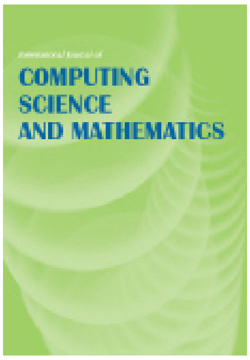


International Journal of Computing Science and Mathematics

ISSN online: 1752-5063 - ISSN print: 1752-5055
https://www.inderscience.com/ijcsm

Short-term power load prediction based on CNN-LSTM model

Jiawen Chen, Chao Cai, Fangbin Yan, Jinfeng Liu

DOI: 10.1504/IJCSM.2025.10070450

Article History:

Received:	11 March 2024
Last revised:	31 May 2024
Accepted:	04 June 2024
Published online:	06 May 2025

Short-term power load prediction based on CNN-LSTM model

Jiawen Chen, Chao Cai and Fangbin Yan

State Grid Hubei Electric Power Co., Ltd.,
Wuhan, Hubei Province, 430072, China
Email: jiawenchen_wh@163.com
Email: chaocai_wh@163.com
Email: fangbinyan_wh@163.com

Jinfeng Liu*

Department of Statistics and Mathematics,
Zhejiang Gongshang University,
Hangzhou, Zhejiang Province, 310018, China
Email: Jinfengliu28@163.com
*Corresponding author

Abstract: The load forecasting of power system is to forecast the load of the system in a future period of time, considering the influence of historical load, economic condition, meteorological condition and social events. Therefore, a CNN-LSTM model is proposed to predict short-term power load fluctuations in the next few days on the basis of the original, in which the convolutional layer and pooling layer in convolutional neural network (CNN) are used to extract features and reduce dimensions, and then the reconstructed data output by CNN is forecasted. The experimental results show that the prediction accuracy and error of CNN-LSTM model are obviously better than that of long short-term memory (LSTM) model, which also shows that CNN-LSTM model is suitable for short-term power load data prediction.

Keywords: short-term power load; CNN; convolutional neural network; LSTM; long short-term memory network; deep learning.

Reference to this paper should be made as follows: Chen, J., Cai, C., Yan, F., and Liu, J. (2025) 'Short-term power load prediction based on CNN-LSTM model', *Int. J. Computing Science and Mathematics*, Vol. 21, No. 1, pp.77–89.

Biographical notes: Jiawen Chen holds a Master's degree, is a Senior Engineer, and is a first class employee of State Grid Hubei Electric Power Co., Ltd. He has been engaged in professional management work such as power grid construction, operation and maintenance for a long time, proficient in modern smart distribution network construction, power grid digitisation and other new technologies.

Chao Cai holds a Doctoral degree, is a Senior Engineer, and is responsible for the operation and maintenance of the distribution network of State Grid Hubei Electric Power Co., Ltd. He has been engaged in the operation and maintenance of the distribution network, equipment technology management, and other work for a long time. His key research directions include the construction and application of modern smart distribution networks, power quality management.

Fangbin Yan holds a Master's degree, an Intermediate Engineer, and is a member of the Electric Power Science Research Institute of State Grid Hubei Electric Power Co., Ltd. He has been engaged in long-term work such as distribution network operation analysis and equipment maintenance management, with a focus on research areas such as distribution network digitisation and distribution automation.

Jinfeng Liu is currently studying applied statistics at the School of Statistics and Mathematics, Zhejiang Gongshang University. His current research direction is machine learning.

1 Introduction

Accurate short-term electricity load forecasting is of great significance for improving energy utilisation efficiency and reducing carbon emissions. Through effective load forecasting, the power industry can significantly improve economic benefits and make positive contributions to environmental protection (2022). As an important driving force for modern society and economic development, the effective utilisation of electricity is crucial. Power load forecasting not only helps to promote efficient utilisation of electricity, but also ensures the safe and stable operation of the power grid, formulates reasonable power grid operation methods, and maintains a balance between supply and demand of the power grid. Accurate load forecasting provides important data support and theoretical basis for power supply companies, enabling them to carry out effective electricity scheduling. Providing sufficient backup capacity in a timely manner in the event of a malfunction or power outage can help ensure the stability of power supply. In addition, load forecasting can clearly reflect the electricity consumption at different time periods, providing users with reasonable electricity planning references. It can be seen that accurately predicting short-term electricity load data is of great significance for the development of China's power system.

This paper introduces a short term load forecasting method based on CNN-LSTM combination model. This method makes full use of the advantages of CNN and long short-term memory (LSTM) models, and combines them to improve the prediction performance. In this model, CNN is responsible for extracting the feature factors of the input data, using its convolution layer and pooling layer to extract the feature and reduce the dimension of the data. Subsequently, the LSTM model receives the output data of CNN and makes use of the capability of the long and short term memory network to predict these data, which can fully tap their advantages in feature extraction and time series prediction, thus improving the accuracy and stability of the short-term power load prediction.

2 Background

2.1 Review of electric load forecasting

For past short-term electricity load forecasting, Bracale et al. (2019) adopted regression analysis. They established a multiple linear regression model by identifying the influencing factors of electricity load changes and constructing relationships between independent and dependent variables. However, it is evident that the influencing factors of electricity load are not linear. Ismail et al. (2008) utilised autoregressive time series models to predict daily peak electricity demand. They selected factors such as weather conditions, holidays, and seasons, and used regression analysis to forecast these indicators, thereby establishing predictions for future electricity peaks based on these forecasted indicators. To address this issue, machine learning-based models such as neural networks and decision trees have been proposed. Deep learning, a complex multi-layer neural network, aims to mimic the human brain's data processing capabilities. The core idea of deep learning is to automatically learn data features through multiple layers of neural network layers and use these features for prediction or classification tasks. Xiao et al. (2009) implemented a load forecasting model based on the backpropagation neural network model, which obtained expected values through forward and backward propagation through the network. Zhou et al. (2021) constructed an LSTM load forecasting model, leveraging LSTM's ability to learn long-distance temporal dependencies. This allows the model to identify load change patterns in the time dimension while recognising the nonlinear effects of factors such as weather and holidays in the factor dimension. This comprehensive analysis enhances the accuracy and comprehensiveness of the load forecasting model. Cui et al. (2020) based their work on the CNN-LSTM model to extract spatially coupled interaction features from data. They input reconstructed data into the LSTM network to mine load temporal features and utilised an adaptive moment estimation optimiser to train the model.

2.2 Research data

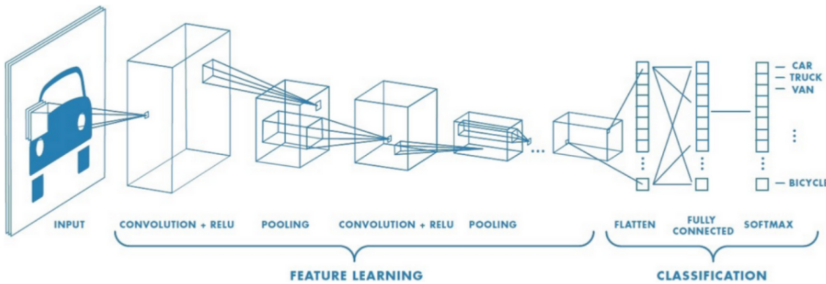
The data used in this study are from the electric power system load data and meteorological data from January 1, 2018, to August 31, 2021, in a certain region. The data are collected at a frequency of 15 minutes, where each data point represents the cumulative load over a 15-minute period. The dataset includes features such as weather conditions, maximum temperature, minimum temperature, daytime wind speed and direction, and nighttime wind speed and direction. The dataset is divided into training and testing sets in an 8 : 2 ratio.

3 Methods

3.1 Convolutional neural network model

The convolutional neural network (CNN) model is based on the neural perceptron and is inspired by the tissue processing of the biological visual cortex. The core components of CNN include input layer, convolution layer, pooling layer and full connection layer. The convolution layer mainly applies a series of filters to the input data to extract features, the pooling layer is used to reduce the computational complexity and the number of parameters, and the fully connected layer is used to achieve classification, detection and other tasks (Rafi et al., 2021). In general, CNNS gradually extract high-level features of images through multi-layer convolution and pooling operations, and perform classification or other tasks through the fully connected layer. Its excellent performance in the field of image processing is due to the efficient capture and learning of image local information and features. The CNN network structure is shown in the Figure 1.

Figure 1 CNN network structure (see online version for colours)



3.2 Long short-term memory model

Long short-term memory networks are a variant of recurrent neural networks commonly used to process sequential data. The core components of the LSTM model are: memory unit is responsible for storing and processing information in the sequence data, forgetting gate decides to retain and discard information, input gate decides whether to update the content in the memory unit, and output gate is responsible for adjusting the output information (Guo et al., 2020).

LSTM model realises the control and adjustment of information in memory unit through three gating mechanisms: forgetting gate, input gate and output gate. Compared with traditional recurrent neural network (RNN). LSTM has significant advantages in solving problems such as gradient disappearance and gradient explosion in the process of modelling long series data. The key elements of LSTM model include memory unit and gating mechanism. Together, they are responsible for handling long-term dependencies in sequence data, so they are widely used in natural language processing, time series prediction, speech recognition and other fields.

The LSTM network mainly includes three parts: the forgetting gate, the input gate and the output gate. The specific structure of LSTM is shown in Figure 2.

$$f_t = \sigma(W_{if} \cdot x_t + b_{if} + W_{hf} \cdot h_t - 1 + b_{hf}) \quad (1)$$

where σ is the sigmoid activation function.

$$i_t = \sigma(W_{ii} \cdot x_t + b_{ii} + W_{hi} \cdot h_t - 1 + b_{hi}) \quad (2)$$

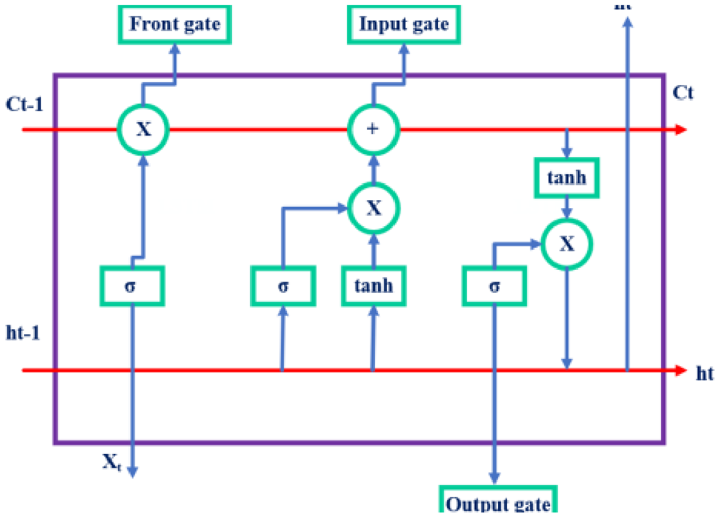
where σ is the sigmoid activation function.

$$o_t = \sigma(W_{ho} \cdot h_{t-1} + b_{ho}) \quad (3)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (4)$$

where σ is the sigmoid activation function.

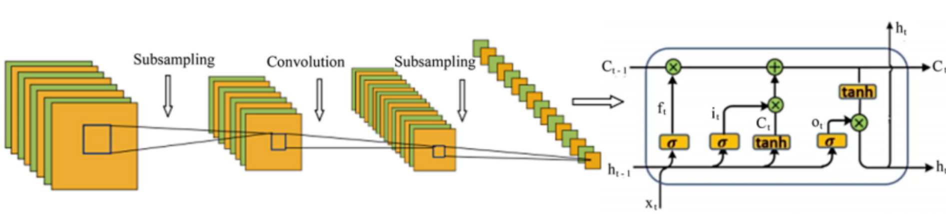
Figure 2 LSTM structure (see online version for colours)



3.3 CNN-LSTM model

The CNN-LSTM combined model integrates the advantages of CNNs and LSTM networks for processing time-series data or sequences with spatial structures. This combined model is designed to effectively capture spatiotemporal features in sequence data, exhibiting strong modelling and generalisation capabilities. The CNN component is responsible for extracting features from input data through convolution and pooling operations, effectively capturing spatial structures in the input data, which is crucial for handling features in time-series data (Rafi et al., 2021). The LSTM component is used to model the feature sequences extracted by CNN. LSTM, a variant of recurrent neural networks suitable for sequence data, captures long-term dependencies in sequence data through gate units and memory cells, thereby better understanding the temporal dependencies of the data. A feature fusion layer or operation is typically added to the CNN-LSTM model to effectively integrate the spatial features extracted by CNN and the temporal features captured by LSTM. The structure of the CNN-LSTM network is illustrated in Figure 3.

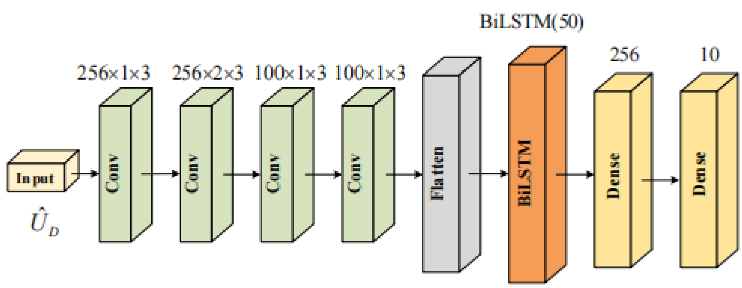
Figure 3 CNN-LSTM network structure (see online version for colours)



3.4 CNN-BiLSTM model

The CNN-BiLSTM model combines two neural network structures: CNN and bidirectional long short-term memory network, which mainly includes word embedding layer, BiLSTM layer and CNN layer (Wu et al., 2021). Bidirectional Long Short-term Memory Network: BiLSTM consists of two independent LSTMS. The input sequence is extracted by the input of two LSTMS forward and backward respectively, and the extracted feature vector is splicing as the final output feature. As a variant of recurrent neural network, BiLSTM captures contextual information in sequences by processing both forward and backward information in time steps. After phase correction, the original signal of CNN-BiLSTM model enters the processing stage of CNN-BiLSTM network. The structure of CNN-BiLSTM network includes input layer, convolutional layer, BiLSTM layer, fully connected layer and output layer. Although independent CNN networks can effectively extract the spatial features of wireless signals, they cannot capture the temporal features of signals. By connecting BiLSTM network with CNN network in series, we can obtain the bidirectional timing feature of signal in time dimension. It is worth noting that the gate structure of BiLSTM effectively alleviates the problem of gradient explosion and gradient disappearance, thus improving the classification accuracy. Our proposed approach based on phase correction and CNN-BiLSTM networks, named PCB networks, provides a novel and efficient method for processing complex signals. The structure of the CNN-BiLSTM network is illustrated in Figure 4.

Figure 4 CNN-BiLSTM network structure (see online version for colours)



4 Prediction model

4.1 Predicting short-term power load process

Short-term power load forecasting is based on historical data to predict the power load for the next few hours or weeks. There are many factors influencing power load, such as weather, holidays, day and night cycles, which are unpredictable. Therefore, short-term power load forecasting is highly complex, requiring high-quality data preprocessing. The main tasks include collecting historical power load data, preprocessing the data, establishing appropriate models, and adjusting model parameters as needed. The specific process of power load forecasting is illustrated in Figure 5.

4.2 Data preprocessing

Due to the original data being recorded at 15-minute intervals, some days have missing power load records. It is evident from Table 1 that there are missing power load data for the years 2019, 2020, and 2021. For power load forecasting, the presence of missing values for certain months can affect future predictions. Prediction of time series data relies on historical data to forecast the future. If there is missing data for a certain period, it means there is a lack of historical information, which can affect future predictions. Therefore, considering the 'forward and backward filling method' or 'interpolation method', in this study, the 'forward and backward filling method' is adopted to interpolate the missing data.

Figure 5 Short-term load forecasting based on CNN-LSTM

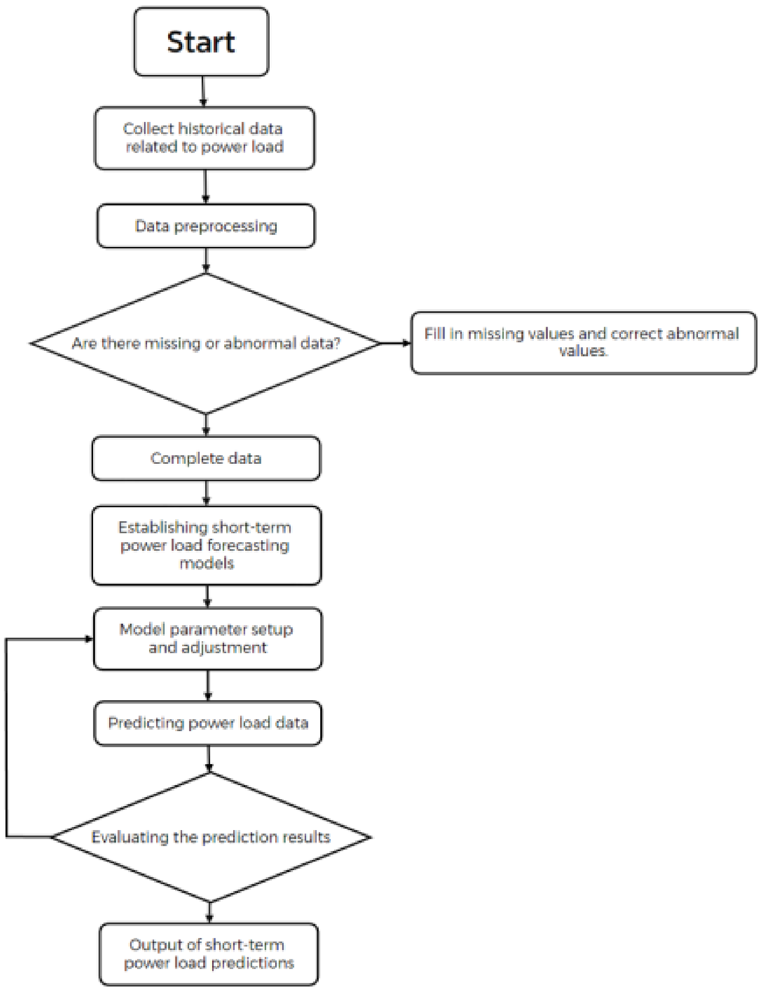
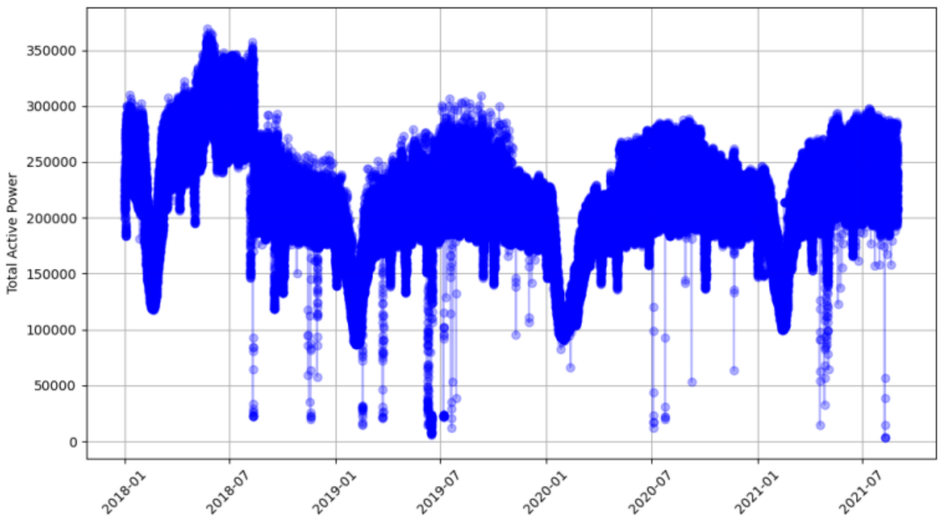


Table 1 The missing power load data

Year	*Record count*	*Actual records*	*Missing number*
2018	35,040	35,040	0
2019	35,040	34,976	64
2020	35,136	35,115	21
2021	35,040	23,025	12,015

Normally refers to data entry errors or values that are logically inconsistent. It generally involves several steps: anomaly detection, anomaly screening, and anomaly handling. Common methods for handling anomalies include box plots, standard deviation method, and the 3 principle (Sullivan et al., 2021). In this paper, the 3 rule is adopted. In statistics, if a data distribution is approximately normal, about 68% of the data values will fall within one standard deviation of the mean; about 95% of the data values will fall within two standard deviations; and about 99.7% of the data values will fall within three standard deviations. Therefore, under the 3 principle, an anomaly is defined as a sample whose deviation from the mean exceeds 3 times the standard deviation, as shown in Figures 6.

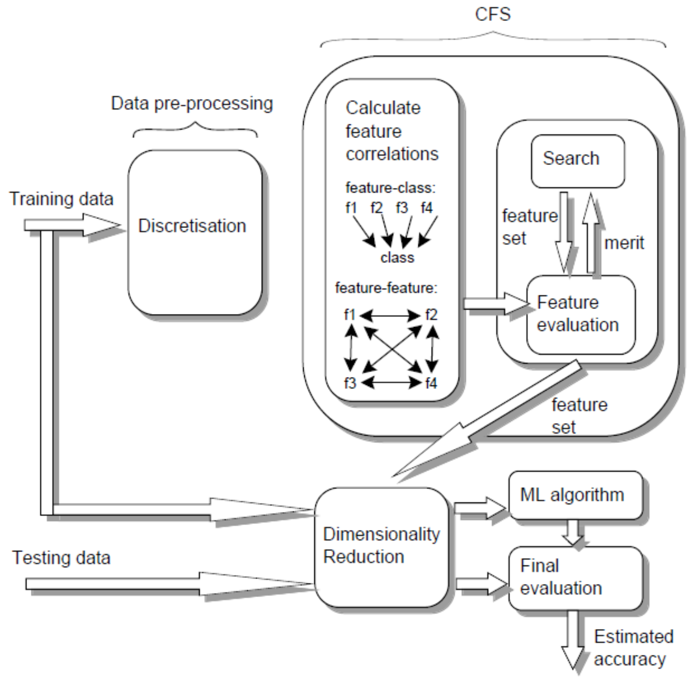
Figure 6 Load data time series line chart (see online version for colours)



4.3 Data feature extraction

Feature selection, also known as feature subset selection or attribute selection, refers to the process of selecting N features from M existing features to optimise a specific indicator of the system. It is the process of selecting the most effective features from the original features to reduce the dimensionality of the dataset, which is different from feature extraction where new features are constructed from the original data. The decision is made to adopt the correlation method of the filtering method for feature selection. In this study, the relationship between multiple variables is investigated, including the relationship between charge quantity and interval time, weather, and industry category. The correlation between independent variables is studied, and features with high correlation are retained. Since the variables are divided into continuous variables and categorical variables, different methods are used to study the relationship between variables: Pearson correlation coefficient and Spearman correlation coefficient. The process of correlation feature selection is shown in Figure 7.

Figure 7 Flowchart of correlation-based feature selection



Meteorological factors are essential factors influencing short-term power loads. However, meteorological datasets contain numerous features. By calculating the Pearson correlation coefficients between each meteorological feature and power loads, we can select features strongly correlated with power loads to input into the model. The Pearson linear correlation coefficients between meteorological data features and loads are shown in Table 2.

Table 2 The correlation coefficients between meteorological features and load are as follows

Meteorological features	*Wind direction*	*Wind speed*	*Temperature*	*Weather*
Pearson correlation	0.12	0.345	–0.53	0.652

From Table 2, it can be observed that wind speed, temperature, and weather have relatively high correlation coefficients, which can be considered as feature factors inputted into the prediction model to participate in short-term load forecasting.

However, wind direction, with a relatively small Pearson correlation coefficient, has minimal impact on electricity load data and can be omitted.

4.4 Model evaluation for prediction

For the CNN-LSTM model and LSTM model established in this study, root mean square error (RMSE) and mean absolute error (MAE) are used for evaluation2. In this context, smaller values of these two indicators indicate smaller errors between actual values and predicted values, thus implying more accurate predictions by the models. The formulas for calculating these two indicators are as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{6}$$

Here, n represents the number of samples, y_i denotes the actual observation, and $\hat{y}_i$ stands for the predicted value,

In order to achieve the best predictive performance of the model, continuous parameter tuning is necessary. For the CNN model, network parameters such as the number of layers, kernel size, and number of filters are optimised (Agga et al., 2021). Similarly, for the LSTM model, parameters such as the number of layers and neurons are adjusted. After a series of optimisation tests, the final configuration of the CNN-LSTM model parameters is shown in Table 3.

Table 3 Parameter configuration of CNN-LSTM model

The types of hyperparameters	*Hyperparameter*	*Parameter setting*
CNN network parameters	Network layer number	2
	Convolution kernel size	3
	Filter number	16
LSTM	Network layer number	2
	Number of neurons	128

LSTM model and CNN-LSTM model are established for power load data, the prediction effects of the two models are shown in Table 4.

Table 4 Prediction error table based on CNN-LSTM model, CNN-BiLSTM model and LSTM model

Model	*MAE*	*RMSE*	*Prediction accuracy*
LSTM	4632.54	5666.36	0.9038
CNN-LSTM	3305.57	4362.73	0.9556
CNN-BiLSTM	4036.10	5092.79	0.9289

From the table, RMSE and MAE of CNN-LSTM model are 4362.73 and 3305.57, and the prediction accuracy is 0.9556. The RMSE and MAE of CNN-BiLSTM were 5092.79, 4036.10, and the prediction accuracy was 0.9289. The RMSE and MAE of LSTM model are 5666.36 and 4632.54, and the prediction accuracy is 0.9038. It can be seen from the above results that both models have higher prediction accuracy, and the CNN-LSTM model is significantly higher than the LSTM model in terms of prediction error or prediction accuracy. The power load prediction curves of LSTM model and CNN-LSTM model are drawn, as shown in Figures 8–10.

Figure 8 LSTM model for power load forecasting curve (see online version for colours)

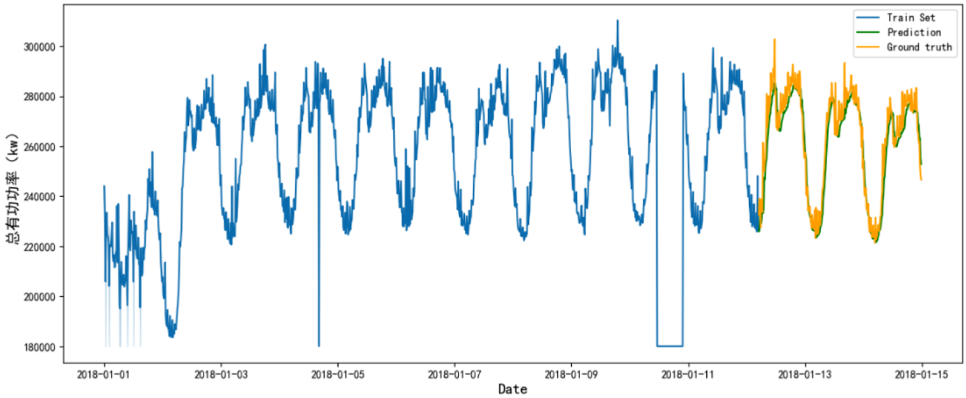


Figure 9 CNN-LSTM model for power load forecasting curve (see online version for colours)

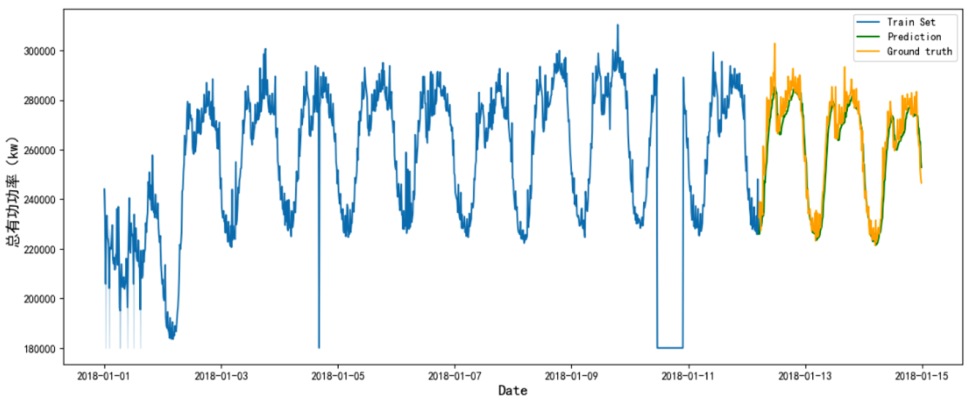
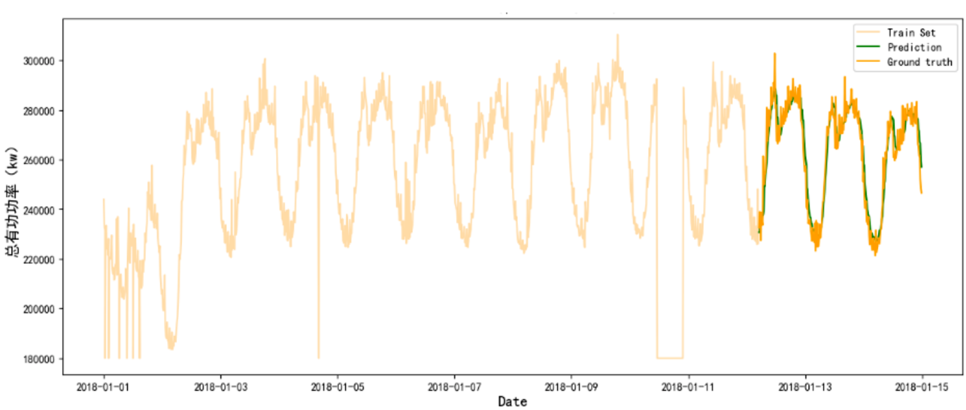


Figure 10 CNN-BiLSTM model for power load forecasting curve (see online version for colours)



From the above figure, it is evident that the predictions generated by the CNN-LSTM model closely match the actual values, indicating a strong predictive performance. A comparison of the RMSE, MAE, and prediction accuracy metrics between the two models reveals that the CNN-LSTM model, which combines CNN and LSTM architectures, delivers superior forecasting results. By leveraging the CNN layer for feature extraction and complexity reduction through pooling layers, and subsequently utilising the LSTM layer for training input to output, the CNN-LSTM model demonstrates its suitability for short-term electricity load forecasting.

5 Conclusion

In order to accurately forecast short-term electricity loads, this study combines machine learning and deep learning models. Building upon the LSTM model, the CNN-LSTM model is proposed by integrating the CNN neural network model. Leveraging the advantages of CNN convolutional neural networks in extracting complex data features, the CNN-LSTM model employs dual layers of 'convolution' and 'pooling' to extract features and reduce data dimensions. The processed data is then input into the LSTM layer, enabling the model to learn temporal data features and ultimately output electricity load predictions. Experimental results demonstrate that the CNN-LSTM model performs exceptionally well in handling complex nonlinear problems, making it particularly suitable for short-term electricity load forecasting? This model effectively predicts load demands in the near future, providing reliable theoretical support for the internal unit start-stop, scheduling, and operational planning within the power sector.

References

Agga, A., Abbou, A., Labbadi, M. and el Houm, Y. (2021) 'Short-term load forecasting: based on hybrid CNN-LSTM neural network', *2021 6th International Conference on Power and Renewable Energy (ICPRE)*, pp.886–891, doi:10.1109/ICPRE52634.2021.9635488.

Bracale, A., Carpinelli, G., De Falco, P. and Hong, T. (2019) 'Short-term industrial reactive power forecasting', *International Journal of Electrical Power & Energy Systems*, Vol. 107, pp.177–185, https://doi.org/10.1016/j.ijepes.2018.11.022

Cui, C., He, M., Di, F., Lu, Y., Dai, Y. and Lv, F. (2020) 'Research on power load forecasting method based on LSTM model', *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, pp.1657–1660, DOI: 10.1109/ITOEC49072.2020.9141684

Guo, X., Zhao, Q., Zheng, D., Ning, Y. and Gao, Y. (2020) 'A short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price', *Energy Reports*, Vol. 6, pp.1046–1053, Retrieved from https://www.sciencedirect.com/science/article/pii/S2352484720315031

Ismail, Z., Jamaludin, F.J. and Jamaludin, F. (2008) 'Time series regression model for forecasting Malaysian electricity load demand', *Asian Journal of Mathematics and Statistics*, Vol. 1.

Jin, Y., Guo, H., Wang, J. and Song, A. (2020) 'A hybrid system based on LSTM for short-term power load forecasting', *Energies*, Vol. 13, p.6241, https://www.mdpi.com/1996-1073/13/23/6241

Rafi, S.H., Nahid-Al-Deeba, M.S.R. and Hossain, E. (2021) 'A short-term load forecasting method using integrated CNN and LSTM network', *IEEE Access*, Vol. 9, pp.32436–32448.

Sullivan, J.H., Warkentin, M. and Wallace, L. (2021) 'So many ways for assessing outliers: what really works and does it matter?', *Journal of Business Research*, Vol. 132, pp.530–543, https://www.sciencedirect.com/science/article/pii/S0148296321002290

Wu, K., Wu, J., Feng, L., Yang, B., Liang, R., Yang, S. and Zhao, R. (2021) 'An attention-based CNN-LSTM-BiLSTM model for short-term electric load forecasting in integrated energy system', *International Transactions on Electrical Energy Systems*, Vol. 31, p.e12637, https://doi.org/10.1002/2050-7038.12637

Xiao, Z., Ye, S-J., Zhong, B. and Sun, C-X. (2009) 'BP neural network with rough set for short term load forecasting', *Expert Systems with Applications*, Vol. 36, pp.273–279, Retrieved from https://www.sciencedirect.com/science/article/pii/S095741740700423X

Zhou, X., Feng, J. and Li, Y. (2021) 'Non-intrusive load decomposition based on CNN-LSTM hybrid deep learning model', *Energy Reports*, Vol. 7, pp.5762–5771, Retrieved from https://www.sciencedirect.com/science/article/pii/S2352484721008064