ORIGINAL ARTICLE

Load Forecasting Based on LSTM Neural Network and Applicable to Loads of "Replacement of Coal with Electricity"

Zexi Chen[1,2] · Delong Zhang[1] · Haoran Jiang[1] · Longze Wang[1] · Yongcong Chen[1] · Yang Xiao[2] · Jinxin Liu[1] ·
Yan Zhang[3,4] · Meicheng Li[1]

Abstract
With the complete implementation of the "Replacement of Coal with Electricity" policy, electric loads borne by urban power systems have achieved explosive growth. The traditional load forecasting method based on "similar days" only applies to the power systems with stable load levels and fails to show adequate accuracy. Therefore, a novel load forecasting approach based on long short-term memory (LSTM) was proposed in this paper. The structure of LSTM and the procedure are introduced firstly. The following factors have been fully considered in this model: time-series characteristics of electric loads; weather, temperature, and wind force. In addition, an experimental verification was performed for "Replacement of Coal with Electricity" data. The accuracy of load forecasting was elevated from 83.2 to 95%. The results indicate that the model promptly and accurately reveals the load capacity of grid power systems in the real application, which has proved instrumental to early warning and emergency management of power system faults.

Keywords Load forecasting · Long short-term memory · "Replacement of Coal with Electricity · Time series · Neural network

1 Introduction

Due to the impact of air pollution and energy shortages, Beijing, China has introduced a "Replacement of Coal with Electricity" policy that encourages household users to use electricity for heating instead of traditional coal-fired heating [1]. Some provinces north of Beijing have a large number of wind farms, which not only provides green power to Beijing but also reduces coal use and air pollution [2]. The major technology adopted in the policy of "Replacement of Coal with Electricity" is the air source heat pump [3], which has been utilized for heating in winter but appreciably increases the electric load at the same time.

These authors contributed equally to this work: Zexi Chen and Delong Zhang.

✉ Yan Zhang
   zhangyan8698@ncepu.edu.cn

   Zexi Chen
   chenzexi@bj.sgcc.com.cn

   Delong Zhang
   zhangdelong@ncepu.edu.cn

   Haoran Jiang
   1182211031@ncepu.edu.cn

   Longze Wang
   1182111018@ncepu.edu.cn

   Yongcong Chen
   120192211901@ncepu.edu.cn

   Yang Xiao
   xiaoyang@bj.sgcc.com.cn

   Jinxin Liu
   120192211823@ncepu.edu.cn

   Meicheng Li
   mcli@ncepu.edu.cn

1   School of New Energy, State Key Laboratory of Alternate Electrical Power System With Renewable Energy Sources, North China Electric Power University, Beijing 102206, China

2   State Grid Beijing Electric Power Company, Beijing 100031, China

3   School of Economics and Management, North China Electric Power University, Beijing 102206, China

4   Beijing Key Laboratory of New Energy and Low-Carbon Development, Beijing 102206, China

With the implementation of the policy, the fault rate has increased, especially in winter the demand for heating is very high. The faults are often severe due to incorrect load forecasting (inadequate estimation of impending accidents) [4, 5]. There are many load forecasting methods, including the linear fitting, regression analysis models, and various nonlinear models. The actual electric load is nonlinear and the load is usually affected by various factors, such as temperature and humidity [6]. Consequently, the forecasting accuracy of the conventional nonlinear forecasting model cannot meet the accuracy requirements of the modern power management system.

In the traditional unidirectional neural network (NN) model [7, 8], there is no connection between the neural units of the hidden layer, and each neural unit doesn't have a recurrent structure [9, 10]. In [11–13], the load forecasting method based on the artificial neural networks are studied and compared. The disadvantage of traditional NN model is that it is impossible to consider the past and future training data on the current output. Feedforward neural network (FNN) is similar to the structure of the traditional neural network model. The difference is that its hidden layer can be a multi-layer structure, but the neural unit has no recurrent structure, so it still cannot avoid the problem of long-term forgetting [14].

Therefore, a recurrent neural network (RNN) is proposed to solve this problem, which is characterized by adding a recurrent structure to the neural unit of the hidden layer. In [15], a path forecasting method is proposed and got good forecasting results. However, there is only one processing function in the neural unit of RNN, which will cause the problem of gradient disappearance or explosion after recurrent training, and will adversely affect the forecasting results [16].

Long short-term memory (LSTM) neural network [17] is a special RNN. The structure of its neural unit contains four processing sections. By processing the input data and the cyclic input data, the problem of gradient disappear or gradient explosion is avoided. LSTM has been used in some forecasting applications. In [18], a trading technology analysis method based on the LSTM is proposed to learn and forecast market behavior. In [19], a short-term load forecasting method is proposed, which obtain good forecasting results. At present, according to our investigation, there are few studies on the application of using LSTM to forecast load. After the introduction of the "Replacement of Coal with Electricity" policy, the residents' load in some areas of Beijing has changed significantly, and the accuracy requirement of load forecasting will be higher.

Therefore, in order to more accurately forecast the load and reduce the loss due to load overload, this paper proposes a method to predict the load using the LSTM neural network model. A load forecasting model based on LSTM is established first, and the structure of its hidden layer neural unit is introduced. The model fully considers the time series characteristics of the power load. Based on this model, the problem of forgetting long-term training data of traditional NN and FNN can be avoided. Finally, this paper experimentally verified the 2016–2017 data in Changping District, Beijing, and analyzed the impact of temperature and wind speed on the prediction results.

The main contributions of this paper are as follows. (1) A load forecasting method based on the LSTM model is proposed, which take many factors, such as temperature, wind force, into account and avoids the shortages of gradient disappearance or explosion. This model can reflect the load capacity of the power grid in a timely and accurate manner. In the actual application, the load forecast accuracy has been increased from 83.2 to 95%. (2) It is the first time that we consider the effect of "Replacement of Coal with Electricity" on the load forecasting. This research will mitigate the influence of "Replacement of Coal with Electricity" on the power system.

The rest of this paper is organized as follows. Section 2 is the load forecasting model based on LSTM. Section 3 is the model verification. Section 4 is the conclusion.

2 LSTM Model of Load Forecasting

LSTM neural network was proposed by Hochreiter & Schmidhuber (1997) and was improved by Graves. It is an optimized RNN neural network and has achieved success in many applications. Compared with conventional RNN, LSTM can learn long short-term information of time series and easily overcome gradient disappearance or explosion problem arising in RNN.

2.1 LSTM Model Unit

The greatest difference between the ordinary neural network and RNN is that each hidden unit of RNN is not independent. They are not only related to each other but also associated with the sequential input which follows the input of load data at the current moment into the unit of the hidden layer. This feature is highly instrumental in processing time series data.
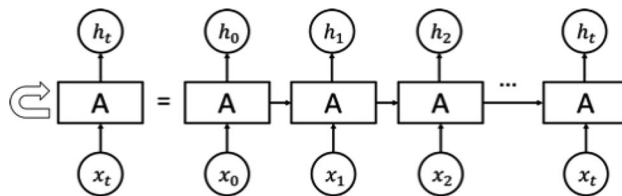

Fig. 1 RNN unfolded view

The unfolded view of a single RNN hidden unit is shown in Fig. 1. x_i is input the neural network module A and h_i is output. In this cycle, the information is passed from the current step to the next step, which means the same neural network is copied multiple times and each neural network module will pass the information to the next one.

Figure 1 shows the procedure repeated at every step of RNN, except for input. Such a training method considerably decreases the parameters that need to be learned in the network and greatly shortens training time while ensuring accuracy at the same time.

However, RNN has its disadvantages. For a standard RNN architecture, the historical data that can be used in practice is rather limited. Moreover, the influence of historical data in the distant future on the output either reduces overwhelmingly or explodes exponentially, which is commonly known as vanishing gradient or exploding gradient problem. LSTM, as a solution to the problem, is an improved RNN. There are four special structures in a single LSTM unit to describe the current state of the LSTM unit. Figure 2 presents three control gates: input, output and forget gates.

The outputs of the three gates relate to a multiplication unit to respectively control the input, output and status units of the network. When an input at the first moment is processed, the output of the network will be affected continuously by it, if the input gate is closed and the forget gate is open. The related formula is as follows:

$$a_c(t) = \sum_i^I x_i(t)w_{ic} + \sum_h^H b_h(t-1)w_{hc} \tag{1}$$

where, $\sum_i^I x_i(t)w_{ic}$ denotes the input of input gate at the moment t and $\sum_h^H b_h(t-1)w_{hc}$ is the input of forget gate at the moment $t-1$. Besides, another formula is described below:
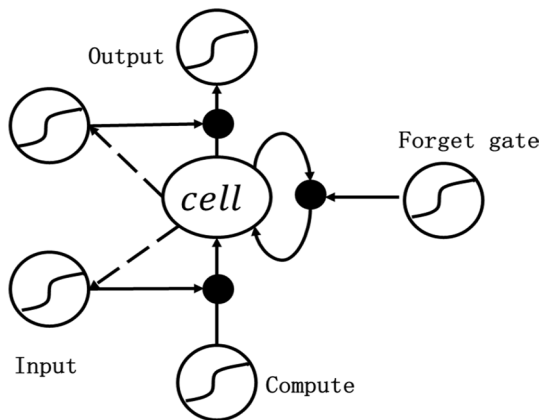
$$s_c(t) = b_\emptyset(t)s_c(t-1) + b_l(t)g\big(a_c(t)\big) \tag{2}$$

where, $b_l(t)g\big(a_c(t)\big)$ represents the product mapped by the forget gate $a_c(t)$ at the moment t. $b_\emptyset(t)s_c(t-1)$ refers to the product of forget gate at the moment t and the output of cell status at the moment $t-1$. $g(\cdot)$ is a mapping function. $s_c(t)$ denotes the output of cell status at the moment t. Fig. 3

In the LSTM model presented in this paper, the data needs to be pre-processed firstly. The input x_i of each neuron incorporates current status and status of $t-1$. In this way, the load on the day t can be predicted by the data of day $t-1$. The model is constructed as shown in Fig. 3.

2.2 Development of LSTM Load Forecasting Model

2.2.1 Data Processing

The data collected in this paper was derived from the electric load data with the introduction of "Replacement of Coal with Electricity" in a certain area of Beijing from 2016 to 2017. Table 1 shows the brief description of the data collected from 10 kV low-voltage power grid of various stations in February 2017, among which Changshuiyu is the most representative example with high fault rate. The load fluctuation was more violent when high loads were connected to the power system. In the experiment of this study, the maximum load of Changshuiyu was 215.338A in the examined month, which was 18.2% higher than the average. Thus, the data collected from Changshuiyu were used as the sample to train the LSTM model.

During the data acquisition process, some data with 0 load value were removed in this paper since the number of days differs in some months. For the remaining data, November 01 was set as the start date and February 28 of the following year as the end. All the data was summarized in the load distribution chart (Fig. 4) below, where the number of days is the abscissa and electric load value is the ordinate.
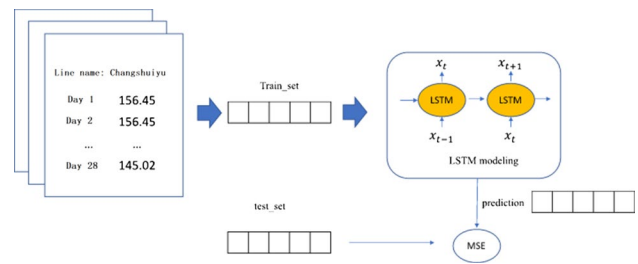

Fig. 2 LSTM compute nodes

Fig. 3 Model architecture

Table 1 Electric loads (A)

| Station name	Line name	Day 1	Day 2	Day 3	…	Day 28
Heying	Yingfang	155.22	155.92	155.22	…	132.54
Taowa	Taoliu	50.98	50.98	50.98	…	59.77
Taowa	Huata	192.49	192.49	192.49	…	139.75
Jundou	Zhenshun	199.69	200.92	202.68	…	348.05
Jundou	Shangsi	151.35	165.41	157.15	…	105.47
Taowa	Changshuiyu	156.45	156.45	156.45	…	145.02
Nankou	Xueshan	327.67	326.08	331.89	…	327.31
Xiangtun	Xinzhuang Village	134.30	131.84	136.76	…	115.67
Shangyuan	Xiying	98.44	93.17	93.17	…	107.23
Shangyuan	Laima Village	163.48	151.18	154.69	…	143.27
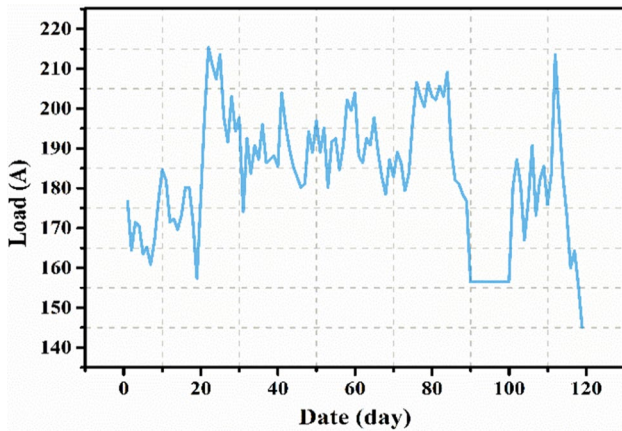
Fig. 4 Time distribution of electric load
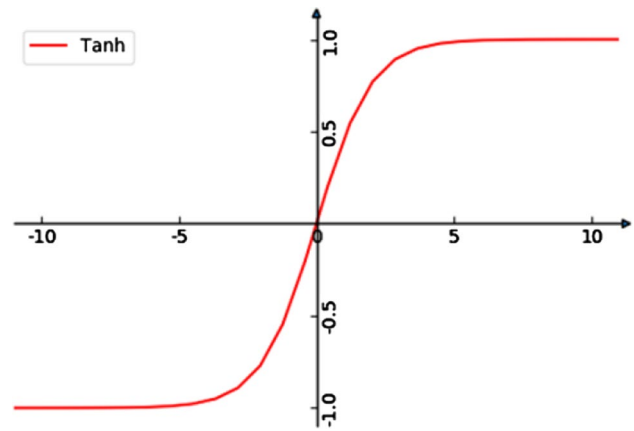
Fig. 5 Tanh function graph

Table 2 Single-step forecasting

Learning frequency (time)	var($t-1$)	var(t)
0	0	176.67
1	176.67	164.36
2	164.36	171.39
3	171.39	170.51
4	170.51	163.48
5	163.48	165.24

Table 3 Standardized dataset

Learning Frequency (time)	var($t-1$)	var(t)
0	−1	−0.3133
1	0.640847	−0.73136
2	0.526516	−0.49261
3	0.591808	−0.5225
4	0.583635	−0.76125
5	0.518343	−0.70148

2.2.2 Data Preparation

The existing data is sequential in chronological order. However, as a sample for supervised learning, the data is required to create the input sequence X and label Y for load forecasting. t denotes the present moment, $(t+1, \ t+n)$ refers to the future time, and $(t-1, \ t-n)$ means the past time. These are intended to predict Y at the time of t. A forecasting model was established in this paper, where the input data, including var($t-1$) and var(t), was used to predict the variable var($t+1$). In this case, the data was processed in Table 2.

The default activation function of LSTM is a hyperbolic tangent (tanh), in which the output values lie between -1 and 1, as shown in Fig. 5. Hence, all the data also needed to be standardized. To ensure fairness in the experiment, the scaling factor must be calculated from the training dataset and must be used to scale the test dataset and any forecasting dataset. This method was intended to avoid the adverse impact of the test dataset on the experiment fairness and to ensure reliable forecasting outcomes of the model. In this paper, datasets were converted and kept within [− 1, 1] using MinMaxScaler. The data obtained is as shown in Table 3.
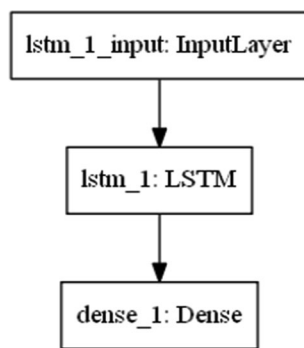
Fig. 6 Diagram of LSTM model architecture

2.3 Development of LSTM Neural Network Model

LSTM is a special type of RNN, which can learn and memorize longer sequences, and does not depend on the pre-specified observed value of window lag as the input. Keras [20] is a deep-learning modeling environment, with CNTK, Tensorflow [21] or Theano [22] as the backend in python. It has the following advantages compared with several common deep-learning frameworks, such as Tensorflow and Caffe:

(1) Keras is designed to support rapid modeling so that users can quickly map the architecture of the required model into the code. It can minimize coding workload, especially for well-established models, thereby speeding up the development process and directing more attention to the design of the model.

(2) Keras is highly modularized, with which users can combine modules randomly to construct a desired model. In Keras, any neural network can be described as a graph or sequence model, in which the components are divided into the following modules: neural network layer, loss function, activation function, initialization method, regularization method, and optimization engine. The user can select, in a reasonable manner, the network required by the module construction.

(3) Keras can switch seamlessly between CPU and GPU, suitable for different applications, especially GPU.

In this paper, LSTM was fast built using Keras. In default status, the LSTM network layer in Keras maintains states between two batches of data. This batch of data represented a fixed number of rows in the training dataset which defines running frequency before the weight of the network is updated. The default state of the LSTM layers between batches of data is reset, thus LSTM can't be presented without states. The reset time of the state for the LSTM layer can be controlled precisely by calling the function reset states. During the network compiling,
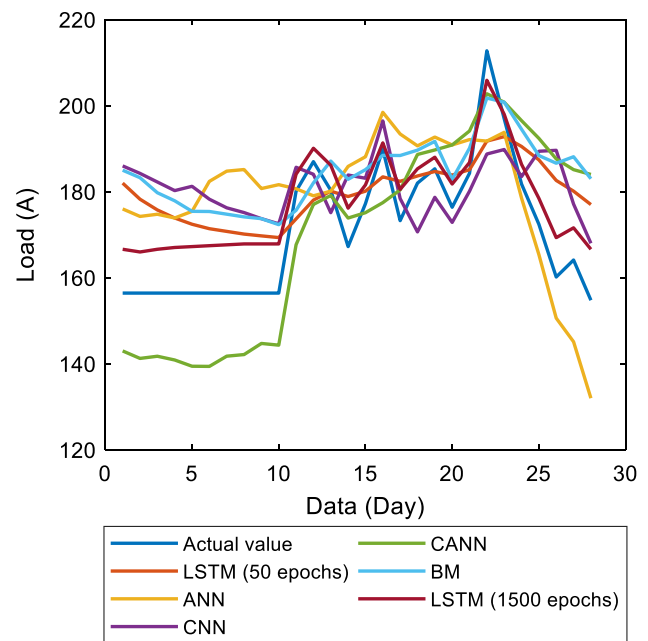

Fig. 7 Fitting effect comparison of different models

"mean-squared-error" served as the loss function because it was very close to the root mean squares error to be calculated. Dropout was used to reduce overfitting [23] and the efficient ADAM optimization algorithm [24] was also adopted. The network architecture was finally built as shown in Fig. 6.

3 Model Verification and Results Discussion

3.1 Data Introduction

In this paper, the first 90 load values of the Changshuiyu line were taken as the training data and the load values during 28 days in February as the test sample. The final evaluation indicator was the root mean square error [25] (RMSE).

$$MSE = \frac{1}{T_2} \sum_{t=T_1+1}^{T_1+T_2} e_t^2 \qquad (3)$$

$$RMSE = \sqrt{MSE} \qquad (4)$$

Compared with mean error, RMSE can detect more data patterns that, in addition to the linear trend, have not been described by the model, such as periodicity. For the load predicted herein, periodicity also exists. It will rise when the utilization rate of electric heating in winter increases with cold weather. The load will slowly return to the lower level when the coldest season ends.

3.2 Model Verification

3.2.1 Verification of LSTM Model

The trained LSTM model was tested using the test dataset. The fitting effect of the test data is shown in Fig. 7, where the number of days is the abscissa and load value is the ordinate, with actual load values in blue and predicated ones in orange.

As shown in Fig. 7, the actual data of the first few days in February was unchanged but not consistent with real values, which might be the result of problematic data processing during data acquisition. However, the trend of the electric load values during this period can still be obtained through forecasting of the training model. When the training number was 50 epochs, there is a difference existed between the predicted and actual values but the overall trend of rising and fall keep consistent. Thus, more complex models can be trained by adding more epochs to acquire more accurate forecasting data (LSTM of 1500 epochs in Fig. 7).

As shown in Fig. 7, the LSTM neural network model can be employed based on the historical training data of 90 days, that is 1500 epochs. Compared with 50 epochs, we can find that the forecasting accuracy of 28 days is improved by using more historical data to train the LSTM model. From the 11th to the 25th day, the predicted values are closed to the actual data. Actually, through analyzing the original data used to train, we find that the training data doesn't include a similar value of the 1st-10th day and the 26th-29th days. The difference between the predicted and actual value is caused by insufficient training data. The improvement in precision can be secured by increasing the complexity of the training model and history training data.

Besides, we also applied different models, such as artificial neural network (ANN), convolutional neural network (CNN), cascading neural networks (CANN) and Boltzmann machine (BM). For showing clearly in the figure, we simulate the 50 epochs for these models. It can be seen that, compared with the LSTM model, the predicted effects of other models are worse. We showed the predicted error of different models in the following section. It also verified that the LSTM model is better than others.

3.2.2 Comparison with Polynomial Models

The polynomial curve fitting method has been adopted in the traditional load forecasting models of which the mathematical model is as below:

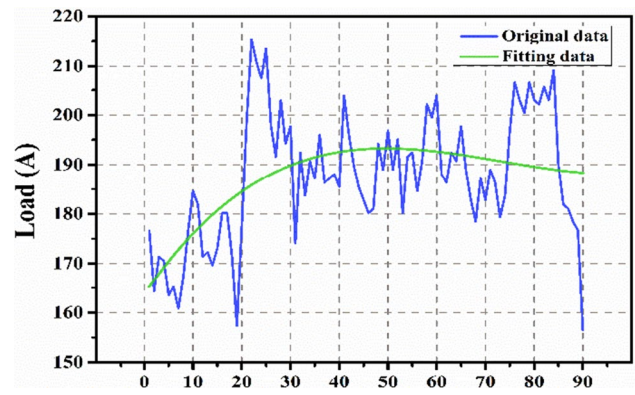$$f(x|p;n) = p_0 x^n + p_1 x^{n-1} + p_2 x^{n-2} + \cdots + p_{n-1} x + p_n \quad (5)$$

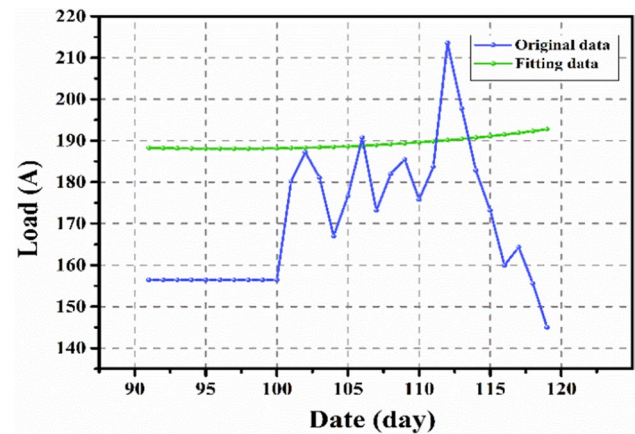

Fig. 8 Fit curves for historical dataset

Fig. 9 Forecasting results for the next 28 days

$$L(p;n) = \frac{1}{2} \sum_{i-1}^{m} \left[f(x|p;n) - y \right]^2 \quad (6)$$

where, $f(x|p;n)$ refers to the model. p and n are the parameters of the model. p is the coefficient of the polynomial f. n represents the degree of the polynomial. $L(p; n)$ is the loss function of the model. The common loss function is the square loss. Traditionally, the parameters p and n, which can minimize the loss function $L(p; n)$, are obtained by the fitting of historical data, and then the values to be predicted are gained through new input. In this study, the dataset adopted in the LSTM model is in the conventional model for direct comparison. The first 90 days is selected as the training set and the next 28 days as the test set. Figure 8 reveals three spline curves of the historical dataset and indicates that simple polynomial curve fitting is not applicable to the characteristics of load variation due to frequent changes of historical data.

Figure 9 demonstrates the comparison between the predicted and actual curves obtained by polynomial curve

fitting. It is not difficult to figure out that the prediction curve cannot correctly reflect the variation of the original load value with time. The RMSE value was 23.45, much higher than the LSTM model. The results of the comparison between LSTM, polynomial model, artificial neural network (ANN), convolutional neural network (CNN), cascading neural networks (CANN) and Boltzmann machine (BM) are shown in the Table 4. The abbreviations are mean absolute error(MAE) and mean absolute percentage error (MAPE) respectively.

Through the predicted error comparison of different models, it can be seen that the neural networks are better than the polynomial fitting model. In these models, the predicted effect of ANN is the worst, the LSTM model is the best. The CANN is the second. It is obviously that the result of the 1,500 epochs is better than the 50 epochs.

3.3 Comparison Among Multiple Factors

Apart from historical data, multiple factors can affect the electric loads, such as weather, temperature and wind force. Electric heating is an example, which is widely used in winter when more residents are facing lower temperature and tend to stay indoors, which may lead to a sudden elevation of

electrical load. In this paper, the data of meteorological factors (including weather, temperature, and wind force) from November 2016 to February 2017 in Changshuiyu area were collected to describe the effect of meteorological factors on electric load.

There are two types of variables in the original data, character and numerical types. The wind force was simplified to numerical values, such as "2, 2, 2, 1, 1…". The range of variables varied greatly, thus it was converted into the interval of 0–1 using the following normalization method for convenient description.

$$x_{\text{normalization}} = \frac{x - \text{Min}}{\text{Max} - \text{Min}} \qquad (7)$$

The Table 5 is the Meteorological Data of Changshuiyu area and Table 6 is the pre-processed data of Table 5.

3.3.1 Effect of Temperature on Electric Load

The temperature will directly affect the time that people spent outdoors. The intention to stay indoors is greater when the environment temperature is lower, which will increase the electric load. To clarify the relationship between temperature and load, the temperature data hereinafter is represented by "1-actual temperature". Figure 10 shows that the tendencies of daily average temperature and daily load are essentially the same, and almost synchronize at the turning points. Moreover, the RMSE (0.227) obtained by calculation indicates that the tendencies of average temperature and electric loads share considerable similarities. When the temperature was considered as maximum and minimum daily temperatures, the RMS values were 0.3350 and 0.3325 respectively. It is found that the minimum daily temperature can more adequately reflect the variation rules of load compared with the maximum. However, both variables are not as accurate as average temperature. The inaccuracy is
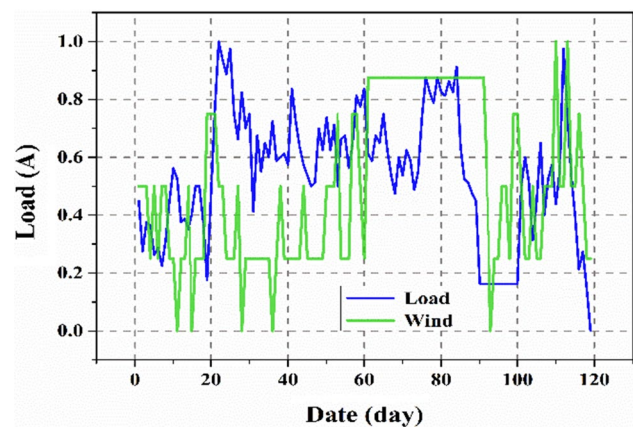
Table 4 Model comparison results

Model Name	RMSE	MAE	MAPE
LSTM (50 epochs)	14.00	11.97	0.0727
LSTM (1,500 epochs)	7.07	6.10	0.0610
Polynomial Fitting	23.45	18.22	0.1240
ANN	17.03	14.85	0.0902
CNN	16.96	14.45	0.0879
CANN	14.40	12.82	0.0773
BM	16.79	14.72	0.0900

Table 5 Meteorological data of changshuiyu area from November 2016 to February 2017 (only showing the data of the first 10 days)

Electric load(A)	Number of days	Date	Max. Daily temperature (℃)	Min. Daily temperature (℃)	Weather description	Wind direction	Wind force
176.67	1	01-NOV-2016	8	−2	Cloudy	Southwest	2
164.36	2	02-NOV-2016	12	1	Clear	Southwest	2
171.39	3	03-NOV-2016	13	2	Haze	South	2
170.51	4	04-NOV-2016	14	4	Haze	West	1
163.48	5	05-NOV-2016	18	2	Clear	South	2
165.24	6	06-NOV-2016	10	3	Cloudy	North	1
160.84	7	07-NOV-2016	11	2	Clear	Northwest	2
167.00	8	08-NOV-2016	11	−1	Clear	South	2
176.67	9	09-NOV-2016	8	2	Haze	Northeast	1
184.58	10	10-NOV-2016	10	1	Haze	South	1

Table 6 Pre-processed data

Electric load(A)	Number of days	date	Max. Daily temperature (℃)	Min. Daily temperature (℃)	Weather description	Wind direction	Wind force
0.450085	1	01-NOV-2016	0.65517241	0.63636364	Cloudy	Southwest	0.5
0.275028	2	02-NOV-2016	0.79310345	0.77272727	Clear	Southwest	0.5
0.375	3	03-NOV-2016	0.82758621	0.81818182	Haze	South	0.5
0.362486	4	04-NOV-2016	0.86206897	0.90909091	Haze	West	0.25
0.262514	5	05-NOV-2016	1	0.81818182	Clear	South	0.5
0.287543	6	06-NOV-2016	0.72413793	0.86363636	Cloudy	North	0.25
0.224972	7	07-NOV-2016	0.75862069	0.81818182	Clear	Northwest	0.5
0.312571	8	08-NOV-2016	0.75862069	0.68181818	Clear	South	0.5
0.450085	9	09-NOV-2016	0.65517241	0.81818182	Haze	Northeast	0.25
0.562571	10	10-NOV-2016	0.72413793	0.77272727	Haze	South	0.25
0.525028	11	01-NOV-2016	0.75862069	0.77272727	Cloudy	North	0
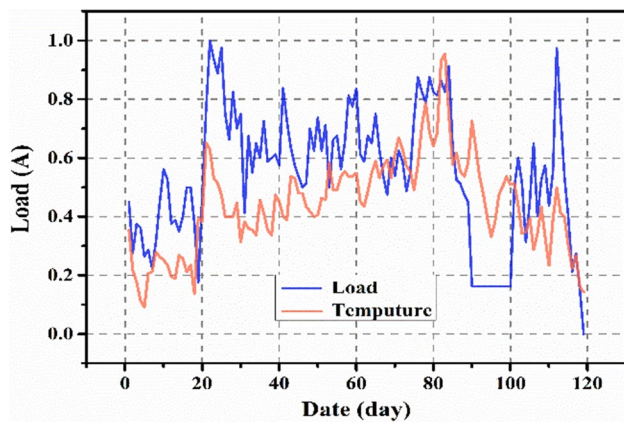
Fig. 10 Relationship between average temperature and load (x: time, y: temperature/load)

Fig. 11 Relationship between wind force and electric load (x: time, y: wind force/load)

caused by the minimum temperature, which only lasts for a short period, such as the early hours of a day, and minimally impacts the frequency of heater use throughout the day. Additionally, the average daily temperature indicates that the overall temperature of the day will have a greater influence on people's decision on travel.

3.3.2 Effect of Wind Force on Electric Load

Wind force herein refers to the average wind force in the area included in the weather data of the day. The presence of drastic variations revealed in Fig. 11 is caused by a few discrete values of wind force, so it is omitted in this study. As shown in Fig. 11, when the wind was relatively stronger around day 23 and day 80, the load also reached a peak. These noticeable time nodes suggest that wind force can provide certain guidance about load forecasting, but the effect is not as obvious as the prediction through average temperature.

3.3.3 Effect of Weather on Electric Load

Unlike temperature and wind force, the weather can't be denoted by numerical variables. There were totally 15 types of weather in Changshuiyu area during the investigated period. The influence of weather on electric load is less significant than that of wind force and temperature, which can be illustrated by snow and cloudy days. Due to travel inconvenience, people tend to stay indoors on snowy days when the temperature is lower compared to cloudy days. As shown in Fig. 12, the load is the abscissa, and frequency of certain weather is the ordinate. It demonstrates that high load occurred more frequently on snowy days. The difference between the two types of weather, however, is not rather evident. In fact, the temperature is generally lower throughout the winter, even on some cloudy or clear days despite colder weather on snow days. In these circumstances, it is difficult to forecast the electric load only with the weather.
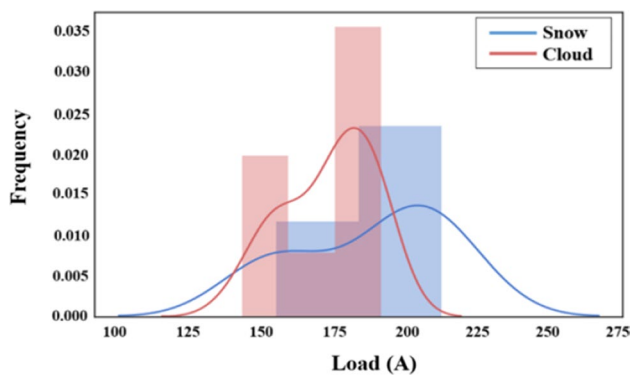
Fig. 12 Effect of weather on electric load (x: load, y: frequency)

4 Conclusion and Prospect

With the implementation of the "Replacement of Coal with Electricity" policy, the connection of more loads to the power system has resulted in increased fluctuation. The forecasting accuracy of the traditional linear fitting model is only 83.2%, far from meeting the actual needs. Against such backdrop, an LSTM model was established in this paper to simplify load forecasting into the prediction about time series data for one-dimensional load, which can be used to precisely predict the load variations in the scenario of considerable load fluctuations.

In the experiment of this study, the load data of Changshuiyu in Changping District of Beijing in 2016 and 2017 was illustrated as an example. The area features violent load fluctuation, with a high incidence of accidents. In this case, the RMSE of the LSTM model is 12.089, which is much lower than 23.45 obtained by the conventional polynomial curve fitting method. The forecasting accuracy has been improved from 83.2% (the precision of the traditional approach) to 95%. Meanwhile, compared with the regression analysis of temperature and weather, the LSTM model is more sensitive to load variation. Hence, it can efficiently assist in overload prevention, which is intensely practical in warning of faults.

Author Contributions Conceptualization, Z.C., D.Z.; methodology, Z.C., D.Z., Y.Z.; software, Z.C., D.Z., H.J.; validation, Z.C., D.Z., H.J., L.W.; formal analysis, Z.C., D.Z., H.J., L.W.; investigation, Y.C., Y.X., J.L.; resources, Y.C., Y.X., J.L.; data curation, H.J.; writing—original draft preparation, Z.C., D.Z., H.J.; writing—review and editing, Z.C., D.Z., H.J., L.W., Y.C., Y.X., J.L., Y.Z. and M.L.; visualization, H.J., L.W.; supervision, Y.Z., M.L.; project administration, Y.Z., M.L.; funding acquisition, Y.Z., M.L. All authors have read and agreed to the published version of the manuscript.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Jia DX, Shan BG (2015) Research on the Replacement of Coal by Electricity in East China. International Conference on Computer Information Systems and Industrial Applications.

2. Wang Q, Luo K, Wu C, Fan J (2019) Impact of substantial wind farms on the local and regional atmospheric boundary layer: case study of zhangbei wind power base in china. Energy 183:1136–1149

3. Lohani SP, Schmidt D (2010) Comparison of energy and exergy analysis of fossil plant, ground and air source heat pump building heating system. Renew Energy 35(6):1275–1282

4. Michishita K, Yokoyama S (2020) Lightning Fault Rate of Power Distribution Line in Wind Farm in Winter Lightning Area. Proceedings of the 21st International Symposium on High Voltage Engineering.

5. Tian W, Lei C, Zhang Y, Li D, & Winter R (2016) Data Analysis and Optimal Specification of Fuse Model for Fault Study in Power Systems. 2016 IEEE Power and Energy Society General Meeting (PESGM). IEEE.

6. Liao NH, Hu ZH, Ma YY, Lu WY (2011) Review of the short-term load forecasting methods of electric power system. Power Syst Prot Control 39(1):147–152

7. Maleki A, Nasseri S, Aminabad MS, Hadi M (2018) Comparison of arima and nnar models for forecasting water treatment plant's influent characteristics. KSCE J Civ Eng 22(9):1–13

8. Taieb SB, Bontempi G, Atiya AF, Sorjamaa A (2011) A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition. Expert Syst Appl 39(7):7067–7083

9. Chenhui W, Xiaoliang Z, Xiaochuan L (2016) Research on load forecasting strategy based on bp neural network under cloud computing architectures. Electric Power Inf Commun Technol 14(11):46–50

10. Xiangqian Y, Minhui L, Linjie R, Ying L (2014) Short-term load forecast based on improved Elman neural network. Electric Power Inf Commun Technol 12(2):39–42

11. Pappas SS, Ekonomou L, Moussas VC, Karampelas P, Katsikas SK (2008) Adaptive load forecasting of the Hellenic electric grid. J Zhejiang Univ Sci A 9(12):1724–1730

12. Ekonomou L, Oikonomou DS (2008) Application and comparison of several artificial neural networks for forecasting the Hellenic daily electricity demand load. 7th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED' 08). World Scientific and Engineering Academy and Society (WSEAS).

13. Ekonomou L, Christodoulou CA, Mladenov V (2016) A short-term load forecasting method using artificial neural networks and wavelet analysis. Int J Power Syst 1:64–68

14. Hu, Y, Ji H, & Song X (2009) To Forecast Short-Term Load in Electric Power System Based on FNN. FSKD 2009, Sixth International Conference on Fuzzy Systems and Knowledge Discovery, Tianjin, China, 14–16 August 2009, 6 Volumes. IEEE.

15. Min K, Kim D, Park J, Huh K (2019) Rnn-based path prediction of obstacle vehicles with deep ensemble. IEEE Trans Veh Technol 68:10252–10256

16. Son N, Yang S, & Na J (2019) Hybrid forecasting model for short-term wind power prediction using modified long short-term memory. Energies, 12.

17. Yongsheng D, Fengshun J, Jie Z, Zhikeng L (2020) A short-term power output forecasting model based on correlation analysis and elm-lstm for distributed pv system. J Electr Comput Eng 2020:1–10

18. Sang C, Pierro MD (2019) Improving trading technical analysis with tensorflow long short-term memory (lstm) neural network. J Financ Data Sci 5(1):1–11

19. Zhuo C, Long-Xiang S (2018) Short-term electrical load forecasting based on deep learning lstm networks. Electron Technol 47(01):39–41

20. Chollet F. keras-team/keras. GitHub; (2015). https://github.com/fchollet/keras.

21. Abadi M et al. TensorFlow: Large-Scale Machine Learning on Hetero-geneous Distributed Systems. arXiv.org; (2016) https://arxiv.org/abs/1603.04467.

22. The Theano development team. Theano: A Python framework for fast computation of mathematical expressions. arXiv.org; (2016). https://arxiv.org/abs/1605.02688.

23. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958

24. Kingma DP, Ba J. Adam: A method for stochastic optimization. 3rd International Conference for Learning Representations 2015.

25. Duchi J, Hazan E, Singer Y (2010) Adaptive sub-gradient methods for online learning and stochastic optimization. J Mach Learn Res 12(7):257–269

Zexi Chen is an engineer of State Grid Beijing Electric Power Company. He is currently pursuing the Ph.D. degree in electrical engineering with the School of New Energy, North China Electric Power University, Beijing, China. His research interest includes the forecasting of renewable energy power generation, the operation and planning of energy storage system.

Delong Zhang is a post-doctoral of the School of New Energy, North China Electric Power University, Beijing, China. His research interest includes the operation and planning of integrated energy systems, energy storage system and machine learning.

Haoran Jiang is currently pursuing the M.S. degree in electrical engineering with the School of New Energy, North China Electric Power University, Beijing, China. His research interest includes photovoltaic power generation and its application.

Longze Wang is currently pursuing the Ph.D. degree in electrical engineering with the School of New Energy, North China Electric Power University, Beijing, China. His research interest includes the integrated energy system and the blockchain.

Yongcong Chen is currently pursuing the M.S. degree in electrical engineering with the School of New Energy, North China Electric Power University, Beijing, China. His research interest includes the operation and planning of integrated energy system.

Yang Xiao is a deputy senior engineer of State Grid Beijing Electric Power Company. His research interest includes renewable energy power generation, the operation and planning of power system.

Jinxin Liu is currently pursuing the M.S. degree in electrical engineering with the School of New Energy, North China Electric Power University, Beijing, China. His research interest includes the operation of the integrated energy systems and the blockchain.

Yan Zhang is an associate professor of the School of Economics and Management, North China Electric Power University, Beijing, China. Her research interest includes the application of the blockchain in the energy system.

Meicheng Li is a professor of the School of New Energy, North China Electric Power University, Beijing, China. His research interest includes new energy generation, Li-ion batteries and photovoltaic materials, the operation of the integrated energy system and the blockchain.