

# Research on power demand forecasting based on attention mechanism and deep learning network

Chenglong Zhang<sup>1</sup>, Li Yao<sup>1</sup>, Jinjin Zhang<sup>2,\*</sup>, Junyong Wu<sup>2</sup>, Baoguo Shan<sup>1</sup>, and Dong Lan<sup>2</sup>

<sup>1</sup> State Grid Energy Research Institute Co. LTD, State Grid, Beijing, China

<sup>2</sup> Electrical Engineering College, Beijing Jiaotong University, Beijing, China

**Abstract.** Combining actual conditions, power demand forecasting is affected by various uncertain factors such as meteorological factors, economic factors, and diversity of forecasting models, which increase the complexity of forecasting. In response to this problem, taking into account that different time step states will have different effects on the output, the attention mechanism is introduced into the method proposed in this paper, which improves the deep learning model. Improved models of convolutional neural networks (CNN) and long short-term memory (LSTM) that combine the attention mechanism are proposed respectively. Finally, according to the verification results of actual examples, it is proved that the proposed method can obtain a smaller error and the prediction performance are better compared with other models.

**Keywords:** Power demand forecasting, Attention mechanism, Convolutional neural networks, Long short-term memory, Multiple influencing factors.

## 1 Introduction

Power load forecasting and power consumption forecasting are included in power demand forecasting. Effective power demand forecasting results can provide a powerful reference for the construction of power equipment and effectively ensure the stability of social power consumption.

In recent years, power demand forecasting methods are divided into two categories, one is the traditional load forecasting method based on statistics, and the other is the artificial intelligence power demand forecasting method. In combination with the actual situation, the randomness of users' electricity consumption become increasingly complicated, which poses a great challenge to power demand forecasting.

In this paper, multi-influencing factors such as historical power load, power consumption, economy and weather are comprehensively considered to the medium and long-term power demand forecasting. Based on this, the improved attention-CNN and attention-LSTM model are proposed. Firstly, data acquisition and pre-processing are completed, and the most influential weather factors are obtained through correlation

analysis. After that, on the basis of CNN and LSTM respectively, attention mechanism is introduced to highlight the timing information that has a key impact on the output, and the improved deep learning prediction models are established. Finally, through the verification, the proposed methods are compared with the deep belief network (DBN), support vector machine (SVM), BP and autoregressive integrated moving average model (ARIMA). The results show that the prediction effect of the proposed method are superior.

## 2 Data preprocessing and correlation analysis

The original power load data, economic data and weather data obtained in this paper are from the data of a regional power grid company from 2010 to 2013. The format of load data and weather data such as temperature, humidity, rainfall are sampling points with 15min intervals every day, that is, 96 sampling points a day. The economic data is the monthly DGP value.

### 2.1 Data preprocessing

The original data has been preprocessed[1], including the correction of bad data such as null data and skip data. Taking into account that different units of load, economic data, and weather data may affect the performance of model training, the data are normalized, and the normalization formula is as shown in (1).

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

In (1),  $x_i^*$  is the normalized sample data;  $x_i$  is the historical data value of a certain sampling point;  $x_{\min}$  is the minimum value of the variable data;  $x_{\max}$  is the maximum value of the variable data.

### 2.2 Correlation analysis between load and influencing factors

Meteorological factors largely affect the accuracy of power demand forecasting [2]. And the summer power load fluctuates and changes significantly due to weather factors particularly. Therefore, taking the summer of 2012 as an example, the Pearson correlation coefficients between power load and weather factors such as temperature, humidity, rainfall, wind speed, and wind direction has been analyzed respectively. The calculation of Pearson correlation coefficient is as (2).

$$\rho_{x,y} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (2)$$

In (2):  $\rho_{x,y}$  is the Pearson correlation coefficient,  $N$  is the number of samples,  $x_i$  is the power load value of a sampling point,  $\bar{x}$  is the average value of the load data,  $y_i$  is the value of each weather influencing factor at the sampling point, and  $\bar{y}$  is the value of each weather influencing factor average value. The Pearson correlation coefficient analysis results of power load and various weather influencing factors in the summer of 2012 are shown in Table 1:

**Table 1.** Correlation coefficient between load and various influencing factors in summer 2012.

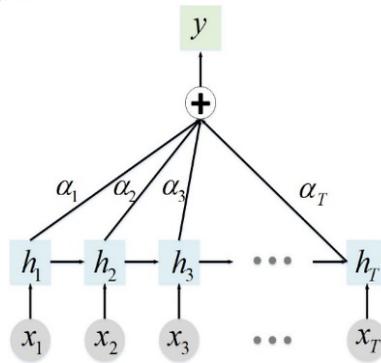
Influencing factors	Correlation coefficient
temperature	0.76
humidity	-0.43
rainfall	0.13
wind speed	0.07
wind direction	0.09

It can be seen from Table 1 that temperature is the weather factor that has the greatest impact on electricity load. And it is taken as one of the input variables of the model. Considering that the granularity of input data of power demand forecasting in this paper is monthly, the original data from 2010 to 2013 is sorted into the form of monthly maximum power load, monthly power consumption, monthly GDP, monthly maximum temperature, and monthly minimum temperature. And the series is taken as the input data set of the model.

### 3 Related theories and methods

#### 3.1 Attention mechanism

The attention mechanism is an attention resource allocation mechanism that simulates the human brain [3]. The model performance can be improved by highlighting important information while reducing or even ignoring the attention to other information and dynamically allocating weights during model training[4]. The structure of the attention mechanism is shown in Figure 1.



**Fig. 1.** Attention mechanism.

The essence of the attention mechanism is a process of weighted summation of variables based on similarity. In Figure 1,  $x_t (t \in [1, T])$  is the input sequence of the network,  $h_t (t \in [1, T])$  is the output of the hidden layer,  $\alpha_t (t \in [1, T])$  is the weight given by the attention mechanism to the hidden layer's output, and  $y$  is the output after the attention mechanism is added to the model. The weights of attention mechanism are calculated from the hidden layer state of the deep learning model, and the calculation are as (3), (4) and (5):

$$score(h_t, h_s) = h_t^T W h_s \tag{3}$$

$$\alpha_{ts} = \frac{\exp(\text{score}(h_t, h_s))}{\sum_{s=1}^S \exp(\text{score}(h_t, h_s))} \quad (4)$$

$$c_t = \sum_s \alpha_{ts} h_s \quad (5)$$

$\text{score}(h_t, h_s)$  is the relationship score, the higher the value, the greater the degree of association. Both  $h_t$  and  $h_s$  are the hidden layer state of the model.  $\alpha_{ts}$  is the attention coefficient corresponding to  $\text{score}(h_t, h_s)$ ;  $c_t$  is the output value of the attention layer that is the weighted summation of the hidden layer state.

### 3.2 Attention-LSTM model

LSTM can extract long-term dependencies in transmission sequences with the special structure of forgetting gate, input gate and output gate [5]. Its structure is shown in Figure 2.

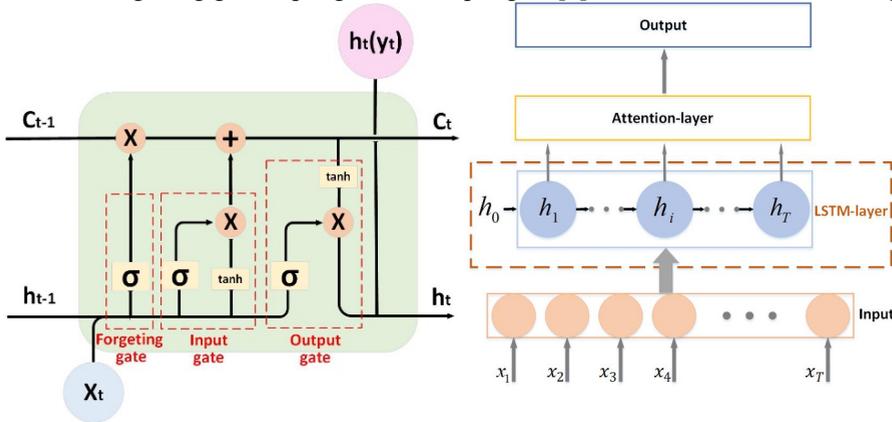


Fig. 2. The structure of LSTM.

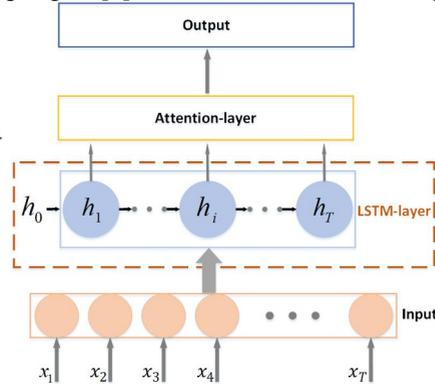


Fig.3. The structure of attention-LSTM.

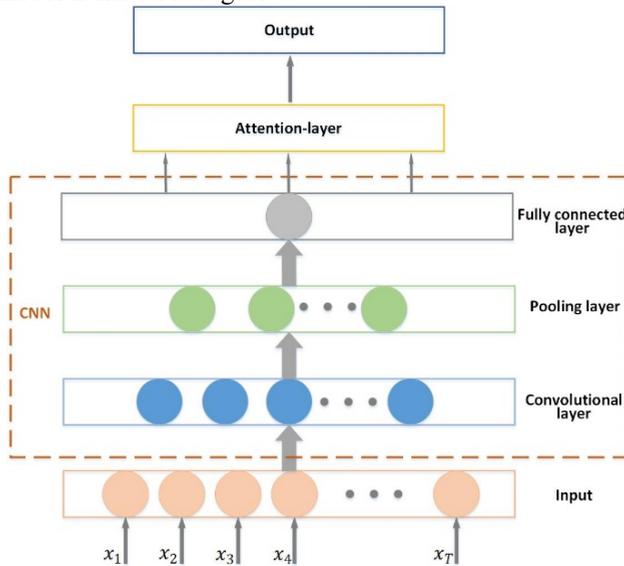
In Figure 2,  $X_t$  is the input vector at the current moment,  $h_{t-1}$  and  $C_{t-1}$  respectively represent the output state of the hidden layer and the internal state of the unit at the previous time,  $h_t$  and  $C_t$  respectively represent the output state of the hidden layer and the internal state of the unit at the current time,  $\sigma$  is the sigmoid activation function,  $\tanh$  is activation function too. It can be seen from the network structure that LSTM can selectively save historical information and input current information for feature extraction of long-term sequences.

The traditional LSTM model transfers the output state of the hidden layer upwards indiscriminately at each moment, without considering its influence on the prediction effect. In this paper, attention mechanism and LSTM are combined to enhance the model's attention to key timing information. The structure of attention-LSTM model is shown in Figure 3.

### 3.3 Attention-CNN model

The CNN model is composed of convolutional layers, pooling layers and fully connected layers [6], which can reduce the difficulty of training while extracting high-level internal features of the information. In this paper, the output state of CNN is given different

dynamic weights, so that important information is highlighted. The structure of the attention-CNN model is shown in Figure 4:



**Fig. 4.** The structure of attention-CNN.

## 4 Comparison and verification

Monthly power data, monthly GDP, monthly high temperature and monthly low temperature of the region from 2010 to 2012 are used as training set data of the models. Based on the two proposed methods, DBN, SVM, BP, ARIMA and the combined model respectively, the data in 2013 are used as the validation set, and the prediction results are obtained.

### 4.1 Evaluation index of prediction results

In this paper, the average absolute percentage error (MAPE) is selected as the evaluation index of the models, and the calculation of MAPE is as (6).

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (6)$$

In (6),  $\hat{y}_i$  is the predicted value,  $y_i$  is the true value,  $n$  is the number of predicted values. Obviously, smaller MAPE means smaller error and higher accuracy.

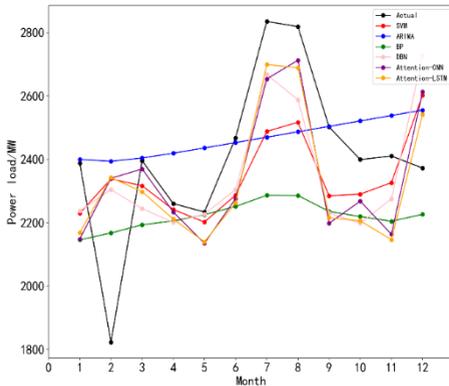
### 4.2 The analysis of result

Based on the attention-LSTM, attention-CNN, DBN, SVM, BP and ARIMA respectively, the prediction results are compared. The results are shown in Table 2.

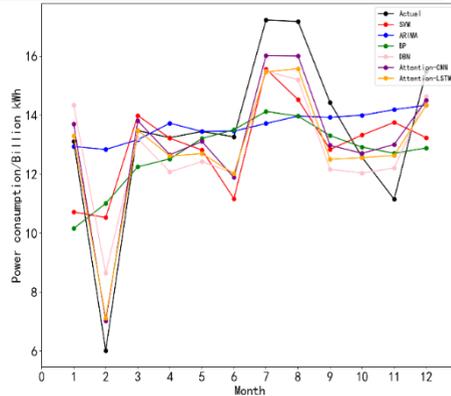
**Table 2.** Comparison of prediction result of various methods.

Model	MAPE of load forecasting	MAPE of power consumption forecasting
SVM	8.10%	16.47%

BP	10.00%	16.80%
ARIMA	7.64%	17.57%
DBN	6.75%	8.14%
attention-LSTM	6.74%	7.42%
attention-CNN	6.49%	7.30%



**Fig. 5.** Comparison of the results of power load forecasting based on various methods.



**Fig. 6.** Comparison of the results of power consumption forecasting based on various method.

It can be seen from the results that whether for power load forecasting or power consumption forecasting, the MAPE of the attention-LSTM and attention-CNN models proposed in this paper are smaller than other comparison methods, and they have better prediction effects. Moreover, the comparison of prediction results of various methods is shown in Figure 5 and Figure 6, where the abscissa represents 12 months and the ordinate represents power load value and power consumption value respectively. By comparison, the prediction effect of the proposed methods are better, and the law of data change can be reflected better.

## 5 Conclusion

Combining historical load data, historical power consumption data, economic data, and related weather data, the medium and long-term power demand forecasting methods based on attention mechanism, attention-LSTM and attention-CNN, are proposed respectively. The attention mechanism combines LSTM and CNN models to dynamically assign weights to the time series, and long-term information features are extracted to perform power demand forecasting. Compared with other methods, it is proved that the proposed methods can get better prediction effect. The optimization of model structure can be studied in the future in order to reduce the prediction error and improve the universality of the model.

I would like to acknowledge assistance and support from the State Grid Corporation of China Headquarters Management Technology Project. The title of the project is "Research and Application Practice on Key Technologies of State Grid New Energy Cloud Smart Operation" (1400-202057409A-0-0-00).

## References

1. J Hu, S S Wen, D F Hu. Analysis, comparison and application of common methods of power load forecasting. Hubei electric power, vol. 32, no.2, pp.13-15, 2008.
2. D X Niu, Z H Gu, M Xing. Research on SVM Short-term Load Forecasting Method Based on Data Mining. Proceedings of the CSEE, vol. 26, no.18, pp.6-12, 2006.
3. S L Qiao. Short-term Traffic Flow Prediction based on deep learning. Qingdao University, Qingdao, 2018.
4. Wang F, Tax D M J. Survey on the attention based RNN model and its applications in computer vision[J]. ar Xiv preprint ar Xiv:1601.06823, 2016.
5. X J Zhu, H L Li, X Q Lu. An improved Attention-Based LSTM feature selection mode. Journal of Beijing Information Science and Technology University, vol. 33, no.2, pp.54-59, 2018.
6. J X Lu, Q P Zhang, Z H Yang. Short-term load forecasting method based on CNN-LSTM hybrid neural network model. Automation of Electric Power Systems, vol. 48, no.3, pp.131-137, 2019.