University of Technology, Hefei, China
[b] Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei University of Technology, Hefei, China
[c] Department of Public Policy, City University of Hong Kong, Kowloon, Hong Kong SAR, China

ARTICLE INFO

Article history:
Received 28 January 2018
Received in revised form
12 July 2018
Accepted 14 July 2018
Available online 17 July 2018

Keywords:
Deep learning
Probability density forecasting
Feature engineering
Power load forecasting

ABSTRACT

Accurate load forecasting is critical for power system planning and operational decision making. In this study, we are the first to utilize a deep feedforward network for short-term electricity load forecasting. Our results are compared to those of popular machine learning models such as random forest and gradient boosting machine models. Then, electricity consumption patterns are explored based on monthly, weekly and temperature-based patterns in terms of feature importance. Also, a probability density forecasting method based on deep learning, quantile regression and kernel density estimation is proposed. To verify the efficiency of the proposed methods, three case studies based on daily electricity consumption data for three Chinese cities for 2014 are conducted. The empirical results demonstrate that (1) the proposed deep learning-based approach exhibits better forecasting accuracy in terms of measuring electricity consumption relative to the random forest and gradient boosting model; (2) monthly, weekly and weather-related variables are key factors that have a great influence on household electricity consumption; and (3) the proposed probability density forecasting method is capable of forecasting high-quality prediction intervals via probability density forecasting.

1. Introduction

The electric power industry plays a key role in the economic development of a country, and its reliable operation contributes considerably to societal wellbeing. However, the main mode of power generation is that of thermal power generation, which has generated many air pollution problems. Statistics show that total global energy consumption levels are increasing [1]. Given this situation, improving electricity consumption efficiency levels has important implications for the sustainable development of society [2]. Smart meters, as import features of smart grids, are currently deployed in millions of households to collect detailed individual electricity consumption data. In such cases, energy big data analytics are effective tools for improving energy consumption behaviors and energy management systems [3]. In smart grids, large amounts and various types of data such as device status, electricity consumption, and user interaction data are collected. Then, many data analysis techniques such as those of optimization, forecasting, classification and clustering can be applied to these large volumes of smart grid big data.

Accurate power load forecasting plays an important role in reducing the amount of electricity consumed from the perspective of Demand Side Management (DSM). On the one hand, a generating station can make more reasonable production plans based on accurate demand forecasts, which is essential to improving energy efficiency levels. On the other hand, power load forecasting is fundamental to developing personalized intervention strategies for limiting household electricity consumption. Recently, a large body of research has focused on electricity consumption forecasting, and the recent introduction of smart grids has contributed to this area of research given their use of more specific data.

This study makes the following contributions to the existing literature. First, a deep neural network is used for load foresting purposes. We investigate the performance of a deep neural network with multi-layer perceptron (MLP) functions in terms of load forecasting as a typical deep learning model. Then, feature engineering is applied to identify the most influential factors. Based on this, data visualizations are employed to explore electricity

* Corresponding author. School of Management, Hefei University of Technology, Hefei, China.
E-mail addresses: zhoukaile@hfut.edu.cn, kailezhou@gmail.com (K. Zhou).

https://doi.org/10.1016/j.energy.2018.07.090
0360-5442/

consumption patterns. Finally, probability density forecasting is applied by combining deep learning model, quantile regression and kernel density estimation methods in applying the loss function.

The remainder of this paper is organized as follows. Section 2 presents a systematic review of the power load forecasting method, and many new methods are summarized and compared. Two deep learning-based methods for point prediction and probability density forecasting are presented in Section 3. Three case studies are given in Section 4 to verify the effectiveness and efficiency of the proposed model. Finally, in Section 5, conclusions are drawn.

2. Literature review

The accurate forecasting of electricity consumption levels can significantly benefit power systems. Currently, several methods support electricity demand forecasting. These can be mainly divided into statistical models and machine learning models according to model properties. Short-, medium- and long-term predictions can be made for forecasting purposes. In terms of forecasting outputs, load forecasting tools can be categorized into point forecasting, interval prediction and probability density forecasting tools.

Several studies have been conducted on forecasting, and numerous state-of-the-art methods and methodologies have been validated for this purpose [4–7]. For example, in terms of statistical models, a multiple regression (MR) model has been used to forecast daily electricity consumption levels of an administration building in London [8]. Five important independent variables, i.e., ambient temperature, solar radiation, relative humidity, wind speed and the weekday index, are used as explanatory variables. However, such models are limited in their instability and have not been used to predict the energy consumption levels of similar buildings. Do et al. [9] modeled for hourly electricity consumption in Germany and utilized temperature, industrial production levels, daylight hours and dummies for days of the week and months of the year to predict electricity consumption levels. Their results show that hourly electricity consumption levels are reflective of various typical profiles and should be separately modeled for each hour. Son and Kim [10] proposed a method based on support vector regression and fuzzy-rough feature selection that uses particle swarm optimization algorithms for the short-term forecasting of electricity demand. A total of 20 influential variables are considered, including monthly electricity consumption levels. El-Shazly [11] analyzed demand for electricity and executed out-of-sample forecasting at the sectorial level using a panel cointegration approach and found that the empirical model generates reliable ex-post forecasts for periods in time nearing the end of their sample period. Econometric models allow for cross-sectional heterogeneity within a dynamic framework that includes information on relevant income levels and on prices of domestic and foreign goods. Yukseltan and Yucekaya [12] established an hourly demand forecasting method for annual, weekly and daily horizons that involves using a linear model that takes into account the harmonics of these variations and the modulation of diurnal periodic variations based on seasonal variations. The approach presents a 3% error margin in the Mean Absolute Percentage Error (MAPE) norm. Al-Hamad and Qamber [13] utilized the multiple linear regression model and adaptive neuro-fuzzy inference system (ANFIS) to predict the Long Term Estimated Load output of the Kingdom of Bahrain. Present peak load, gross domestic product (GDP) and yearly population data were used as input variables. From peak loads estimated via multi linear regression and the ANFIS, the results show that long-term peak load demand forecasting results based on Neuro-Fuzzy and multiple linear regression methods present an acceptable error percentage. In turn, peak load estimation is essential to determining future demand in a given country [14].

A machine learning model must first convert a time series prediction problem into supervised learning problems. In turn, a series of methods are applied for predictions such as the ANN, SVM and extended models [15–18]. For example, electricity consumption is considered a function of socio-economic indicators such as population, gross national product, imports and exports [19]. A separate SVR model is trained for each input variable based on their current and past values, and such models are combined to yield consumption prediction values. Yildiz et al. [18] presented a review of different electricity load forecasting methods and showed that Artificial Neural Networks with Bayesian Regulation Back-propagation tools perform the best among almost all related models in predicting overall campus loads rather than single building loads. Amina et al. [20] presented a novel fuzzy wavelet neural network model and validated its performance in predicting power consumption on the Greek island of Crete. Xiong et al. [21] outlined a novel framework that integrates bivariate empirical mode decomposition (BEMD) and support vector regression (SVR) tools extended from the well-established empirical mode decomposition (EMD) method for the interval forecasting of power loads. The experimental study shows that the proposed modeling framework can improve prediction performance significantly and statistically outperforms some conventional forecasting methods presented in the literature. Ren et al. [22] proposed a number of single hidden layer network configurations with random weights (RWSLFN) and used these for short-term time series forecasting. Their results show that RWSLFNs with direct input-output connections statistically significantly outperform RWSLFN configurations without direct input–output connections. A machine-based model can achieve a higher degree of accuracy than traditional statistical models. However, machine learning models are easily overfit and are difficult to interpret when exploring key factors.

Existing methods for long-term time series prediction rely mainly on the use of iterated predictors, direct predictors and Multi-Input Multi-Output (MIMO) predictors more recently. Multi-step forecasting can be executed recursively by iterating a one-step model or by directly using a specific model for each horizon. For example, Dudek [23] compared several tools for short-term load forecasting based on neural networks such as MLP, radial basis function neural networks, generalized regression neural networks, fuzzy counter propagation neural networks, and self-organizing maps. They also compared neural network models to other popular forecasting methods (e.g., ARIMA and exponential smoothing). They found that the best results were achieved when employing generalized regression neural networks and one-neuron perceptron networks learned locally. Wang et al. [24] proposed a decomposition approach to modeling electricity demand and variability in medium- and long-term forecasting. The method divides historical time series into a number of components according to seasons and days of the week.

Interval prediction methods mainly focus on constructing high-quality prediction intervals (PIs) with a higher coverage probability. As the most traditional prediction method, probability density forecasting presents uncertainties through the use of probability distributions based on a prediction interval. For example, De Gooijer and Hyndman [25] presented a review of time series forecasting methods and found that probabilistic forecasting has become more popular in recent years. Unlike other commonly used methods, probabilistic forecasting generates the probability distribution of forecasting rather than a point. He et al. [26] proposed a probability density forecasting method based on the use of quantile regression neural networks using a triangle kernel function (QRNNT). Complete probability density curves were obtained, showing that the QRNNT model is capable of forecasting high-quality prediction intervals (PIs) with higher levels of coverage probability. Li et al. [27] proposed two Bayesian quantile regression models and a computationally efficient approach to forecasting the quantiles of electricity loads.

Several methods that improve prediction accuracy levels are available (e.g., ensemble learning, outlier removal, and feature engineering). For example, Sun et al. [28] employed a support vector regression model based on outlier detection methods to predict electricity consumption levels. Burger and Moura [29] presented a gated ensemble learning method for short-term electricity demand forecasting and showed that the combination of multiple models yields better results than the use of a single model. Silva [30] extracted several time- and weather-related features from a feature engineering standpoint and then built gradient-boosted decision rh ntrees and linear regression models. The model was ranked first out of several participating teams. Hassan et al. [31] applied various ensemble algorithms using individual neural network (NN) models as a basic learner for forecasting purposes and examined their efficiency levels. Their experimental results show that the aggregation algorithm generally outperforms single NN models. De Felice et al. [32] applied daily load forecasting for Italy by means of statistical modeling with the aim of exploring the influence of temperature. They proposed a numerical weather prediction model that serves as an indicator of the extent to which weather-related variables can predict electricity loads. They found that the use of weather data leads to an improvement in performance, especially for high temperature areas where the use of electricity is more heavily influenced by temperature. Ben Taieb and Hyndman [33] applied separate models for each hourly period and employed component-wise gradient boosting forest mating each model using univariate penalized regression splines as base learners. The models measure changes in electricity demand based on the time of year, the day of the week, the time of day, public holidays, current and past temperatures, and past demand, with the main predictors used being current and past temperatures and past demand. Their results were ranked fifth of 105 participating teams. Liu et al. [34] proposed a new integrated model that employs SWEMD (sliding window empirical mode decomposition), feature selection and a hybrid forecasting engine for small-scale load prediction. The model is compared to several well-established counterparts described in the literature.

The ongoing development of deep learning (DL) methods has resulted in the development of powerful tools that can manage large datasets and that often outperform traditional machine learning methods such as SVM and random forest (RF) methods used in many fields. Several recent studies have investigated power load forecasting by means of deep learning. For example, Tong et al. [35] proposed a deep learning-based model that works over two stages by refining features based on stacked denoising auto-encoders (SDAs) drawn from historical electricity demand data and related temperature-related variables.

This statistical model offers explanatory power but less forecasting ability than machine learning models. Both statistical and machine learning methods play a key role in specific cases. Long-term forecasts are more in line with actual demand. However, long-term forecasts are more difficult to perform than short-term forecasts. Recently, deep learning methods have received more attention in the field of image processing and natural language. More flexible neural network architectures can be used for power load forecasting. For instance, the LSTM (long-term memory model) is suitable to use for MIMO (multiple input and multiple output) tasks.

3. Deep learning-based model

Deep learning methods, which have been developed rapidly in recent years, are a broad class of models that are more complex than traditional shallow neural networks. They mainly include MLP, deep belief, convolution neural and recurrent neural networks [36]. Compared to traditional single hidden layer neural networks, deep neural networks can extract features and can manage numerous sequences and large image datasets. For example, convolution neural networks perform very well in the area of image recognition, performing similar to recurrent neural networks in the field of natural language processing [37]. This is mainly attributable to the fact that deep neural networks include more hidden layers and flexible structures. The use of more complex networks renders training more difficult (e.g., vanishing and exploding gradients). Hinton proposed that the use of pre-training can help facilitate deep neural network learning [38], which has considerably accelerated the development of deep learning methods. In recent years, the use of big data, cloud computing and related new technologies has also contributed to deep neural network training.

In this study, we used deep learning-based models for short-term electricity consumption forecasting to identify key influencing factors and for probability density predictions.

3.1. Deep learning-based framework for electricity consumption forecasting

A block diagram of the proposed short-term power load forecasting framework based on deep learning is shown in Fig. 1.

The framework involves data preprocessing, deep learning, model evaluation and parameter tuning. In the data pre-processing stage, missing values and outliers are removed. A time series is converted into a supervised learning problem. In the deep learning stage, initial neural networks (e.g., the number of hidden layers and the number of neurons in each hidden layer) are determined. Next, training data, validation sets and test data are determined. Finally, a deep neural network is trained. The performance of a model is evaluated by means of several related indicators (e.g., the MAE, MPRE and MAPE). Parameters used in the model are adjusted to enhance its performance. The deep learning model of this framework is referred to as MLP.

3.1.1. Model representation for the MLP

The MLP is a typical deep neural network model [36] serving as a feedforward neural network in which information is propagated from front to back. Fig. 2 shows a typical shallow layer neural network with a single hidden layer, and Fig. 3 shows a MLP network with three hidden layers.

From the first to the second layer, we have

$$z_l^2 = w_{l0}^{(1)} + \sum_{j=1}^{p} w_{lj}^{(1)} x_j \quad (1)$$

$$a_l^{(2)} = g^{(2)}\left(z_l^2\right) \quad (2)$$

where x_j denotes input variables. $w_{lj}^{(1)}$ is the weight between unit l and unit j. Intercept term $w_{l0}^{(1)}$ is a bias item. g is a nonlinear transformation function such as sigmoid function $g(t) = 1/(1 + e^{-t})$.

More generally, we have a transition from layer $k-1$ to layer k:

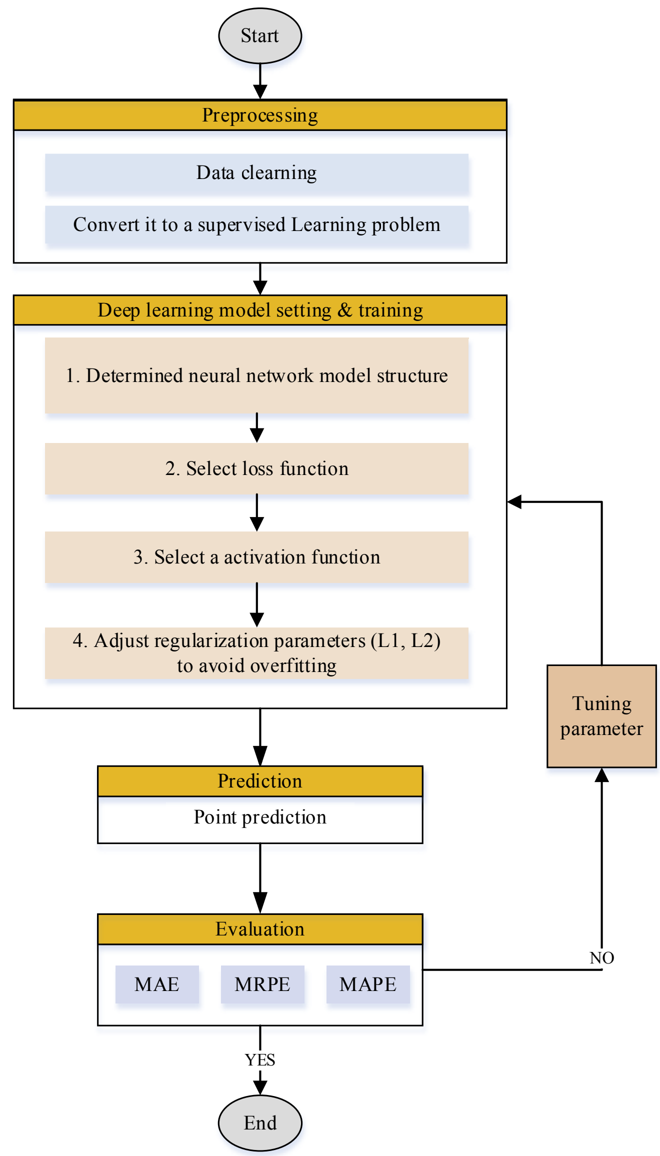

Fig. 1. Block diagram of the proposed short-term power load forecasting method based on deep learning.

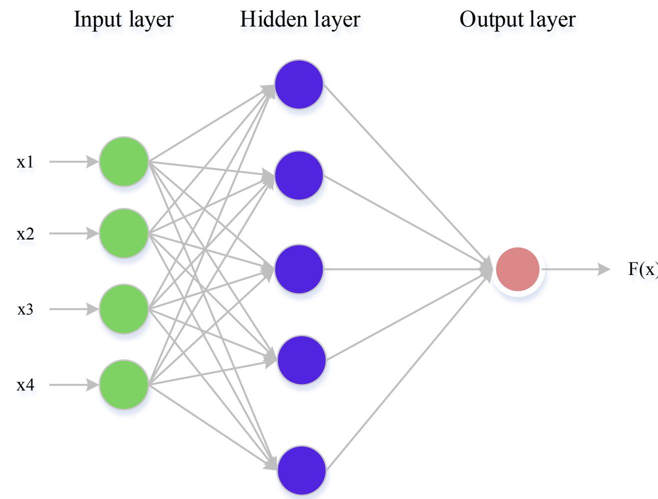

Fig. 2. Traditional shallow layer neural network.

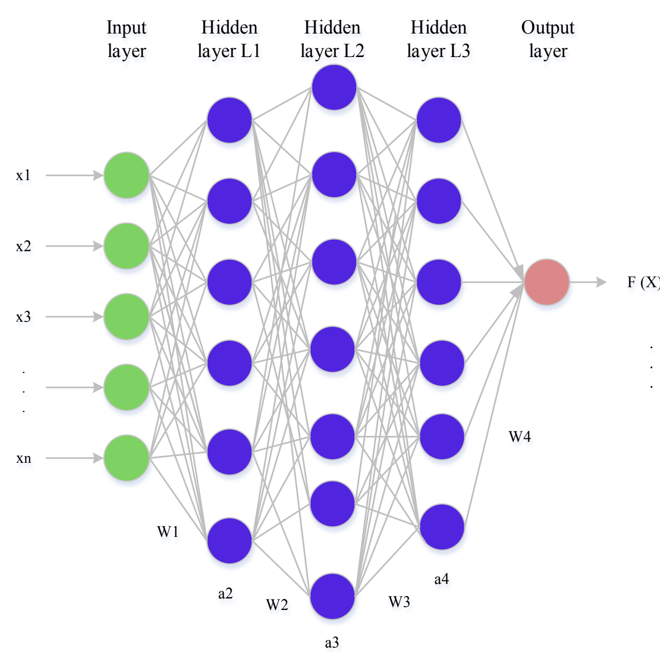

Fig. 3. Deep neural networks with three hidden layers.

$$z_l^{(k)} = w_{l0}^{(k-1)} + \sum_{j=1}^{p_{k-1}} w_{lj}^{(k-1)} a_j^{(k-1)} \tag{3}$$

$$a_l^{(k)} = g^{(k)}\left(z_l^k\right) \tag{4}$$

In this case, l refers to the $k-1$ layer and lj refers to unit l of layer k and to unit j of layer $k-1$. In this case, the final output is $F(x_1, x_2, \ldots, x_{p^{(0)}}) = h(w_0^{(k)} + \sum_{l=1}^{p_k} w_l^{(k)} a_l^{(k)})$. h is another transformation function (e.g., ReLU in deep learning settings).

3.1.2. Parameter learning

As we show above, we derive a loss function through neural network propagation. A regularization term $J(w)$ is also used to prevent overfitting. The objective function is as follows.

$$\min\left\{\frac{1}{n}\sum_{i=1}^{n} L[y_i, f(x; w)] + \lambda J(w)\right\} \tag{5}$$

where $J(w) = \frac{1}{2}\sum_{k=1}^{K-1}\sum_{j=1}^{p_k}\sum_{l=1}^{p_{k+1}}\{w_{lj}^k\}^2$.

Here, we use a backpropagation algorithm [39] to obtain the gradient (see Table 1).

Furthermore, the maximal relative percentage error (MRPE), mean absolute percentage error (MAPE), and mean absolute error (MAE) as evaluation indicators are written as follows.

$$MRPE = \max\left|\frac{\hat{X}^t - X^t}{X^t}\right| \times 100\%, t = 1, 2, \ldots, N \tag{6}$$

Table 1
BP algorithm.

Algorithm: Backpropagation

1. For a given pair (x, y), perform a "feedforward pass" while computing activations $a_l^{(k)}$ at each of layer $L_2, L_3, \ldots, L_K$; i.e., compute $f(x; W)$ at x from the current W while incrementally saving each intermediary quantity.
2. For each output unit l of layer L_K, compute

$$\delta_l^K = \frac{\partial L[y, f(x, W)]}{\partial z_l^K} = \frac{\partial L(y, f(x, W))}{\partial a_l^{(K)}} \overset{\bullet}{g}^{(K)}(z_l^{(K)})$$

where $\overset{\bullet}{g}$ denotes the derivative of $g(z)$ with respect to z.

3. For layers $k = K-1, K-2, \ldots, 2$ and for each node l of layer k, set

$$\delta_l^{(k)} = (\textstyle\sum_{j=1}^{p_{k+1}} w_{jl}^{(k)} \delta_j^{(k+1)}) \overset{\bullet}{g}^{(k)}(z_l^{(k)})$$

4. Partial derivatives are given by:

$$\frac{\partial L(y, f(x, W))}{\partial w_{lj}^{(k)}} = a_j^{(k)} \delta_l^{k+1}$$

$$MAE = \frac{1}{N} \sum_{t=1}^{N} \left| \widehat{X}^t - X^t \right| \times 100\% \tag{7}$$

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{\widehat{X}^t - X^t}{X^t} \right| \times 100\% \tag{8}$$

in which $\widehat{X}^t$ is the predicted value for the t period. X^t is the real value of electricity consumption. N is the total number of predictions made.

3.2. Deep learning-based method for short-term power probability density forecasting

Probability density forecasting can be used to generate high-quality prediction intervals (PIs) with a higher coverage probability. In this section, we propose a deep learning-based method for probability density forecasting (see Fig. 4) in which deep neural networks and quantile regressions are combined through a loss function. We first obtain quantile points by combining deep neural networks with quantile regressions. Second, probability density forecasting is executed through kernel density estimations.

3.2.1. Quantile regression

Quantile regression, as proposed by Koenker and Bassett [40], represents an important extension of the traditional regression model. A traditional regression model is mainly used to directly analyze the relationship between independent variables and mean values of dependent variables. However, quantile regressions can be used to determine the relationship between independent variables and the quantiles of dependent variables. Compared to traditional least squares regressions, quantile regressions can identify more comprehensive relationships and thus offer more explanatory power when applied to non-normal data.

Suppose that there is a random variable Y characterized by its distribution function.

$$F(y) = P(Y \le y) \tag{9}$$

where for any quantile $0 < \tau < 1$.

$$Q(\tau) = \inf\{y : F(y) \ge \tau\} \tag{10}$$

is referred to as the τ-th quantile of Y.

The objective function of a quantile regression, which is determined by a loss function, is as follows.

$$\min_{\beta \in R^p} \sum_{i=1}^{n} \rho_\tau\left(y_i - x_i^T \beta\right) \tag{11}$$

where $\rho_\tau(u)$ is the loss function defined as

$$\rho_\tau(\mu) = \mu(\tau - I(\mu)) \tag{12}$$

where $I(u) = \begin{cases} 0, u \ge 0 \\ 1, u < 0 \end{cases}$. This is a convex and piecewise indicator function. A simplified expression of the linear quantile regression model is written as follows.

$$Q_y(\tau|x) = x^T \beta(\tau), \tau \in (0, 1) \tag{13}$$

where $Q_y(\tau|x)$ is the conditional τ quantile of response variables y under explanatory variables x. $\beta(\tau)$ is the regression coefficient vector. Parameter $\beta(\tau)$ is estimated as follows.

$$\beta(\tau) = \underset{\beta \in R^p}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau\left(y_i - x_i^T \beta\right) \tag{14}$$

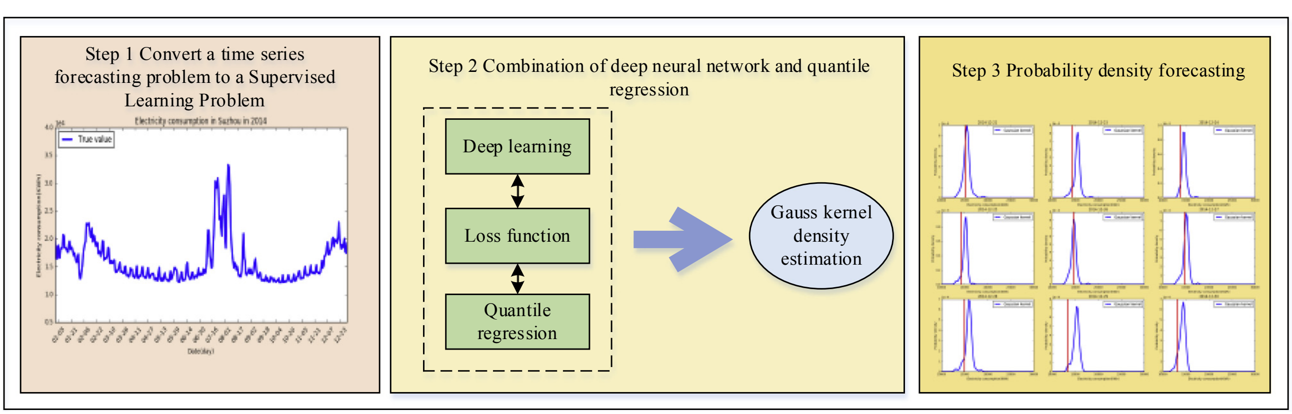

Fig. 4. Block diagram of the proposed short-term power load probability density forecasting method.

It is evident that the effects of explanatory variables on the response of conditional quantiles at different quantile points can be determined according to parameter $\beta(\tau)$.

3.2.2. Kernel density estimation

The probability density estimation approach is a classical statistical method used to estimate a distribution function according to a sample distribution. It is a modern nonparametric method. Kernel density estimations are commonly used to estimate probability densities [41]. They determine densities by means of a kernel function that is a combination of local data and probability densities. The most commonly used kernel function is the Gaussian kernel. A conditional quantile is equivalent to a conditional density value [42]. Therefore, a conditional probability density value can be estimated from a conditional fractal number. Let

$$Z_\tau = \hat{Q}_Y(\tau|X) \quad \tau = (1, 2, ..., n) \tag{15}$$

where n is equal to the number of quantiles.

Let $(Z_1, Z_2, ...Z_T)$ be an independent variable and identical quantile function as that drawn from the same distribution with a density function of f. Its kernel density estimator is

$$\hat{f}_n(Z) = \frac{1}{T}\sum_{\tau=1}^{T} K_h(Z - Z_t) = \frac{1}{Th}\sum_{\tau=1}^{T} K\left(\frac{Z - Z_t}{h}\right) \tag{16}$$

where h is the bandwidth value. $K(x)$ is a kernel function. The Gaussian kernel is written as follows:

$$K(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-u^2/2\right) \tag{17}$$

When estimating $\hat{f}$, the bandwidth is determined according to the MISE defined as follows.

$$h = \text{argmin}(MISE(f_h)) \tag{18}$$

where

$$MISE(f_h) = E_f\left(\hat{f} - f\right)^2 = E_f\left[\int \hat{f}^2 - 2\int \hat{f}*f + \int f^2\right] \tag{19}$$

$$Var\left(\hat{f}_h\right) = \frac{1}{Th}\|K\|_2^2 f + o\left(\frac{1}{Th}\right) \text{as } Th \to +\infty \tag{20}$$

In this case,

$$\begin{aligned} MISE\left(\hat{f}_h\right) &= \frac{1}{nh}\|K\|_2^2\int f(x)dx + \frac{h^4}{4}(u_2(K))^2\int\left(f''(x)\right)^2dx \\ &\quad + o\left(\frac{1}{nh}\right) + o\left(h^4\right) \\ &= \frac{1}{nh}\|K\|_2^2 + \frac{h^4}{4}(u_2(K))^2\left\|f''\right\|_2^2 + o\left(\frac{1}{nh}\right) \\ &\quad + o\left(h^4\right) \text{As} n \to 0, nh \to \infty. \end{aligned} \tag{21}$$

When disregarding higher-order terms, an approximate formula for the MISE called the AMISE can be written as

$$AMISE\left(\hat{f}_h\right) = \frac{1}{nh}\|K\|_2^2 + \frac{h^2}{4}(\mu_2(K))^2\left\|f''\right\|_2^2 \tag{21}$$

In this case,

$$h_{opt} = \left(\frac{\|K\|_2^2}{\|f''\|_2^2(u_2(K))^2 n}\right)^{1/5} \sim n^{-1/5} \tag{22}$$

This summarizes the direct plug-in approach to bandwidth selection.

4. Case study

4.1. Data description

To verify the feasibility of the proposed method, electricity consumption data collected from three cities in Jiangsu province, China, are treated as a training set. In the preprocessing phase, outliers and missing values are removed for the sake of efficiency. Finally, 3000 users are randomly selected. We forecast their total electricity consumption from a deep learning-based framework, random forest [43] and gradient boosting machine [44]. Rather than making time series predictions using a machine learning model, the time series forecasting problem is first converted into a supervised learning problem via machine learning. We use 349 samples for in-sample data and 10 for out-sample data for our test, and 70% of the in-sample data are used for training, while 30% are used for validation.

4.2. Feature engineering

Given the various factors that affect residential electricity consumption, we employ feature engineering to identify influencing factors. We divide these factors into four categories. We first considered date-related factors (e.g., monthly, daily, seasonal data). Monthly, seasonal and yearly electricity consumption patterns were measured. Second, we measured air-quality-related factors (e.g., PM2.5, SO_2, CO, etc.). Global air pollution levels have recently attracted more attention. Air quality levels affect human engagement in outdoor activities and thus alter electricity consumption patterns. Third, we considered weather-related factors such as rainfall levels, daily temperatures, etc. Fourth, we considered economic factors, as economic patterns affect resident living standards. We use the CPI as a target factor. All influencing factors considered are listed in Table 2.

4.3. Nanjing case study

In this section, electricity consumption levels for Nanjing for January 1, 2014, to December 31, 2014, are predicted. Parameter settings used for the deep learning model, for the random forest and for gradient boosting machine learning are shown in Table 3. Hidden layers in deep neural network models have a 200-200-200 structure with 200 nodes found in each hidden layer. We employ regularization parameter L1 = 0.001, L2 = 0.001 to prevent overfitting. The learning rate is set as 0.1, and the maximum number of iterations is set as 1000. For the random forest, the number of trees is set as 200, and the maximum tree depth is set as 20. The same settings are suitable for use in a gradient boosting machine.

4.3.1. Point prediction

Electricity consumption levels for Nanjing for June 1, 2014, to December 31, 2014, are shown in Fig. 5. Three pronounced fluctuations are observed in February, August and December. Fluctuations occurring in February are attributed to the occurrence of the Spring Festival. The Spring Festival is China's largest festival, and thus, residents' lifestyles change considerably during this period, leading to a significant change in electricity consumption levels.

Table 2
Features used for forecasting.

ID	Name	Type	ID	Name	Type
1	Month	Categorical	18	Mean difference in temperature (pre 1–3)	Numeric
2	Day of week	Categorical	19	Min average temperature (pre 1–3)	Numeric
3	Electricity consumption [*t*-*n*]	Numeric	20	Mean average temperature (pre 1–3)	Numeric
4	Max electricity consumption (pre 1–7)	Numeric	21	AIR QUALITY INDEX	Numeric
5	Min electricity consumption (pre 1–7)	Numeric	22	PM2.5 [*t*-*n*]	Numeric
6	Mean electricity consumption (pre 1–7)	Numeric	23	Max PM2.5 (pre 1–3)	Numeric
7	Highest temperature [*t*-*n*]	Numeric	24	Min PM2.5 (pre 1–3)	Numeric
8	Max highest temperature (pre 1–3)	Numeric	25	Mean PM2.5 (pre 1–3)	Numeric
9	Min highest temperature (pre 1–3)	Numeric	26	PM10	Numeric
10	Mean highest temperature (pre 1–3)	Numeric	27	SO2	Numeric
11	Lowest temperature [*t*-*n*]	Numeric	28	NO2	Numeric
12	Min lowest temperature (pre 1–3)	Numeric	29	CO	Numeric
13	Max lowest temperature (pre 1–3)	Numeric	30	O3	Numeric
14	Mean lowest temperature (pre 1–3)	Numeric	31	CPI	Numeric
15	Difference in temperature [*t*-*n*]	Numeric	32	Rain	Categorical
16	Max difference in temperature (pre 1–3)	Numeric	33	Average temperature [*t*-*n*]	Numeric
17	Min difference in temperature (pre 1–3)	Numeric			

Note: $n = 1, 2, 3$.

Table 3
RF, GBM and DL parameters.

Models	Parameters	Value
RF	The number of trees	200
	The maximum tree depth	20
	Early stopping based on stopping_metric convergence	2
	Relative tolerance of the metric-based stopping criterion	0.001
GBM	The number of trees	200
	The maximum tree depth	20
	Early stopping based on stopping_metric convergence	2
	Relative tolerance of the metric-based stopping criterion	0.01
	Learning rate	0.2
DL	The number of net layers	5
	The number of hidden layers	3
	The number of layer nodes	[1,45,200,200,200]
	L1	0.001
	L2	0.001
	The maximum number of iterations	1000
	Activation function	Rectifier
	Learning rate	0.1

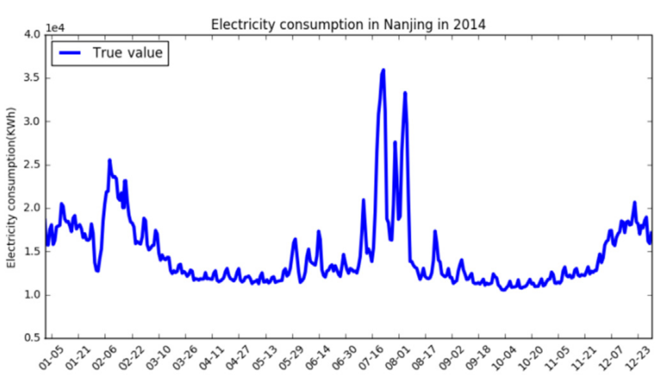

Fig. 5. Electricity consumption in Nanjing from January 1, 2014, to December 31, 2014.

Electricity consumption levels also change with extreme temperature fluctuations. Fluctuations occurring in August are mainly caused by high temperatures while fluctuations occurring December are also related to temperature changes.

MAPE, MRPE and MAE results for RF, GBM and MLP are shown in Table 4. Based on Table 4, MLP MAPE, MRPE, and MAE values are measured as 3%, 5% and 620, respectively. This clearly shows that MLP forecast errors are less significant than those of other models. A comparison of predictions made from the proposed method and from existing methods is presented in Fig. 6.

4.3.2. *Variable importance*

Levels of variable importance according to deep learning, random forest and gradient boosting machine methods are shown in Tables 5–7, respectively.

According to Table 5, monthly and weekly data are the most important variables for the deep learning model. Electricity consumption levels of the day before, temperature-related factors and air-quality-related factors are less important. However, according to Tables 6 and 7, electricity consumption levels of the day before and temperature-related factors are most the important according to the random forest model and gradient boosting machine. There is no need to determine which factors are the most important, as

Table 4
Forecasting errors of three algorithms measuring electricity consumption in Nanjing.

Models	MAPE (%)	MRPE (%)	MAE (KWh)
RF	0.05925	0.170209	1060.406
GBM	0.043692	0.077298	775.1881
MLP	0.035504	0.057273	620.0159

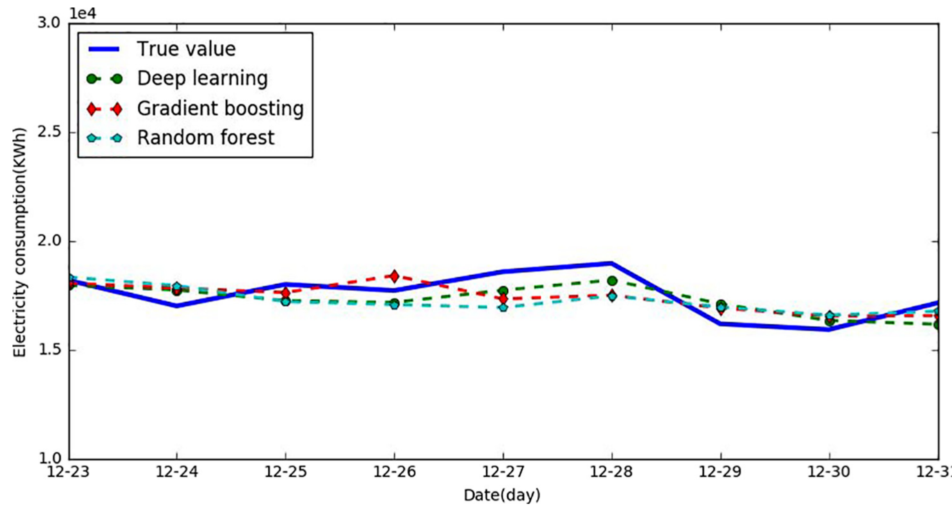

Fig. 6. Forecasts derived from different methods employing daily electricity consumption data.

Table 5
Top 20 most important variables for Nanjing derived from the deep learning model.

Rank	Feature	Importance	Rank	Feature	Importance
1	December	1.00	11	Wednesday	0.86
2	June	0.96	12	July	0.84
3	March	0.96	13	May	0.84
4	Monday	0.95	14	April	0.81
5	Tuesday	0.93	15	January	0.78
6	Sunday	0.91	16	November	0.76
7	Thursday	0.90	17	rain(1)	0.72
8	August	0.89	18	February	0.72
9	Saturday	0.89	19	September	0.69
10	Friday	0.87	20	October	0.57

Table 6
Top 20 most important variables for Nanjing derived from the random forest model.

Rank	Variable	Importance
1	Electricity consumption (t-1)	1.00
2	Electricity consumption (t-2)	0.32
3	Average temperature (t)	0.25
4	Lowest temperature (t-1)	0.20
5	Max electricity consumption (pre 1–7)	0.19
6	Min electricity consumption (pre 1–7)	0.17
7	Lowest temperature (t)	0.14
8	Mean electricity consumption (pre 1–7)	0.09
9	Average temperature (t-1)	0.05
10	Highest temperature (t)	0.04
11	Mean highest temperature (pre 1–3)	0.04
12	Min lowest temperature (pre 1–3)	0.03
13	O3	0.03
14	Mean average temperature (pre 1–3)	0.02
15	Month	0.02
16	Max average temperature (pre 1–3)	0.02
17	Day of the week	0.01
18	Difference in temperature (t-1)	0.01
19	Electricity consumption (t-7)	0.01
20	Electricity consumption (t-3)	0.01

the mechanisms of the each machine model differ. In fact, these factors do influence electricity consumption. Based on this, we explore the relationship between these factors and electricity demand through data visualization.

The relationship between high local temperatures and electricity consumption for June to August is shown in Fig. 7. The blue line shown in the figure denotes the average electricity

Table 7
Top 20 most important variables for Nanjing derived from the gradient boosting model.

Rank	Feature	Importance
1	Electricity consumption (t-1)	1.0000
2	Average temperature (t)	0.1276
3	Highest temperature (t)	0.0400
4	Electricity consumption (t-2)	0.0132
5	Lowest temperature (t)	0.0130
6	Month	0.0124
7	Day of the week	0.0117
8	Electricity consumption (t-7)	0.0068
9	O3	0.0050
10	Max difference in temperature	0.0042
11	Min electricity consumption (pre 1–7)	0.0035
12	Average temperature (t-3)	0.0026
13	Highest temperature (t-1)	0.0026
14	Difference in temperature (t)	0.0024
15	Min average temperature	0.0023
16	Mean average temperature (pre 1–3)	0.0022
17	Max electricity consumption (pre 1–7)	0.0016
18	CO	0.0014
19	Average temperature (t-1)	0.0010
20	PM10	0.0007

consumption level, the purple line is the temperature curve, and vertical dashed lines denote days for which high local temperatures were recorded. We find that for Nanjing from June to August, high local temperatures are always accompanied by high local levels of electricity consumption (see the vertical dashed line shown in Fig. 7). Temperature has a considerable influence on residential electricity consumption. The months of June, July and August are considered summer season months according to the Chinese lunar calendar. High temperatures reduce resident engagement in outdoor activities while increasing their engagement in indoor activities, inevitably leading to an increase in electricity consumption. As temperatures increase, many families use more home appliances such as fans and air conditioning units, also increasing electricity consumption.

A weekly electricity consumption heatmap is shown in Fig. 8. The horizontal axis covers days of the week, while the vertical axis covers months of the year, and each cell represents the average electricity consumption level for a given day of the week. The darker a cell is, the more electricity is consumed. Here, we find that more electricity is consumed in July and August and on the

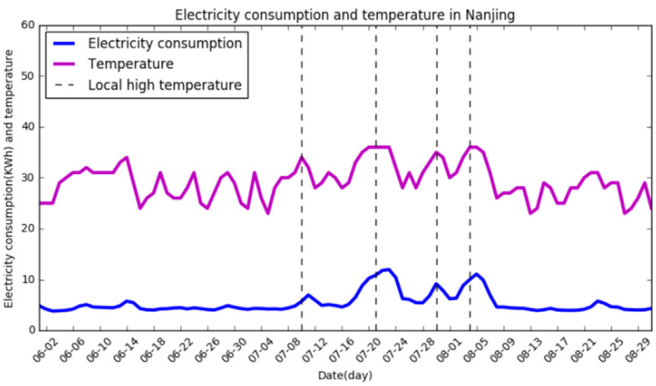

Fig. 7. The relationship between high local temperatures and electricity consumption.

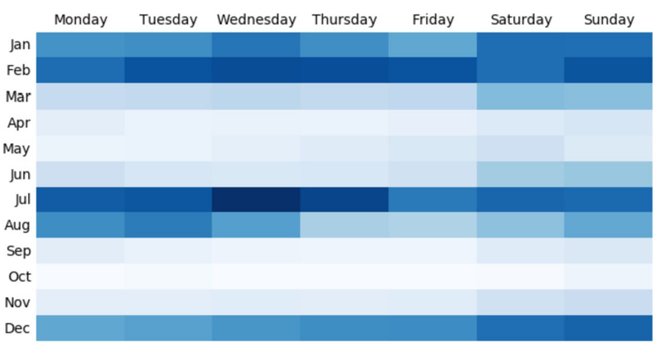

Fig. 8. Heatmap for weekly electricity consumption.

weekends. In July and August, the most electricity is consumed on a single day and not on weekends, reflecting the influence of high temperatures on residents' patterns of life.

Fig. 9 presents a monthly electricity consumption bar plot. We find that electricity consumption levels are significantly higher during December, January, February, July and August. Factors leading to an increase in electricity consumption vary. However, temperature changes and the occurrence of festivals always lead to an increase in electricity consumption. For instance, low temperatures combined with the occurrence of the Spring Festival in February, China's most magnificent traditional festival, lead to an increase in electricity consumption in this month. The months of July and August are the hottest of the year. In high temperature periods, residents' lifestyle patterns change significantly, leading to an increase in electricity consumption.

4.3.3. Probability density forecasting

Probability density forecasting can be used to derive probability distributions based on a prediction interval that captures prediction uncertainties. Policy makers can generate more reasonable plans according to probability distributions. Fig. 10 shows probability density forecasts for Nanjing for December 22, 2014, to December 30, 2014, based on deep learning probability density values.

Fig. 10 shows that the probability distribution of electricity consumption for December 22, 2014, to December 30, 2014, can be estimated through deep neural network quantile regression. Actual electricity consumption values are also shown in the probability density function diagram. All actual values are positioned in the center of the probability density curve except for the value for the ninth day. Actual values are positioned close to the maximum probability point in the first, second, fourth, fifth, sixth and seventh subplots. Fig. 10 shows that the median and mode of probability density curves for daily electricity consumption are positioned within the predicted lower and upper bounds in most cases.

4.4. Lianyungang case study

4.4.1. Point prediction

Electricity consumption in Lianyungang from June 1, 2014, to December 31, 2014, is illustrated in Fig. 11. Here, we find less significant fluctuations in electricity consumption occurring around the Spring Festival relative to those found for Nanjing. The trends are similar to those found for Nanjing for other time periods.

MAPE, MRPE and MAE results for RF, GBM and MLP are shown in Table 8. MAPE, MRPE, MAE of MLP values are measured as 1%, 3% and 243, respectively. This clearly shows that MLP forecasting errors are superior to those of the other models. Electricity consumption levels for Lianyungang for January 1, 2014, to December 31, 2014, are shown in Fig. 11 and a comparison of predictions made using the proposed versus existing methods is given in Fig. 12.

4.4.2. Variable importance

According to Table 9, similar to what was found for Nanjing, the

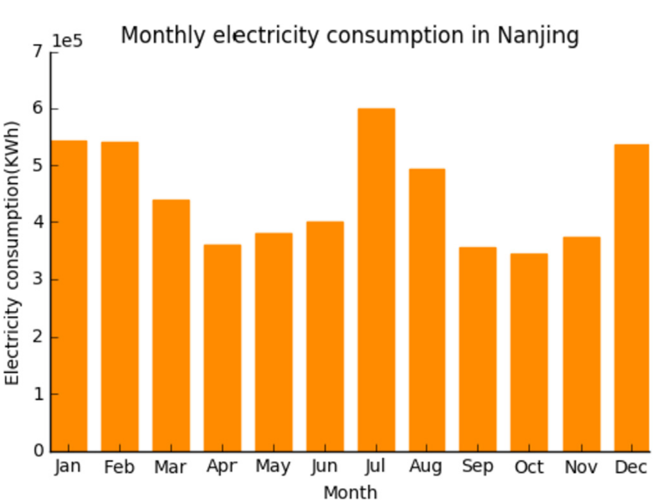

Fig. 9. Monthly electricity consumption levels for Nanjing for 2014.

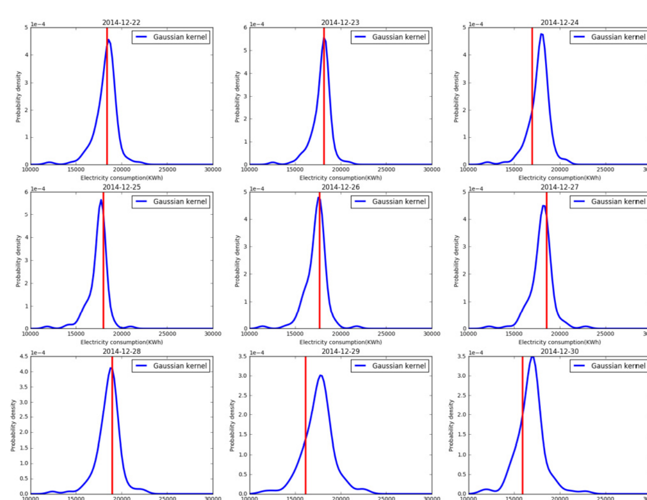

Fig. 10. Probability density functions for Nanjing for December 22, 2014, to December 30, 2014.

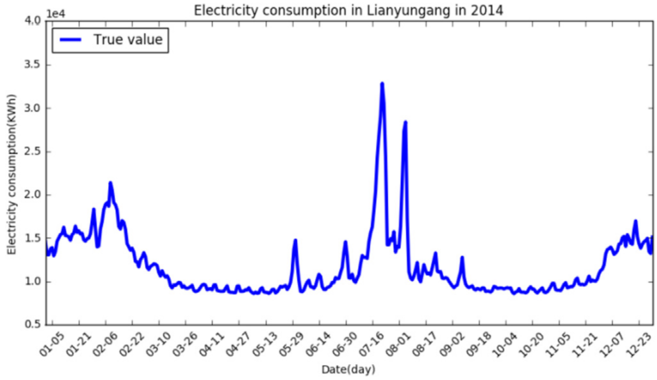

Fig. 11. Electricity consumption in Lianyungang from January 1, 2014, to December 31, 2014.

Table 8
Forecasting errors of the three algorithms for electricity consumption in Lianyungang.

Models	MAPE (%)	MRPE (%)	MAE (KWh)
RF	0.067896	0.153869	948.1974
GBM	0.048661	0.155191	706.3035
MLP	0.016992	0.034206	243.3405

month of the year and day of the week are the most important variables considered according to the deep learning model. Electricity consumption for the day before, temperature-related factors and air-quality-related factors are less important. Levels of variable importance based on the random forest and gradient boosting machine models are shown in Tables 10 and 11, respectively. According to Tables 10 and 11, the random forest and gradient boosting machine models focus more on electricity consumption levels for the day before and temperature-related factors. Here, we explore the relationship between these factors and electricity demand via data visualization.

The relationship between high local temperatures and electricity consumption is shown in Fig. 13. We find that from June to August in Lianyungang, extreme temperatures are always accompanied by high levels of local electricity consumption (see the vertical dashed line in Fig. 13). Temperatures have an important influence on residential electricity consumption. High temperatures reduce resident engagement in outdoor activities while increasing their engagement in indoor activities, inevitably leading to an increase in electricity consumption. As temperatures increase, many families use more home appliances such as fans and air conditioning units, also increasing electricity consumption.

Table 9
Top 20 most important variable for Lianyungang derived from the deep learning model.

Rank	Feature	Importance	Rank	Feature	Importance
1	May	1.00	11	November	0.62
2	Tuesday	0.82	12	September	0.59
3	January	0.81	13	Sunday	0.59
4	February	0.77	14	June	0.58
5	Wednesday	0.76	15	Rain(1)	0.57
6	July	0.71	16	April	0.50
7	August	0.68	17	Friday	0.41
8	Monday	0.67	18	October	0.38
9	December	0.66	19	Thursday	0.36
10	Saturday	0.66	20	March	0.27

Table 10
Top 20 most important variables for Lianyungang derived from the random forest model.

Rank	Feature	Importance
1	Electricity consumption (t-1)	1.00
2	Electricity consumption (t-2)	0.32
3	Average temperature (t)	0.25
4	Lowest temperature (t-1)	0.20
5	Max electricity consumption (pre 1–7)	0.19
6	Min electricity consumption (pre 1–7)	0.17
7	Lowest temperature (t)	0.14
8	Mean electricity consumption (pre 1–7)	0.09
9	Average temperature (t-1)	0.05
10	Highest temperature (t)	0.04
11	Mean highest temperature (pre 1–3)	0.04
12	Min lowest temperature (pre 1–3)	0.03
13	O3	0.03
14	Mean average temperature (pre 1–3)	0.02
15	Month	0.02
16	Max average temperature (pre 1–3)	0.02
17	Day of the week	0.01
18	Difference in temperature (t-1)	0.01
19	Electricity consumption (t-7)	0.01
20	Electricity consumption (t-3)	0.01

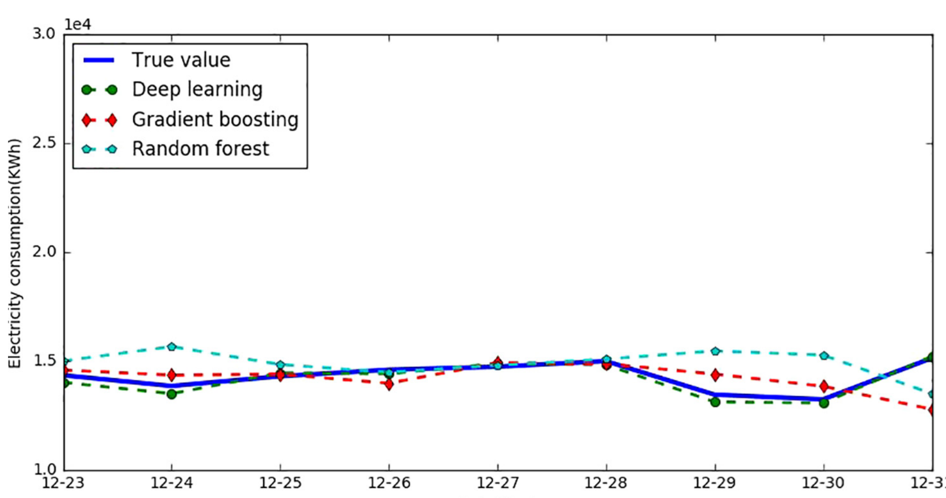

Fig. 12. Forecasting values of different methods for daily electricity consumption data.

Table 11
Top 20 most important variables for Lianyungang derived from the gradient boosting machine.

Rank	Feature	Importance
1	Electricity consumption (t-1)	1.0000
2	Average temperature (t)	0.1276
3	Highest temperature (t)	0.0400
4	Electricity consumption (t-2)	0.0132
5	Lowest temperature (t)	0.0130
6	Month	0.0124
7	Day of the week	0.0117
8	Electricity consumption (t-7)	0.0068
9	O3	0.0050
10	Max difference in temperature	0.0042
11	Min electricity consumption (pre 1–7)	0.0035
12	Average temperature (t-3)	0.0026
13	Highest temperature (t-1)	0.0026
14	Difference in temperature (t)	0.0024
15	Min average temperature	0.0023
16	Mean average temperature (pre 1–3)	0.0022
17	Max electricity consumption (pre 1–7)	0.0016
18	CO	0.0014
19	Average temperature (t-1)	0.0010
20	PM10	0.0007

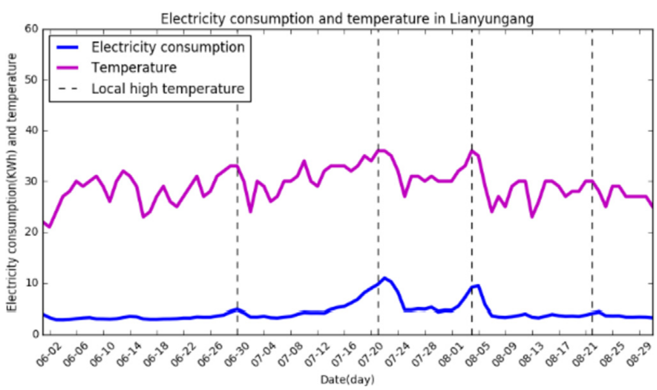

Fig. 13. Relationship between high local temperatures and electricity consumption.

A heatmap of weekly electricity consumption is shown in Fig. 14. In July and August, the most electricity is consumed on a single day and not on weekends. However, the highest electricity consumption levels appear on the weekend in other months, reflecting the influence of high temperatures on people's lifestyle patterns.

Monthly electricity consumption levels are shown in Fig. 15. We find that electricity consumption levels are significantly higher during December, January, February, July and August. Various factors can lead to an increase in electricity consumption. However, temperature changes and the occurrence of festivals always lead to an increase in electricity consumption. For instance, low temperatures combined with the occurrence of the Spring Festival in February, China's most magnificent traditional festival, lead to an increase in electricity consumption in this month. The months of July and August are the hottest of the year. In high temperature periods, residents' lifestyle patterns change significantly, leading to an increase in electricity consumption.

4.4.3. Probability density prediction

Probability density forecasts for Lianyungang for December 22, 2014, to December 30, 2014, based on deep learning are shown in Fig. 16.

In Fig. 16, all actual values are distributed at the center of the probability density curve except for values for the test samples. Actual values are positioned close to the maximum probability point in the first, third, fifth and sixth subplots. Fig. 16 shows that both the median and mode of the probability density curve for daily electricity consumption lie within the predicted lower and upper bounds in most cases.

4.5. Suzhou case study

4.5.1. Point prediction

Electricity consumption in Lianyungang for January 1, 2014, to December 31, 2014, is illustrated in Fig. 17. We find a pronounced decline in consumption before the occurrence of the Spring Festival, reflecting the main difference observed between Suzhou and the other two cities.

MAPE, MRPE and MAE results for RF, GBM and MLP are shown in Table 12. Based on Table 12, DL MAPE, MRPE and MAE values are measured as 3%, 6% and 594, respectively. This clearly shows that MLP forecasting errors are superior to those of the other models. A comparison of predictions made from the proposed method and from existing methods is presented in Fig. 18.

4.5.2. Variable importance

According to Table 13, the month of the year and day of the week are the most important variables used in the deep learning model. Electricity consumption for the day before, temperature-related factors and air-quality-related factors are less important. However, according to Tables 14 and 15, electricity consumption for the day before and temperature-related factors are the most important variables used by the random forest and gradient boosting machine

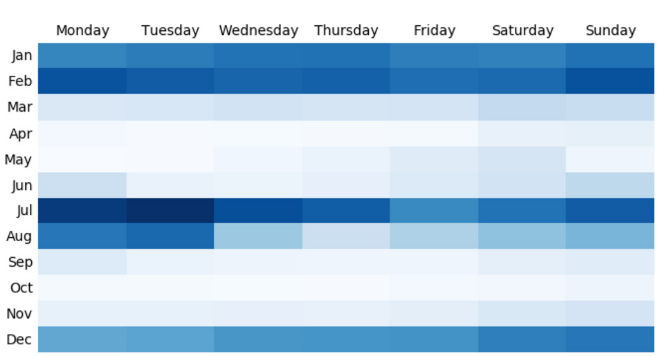

Fig. 14. Heatmap of weekly electricity consumption.

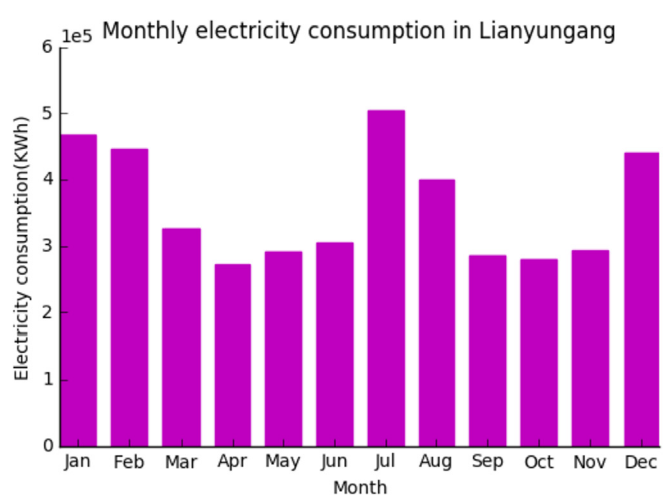

Fig. 15. Monthly electricity consumption in Lianyungang for 2014.

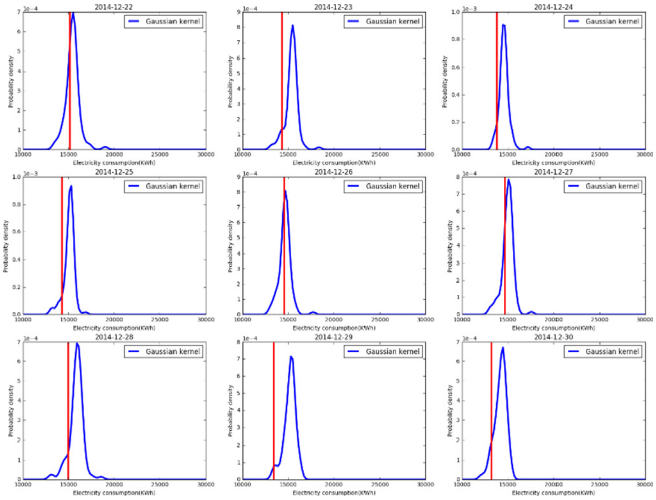

Fig. 16. Diagram of the probability density functions for Lianyungang for December 22, 2014, to December 30, 2014.

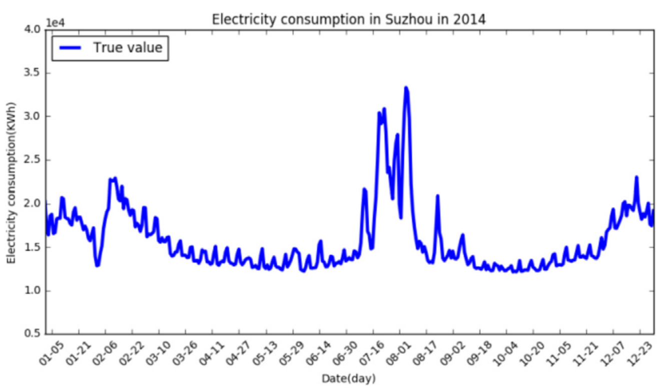

Fig. 17. Electricity consumption in Suzhou from January 1, 2014, to December 31, 2014.

Table 12
Forecasting e,s of three algorithms measuring electricity consumption in Suzhou.

Models	MAPE (%)	MRPE (%)	MAE (KWh)
RF	0.038666	0.110005	726.6495
GBM	0.043417	0.093651	823.9742
MLP	0.03203	0.067121	594.8447

models. There is no need to determine which factors are the most important because the mechanisms of each machine model differ. These factors influence electricity consumption. We now explore the relationship between these factors and electricity demand via data visualization.

The relationship between local high temperatures and electricity consumption is shown in Fig. 19. We find that in Suzhou from June to August, high local temperatures are always accompanied by high levels of local electricity consumption (see the vertical dashed line in Fig. 19). Temperature has an important influence on residential electricity consumption. The months of June, July and August are considered summer season months according to the Chinese lunar calendar. The onset of high temperatures limits resident engagement in outdoor activities while increasing engagement in indoor activities, leading to an increase in indoor electricity consumption. The onset of high temperatures also causes many families to purchase air conditioning appliances, further increasing electricity consumption.

From Fig. 20, we find that electricity consumption is highest on weekends and in July and August. In July and August, the most electricity is consumed on a single day and not on weekends, reflecting the influence of high temperatures on lifestyle patterns.

Fig. 21 shows that electricity consumption levels are significantly higher in December, January, February, July and August. Factors that lead to an increase in electricity consumption vary. However, temperature changes and the occurrence of festivals always lead to an increase in electricity consumption. For instance, low temperatures combined with the occurrence of the Spring Festival in February, China's most magnificent traditional festival, lead to an increase in electricity consumption in this month. The months of July and August are the hottest of the year. In high temperature periods, residents' lifestyle patterns change significantly, leading to an increase in electricity consumption.

4.5.3. Probability density prediction

Probability density forecasting for Suzhou for December 22, 2014, to December 30, 2014, based on a deep learning model is shown in Fig. 22.

In Fig. 22, actual values are distributed in the center of the probability density curve but not for December 29. Actual values reflect maximum probability points, especially for the first, fourth, fifth and seventh subplots. Fig. 22 shows that both the median and mode of the probability density curve for daily electricity consumption lie within the predicted lower and upper bounds in most cases.

5. Conclusions

Accurate electricity consumption forecasts can offer critical guidance to power systems and energy suppliers. Over the last two years, with the development of big data cloud computing technologies, artificial intelligence (especially in the field of deep learning) has been widely applied to all areas of life. Based on traditional power load forecasting methods, we propose a deep learning-based approach to power load forecasting. We first identify several factors that may affect electricity consumption levels such as weather-related, air-quality-related, economic, date-related and electricity consumption factors. Second, three popular machine learning models (deep learning, random forest, gradient boosting machine models) are used for electricity consumption forecasting and feature selection, and the results are compared. Furthermore, we explore electricity consumption patterns in relation to temperatures, months of the year and days of the week. Finally, the probability density of electricity consumption is predicted via deep neural network quantile regression. To verify the robustness of the method proposed in this paper, three case studies involving the same parameter settings are presented.

Our experimental results demonstrate that (1) the deep leaning method performs better than random forest and gradient boosting machine methods in terms of prediction errors. (2) Regarding feature selection concerns, the most important variables are monthly, weekly and temperature-related variables according to the deep learning model. However, the random forest and gradient boosting machine models focus more heavily on pre-related electricity consumption levels. Our visual charts show that electricity consumption patterns can be observed in terms of monthly, weekly and temperature-related patterns. Electricity consumption levels increase significantly in June and February and over 7–8 months. Household electricity consumption levels also fluctuate

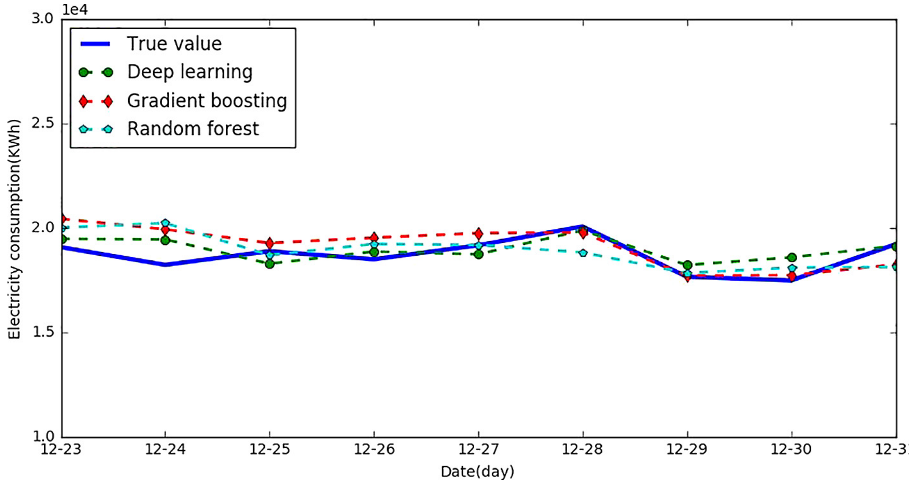

Fig. 18. Forecasting values derived from different methods using daily electricity consumption data.

Table 13
Top 20 most important variables for Suzhou derived from the deep learning model.

Rank	Feature	Importance	Rank	Feature	Importance
1	Sunday	1.00	11	March	0.56
2	June	0.81	12	Friday	0.53
3	August	0.78	13	September	0.48
4	July	0.78	14	Tuesday	0.43
5	May	0.75	15	February	0.39
6	Saturday	0.66	16	January	0.37
7	Monday	0.65	17	April	0.37
8	December	0.64	18	Highest temperature (t)	0.35
9	Wednesday	0.61	19	Rain (0)	0.35
10	Thursday	0.57	20	Rain (1)	0.31

Table 14
Top 20 most important variables for Suzhou derived from the random forest model.

Rank	Feature	Importance
1	Electricity consumption (t-1)	1.0000
2	Average temperature (t)	0.2907
3	Mean electricity consumption (pre 1–7)	0.2590
4	Average temperature (t-1)	0.1789
5	Max electricity consumption (pre 1–7)	0.1378
6	Electricity consumption (t-2)	0.1360
7	Mean average temperature (pre 1–3)	0.0699
8	Electricity consumption (t-3)	0.0551
9	Lowest temperature (t-1)	0.0371
10	Min average temperature	0.0229
11	Day of the week	0.0160
12	Highest temperature (t)	0.0157
13	Lowest temperature (t)	0.0154
14	Month	0.0150
15	Average temperature (t-3)	0.0141
16	Difference in temperature (t)	0.0130
17	Min electricity consumption (pre 1–7)	0.0095
18	Max highest temperature (pre 1–3)	0.0075
19	Max average temperature (pre 1–3)	0.0074
20	Mean average temperature (t-2)	0.0074

Table 15
Top 20 most important variables for Suzhou derived from the gradient boosting machine.

Rank	Feature	Importance
1	Electricity consumption (t-1)	1.0000
2	Highest temperature (t)	0.1411
3	Average temperature (t-1)	0.0668
4	Average temperature (t)	0.0487
5	Lowest temperature (t-1)	0.0270
6	Day of the week	0.0173
7	Mean highest temperature (pre 1–3)	0.0122
8	Month	0.0072
9	Max electricity consumption (pre 1–7)	0.0062
10	Electricity consumption (t-7)	0.0055
11	Mean electricity consumption (pre 1–7)	0.0047
12	Electricity consumption (t-4)	0.0047
13	Highest temperature (t-1)	0.0045
14	Lowest temperature (t)	0.0031
15	Min difference in temperature (pre 1–3)	0.0022
16	Max average temperature (pre 1–3)	0.0021
17	Electricity consumption (t-5)	0.0014
18	CO	0.0013
19	Min electricity consumption (pre 1–7)	0.0012
20	Max highest temperature (pre 1–3)	0.0012

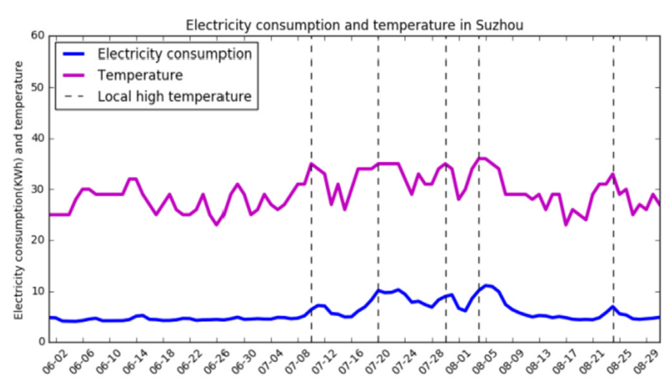

Fig. 19. Relationship between high local temperatures and electricity consumption.

considerably throughout the week, with weekend levels being higher than workday levels. Consumption levels are also higher in July and August, as high summer temperatures have the most significant effects on electricity consumption. (3) In terms of probability density predictions, the combination of deep neural network, quantile regression and kernel density estimation models is proposed. Our results show that most actual values are distributed in the center of the probability density curve. We thus recommend that electricity consumption forecasting methods be improved through feature engineering and deep learning in the current era of big data.

In the future, we will continue to investigate deep learning-

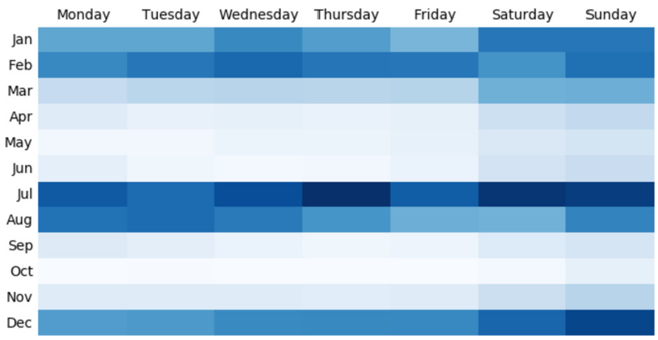

Fig. 20. Heatmap for weekly electricity consumption.

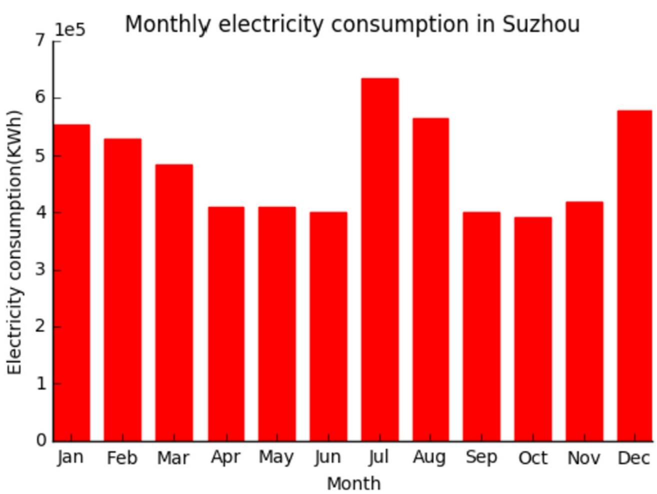

Fig. 21. Monthly electricity consumption in Suzhou for 2014.

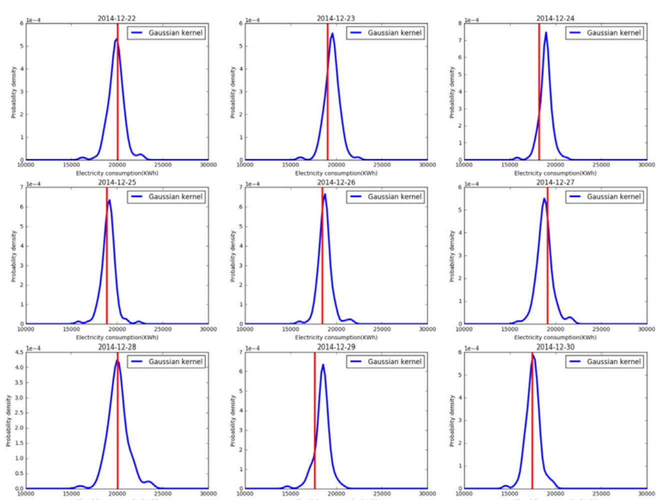

Fig. 22. Probability density function for Suzhou for December 22, 2014, to December 30, 2014.

based forecasting tools such as convolutional neural and recurrent neural networks, which are suited to larger and more complex datasets. More significant influencing factors (e.g., locations) will also be used to improve the accuracy of forecasting methods.

Acknowledgments

The present study was supported by the National Natural Science Foundation of China under grant nos. 71501056, 71521001 and 71690235; the Hong Kong Scholars Program under grant no. 2017-167; the Anhui Science and Technology Major Project under grant no. 17030901024; the China Postdoctoral Science Foundation under grant no. 2017M612072; grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 11271716 and CityU 21209715); and the Fundamental Research Funds for the Central Universities under grant no. JZ2018HGPA0271.

References

[1] Mocanu E, Nguyen PH, Gibescu M, Kling WL. Deep learning for estimating building energy consumption. Sustain Energy, Grids Netw 2016;6:91–9.
[2] Guo Z, Zhou K, Zhang C, Lu X, Chen W, Yang S. Residential electricity consumption behavior: influencing factors, related theories and intervention strategies. Renew Sustain Energy Rev 2018;81:399–412.
[3] Zhou K, Fu C, Yang S. Big data driven smart energy management: from big data to big insights. Renew Sustain Energy Rev 2016;56:215–25.
[4] Bianco V, Manca O, Nardini S. Electricity consumption forecasting in Italy using linear regression models. Energy 2009;34(9):1413–21.
[5] Cao G, Wu L. Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting. Energy 2016;115:734–45.
[6] Chen Y, Xu P, Chu Y, Li W, Wu Y, Ni L, et al. Short-term electrical load forecasting using the support vector regression (SVR) model to calculate the demand response baseline for office buildings. Appl Energy 2017;195:659–70.
[7] Feng Y, Ryan SM. Day-ahead hourly electricity load modeling by functional regression. Appl Energy 2016;170:455–65.
[8] Amber KP, Aslam MW, Hussain SK. Electricity consumption forecasting models for administration buildings of the UK higher education sector. Energy Build 2015;90:127–36.
[9] Do LPC, Lin K-H, Molnár P. Electricity consumption modelling: a case of Germany. Econ Modell 2016;55:92–101.
[10] Son H, Kim C. Short-term forecasting of electricity demand for the residential sector using weather and social variables. Resour Conserv Recycl 2017;123:200–7.
[11] El-Shazly A. Electricity demand analysis and forecasting: a panel cointegration approach. Energy Econ 2013;40:251–8.
[12] Yukseltan E, Yucekaya A, Bilge AH. Forecasting electricity demand for Turkey: modeling periodic variations and demand segregation. Appl Energy 2017;193:287–96.
[13] Al-Hamad MY, Qamber IS. ANFIS and multi linear regression to estimate the LTLF for the kingdom of bahrain. Int J Mech Mater Sci Eng 2016;5:248–61.
[14] Qamber IS. Peak load estimation studies in several countries, vol. 1. Engineering Press; 2017. 2.
[15] Ertugrul ÖF. Forecasting electricity load by a novel recurrent extreme learning machines approach. Int J Electr Power Energy Syst 2016;78:429–35.
[16] Jurado S, Nebot À, Mugica F, Avellana N. Hybrid methodologies for electricity load forecasting: entropy-based feature selection with machine learning and soft computing techniques. Energy 2015;86:276–91.
[17] Yang Z, Ce L, Lian L. Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. Appl Energy 2017;190:291–305.
[18] Yildiz B, Bilbao JI, Sproul AB. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renew Sustain Energy Rev 2017;73:1104–22.
[19] Kavaklioglu K. Modeling and prediction of Turkey's electricity consumption using support vector regression. Appl Energy 2011;88(1):368–75.
[20] Amina M, Kodogiannis VS, Petrounias I, Tomtsis D. A hybrid intelligent approach for the prediction of electricity consumption. Int J Electr Power Energy Syst 2012;43(1):99–108.
[21] Xiong T, Bao Y, Hu Z. Interval forecasting of electricity demand: a novel bivariate EMD-based support vector regression modeling framework. Int J Electr Power Energy Syst 2014;63:353–62.
[22] Ren Y, Suganthan PN, Srikanth N, Amaratunga G. Random vector functional link network for short-term electricity load demand forecasting. Inf Sci 2016;367–368:1078–93.
[23] Dudek G. Neural networks for pattern-based short-term load forecasting: a comparative study. Neurocomputing 2016;205:64–74.
[24] Wang C-h, Grozev G, Seo S. Decomposition and statistical analysis for regional electricity demand forecasting. Energy 2012;41(1):313–25.
[25] De Gooijer JG, Hyndman RJ. 25 years of time series forecasting. Int J Forecast 2006;22(3):443–73.
[26] He Y, Xu Q, Wan J, Yang S. Short-term power load probability density forecasting based on quantile regression neural network and triangle kernel function. Energy 2016;114:498–512.

[27] Li Z, Hurn AS, Clements AE. Forecasting quantiles of day-ahead electricity load. Energy Econ 2017;67:60–71.
[28] Sun S, Li G, Chen H, Guo Y, Wang J, Huang Q, et al. Optimization of support vector regression model based on outlier detection methods for predicting electricity consumption of a public building WSHP system. Energy Build 2017;151:35–44.
[29] Burger EM, Moura SJ. Gated ensemble learning method for demand-side electricity load forecasting. Energy Build 2015;109:23–34.
[30] Silva L. A feature engineering approach to wind power forecasting. Int J Forecast 2014;30(2):395–401.
[31] Hassan S, Khosravi A, Jaafar J. Examining performance of aggregation algorithms for neural network-based electricity demand forecasting. Int J Electr Power Energy Syst 2015;64:1098–105.
[32] De Felice M, Alessandri A, Ruti PM. Electricity demand forecasting over Italy: potential benefits using numerical weather prediction models. Elec Power Syst Res 2013;104:71–9.
[33] Ben Taieb S, Hyndman RJ. A gradient boosting approach to the Kaggle load forecasting competition. Int J Forecast 2014;30(2):382–94.
[34] Liu Y, Wang W, Ghadimi N. Electricity load forecasting by an improved forecast engine for building level consumers. Energy 2017;139:18–30.
[35] Tong C, Li J, Lang C, Kong F, Niu J, Rodrigues JJPC. An efficient deep model for day-ahead electricity load forecasting with stacked denoising auto-encoders. J Parallel Distr Comput 2018;117:267–73.
[36] Goodfellow I, Bengio Y, Courville A. Deep learning. The MIT Press; 2016.
[37] Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J. LSTM: a search space Odyssey. IEEE Trans Neural Netw Learn Syst 2017;28(10): 2222–32.
[38] Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. Neural Comput 2006;18(7):1527.
[39] Rumelhart D, Hinton G, Williams R. Learning representations by back-propagating errors. Nature 1986;323(6088):533–6.
[40] Koenker R. Quantile regression. New York: Cambridge University Press; 2005.
[41] Härdle W, Müller M, Sperlich S, Werwatz A. Nonparametric and semi-parametric models. Springer; 2004.
[42] Xu Q, Jiang C. Quantile partial adjustment model and its application. Quant Tech Econ 2011;28(8):115–33.
[43] Breiman L. Random foret. Mach Learn 2001;45:5–32.
[44] Friedman JH. Greedy Function Approximation:A gradient boosting machine. Ann Stat 2001;29(5):1189–232.