Application of CNN-LSTM based hybrid neural network in power load forecasting

Lixiong Gong
Hubei University of Technology
Yinkang Chao (2577516309@qq.com)
Hubei University of Technology
Xiao Huang
Hubei University of Technology
Jialin Chen
Hubei University of Technology

Research Article

Keywords: LSTM, CNN, Hybrid model, Energy consumption prediction

Posted Date: June 10th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1732766/v1

Abstract

Effective prediction of electrical energy consumption, rational formulation of corresponding safety measures, and improvement of the accuracy of power load time series prediction are important guidelines for improving the application and management of electrical energy. In order to accurately predict the electric energy consumption and enhance the applicability of the model. In this paper, we propose a convolutional neural network (CNN) based on electric energy consumption data combined with a long-term short-term memory recurrent neural network (LSTM) for electric energy consumption prediction model, selecting electric energy consumption time series with large samples and large time span. consumption time series, including model structure design, model training, model prediction, and model optimization, to implement the prediction algorithm. By using the minimum objective function as the optimization objective, the Adam optimization algorithm is used to continuously update the weights of the neural network and to tune the network layers and batch size to select the best. The number of layers and batch size are used as parameters of the power consumption prediction model. Finally, the optimized CNN-LSTM prediction model is invoked to predict the electricity consumption in the next time period using the electricity load data of Interconnection LLC (PJM) under the Regional Transmission Organization (RTO) in the United States as an example. The results show that the combined model can achieve 98.94% prediction accuracy and 0.0066 mean absolute error (MAE), all of which are better than other basic models, proving that the combined prediction model has better performance in terms of power load prediction accuracy.

1. Introduction

Energy transformation is the current development trend of the energy industry, and an important aspect of it is to improve the efficiency of energy utilization [1]. On the one hand, the problems of non-renewable energy depletion and environmental pollution are becoming increasingly prominent and become new challenges for the development of the energy industry [2]; on the other hand, various traditional energy systems such as electricity, gas and heat have the problem of poor energy coupling, and each energy system is relatively independent and managed and operated by different energy companies [3]; in addition, the energy field shows the development trend of high information-physical integration, and data-driven machine Learning is an important aspect of AI in the energy sector [4]. In recent years, we have observed a convergence of these approaches due to advances in computing power, innovations in algorithms, and the availability of sufficient data. The main beneficiaries of this convergence are data-driven dimensionality reduction methods [5–8], model identification procedures [9–13] and forecasting techniques [14–22], which aim to provide accurate short-term predictions while capturing long-term statistics of these systems. Successful prediction methods address the highly nonlinear energy transfer mechanisms between patterns that cannot be effectively captured by downscaling methods. At present, deep learning has achieved fruitful results in the field of time series prediction, and applying data-driven machine learning to make good predictive analysis of power resource consumption and make specific

planning measures is of great guidance and practical significance to maintain the health, stability, and development of the national economy.

Electricity load forecasting plays an important role in the control of smart grids, power security, market operation and the development of rational dispatching plans, which not only helps the safe and reliable operation of power systems, but also reduces the waste of resources and improves economic efficiency, and is one of the main theoretical bases for the power sector to make production development plans [23]. In previous studies, scholars at home and abroad have developed many methods to improve the accuracy of short-term forecasting, which are mainly classified into statistical-based methods and artificial intelligence-based methods. Statistical methods are based on mathematical models, including multiple linear regression analysis [24], stochastic time series [25], and autoregressive integrated moving average models [26], etc., which usually target linear relationship models, while the forecasted load is more random and has obvious nonlinear characteristics, so the fitting ability in nonlinear forecasting is not strong and the performance is poor. Short-term power load forecasting methods based on artificial intelligence such as support vector machines (SVM) [27], artificial neural network models [28], expert system models [29], Bayesian neural networks [30], BP (backpropagation) neural networks [31], etc. have good nonlinear function fitting The shallow machine learning algorithms with good nonlinear function fitting ability have improved the prediction accuracy compared with traditional methods, but the high accuracy short-term prediction is still a challenge due to the complexity and time-series of load data. In recent years, the rapid development of deep learning has led to its wide application in the field of load forecasting. Deep confidence networks [32], deep neural networks [33], and convolutional neural networks [34–35] have been applied to short-term load forecasting problems, and the accuracy of load forecasting has been improved, but the time-series characteristics of load data are ignored. Recurrent neural network (RNN) introduces recurrent structure to extract the time-series characteristics to fit the time series data [36], but it is prone to the gradient disappearance problem when processing the time series. Long short-term memory network (LSTM) solves this problem by incorporating memory cells, which can learn more complete information of long and short-term patterns in historical sequence data and performs better in long time level prediction based on past load data [37]. In addition, in the literature [38], by adding an attention mechanism to the network structure to assign different weights to the hidden layer units, the model is more likely to learn the long-range interdependencies in the sequences and improve the accuracy of the model. In response to the problems of non-stationarity and complexity of load sequences, the difficulty of tuning the machine learning algorithm and the large data demand, the combination of data decomposition and machine learning was applied to load prediction [39].

Based on the superiority of neural network models, domestic and foreign scholars have achieved good results in applying neural networks for prediction, but in the research of one-dimensional power prediction, few people would use combinatorial algorithms for more accurate prediction, we then try to propose a combinatorial algorithm for accurate prediction of one-dimensional data on this basis. First, the real processed historical load time series are processed for data; then, by using the minimum objective function as the optimization objective, the Adam optimization algorithm is used to continuously update the weights of the neural network, comparing the prediction effect of four models CNN, LSTM,

RNN, XGBoot, selecting the two models with the best prediction effect, and tuning the network layers and batch size to choose the best. Finally the best tuned single example models are combined to construct the combined prediction model. The number of layers and batch size are used as the parameters of the power consumption prediction model, and the results show that the proposed prediction model has higher prediction accuracy and better prediction performance than other prediction networks in short-term load prediction.

2. Model

2.1 Convolutional neural network(CNN)

CNNs are built by mimicking biological visual perception mechanisms and are capable of both supervised and unsupervised learning. the sharing of convolutional kernel parameters in the implicit layers and the sparsity of connections between layers allow CNNs to extract deep local features from high-dimensional data with small computational effort and to obtain effective representations through convolutional and pooling layers. the structure of a CNN network consists of two convolutional layers and a spreading operation. Each convolutional layer contains one convolutional operation and one pooling operation. The CNN structure is shown in Fig. 1.

2.2 Long short-term memory network(LSTM)

A recurrent neural network (RNN) can be thought of as a neural network that passes in time with a depth that is the length of time. For moment t, the gradient it generates disappears after a few layers of propagation to history on the time axis and cannot affect the too distant past. To solve the problem of temporal gradient disappearance, the field of machine learning has developed the long and short time memory unit LSTM, which implements the temporal memory function and prevents gradient disappearance by means of gate switches. LSTM Networks is a kind of recurrent neural networks, and the algorithm was first published by Sepp Hochreiter and Jurgen Schmidhuber in Neural computation. Later, the internal structure was gradually improved through continuous improvement. It will perform better than general RNNs in processing and predicting data related to time series. At present, LSTM Networks have been widely used in robot control, text recognition and prediction, speech recognition, protein homology detection, and other fields. Based on the excellent performance of LSTM Networks in these areas, this paper aims to investigate whether LSTM can be applied to the prediction of electric energy consumption time series.

The most basic LSTM cell consists of three gates (input, forget, output) and a cell. Gate uses a sigmoid activation function, while input and cell state are usually transformed using tanh. the cell of the LSTM can be defined using the following equation.

Gates :

$$i_t = g\ W_{xi}x_t + W_{hi}h_{t-1} + b_i$$

1

$$f_t = g\ W_{xf}x_t + W_{hf}h_{t-1} + b_f$$

2

$$o_t = g\ W_{xo}x_t + W_{ho}h_{t-1} + b_o$$

3

Input transformation:

$$\tilde{c}_t = tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_{\tilde{c}_t}\right)$$

4

Status Update:

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t$$

5

$$h_t = o_t tanh(c_t)$$

6

Thanks to the gating mechanism, the cell can be kept informed for a period of time while working and keep the internal gradient undisturbed by adverse changes during training. va-nilla LSTM has no forget gate and adds the cell state without changes during updates (it can be seen as a recursive link with a constant weight of 1), often referred to as a Constant Error Carousel(CEC). It is so named because it solves the severe gradient disappearance and gradient explosion problems during RNN training, thus making it possible to learn long-term relationships. The LSTM cell structure is schematically shown in Fig. 2.

2.3 Multi-layer CNN-LSTM combined prediction model

In load prediction, the load time series are complex and not smooth, so it is difficult to build a single model to capture all the features of the signal for training and making accurate prediction. Based on the above reasons, we selected two individual models with the most accurate prediction results from the analysis of individual models such as LSTM, CNN, RNN, and XGBoot to create a combination model with more accurate prediction results than individual models. The proposed power load forecasting process in this paper is shown in Fig. 3.

Step 1 : Data acquisition and processing. Raw loads are used as input, data are pre-processed, mean values are used to fill in vacant data and replace abnormal data, and the data are normalized.

Step 2: Build feature equations. In order to make the model for effective simulation and validation, feature datasets and labels will be constructed first, and then the data will be sliced based on the new feature datasets and label sets to obtain the training and test sets of the data, and finally the batch data will be created based on the training and test sets, where the batch data size will be defined and judged based on the dataset type so that the test and training batches can be constructed.

Step 3: Build the model. Use algorithmic statements for LSTM, CNN, RNN, XGBoot model construction, define their own model parameters as well as the model hierarchy, so as to complete the initial construction of the model.

Step 4: Compilation, training and validation of the model. After initializing the parameters of LSTM, CNN, RNN, and XGBoot prediction models by the above steps, the models are compiled and run, and the optimal two training models are found by comparing and analyzing the r^2 values.

Step 5:Model fusion. Based on the two optimal models that have been screened above, the parameters are tuned to determine the optimal parameters, and then the models are compiled, and the optimal single case training model is found by comparing and analyzing the r^2 values by changing the number of implied layers of the model and other parameter adjustments, and finally the optimized single case models are combined and the prediction effect is verified.

The number of layers of neural network mainly depends on the complexity of the learning target, theoretically increasing the number of network layers can improve the model nonlinear fitting ability, but the complexity of the model and training time will also increase, when the number of hidden layers is too large, the speed of update iteration will be reduced, the convergence effect and efficiency decreases, and the accuracy will not improve, so choose the solution with better effect and less time, this paper has been experimentally verified. In this paper, it is verified that setting 2 CNN layers fusion and then 2 LSTM implicit layers can well balance the prediction accuracy and learning efficiency. There are many kinds of gradient-based optimization algorithms, but the gradient-based adaptive momentum estimation algorithm is chosen in this paper, which can dynamically adjust the learning rate of each parameter, so that the learning rate of each iteration has a certain range, and the parameter changes are relatively stable. To evaluate the accuracy of the prediction model, R^2 value, Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are selected as evaluation indexes of this model to measure the accuracy of prediction.

The evaluation indexes are mathematically expressed as:

$$R^2 = \frac{\sum \left(\widehat{yi} - \bar{yi}\right)^2}{\sum \left(\bar{yi} - \widehat{yi}\right)^2} \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\widehat{yi} - yi\right)^2} \quad (8)$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left|\widehat{yi} - yi\right| \quad (9)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{\widehat{yi} - yi}{yi} \right| \quad (10)$$

Where: N is the number of samples; yi, $\widehat{yi}$, $\bar{y}\ i$ are the actual load, the predicted load and the mean value of the actual fit at the ith sampling point of the prediction, respectively. R^2 is an index to evaluate the goodness of the regression model, which can visually represent the fitting effect of the prediction model, MAE can reflect the actual situation of the prediction error, RMSE, as a comprehensive index of error analysis, reflects the accuracy of the prediction, MAPE evaluates the degree of fluctuation of the model prediction error, and reflects the robustness and stability of the model.

3. Model Analysis And Discussion

3.1 Data selection and processing

In order to evaluate the performance of the proposed CNN-LSTM prediction model, the electric load data of Interconnection LLC (PJM) under the Regional Transmission Organization (RTO) in the United States are selected to analyze the prediction model in this paper. A total of 66,497 electric load samples were collected from 1:00 on December 31, 2011 to 24:00 on January 2, 2018, with a 1-hour collection interval. The data set is pre-processed and normalized, and the processed data set is shown in Fig. 4 below. The model can accurately predict load values by learning the cyclical pattern of historical changes in load data at the same point in time without factors such as day type and climate, which reduces the cost of data storage and simplifies the application process of large-scale load forecasting.

After preprocessing and normalization of the dataset, in order to facilitate the construction and training of the model, this paper will continue to build the feature project, complete the construction of the feature dataset and label set by establishing different functional functions, and slice the test set and training set based on the new feature dataset and label set. In this paper, we construct time series data for testing and training, choose time series segmentation to segment the data set, slice the first 90% of the data set into training set and the last 10% into testing set, a total of 59829 data for training and 6648 data for testing, and set the sequence length to 20.

3.2 Model and parameter selection

After the feature equations are established, the algorithms are used to construct the LSTM basic model, CNN basic model, RNN basic model and XGBoot model, and to compile and train the models. Finally, through the comparative analysis of the prediction effects of the above basic models, two optimal models suitable for the data prediction of this paper are found, and then the combination of models is optimized. Due to the huge amount of data, if the data are all displayed, the graphical comparison is not clear, so in the following three prediction models of neural networks in this paper intercepted the real value and predicted value of the first 500 groups of data, in XGBoot prediction model selected from January to February 2017, the comparison chart is as follows.

The evaluation coefficients of each model are shown in Table 1 below.

Table 1
Comparison table of model evaluation coefficients

Model	R ^ 2 value	RMSE	MAE	MAPE
LSTM	0.9709	0.0215408	0.0132779	6.762935
CNN	0.9792	0.0182011	0.01077140	5.920370
RNN	0.8055	0.8015393	11.63799	15.195008
XGBoot	0.6135	181.32491	920.49845	7.866397

By analyzing Table 1, it is obvious that in predicting the accuracy of the data in this paper, the LSTM model and CNN model have a better prediction effect, and the accuracy of prediction can be as high as 97% or more. Therefore, in the following, the LSTM model and CNN model will be analyzed separately and the model combination will be optimized in order to achieve better prediction results.

3.3 Integrated Model Discussion and Optimization

3.3.1 Effect of the number of implied layers of LSTM model on prediction results

After segmenting the data, the segmented training dataset is fed into the LSTM model for parameter rate determination and the model is trained. The framework used in this paper is the TF2.0 (TensorFlow2.0) based on python 3.7 to build the LSTM prediction model, which is an open source software library for numerical computation based on data flow diagrams using nodes to impose mathematical operations and lines to represent the input/output relationships between nodes, and after several experiments, the detailed parameters of the final model construction are shown in Table 2.

Table 2
Parameter table of LSTM neural network model construction.

Batch size	Loss function	Optimization algorithm	vector	Monitoring indicators
32	MAE	Adam	0.05	MSE

By adjusting the number of implied layers and other parameters of the LSTM model, the LSTM models with one, two and three implied layers were used to train and predict the data of this paper, and finally the optimal LSTM model suitable for processing the data of this paper was selected by comparing the prediction results. The model prediction pairs are shown in Fig. 7, and the evaluation indexes are shown in Table 3.

Table 3
Table of evaluation coefficients of different layers of LSTM model.

Model	R ^ 2 value	RMSE	MAE	MAPE
LSTM-1	0.9709	0.0215408	0.0132779	6.762935
LSTM-2	0.9851	0.0143038	0.0073448	3.587704
LSTM-3	0.9813	0.0172336	0.00997243	4.990608

Through the analysis of Table 3, it can be found that when the LSTM model is trained to predict the data in this paper, the prediction effect when the implied layer is 2 layers is significantly higher than that of the other two models, and the prediction accuracy can reach 98.51%, and the mean absolute error (MAE) reaches 0.0073448.

3.3.2 Analysis and discussion of CNN models

In this paper, considering the dimensionality of the current CNN model training data when creating the CNN model, the one-dimensional convolutional neural network model and the two-dimensional convolutional neural network model were used to predict the data of this paper, and then the layers of the two CNN models were optimized again separately, in order to find the optimal CNN model for training prediction of the data of this paper.

The prediction effects of the one-dimensional convolutional neural network model (CNN(Conv1D)) with different convolutional layers are shown in Fig. 8 below, and the corresponding evaluation coefficients are shown in Table 4.

Table 4
Table of evaluation coefficients of 1D CNN models with different convolutional layers

Model	R ^ 2 value	RMSE	MAE	MAPE
CNN(Conv1D)-1	0.9792	0.0182011	0.01077140	5.920370
CNN(Conv1D)-2	0.9815	0.0171658	0.00977320	5.689096
CNN(Conv1D)-3	0.9845	0.0156910	0.0082690	4.358851

By observing Fig. 8 and analyzing Table 4, it is obvious that the prediction effect of the one-dimensional convolutional neural network model (CNN(Conv1D)) is improved with the increase of the number of layers when processing the data in this paper. Based on this conclusion, this paper experimentally found that when the number of layers of the (CNN(Conv1D)) model reached three, the prediction accuracy could already reach 98.45%, and the mean absolute error (MAE) reached 0.0082690, which is similar to the prediction accuracy of the optimal model screened by LSTM above, but in order to save the model training time, this paper tentatively put the one-dimensional convolutional neural network model (CNN(Conv1D)) of the optimal model as CNN(Conv1D)-3, and then let it be combined with the LSTM-2

model. If the prediction effect can be improved on the original basis of the two basic models, it means that the two models can be further combined, and this paper will continue to increase the number of implicit layers of (CNN(Conv1D)) to get the optimal 1D convolutional neural network model.

After discussing the prediction effect of the one-dimensional convolutional neural network model (CNN(Conv1D)) with different convolutional layers, this paper continues to investigate the prediction effect of the two-dimensional convolutional neural network model (CNN(Conv2D)) with different convolutional layers, as shown in Fig. 9 below, and the corresponding evaluation coefficients are shown in Table 5.

Table 5
Table of evaluation coefficients of 2D CNN models with different convolutional layers

Model	R ^ 2 value	RMSE	MAE	MAPE
CNN(Conv2D)-1	0.9798	0.0179456	0.01026735	5.5656215
CNN(Conv2D)-2	0.9829	0.0164944	0.00879986	4.4783502
CNN(Conv2D)-3	0.9793	0.0181290	0.01120285	5.0469165

According to the analysis of Table 5, it is obvious that the prediction effect of the two-dimensional convolutional neural network model (CNN(Conv2D)) in processing the data of this paper, the model is optimal when the hidden layer is two layers. The prediction accuracy can already reach 98.29%, and the mean absolute error (MAE) reaches 0.00879986, so we set the optimal model of the two-dimensional convolutional neural network model (CNN(Conv2D)) as CNN(Conv1D)-2, and then construct the model in combination with the LSTM-2 model.

3.4 Forecast results and analysis

Through the model analysis above, it can be obtained that the LSTM model has the best prediction effect when the number of implied layers is 2 when processing the data in this paper, and the CNN(Conv1D) model and CNN(Conv2D) model have the highest prediction accuracy when the number of implied layers reaches 3 and 2, respectively. To further analyze the optimal combination models, this paper combined the CNN(Conv1D) three-layer model with the LSTM two-layer model and the CNN(Conv2D) two-layer model with the LSTM two-layer model, respectively, and then compared the prediction results of individual models separately. The prediction effect is shown in Fig. 10 below, and the corresponding evaluation coefficient is shown in Table 6.

Table 6
Single model and combined model predict effect evaluation coefficient.

Model	R ^ 2 value	RMSE	MAE	MAPE
CNN(Conv1D)-LSTM(3 + 2)	0.9819	0.0169567	0.00941868	4.8474206
CNN(Conv2D)-LSTM(2 + 2)	0.9894	0.01297160	0.0066239	3.2421672
CNN(Conv2D)-2	0.9829	0.0164944	0.00879986	4.4783502
CNN(Conv1D)-3	0.9845	0.0156910	0.0082690	4.358851
LSTM-2	0.9851	0.0143038	0.0073448	3.587704

The analysis of the above table shows that the prediction accuracy of the combined model of CNN(Conv1D) and LSTM can reach 98.19%, but compared with the original single base model, the prediction accuracy of LSTM can reach 98.51% and the prediction accuracy of CNN(Conv1D) model can reach 98.45%. In comparison, the prediction effect of the combined model is still reduced, and it is obvious that the combined model of CNN(Conv1D) and LSTM does not achieve the effect of optimization and improvement, indicating that the prediction effect of the combined CNN(Conv1D) model and LSTM model does not improve after processing the data in this paper, and the two are not suitable for combined improvement.

The prediction accuracy of the combined model of CNN(Conv2D) and LSTM can reach 98.94%, compared with 98.51% of the original LSTM model, the combined model improves the prediction accuracy by 0.43%, and also improves the accuracy by 0.65% compared with the original CNN(Conv2D) model. This indicates that the combined model of CNN(Conv2D) and LSTM has better prediction effect than the original single model and can analyze and predict the electric load data more accurately.

4. Conclusion

Electricity load forecasting is of great importance to the economic operation of power grids. In this paper, a combined forecasting model is applied, which mainly consists of a combination of convolutional neural network model (CNN) and LSTM neural network, and the following conclusions are obtained:

(1) the pre-processed raw electricity load data are simply split by the algorithm, and then the next set of electricity load values are predicted by using different kinds of neural networks, combined with Adam optimization algorithm is used to continuously update the weights of the neural network, and the load smoothing process can greatly improve the prediction performance of the model.

(2) By adjusting different layers and other parameters in the LSTM neural network model and CNN, the prediction accuracy of the model can be well improved and the problem of gradient explosion can be avoided to the greatest extent.

(3) Experiments are conducted through the actual power load data, and the effectiveness of the model is evaluated respectively. The prediction performance of the model is better than other models and the prediction results are better compared with the prediction models of comparative data processing methods or networks. In summary, the combined forecasting model proposed in this paper provides a new method for electric load forecasting, which can obtain better forecasting results and is excellent in dealing with periodic, trending and stochastic nonlinear loads.

Declarations

Data Availability

Previously reported data were used to support this study and are available at https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption

Conflicts of Interest

There is no conflict of interest regarding the publication of this paper.

Funding Statement

This work was supported by the Ministry of Education, Humanities and social science research projects of China [No. 15YJCZH049] and the Research on Multi-dimensional Evaluation and Optimization of Manufacturing Energy Efficiency under the Background of "Industry 4.0" and "Made in China 2025" of China [No. BSQD2019004]

Acknowledgments

Financial support from the following sources is gratefully acknowledged by the Ministry of Education, Humanities and social science research projects [No. 15YJCZH049] and the Research on Multi-dimensional Evaluation and Optimization of Manufacturing Energy Efficiency under the Background of "Industry 4.0" and "Made in China 2025" of China [No. BSQD2019004].

Author contributions

All authors contributed to the study conception and design. Professor Lixiong Gong provided research direction and writing guidance for this paper. Material preparation, data collection and analysis were performed by Xiao Huang and Jialin Chen. The first draft of the manuscript was written by Yinkang Chao and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethical recognition and consent to participate

All authors of this article are willing to participate and abide by their ethical responsibilities

Publish a consent form

All authors are known and agree to publication

References

1. Jette,S..(2011). The third industrial revolution– how lateral power is transform-ing energy, the economy, and the world. survival, 2(2), 67–68.
2. Cheng,Y., Zhang,N., Lu,Z., & Kang,C.. (2019). Planning multiple energy systems toward low-carbon society: a decentralized approach. Smart Grid, IEEE Transactions on, 10(5), 4859–4869.
3. CHENG Haozhong.,HU Xiao.,WANG Li.,et al..(2019).Review on research of regional integrated energy system planning[J].Automation of Electric Power Systems, 43(7): 1–13(in Chinese).
4. WEI Zhinong.,ZHONG Leilei.,XUE Mingfeng.,et al.. (2019).Linearization flow calculation for integrated electricityheat energy system based on data-driven[J].Electric Power Automation Equipment,39(08):31–37.DOI:10.16081/j.epae.201908020.
5. Williams MO, Kevrekidis IG, Rowley CW. (2015). A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. J. Nonlinear Sci. 25, 1307–1346. (doi:10.1007/s00332-015-9258-5)
6. Tu JH, Rowley CW, Luchtenburg DM, Brunton SL, Kutz JN. (2014). On dynamic mode decomposition: theory and applications. J. Comput. Dyn. 1, 391–421. (doi:10.3934/jcd.2014.1.391)
7. Kutz JN, Fu X, Brunton SL. (2016).Multiresolution dynamic mode decomposition. SIAM.J.Appl. Dyn. Syst. 15, 713–735. (doi:10.1137/15M1023543)
8. Arbabi H, Mezic I. 2017 Ergodic theory, dynamic mode decomposition and computation of spectral properties of the Koopman operator. SIAM. J. Appl. Dyn. Syst. 16, 2096–2126. (doi:10.1137/17M1125236)
9. Brunton SL, Proctor JL, Kutz JN. 2016 Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc. Natl Acad. Sci. USA 113, 3932–3937. (doi:10.1073/pnas.1517384113)
10. Duriez T, Brunton SL, Noack BR. 2016 Machine learning control: taming nonlinear dynamics and turbulence. Berlin, Germany: Springer.
11. Majda AJ, Lee Y. 2014 Conceptual dynamical models for turbulence. Proc. Natl Acad. Sci. USA 111, 6548–6553. (doi:10.1073/pnas.1404914111)
12. Schaeffer, & Hayden. *(*2017*). Learning partial differential equations via data discovery and sparse optimization. Proceedings of the Royal Society of London, 473(2197), 20160446.*

13. Farazmand M, Sapsis TP. 2016 Dynamical indicators for the prediction of bursting phenomena in high-dimensional systems. Phys. Rev. E 94, 032212. (doi:10.1103/PhysRevE.94.032212)
14. Lee, Y., & Majda, A. J.. *(*2016*). State estimation and prediction using clustered particle filters*. Proceedings of the National Academy of Sciences of the United States of America, *113*(*51*), 14609.
15. Comeau D, Zhao Z, Giannakis D, Majda AJ. 2017 Data-driven prediction strategies for low-frequency patterns of North Pacific climate variability. Clim. Dyn. 48, 1855–1872.(doi:10.1007/s00382-016-3177-5)
16. Konstantinos, Tatsis, Vasilis, Dertimanis, Imad, & Abdallah, *et al. (*2017*). A substructure approach for fatigue assessment on wind turbine support structures using output-only measurements*. Procedia Engineering. 199, 1044–1049. (doi:10.1016/j.proeng.2017.09.285)
17. Quade M, Abel M, Shafi K, Niven RK, Noack BR. 2016 Prediction of dynamical systems by symbolic regression. Phys. Rev. E. 94, 012214. (doi:10.1103/PhysRevE.94.012214)
18. Mirmomeni M, Lucas C, Araabi NB, Moshiri B, Bidar MR. 2011 Recursive spectral analysis of natural time series based on eigenvector matrix perturbation for online applications. IET Signal Process. 5, 512–526. (doi:10.1049/iet-spr.2009.0278)
19. Mirmomeni M, Lucas C, Araabi NB, Moshiri B, Bidar MR. 2011 Online multi-step ahead prediction of time-varying solar and geomagnetic activity indices via adaptive neurofuzzy modeling and recursive spectral analysis. Sol. Phys. 272, 189–213. (doi:10.1007/s11207-011-9810-x)
20. Ye L, Liu P. 2011 Combined model based on EMD-SVM for short-term wind power prediction.Commun. Nonlinear Science Numer. Simul. 31, 102–108.
21. Cousins W, Sapsis TP. 2014 Quantification and prediction of extreme events in a one-dimensional nonlinear dispersive wave model. Phys. D 280–281, 48–58. (doi:10.1016/j.physd.2014.04.012)
22. Cousins W, Sapsis TP. 2016 Reduced order precursors of rare events in unidirectional nonlinear water waves. J. Fluid Mech. 790, 368–388. (doi:10.1017/jfm.2016.13. 368)
23. LIAO Nihuan, HU Zhihong, MA Yingying, et al. Review of the short-term load forecasting methods of electric power system [J].Power System Protection and Control, 2011, 39(1): 147–152.
24. VAGHEFI A,JAFARI M A, BISSE E, et al. Modeling and forecasting of cooling and electricity load demand [J]. Applied 105\http://www.cepc.com.cn Energy, 2014, 136: 186–196.
25. CHEN Hong.Implementation of a short-term forecasting system[J]. Automation of Electric Power Systems, 1997,21(12):58–60.
26. LEE C M, KO C N. Short-term load forecasting using liftingscheme and ARIMA models[J]. Expert Systems With Applications,2011, 38(5): 5902–5911.
27. JIANG Min, GU Dongjian, KONG Jun, et al. Short-term load forecasting model based on online sequential extreme support vector regression [J]. Power System Technology, 2018, 42 (7):2240–2247.
28. DAGDOUGUI H, BAGHERI F, LE H, et al. Neural network model for short-term and very-short-term load forecasting in district buildings[J]. Energy and Buildings, 2019, 203: 109408.

29. GHANBARI A, KAZEMI S M R, MEHMANPAZIR F, et al. A cooperative ant colony optimization-genetic algorithm approach for construction of energy demand forecasting knowledge-based expert systems[J]. Knowledge-Based Systems, 2013, 39: 194–206.
30. SHI Huifeng, NIU Dongxiao, LU Yanxia. The short-term load forecasting model based on Bayesian neural network [J]. Chinese Journal of Management Science, 2012, 20(4): 118–124.
31. ZHOU Dianmin, GUAN Xiaohong, SUN Jie, et al. A short-term load forecasting system based on BP artificial neural network [J].Power System Technology, 2002, 26(2): 10–13,18.
32. KONG Xiangyu, ZHENG Feng, E Zhijun, et al. Short-term load forecasting based on deep belief network [J]. Automation of Electric Power Systems, 2018, 42(5): 133–139.
33. HOSSEN T, PLATHOTTAM S J, ANGAMUTHU R K, et al.Short-term load forecasting using deep neural networks (DNN)[C]//2017 North American Power Symposium (NAPS).Morgantown, WV, USA: IEEE, 2017: 1–6.
34. YAO Chengwen, YANG Ping, LIU Zejian. Load forecasting method based on CNN-GRU hybrid neural network [J]. Power System Technology, 2020, 44(9): 3416–3424.
35. ZHAO Bing, WANG Zengping, JI Weijia, et al. A short-term power load forecasting method based on attention mechanism of CNN-GRU[J]. Power System Technology, 2019, 43(12): 4370–4376.
36. RAHMAN A, SRIKUMAR V, SMITH A D. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks [J]. Applied Energy, 2018, 212:372–385.
37. KONG W C, DONG Z Y, JIA Y W, et al. Short-term residential load forecasting based on LSTM recurrent neural network[J]. IEEE Transactions on Smart Grid, 2019, 10(1): 841–851.
38. PENG Wen, WANG Jinrui, YIN Shanqing. Short-term load forecasting model based on attention-LSTM in electricity market[J]. Power System Technology, 2019, 43(5): 1745–1751.
39. Kang Chongqing. Load Forecasting of Power System [M]. Beijing: China Electric Power Press,2017. (in Chinese)

Figures

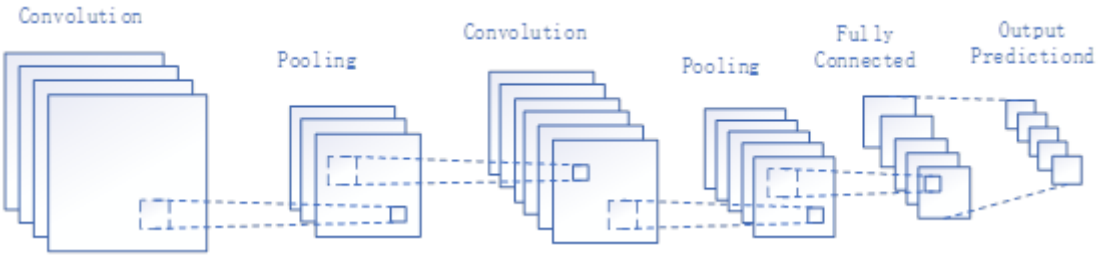

Figure 1

CNN structure

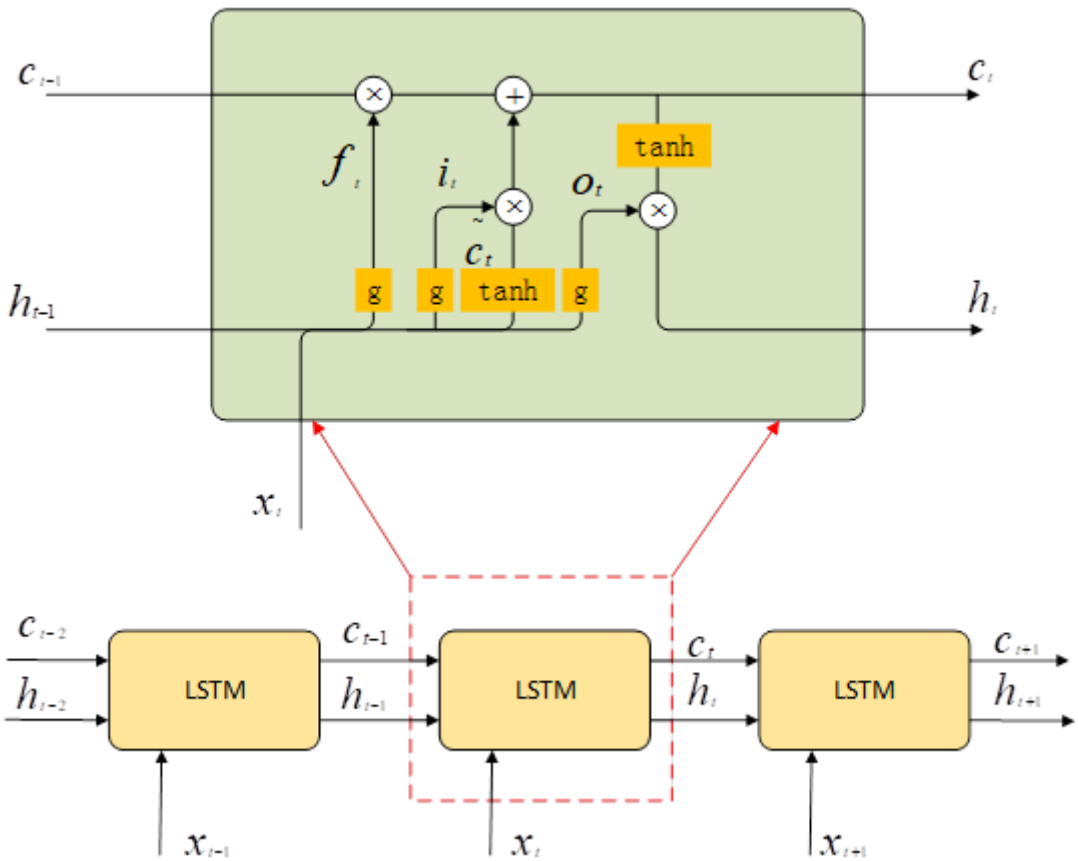

Figure 2

Schematic diagram of LSTM unit structurec

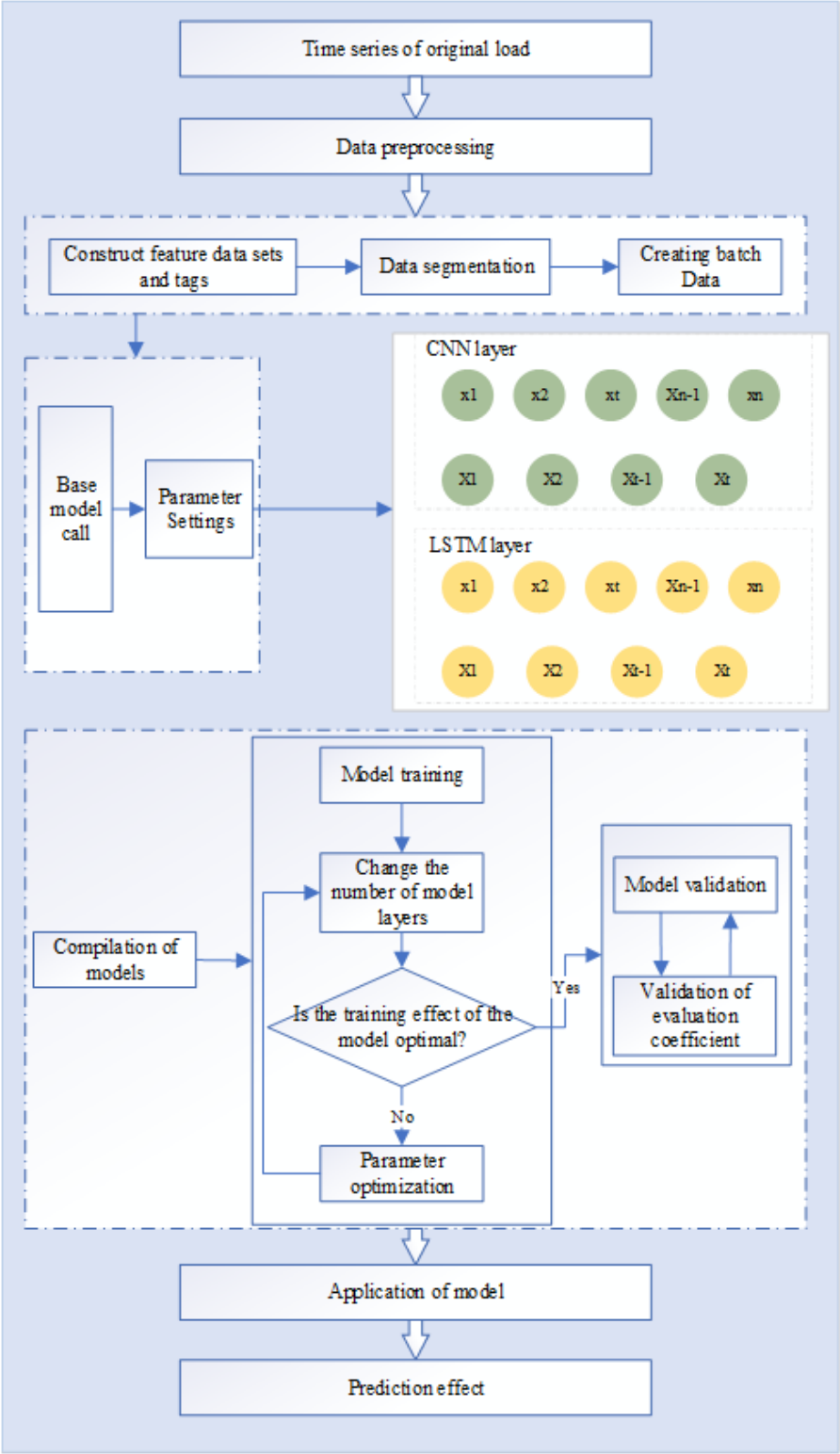

Figure 3

Flow chart of power load forecasting

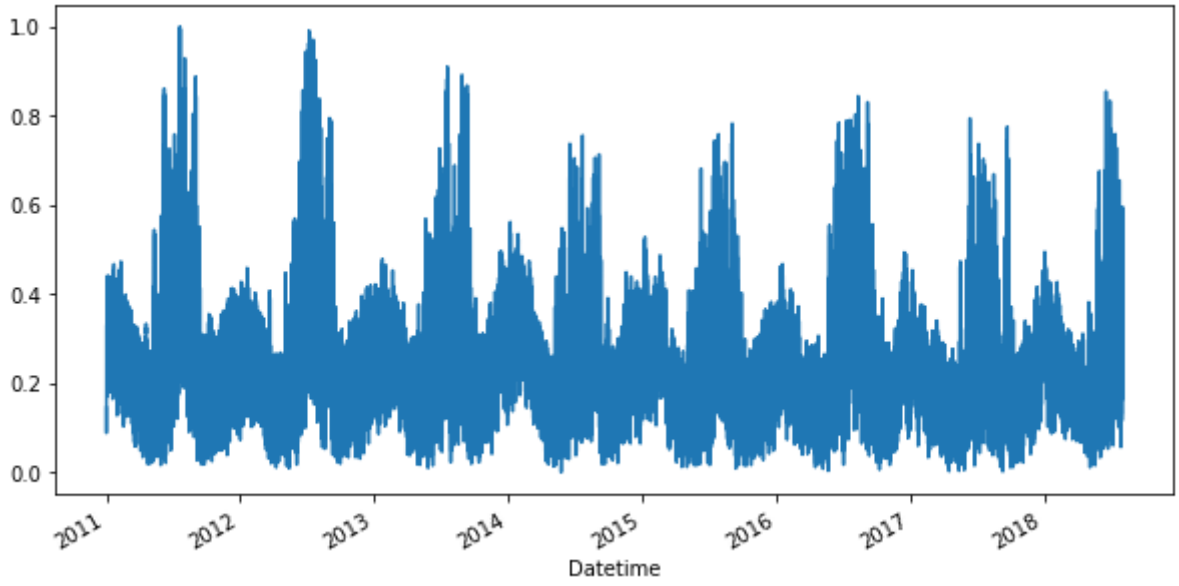

Figure 4

Data set of 2011-2018

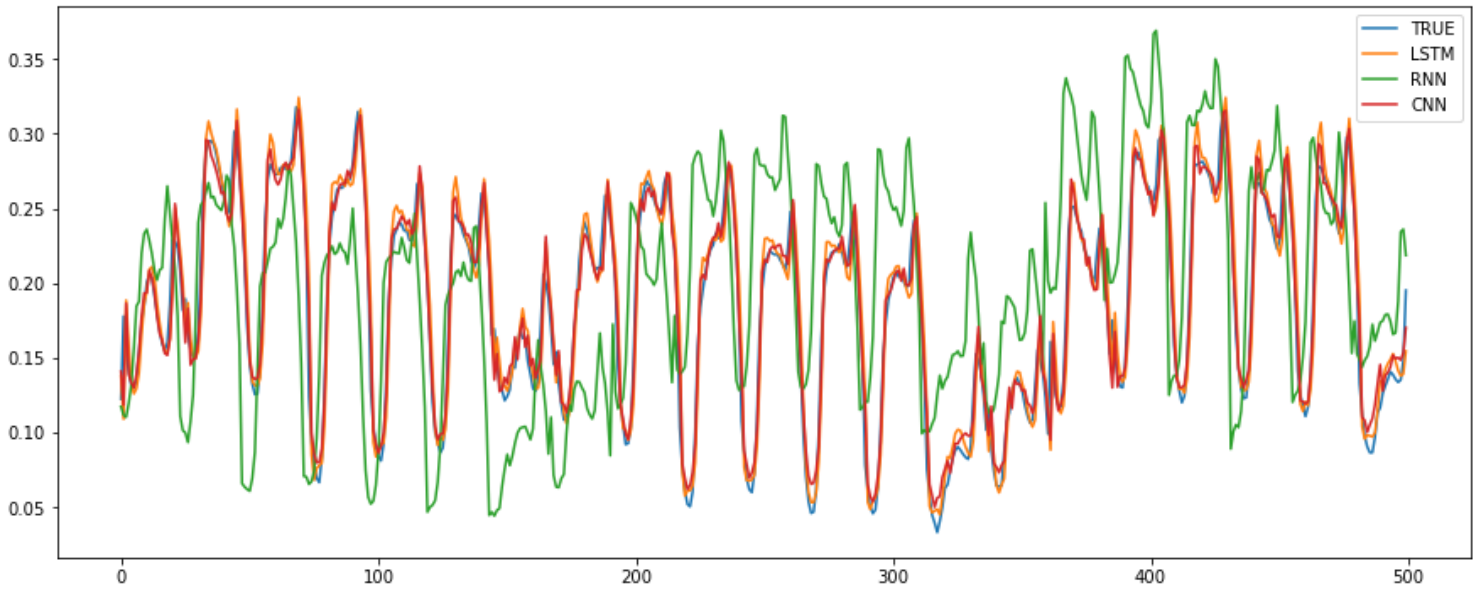

Figure 5

Comparison of prediction effects of CNN, LSTM and RNN

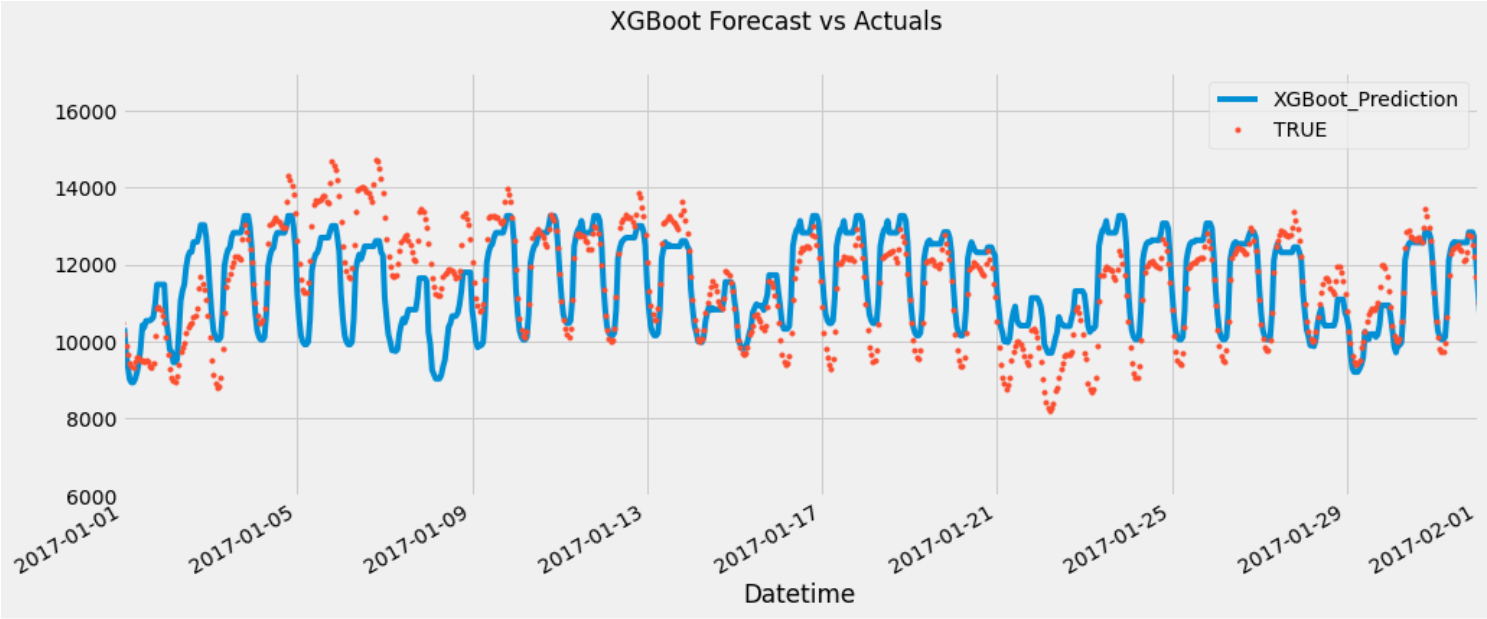

Figure 6

Comparison diagram of XGBoot prediction effect

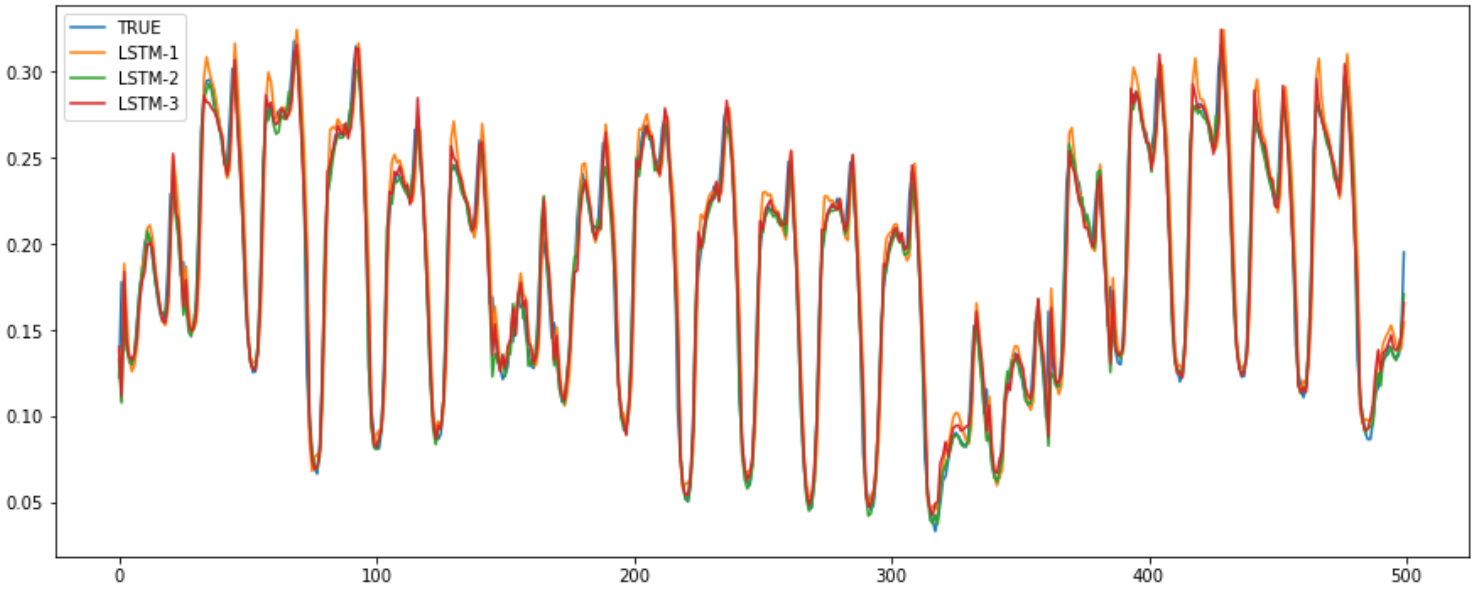

Figure 7

LSTM model effect comparison diagram of different layers

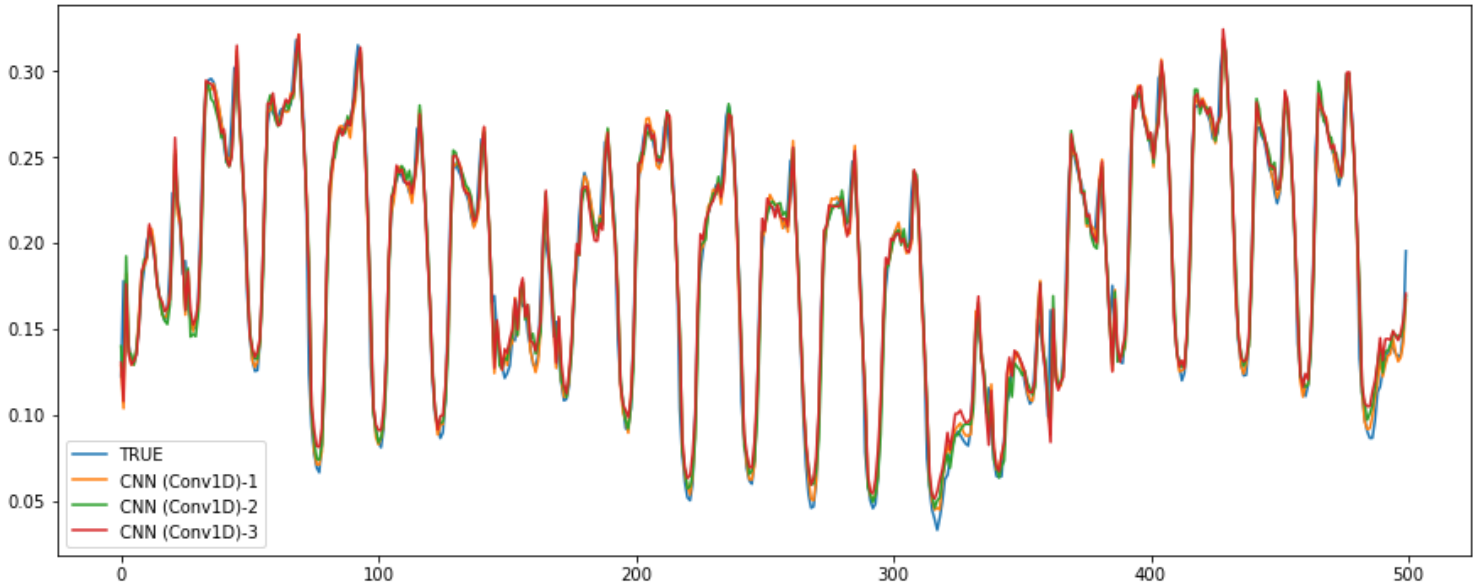

Figure 8

Comparison of the effect of 1D CNN models with different convolutional layers

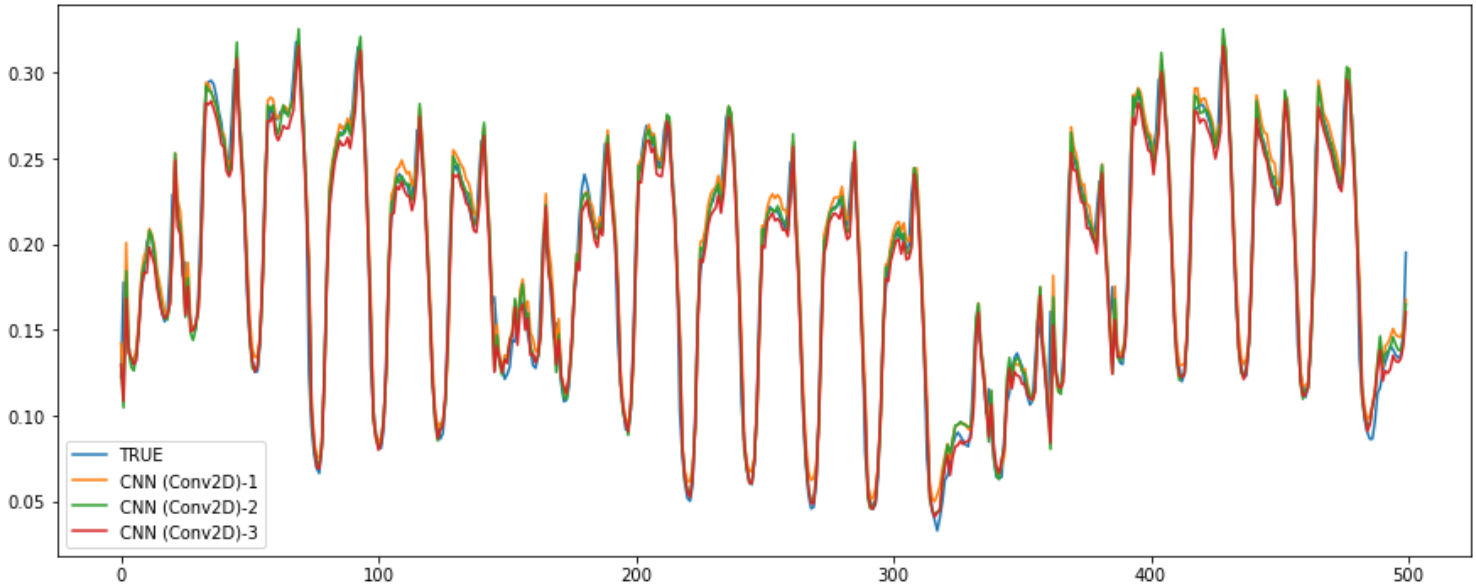

Figure 9

Comparison of the effect of 2D CNN models with different convolutional layers

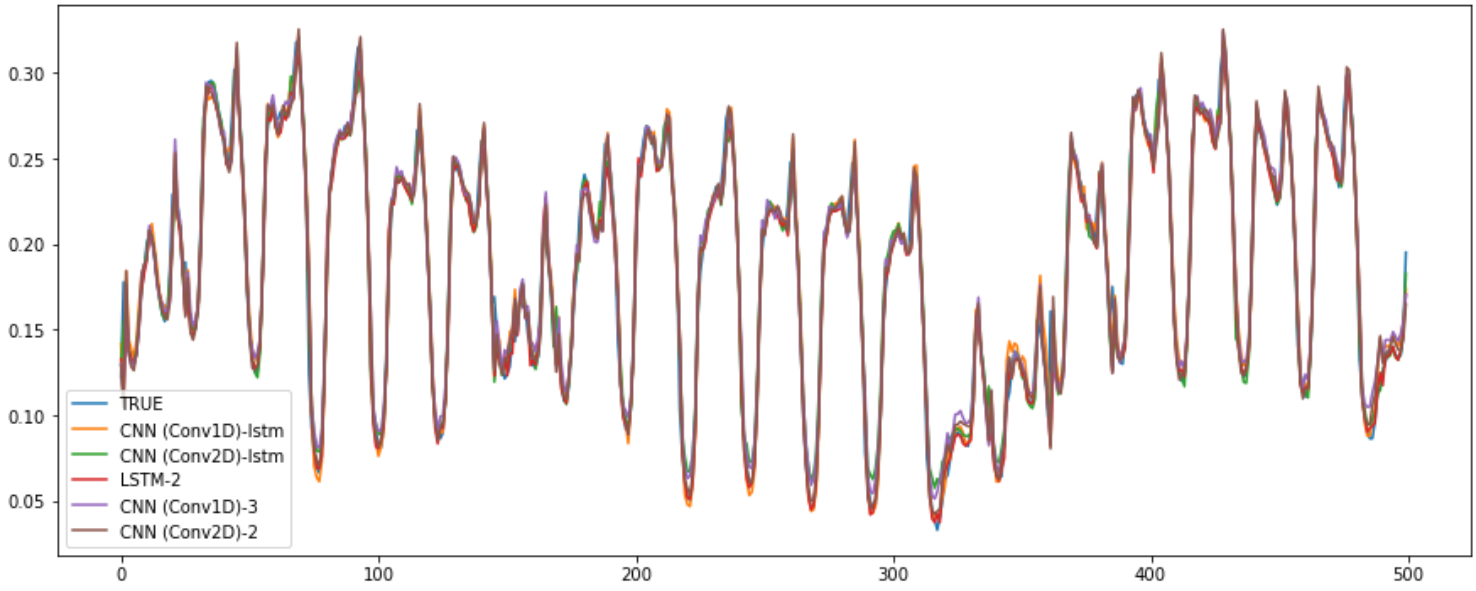

Figure 10

Comparison of prediction effect between single model and combination model