ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20230708000790
Original scientific paper

Power Load Forecasting Based on LSTM Deep Learning Algorithm

Dalei WU*, Shuhua LIANG, Changji CHEN, Yupei CHEN, Pishi WANG, Zhiyuan LONG

Abstract: In recent years, the scale of China's power grid has been expanding, and the electricity consumption load has been rising year by year. Load forecasting plays a crucial role in ensuring the efficient coordination of power generation, transmission, and distribution in intelligent power systems. It holds immense significance in the planning, operation, control, and scheduling of new power systems. However, many current forecasting models do not take into account the temporal relationship of electricity consumption data, and therefore the models do not perform very well. In order to improve the accuracy of electricity consumption prediction, a Long Short Term Memory neural network model is proposed. We collect a series of electricity consumption data based on a fixed time interval, and use the daily collected data as a time series, and use Long Short Term Memory to build a simulated ensemble for the time series. Considering the impact of different acquisition intervals on the prediction results, we conducted experiments on load prediction with different sampling intervals. Our experiments were all performed on the data provided by the test questions of the 9th Electrician Mathematical Modeling Contest 2016.

Keywords: electricity consumption forecast; long Short Term Memory; time series data

1 INTRODUCTION

Due to the development of social economy and the increasing improvement of people's living standard, high-power electrical equipment is more and more widely used in daily life. In order to ensure the normal operation of the general public, factories, etc., power companies must ensure the stability of power supply and provide a higher quality of power supply. As a very important step in the power grid, power load forecasting is a prerequisite to ensure the stability and reliability of the entire regional electrical network. Therefore, power load forecasting is one of the foundations to ensure the normal supply of electricity and the importance of power load forecasting is becoming more and more prominent [1]. In the past, the traditional methods of electricity load forecasting were mainly based on time series analysis and statistical methods, such as multiple linear regression, exponential smoothing [4], Kalman filter [5], and time series method [6], which are small in computation, simple in structure, rapid in prediction, and very suitable for load series forecasting with strong time series but weak in volatility and randomness. These methods are good at predicting short-term power loads, but with the growing size of power systems and the diversification of power demand, these methods as well cannot meet the complex nonlinear problems and long-term dependent forecasting needs in today's power systems [7]. In order to solve these problems, some scholars have started to use machine learning methods for power load forecasting. Support vector machines (SVMs) can handle high-dimensional data well, and Wan et al [9] used SVMs to forecast short-term electricity load in a region with certain results; Fan et al [18] used Grey Model (GM) to fit the growth trend of electricity load, and then used Markov Chain (MC) to correct the forecast results; Feng [10] and Li [11] used Random Tree algorithm to forecast the electric load, but this method will have obvious overfitting defects on larger data. With the development of deep learning in recent years, it has achieved good results in various fields [2-7]. Some scholars have started to use deep learning methods to forecast electric loads, and in theory, neural network-based algorithms have strong nonlinear fitting ability and can fit arbitrary nonlinear functions, so simple network models based on back propagation [12, 13] have achieved more mature results in the field of short-term electric load forecasting. However, the previous methods usually just take the collected data as input without considering the time-series relationship between the data, and thus still do not fully exploit the connection between the data. Data that consider time-series relationships are called time-series data. Time series data is a series of observations obtained at a certain time interval. The first network used by scholars to solve time series data is Recurrent Neural Network (RNN). In an RNN, the current output of a sequence is related to the previous output as well. This takes the form of the network memorizing the previous information and applying it to the computation of the current output. The nodes between the hidden layers are no longer connectionless, and the inputs to the hidden layers include not only the output of the input layer but also the output of the hidden layer at the previous moment. Theoretically, RNNs are able to process sequence data of any length, but when two words are far apart, the gradient will disappear or explode after many stages of propagation. So traditional RNN networks suffer from long-term dependency problem. To better handle time series data, LSTM networks have been proposed. LSTM [8], a special type of RNN network, has been widely used in the field of time series forecasting because of its ability to effectively solve the long-term dependence problem. LSTM can not only cope with nonlinear problems and long-term dependence problems that cannot be solved by traditional algorithms, but also make full use of historical data in the power system to improve the accuracy of forecasting. Electricity load data collected by electric utilities at different time periods can well constitute a time series dataset, which is very suitable for the input requirements of LSTM. In this paper, we apply LSTM networks to electric load forecasting and propose an electric load forecasting model based on LSTM deep learning algorithm. To verify the effectiveness of the method, we experimentally validate the model on a public dataset.

2 LSTM

LSTM (Long Short Term Memory) is a special RNN network, which was proposed by Hochreiter and

Schmidhuber in 1997. Compared with the traditional RNN network, LSTM network can effectively solve the problem of long-term dependence. It can store and access long-term memory and filter unimportant information while maintaining long-term memory. LSTM network consists of three gated units: The output gate controls the updating of new input information, and its calculation formula is Eq. (1). The forgetting gate controls the forgetting of historical information, and its calculation formula is Eq. (2) and Eq. (3). The output gate controls the information output by the LSTM network, and its calculation formula is Eq. (5) and Eq. (6). The LSTM network structure is shown in Fig. 1.

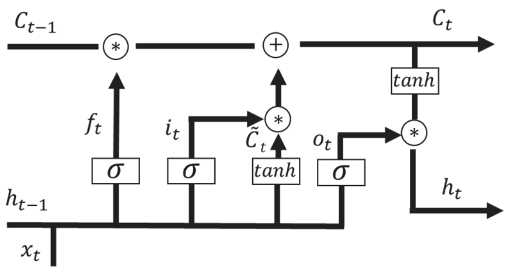

Figure 1 The structure of LSTM

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x\right] + b_f\right) \tag{1}$$

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{2}$$

$$\tilde{C}_t = tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \tag{3}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{4}$$

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{5}$$

$$h_t = o_t \cdot \tan h\left(C_t\right) \tag{6}$$

where, σ represents sigmoid function, tan$h$ represents tan$h$ activation function, W_f, W_i, W_c, W_o represent different weight matrices respectively. b_f, b_i, b_c, b_o represent different biases; means multiplying elements. Eq. (4) is used to update the output state, C_t, f_t, i_t, $\tilde{C}_t$, o_t are the intermediate variables. The process of LSTM state update and output information is as follows: (1) update the input gate according to the previous moment's input and output, and the forgetting gate controls the retention ratio of useful historical information; (2) update the state according to the current input, previous moment's output, and historical memory information; (3) output the current information under the control of the output gate.

3 ELECTRICITY CONSUMPTION FORECASTING PROCESS BASED ON LSTM MODEL

The LSTM model is used to forecast the power load, as shown in Fig. 2, which can be divided into the following three steps:

(1) Data preprocessing: fill missing values first, and then divide training set and test set.

(2) Model training: The data from the training set is fed into the model for training. The data sampled at different moments of each day are formed into a time-series sequence in chronological order and fed into the LSTM, which is used to model the data as a time series, and finally the predicted data are output through two fully connected layers. Computational losses are used for back propagation to update the model parameters.

(3) Result evaluation: The trained model is predicted on the test set, and the fit between the predicted and true values is measured using performance evaluation metrics.

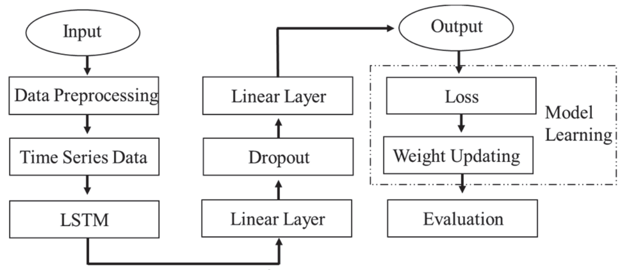

Figure 2 Using LSTM models to predict electrical loads

4 EXPERIMENTAL SETUP

4.1 Data Preprocessing

Electricity load data for Area 1 and Area 2 from January 1, 2009 to January 10, 2015 (one sampling point every 15 min, 96 points per day, in MW). For the dataset, we did the following:

(1) Missing value processing

In the data, missing values refer to data that is not present or unknown in some records or variables. To fill in the missing values, we use the mean substitution method to fill in the missing values data by replacing them with the mean.

(2) Feature scaling

A depth network is more sensitive to the input data. If there is a large gap in the scale between the input features, it can impact the network convergence, so we normalize the data by dividing all the features between 0 and 1.

(3) Splitting the dataset

In order to ensure that the data distribution of the validation set is as consistent as possible with that of the training set, we split the training and validation sets into intervals of 10, using 9 consecutive days for training data and 1 day for testing data to ensure that the data distribution of the validation set is consistent with that of the training set.

4.2 Model Parameter Settings

The experiment is implemented in Python, and the model framework is built based on PyTorch deep learning tool. The training process uses the ReLU activation function and mean square error as the loss function. The batch size is 128; the epoch is 30. The SGD optimizer with driving volume is selected. The learning rate and momentum parameters are set to 0.001 and 0.9, respectively. The hidden layer dimension of each LSTM and the dimension of LSTM are both 256. After the input data is processed by LSTM, we select the final output as the input of our subsequent fully connected network. Our fully connected network consists of two layers, each of which adopts relu activation function. The first layer of the network maps 256 dimensions to 256 dimensions, while the second layer of the network maps 256 dimensions to 1 dimension. To prevent overfitting, we use the Dropout method between the fully connected layers.

4.3 Metrics

In this paper, four of most common evaluation indexes are used to measure the strength of the model's regression ability: mean square error (*MSE*), root mean square error (*RMSE*), mean absolute error (*MAE*) and mean absolute percentage error (*MAPE*). Their formulas are given in Eq. (7) to Eq. (10).

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)^2 \tag{7}$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)^2} \tag{8}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|\hat{y}_i - y_i\right| \tag{9}$$

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{10}$$

where, y_i represents the actual electricity consumption, $\hat{y}_i$ denotes the predicted value of our model, and m represents the number of predicted samples.

5 EXPERIMENTAL RESULTS AND ANALYSIS

The experimental data was collected at a sampling frequency of 15 minutes, resulting in a total of 96 time points gathered within a single day. To further explore the influence of sampling frequency on the results, we subsequently changed the sampling frequency to 30 minutes and 45 minutes. Tab. 1 to Tab. 3 and Fig. 3 to Fig. 5 display the experimental results of our method and other methods under different sampling frequencies. Due to substantial differences in data, we performed log operations on all results. After this operation, the higher the value shown in the bar chart in Fig. 3 to Fig. 5, the better the effect.

Table 1 Performance of different models at a sampling frequency of 15 minutes

Model	*MSE*	*RMSE*	*MAE*	*MAPE*
SVM	2.04	1.03	0.11	0.08
Decision tree	3.68	1.85	2.52	1.06
Random Forest	3.81	1.92	2.52	1.14
Ours	3.92	0.96	2.63	1.79

Table 2 Performance of different models at a sampling frequency of 30 minutes

Model	*MSE*	*RMSE*	*MAE*	*MAPE*
SVM	2.04	1.00	0.10	–0.03
Decision tree	3.69	1.85	2.52	0.99
Random Forest	3.82	1.92	2.52	0.74
Ours	3.85	1.92	2.58	1.33

Table 3 Performance of different models at a sampling frequency of 45 minutes

Model	*MSE*	*RMSE*	*MAE*	*MAPE*
SVM	1.95	0.97	0.08	–0.56
Decision tree	3.25	1.61	2.30	–0.22
Random Forest	3.61	1.79	2.39	–0.07
Ours	3.84	1.80	2.47	0.37

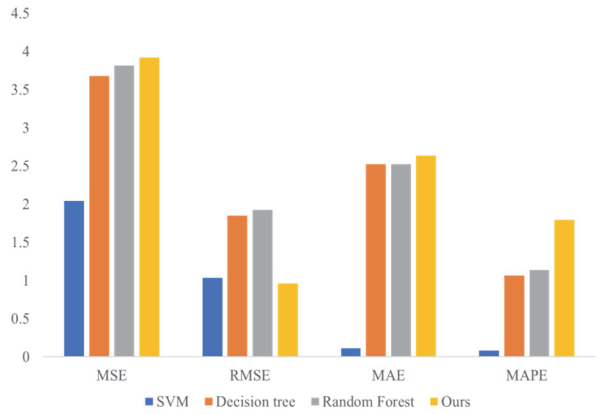

Figure 3 Performance of different models at a sampling frequency of 15 minutes

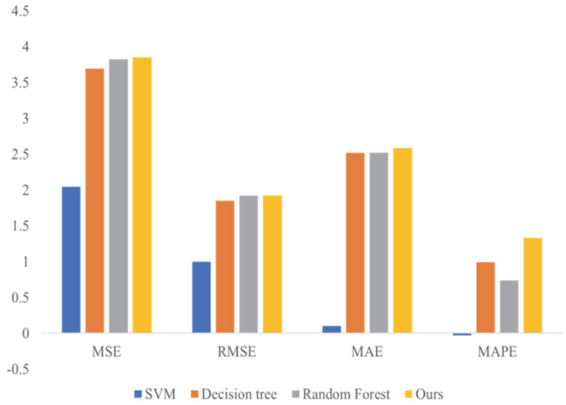

Figure 4 Performance of different models at a sampling frequency of 30 minutes

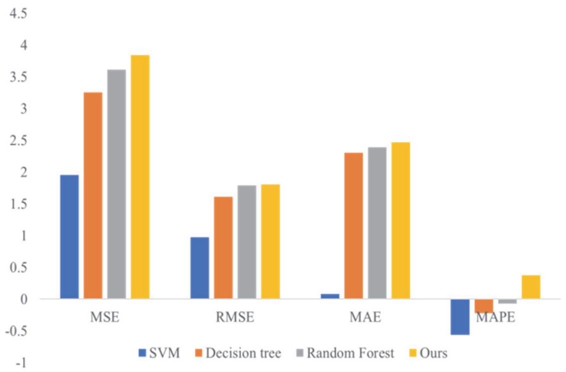

Figure 5 Performance of different models at a sampling frequency of 45 minutes

Upon examination of Tab. 1, we observe that the model's performance remains relatively consistent, regardless of whether or not the connections between data at different time points are considered when utilizing a larger dataset. This finding suggests that there is ample data available to generate a fitting curve that accurately captures the underlying patterns, even without explicitly considering the temporal dependencies between the data points. By examining Tab. 2 and Tab. 3, we can infer that our method, which incorporates time series data, exhibits enhanced stability as the sampling interval increases. This stands in contrast to other methods that do not leverage the temporal nature of the data. Based on our analysis, we are confident that residential electricity consumption does not undergo abrupt changes within short time intervals. This understanding forms the basis for utilizing time series data in curve fitting. By incorporating previous data, we introduce constraints that result in a smoother curve fitting process for subsequent data points. This approach takes advantage of the gradual and continuous nature of electricity consumption patterns in residential settings.

6 CONCLUSION

Load forecasting is a necessary prerequisite for the rational arrangement of power generation, transmission and distribution in intelligent power systems, and is of great significance for the scheduling of power systems, etc. Deep learning has achieved good results in various fields in recent years, so it is also widely used in power systems. In this paper, we propose a deep learning method based on LSTM to predict the electricity load. The method inputs the daily sampled data into the network as a time series in time order, uses the powerful parsing ability of LSTM for time series to analyze the data, and then we select the final output of the LSTM network as the feature input to the fully connected network for regression prediction. The experimental results show that our method is effective in MAE, MAPE, MSE and RMSE metrics. Nevertheless, further testing of our method across multiple regions and with more complex datasets is crucial. Additionally, future work will involve optimizing the model based on the outcomes of these tests. These efforts will contribute to enhancing the reliability and applicability of our approach, ultimately advancing the field of research in this area.

Acknowledgments

This work was supported by Science and Technology Project of China Southern Power Grid Limited Liability Company (070000KK52200021).

7 REFERENCES

[1] Nejat, P., Jomehzadeh, F., Taheri, M. M., Gohari, M., & Majid, M. Z. A. (2015). A global review of energy consumption, CO2 emissions and policy in the residential sector (with an overview of the top ten CO2 emitting countries). *Renewable and sustainable energy reviews, 43*, 843-862. https://doi.org/10.1016/j.rser.2014.11.066

[2] Yuan, L., Chen, Y., Wang, T., Yu, W., Shi, Y., Jiang, Z. H., Francis, E. H., Feng, J., & Yan, S. (2021). Tokens-to-token vit: Training vision transformers from scratch on imagenet. *Proceedings of the IEEE/CVF international conference on computer vision*, 558-567. https://doi.org/10.1109/ICCV48922.2021.00060

[3] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779-788. https://doi.org/10.1109/CVPR.2016.91

[4] Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., Torr, P. H. S., & Zhang, L. (2021). Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6881-6890. https://doi.org/10.1109/CVPR46437.2021.00681

[5] MacLeod, M., Woodfine, D. G., Mackay, D., McKone, T., Bennett, D., & Maddalena, R. (2001). BETR North America: A regionally segmented multimedia contaminant fate model for North America. *Environmental Science and Pollution Research, 8*, 156-163. https://doi.org/10.1007/BF02987379

[6] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting, 37*(4), 1748-1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

[7] Wang, L., Tong, Z., Ji, B., & Wu, G. (2021). TDN: Temporal difference networks for efficient action recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1895-1904. https://doi.org/10.1109/CVPR46437.2021.00193

[8] Hochreiter, S & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[9] Zeng, J. & Qiao, W. (2011, March). Support vector machine-based short-term wind power forecasting. *2011 IEEE/PES power systems conference and exposition*, 1-8. https://doi.org/10.1109/PSCE.2011.5772573

[10] Feng, Z., Wang, Y., Yuan, B., Feng X., Yao, Z., & Wu Z. (2021). Short-term power load forecasting based on random forest and improved local forecasting. *Water conservancy and hydropower technology, S2*, 300-305.

[11] Li, Y., Jia, Y. J., Li, L., Hao, J., & Zhang, X. (2020). Short-term electric load forecasting based on random forest algorithm. Power Syst. *Prot. Control, 48*, 117-124.

[12] Wang, S., Wang, J., Wang, Y., & Ma, W. (2019). Short-term power load forecasting by BP neural network based on improved Genetic algorithm. *Foreign electronic measurement technology,01*, 15-18.

[13] Aoyang, H., Shengqi, Z., Xuehui, J., & Zhisheng, Z. (2021). Short-term load forecasting model based on RBF neural network optimized by artificial bee colony algorithm. *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 486-489. https://doi.org/10.1109/ICBAIE52039.2021.9390043

[14] Qiu, X., Ren, Y., Suganthan, P. N., & Amaratunga, G. A. (2017). Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. *Applied soft computing, 54*, 246-255. https://doi.org/10.1016/j.asoc.2017.01.015

[15] Yan, Y. & Zhang, Z. (2021). Cooling, heating and electrical load forecasting method for integrated energy system based on SVR model. *2021 6th Asia Conference on Power and Electrical Engineering (ACPEE)*, 1753-1758. https://doi.org/10.1109/ACPEE51499.2021.9436990

[16] Deng, D., Li, J., Zhang, Z., Teng, Y., & Huang, Q. (2020). Short-term electric load forecasting based on EEMD-GRU-MLR. *Power System Technology, 44*(2), 593-602.

[17] Liu, Y. & Zhao, Q. (2021). Ultra-short-term power load forecasting based on cluster empirical mode decomposition of CNN-LSTM. *Power System Technology, 45*(11), 4444-4451.

[18] Fan, Y. N. (2012). Short-term power load forecasting based on Markov chain. *Journal of Qinghai University (Natural Science Edition), 2012*(3), 11-14.

Contact information:

Dalei WU, SN ENGR
(Corresponding author)
Electric Energy Metering Center,
Hainan Power Grid Co., Ltd., Hainan, China
E-mail: wdl198609@126.com

Shuhua LIANG, Engineer
Electric Energy Metering Center,
Hainan Power Grid Co., Ltd., Hainan, China
E-mail: 753378106@qq.com

Changji CHEN, Engineer
Wuzhishan Power Supply Bureau of Hainan,
Power Grid Co., Ltd., Hainan, China
E-mail: 285264@qq.com

Yupei CHEN, Engineer
Electric Energy Metering Center,
Hainan Power Grid Co., Ltd., Hainan, China
E-mail: 651917991@qq.com

Pishi WANG, Engineer
Electric Energy Metering Center,
Hainan Power Grid Co., Ltd., Hainan, China
E-mail: 444592607@qq.com

Zhiyuan LONG, Engineer
Electric Energy Metering Center,
Hainan Power Grid Co., Ltd., Hainan, China
E-mail:1219885823@qq.com