Research on photovoltaic ultra short-term power prediction algorithm based on attention and LSTM

Aiyun Yan, Jinbo Gu¹, Yahui Mu, Jingjiao Li, Shuowei Jin and Aixia Wang
No.11, Lane 3, Wenhua Road, Heping District, Shenyang City, Liaoning Province

¹ Email: 961664112@qq.com

Abstract. Based on the actual monitoring historical data of photovoltaic power station, combined with the actual engineering demand of photovoltaic micro-grid on the user side, the lightweight algorithm of ultra short-term photovoltaic power prediction is studied, which is conducive to improving the operation efficiency and economy of power system. In this paper, the ultra short-term power prediction of photovoltaic power station is carried out by combining the LSTM algorithm with attention mechanism. Firstly, Pearson correlation coefficient method is used to reduce the dimension of the data set. The data with low correlation between weather variables and power to be predicted and historical power are eliminated, and the algorithm model structure is simplified. Then, the attention mechanism is combined with LSTM network to improve the effectiveness of the prediction model for long time series input. The proposed model is trained and compared with the data of a photovoltaic power station. The results show that the model achieves good experimental results in different weather conditions, and can effectively improve the prediction accuracy.

1. Introduction

The power of photovoltaic power generation will fluctuate due to weather changes, light changes and other factors, and has a certain periodicity and randomness [1]. Large power fluctuation will impact the power grid balance and affect the operation of power grid systems [2-4]. Therefore, the short-term prediction of photovoltaic power can help power grid departments plan and dispatch, avoid risks, and improve the security of power systems [5].

Photovoltaic power prediction is a hot issue in recent years, scholars at home and abroad have done a lot of research work. Most of the recent researches on power prediction are based on historical data statistics [6]. The common prediction methods are based on radial basis function neural network (RBF) [7], the prediction method based on support vector machine (SVM) [8], and other prediction methods based on historical data, including Markov chain [9], Bayesian theory [10], grey theory [11]. The above methods do not make full use of the data before the prediction time, and a large number of valuable historical data are discarded, so the improvement of prediction accuracy is limited.

Due to the function of preserving historical information, long-term and short-term (LSTM) neural network is more effective than traditional neural network when the input is time series, so it has been widely used in recent years. The long and short term memory neural network was first proposed by Hochreite and Schmiduber [12]. LSTM neural network overcomes the problems of long-term dependence, gradient disappearance and explosion in traditional cyclic neural network [13-14]. In reference [15], LSTM neural network is applied to wind power prediction, and the prediction accuracy is better than SVM. However, when learning the input sequence in the traditional codec LSTM model,
all input sequences are first encoded into a fixed length vector, and the decoding process is limited by the representation of the vector, which also limits the performance of the LSTM model.

Therefore, in order to overcome the problems of LSTM, this paper proposes a LSTM model combined with attention mechanism. Firstly, the Pearson correlation coefficient method is used to analyze the weather factors that have great influence on the future power, and then the combination of the attention mechanism and LSTM is used as the prediction model, which makes the prediction model pay more attention to the key time series and achieve the purpose of improving the prediction accuracy. The data set selects the real historical data of a photovoltaic power station. Compared with the results of traditional LSTM, PSO-BP (particle swarm optimization - back propagation) [4] prediction model, it is verified that this model has higher prediction accuracy.

2. Theoretical basis

2.1. LSTM neural network

The traditional recurrent neural network (RNN) performs well in dealing with the input of short time series. In the traditional RNN model, all hidden layers before a certain time will not affect the update of weight array \( \mathbf{w} \). This is the so-called gradient vanishing problem. LSTM neural network can solve the shortcomings of RNN. At the same time, it is widely used in speech recognition, language translation and image processing, and has more obvious advantages than traditional RNN network.

On the basis of the original RNN, the LSTM memory network adds an additional unit to the hidden layer that can store the long-term state. The internal structure of LSTM unit is shown in Figure 1 [12].

![Figure 1. Internal structure of LSTM unit.](image)

There are three inputs to the LSTM cell, which are the input vector \( \mathbf{X}_t \) at the current time, the cell state \( \mathbf{C}_{t-1} \) at the previous time and the hidden layer state \( \mathbf{h}_{t-1} \) at the previous time. The output is the current time unit state \( \mathbf{C}_t \) and the current time hidden layer state \( \mathbf{h}_t \). There are three gates in the system: forgetting gate \( f_t \), input gate \( i_t \) and output gate \( o_t \) to control the abandonment and inheritance of information. The calculation formula of forgetting gate, input gate and output gate is as follows [14].

\[
  f_t = \sigma(W_f \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + b_f) \quad (1)
\]

\[
  i_t = \sigma(W_i \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + b_i) \quad (2)
\]

\[
  o_t = \sigma(W_o \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + b_o) \quad (3)
\]

\( W_f, W_i, W_o \) are the weight matrix of forgetting gate, input gate and output gate respectively, \( [\mathbf{h}_{t-1}, \mathbf{x}_t] \) indicates that the two vectors are connected into a longer vector. \( b_f, b_i \) and \( b_o \) are the bias vectors of forgetting gate, input gate and output gate respectively. \( \sigma \) is the Sigmoid activation function.

The current moment memory cell state \( \mathbf{C}_t \) is calculated by the previous time cell state \( \mathbf{C}_{t-1} \) and the current input intermediate cell state. Tanh is the hyperbolic tangent activation function and \( * \) is the
Hadamard product. The calculation formula is as follows: The disadvantage of these methods is that the historical data before the prediction time is not fully utilized, and a large number of valuable historical data are ignored, so the improvement of prediction accuracy is limited. follows:

\[ C_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b) \]  
\[ C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \]  
\[ \tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b) \]  
\[ h_t = o_t \cdot \tanh(C_t) \]

Finally, the current output \( h_t \) of LSTM unit is:

It can be seen from the above formula that LSTM structure has memory function, which can save certain long-term input. So LSTM has good performance in processing long-term sequence input.

2.2. BPTT training algorithm of LSTM
Similar to feedforward neural network, the training of LSTM network also adopts the back-propagation algorithm of error. Because LSTM deals with sequence data, the error of the whole time series needs to be propagated back when using the error back-propagation algorithm. The state of the current LSTM unit will be affected by the state of the LSTM unit at the previous moment. At the same time, in the error back-propagation calculation, the error of hidden layer \( h_t \) not only includes the error of current time \( t \), but also includes the error of all time after time \( t \), which is called back propagation through time (BPTT) [15].

In the error back-propagation algorithm of LSTM, the error of hidden layer \( h_{t-1} \) is determined by \( h_t \), the state of LSTM cell \( C_{t-1} \) is determined by \( C_t \), and the error of \( C_t \) is composed of two parts, one is \( C_{t+1} \), the other is \( h_t \). Therefore, when calculating the \( C_t \) back-propagation error, we need two known quantities, namely \( h_t \) and \( C_{t+1} \). While \( h_t \) needs to consider \( h_{t+1} \) when updating. In this way, the gradient at any time can be calculated backward from time \( t \), and the weight coefficient can be updated by using random gradient descent.

3. Short term photovoltaic power prediction method based on attention LSTM model

3.1. Power prediction model based on attention LSTM
In the traditional encode-decode model, the encoder encodes the input sequence \( X_t \) into the hidden vector \( h \) with a fixed growth degree, and gives the corresponding weight to the hidden vector. Therefore, when the input sequence becomes longer, the weight of components is the same, and the model does not differentiate the input sequence \( X_T \), so the prediction performance of the model will decline. Attention mechanism is used to improve the effect of encode-decode model. Its essence is to imitate the thinking activity of human brain when observing things. When something important often appears in a certain scene, the human brain will learn and focus on that part when seeing similar scenes [16].

The memory function of LSTM can be maintained for a long time. In practical application, when the data set is multidimensional and multivariate, its performance is poor. Attention LSTM model gives different weights to input characteristics and emphasizes the influence of key factors, which helps LSTM model to judge accurately without increasing the storage and calculation cost of the model.

Figure 2 shows the structure of power prediction model based on attention LSTM. Input vector \( X_{t-n} \) are \( n \) multidimensional eigenvectors before the time to be predicted. Several intermediate states of the input vector are obtained by processing the implicit layer in LSTM. The attention coefficient is obtained by calculating the hidden layer \( h_{j+1} \) of another LSTM network in Decoder and \( h_j \) of encoder. The calculation formula is as follows:

\[ e_{ij} = \nu \tanh(W \cdot h_j + U \cdot h_{j+1} + b) \]
\[ a_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{n-1} \exp(e_{ik})} \]  

(8)

\[ C = \sum_{j=1}^{n-1} a_{ij} h_j \]  

(9)

e_{ij} is the fraction of the relationship between \( h'_{i,1} \) and \( h_j \). The higher the value, the greater the correlation. \( a_{ij} \) is the attention coefficient corresponding to \( e_{ij} \). After that, the attention coefficients are assigned to different intermediate states \( h_j \), and the vector \( C \) input to Decoder is obtained by summation. Finally, vector \( C \) calculates the predicted power at the next moment through the full connection layer.

The combination of Attention mechanism can improve the selection of historical features and screen out more valuable information. Therefore, the model can judge the specific important time series information through the Attention mechanism, and increase the influence weight of this part of information on the results, so as to improve the prediction accuracy.

3.2. Data analysis and processing

The experimental data in this paper use the historical power generation data set of a photovoltaic power station in Northwest China from September 2017 to February 2018. The data set was recorded every 15 minutes with 96 recording points per day, with a total of 17376 data.

The data format is \( X_t = [x_1, x_2, ..., x_8, x_9] \). Where \( x_i \) is the generating power \( x_1 \), humidity \( x_2 \), air pressure \( x_3 \), ambient temperature \( x_4 \), photovoltaic panel temperature \( x_5 \), total radiation intensity \( x_6 \), scattering radiation intensity \( x_7 \), direct radiation intensity \( x_8 \) and wind speed \( x_9 \) at time \( t \), respectively.

When there are many variable dimensions and there are variables with low correlation with the prediction results, the performance of the model will be negatively affected and the prediction accuracy of the model will be reduced. Therefore, it is necessary to analyse the data set, eliminate the valueless variables and reduce the data dimension.

In this paper, Pearson correlation coefficient method is used to test the correlation between different variables in historical data. The Pearson correlation coefficient formula of two-dimensional variables is as follows:

\[ r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \]  

(10)

Where \( \bar{x}, \bar{y} \) are the average values of variables \( x \) and \( y \).

Through Pearson correlation analysis, the correlation between the input variables of the previous time and the generation power at the next moment is compared, and the appropriate input variables are selected to reduce the dimension of input data and improve the performance of the model. The results of correlation analysis are shown in Table 1.

| Influence factor               | Correlation | Influence factor               | Correlation |
|-------------------------------|-------------|-------------------------------|-------------|
| Predicted time power          | 1.000       | Photovoltaic cell temperature  | 0.741       |
| Previous moment power         | 0.972       | Total radiation intensity     | 0.912       |
| Humidity                      | -0.254      | Scattering radiation intensity| 0.720       |
| Pressure                      | 0.016       | Direct radiation intensity    | 0.845       |
| Ambient temperature           | 0.284       | Wind speed                    | 0.105       |
According to the results of correlation analysis, there are five variables which have strong correlation with the power at the prediction time, namely, the power at the previous moment, the temperature of the photovoltaic panel at the previous moment, the total radiation intensity at the previous moment, the scattered radiation intensity at the previous moment and the direct radiation intensity at the previous moment. Therefore, the dimension of the input sequence is 5.

The length of input sequence can be determined by autocorrelation analysis of historical power. The autocorrelation of the output power is shown in Table 2.

Table 2. Analysis of output power autocorrelation.

| Time lag | Correlation | Time lag | Correlation | Time lag | Correlation |
|----------|-------------|----------|-------------|----------|-------------|
| 1        | 0.972       | 11       | 0.514       | 21       | -0.062      |
| 2        | 0.945       | 12       | 0.453       | 22       | -0.109      |
| 3        | 0.916       | 13       | 0.392       | 23       | -0.152      |
| 4        | 0.881       | 14       | 0.330       | 24       | -0.191      |
| 5        | 0.839       | 15       | 0.269       | 25       | -0.227      |
| 6        | 0.793       | 16       | 0.209       | 26       | -0.259      |
| 7        | 0.742       | 17       | 0.150       | 27       | -0.287      |
| 8        | 0.689       | 18       | 0.093       | 28       | -0.311      |
| 9        | 0.634       | 19       | 0.038       | 29       | -0.331      |
| 10       | 0.575       | 20       | -0.013      | 30       | -0.347      |

From the analysis of autocorrelation results, the autocorrelation of output power decreases with the increase of delay time. The historical power data of the first three hours before the prediction time has strong correlation with the predicted point power, so the length of input sequence 12 is more appropriate. According to the previous judgment, the input sequence format is 12×5, and each input data is 12 groups of 5-Dimensional data before the predicted power point.

4. Experimental results and analysis

There are 17876 pieces of data in the data set. The first 9000 pieces of data are selected as the training set, and the last 8376 pieces of data are used as the test set.

The parameter batchsize=182 of the Attention-LSTM prediction model means that 182 samples are taken from the training set for each training, the time step value is timesteps=12, and the number of iterations is set to epoch=300.

In this paper, two other prediction models, LSTM prediction model and PSO-BP prediction model, are selected to make comparative prediction under the same data set, and then the prediction results are compared with the prediction results of Attention-LSTM model, so as to test the prediction effect of the three models.

On December 28, 2017, the power prediction results are shown in Figure 3.

Figure 3. Power prediction results on December 28.
In sunny conditions, it can be seen that at noon, the power predicted by PSO-BP has obvious deviation compared with the actual power. As can be seen from Figure 3 (2), the prediction results of PSO-BP on sunny days are obviously inferior to those of LSTM prediction model and LSTM prediction model combined with Attention mechanism.

The power at noon under cloudy weather on January 17, 2018 is shown in Figure 4 (1). The results predicted by the attention LSTM model are better than LSTM, closer to the real value, and less volatile. The results of PSO-BP model are the worst. It is obvious that the prediction curve of PSO-BP deviates from the real value. The performance of LSTM model is similar to that of attention LSTM, both of which are close to the true power value. The power prediction under rainy and snowy weather on February 6, 2018 is shown in Figure 4 (2). It can be seen from the forecast results that the prediction accuracy of the three models has decreased significantly and the prediction performance of the models has become worse when the weather is bad. Compared with the other two models, the prediction effect of attention LSTM model is more stable, and the deviation between the predicted value and the real value is small when the weather changes suddenly.

The comparison of the prediction results (9:00 a.m. to 5:00 p.m.) of the attention LSTM, LSTM and PSO-BP prediction models under three weather conditions is shown in Table 3.

| Date               | Prediction error | Att-LSTM | LSTM   | PSO-BP |
|--------------------|------------------|----------|--------|--------|
| 28th Sunny         | RMSE/MW          | 2.11     | 2.19   | 2.33   |
|                    | MAPE/%           | 5.93     | 6.09   | 7.82   |
|                    | TIC              | 0.031    | 0.035  | 0.037  |
| 17th Cloudy        | RMSE/MW          | 2.82     | 3.52   | 3.86   |
|                    | MAPE/%           | 7.36     | 9.25   | 10.62  |
|                    | TIC              | 0.046    | 0.057  | 0.064  |
| 6th Rain and snow  | RMSE/MW          | 2.95     | 3.74   | 4.30   |
|                    | MAPE/%           | 21.06    | 24.85  | 27.63  |
|                    | TIC              | 0.091    | 0.126  | 0.144  |

The RMSE and MAPE error evaluation indexes are used to evaluate the advantages and disadvantages of the three models. The data set is to predict 8876 pieces of data in December 2017, January and February 2018. As the average absolute percentage (MAPE) error in the morning and evening is easy to cause large deviation and decrease in accuracy, only the power generation data from 9:00 a.m. to 5:00 p.m. is counted in the calculation. The comparison results are shown in Table 4.

From the error comparison of the three different months, we can see that the hill inequality coefficient of the prediction results based on the Attention-LSTM model is smaller than that of the
other two comparison models. The average value of $e_{\text{RMSE}}$ is 0.26MW less than LSTM and 0.38MW lower than PSO-BP. Compared with LSTM and PSO-BP, $e_{\text{MAPE}}$ decreased by 0.947% and 3.5635% respectively. The experimental results show that the Attention-LSTM model improves the accuracy of power prediction.

Table 4. Comparison of prediction errors between different models in different months.

| Month | Prediction error | Att-LSTM | LSTM | PSO-BP |
|-------|------------------|----------|------|--------|
|       | RMSE/MW          | 2.97     | 3.18 | 3.10   |
| 17.12 | MAPE/%           | 8.56     | 9.81 | 11.25  |
|       | TIC              | 0.045    | 0.047| 0.047  |
| 18.1  | RMSE/MW          | 3.46     | 3.77 | 3.77   |
|       | MAPE/%           | 8.87     | 9.28 | 12.86  |
|       | TIC              | 0.047    | 0.048| 0.046  |
| 18.2  | RMSE/MW          | 4.17     | 4.49 | 4.86   |
|       | MAPE/%           | 15.2     | 16.39| 19.22  |
|       | TIC              | 0.058    | 0.055| 0.064  |

5. Conclusions

In this paper, an ultra short-term power prediction model of photovoltaic power generation based on long-term and long-term memory neural network combined with attention mechanism is proposed, which improves the accuracy of photovoltaic power generation prediction and is practical. The main work of this paper is as follows:

1) Firstly, Pearson correlation coefficient method is used to analyse the data set, and the characteristic dimension and time step of the input data are determined. The influence of irrelevant variables is eliminated, which reduces the complexity of the model and simplifies the calculation process.

2) The LSTM model combined with Attention mechanism is proposed for photovoltaic power prediction, which improves the sensitivity of the prediction model to historical data at different times, and strengthens the ability of feature extraction of the model. Compared with LSTM and PSO-BP model, Attention-LSTM prediction model has higher fitting degree for real power curve, and $e_{\text{RMSE}}$ is reduced by 0.26MW and 0.38MW respectively. $e_{\text{MAPE}}$ decreased by 0.95% and 3.56% respectively, and the improvement effect was obvious. The model has achieved good experimental results in different weather conditions.

After verification, the proposed LSTM photovoltaic power prediction model combined with attention mechanism effectively improves the accuracy of photovoltaic power prediction.

References

[1] Ping Long 2004 The research of supervision system for stand-alone wind/photovoltaic generating station [D] Beijing: Chinese Academy of Sciences Master's Degree Thesis
[2] Zhi Li, Zhinong Wei 2016 Voltage stability bifurcation of large-scale grid-connected PV system [J] Electric Power Automation Equipment 36(1) 17-23
[3] Chen Wei, Ai Xin 2013 Influence of grid-connected photovoltaic system on power network [J] Electric Power Automation Equipment 33(2) 26-32
[4] Elbaset A A, Hassan M S, and Ali H 2017 Performance analysis of grid-connected PV system Power Systems Conference
[5] Bin Wang 2017 The Power Prediction of Photovoltaic Power Station Based on Neural Network [D] North China Electric Power University Master's Degree Thesis
[6] Anshou Li, Qi Chen, Zicai Wang 2016 Review of Power Forecast Methods for Photovoltaic Generating System [J] Electric Drive 46(06) 93-96
[7] Mellit A, Massi P A, Lunghi V 2014 Short-term forecasting of power production in a large-scale
photovoltaic plant [J]. Solar Energy 105 401-413
[8] Ran Li, Guangmin Li 2008 Photovoltaic power generation output forecasting based on support vector machine regression technique [J] Electric Power 41(2) 74-78
[9] Ding Ming, Xu Ningzhou 2011 A Method to Forecast Short-Term Output Power of Photovoltaic Generation System Based on Markov Chain [J] Power System Technology 35(1) 152-157
[10] Bracale A, Caramia P, Carpinelli G, et al. 2013 A Bayesian method for short-term probabilistic forecasting of photovoltaic generation in smart grid operation and control [J] Energies 2013(6) 733-747
[11] Wang Shouxiang, Zhang Na 2012. Short-term Output Power Forecast of Photovoltaic Based on a Grey and Neural Network Hybrid Model [J] Automation of Electric Power Systems 36(19) 37-41
[12] Hochreiter S, Bengio Y, Frasconi P, et al. 2001 Gradient flow in recurrent nets: the difficulty of learning long-term dependencies [C]//A Field Guide to Dynamical Recurrent Networks, Wiley-IEEE Press 2001 237-243
[13] Bengio Y, Simard P, Frasconi P 1994 Learning long-term dependencies with gradient descent is difficult [J]. IEEE Transactions on Neural Networks 5(2) 157-166
[14] Kolen J, Kremer S, Frasconi P, et al. 2001 Gradient flow in recurrent nets: the difficulty of learning long-term dependencies [M] Wiley-IEEE Press, New York 2001 237-243
[15] Qiaomu Zhu, Hongyi Li, Ziqi Wang 2017 Short-term wind power forecasting based on LSTM [J]. Power System Technology 41(12) 3797-3802
[16] Peng Wen, Wang Jinrui, Yi Shanqing 2019 Short-term load forecasting model based on Attention-LSTM in electricity market [J] Power System Technology 43(5) 1746-1751