Short-term Wind Power Prediction Method Based on Wavelet Packet Decomposition and Improved GRU

X R ZU¹ and R X SONG¹

¹School of Control and Computer Engineering, North China Electric Power University, Beijing, 102206, PR China
E-mail: Songruixiao@eplian.com

Abstract. Accurately predicting wind power in wind farms is of great significance to ensure the safe and stable operation of the power system. A novel short-term wind power forecasting method (WPD-GRU-SELU) based on wavelet packet decomposition (WPD) and improved gated recurrent unit (GRU) is proposed. Firstly, this method uses WPD to decompose the time series of wind power into several sub-sequences with different frequencies. Then the sub-sequences of different frequency components are predicted by using the improved GRU neural network, which uses the scaled exponential linear units (SELU) as the activation function to squash the hidden states to calculate the output. Finally, the output datum of GRU neural networks are reconstructed to obtain the complete wind power predicting results. Experiments illustrate that the WPD-GRU-SELU model have a more accurate forecast to the short-term wind power prediction compared with other RNN models.

1. Introduction

Wind energy is an excellent alternative to fossil energy, and it is a good choice for solving current environmental problems and resource crisis. It is one of the most promising renewable energy sources in recent decades. With the large-scale integration of wind power into electric grid, the impact of wind power fluctuation and intermittence on the security and stability of power system has increased dramatically. According to the prediction results of the output power of the wind power plant, the power grid can arrange the power generation plan and stop overhaul, which is beneficial to improve the security, reliability and economic benefit of the system. Therefore, the wind power forecast is of great significance to the large-scale integration of wind power into electric grid.

The wind power forecasting approaches can be divided into four categories according to the different model mechanism: physical method, statistical method, artificial intelligence method and combined prediction method. The physical method is based on the weather forecast information, by analysing the process of weather evolution to predict wind speed and wind direction of the wind farm, and then according to the power curve of the wind farm to predict the output power value. Physical methods do not require statistical historical data, but the establishment of their physical models is very complex, so it is usually used as an input for empirical models. Statistical analysis is based on historical data to establish a linear or nonlinear mapping between the input and output of the system. Artificial intelligence methods, inspired by natural laws or biological intelligence methods, are designed to solve computer programs, including fuzzy algorithms that mimic the concept of fuzziness in human thinking, imitate the evolutionary algorithms of biological evolution and swarm intelligence, imitate the structure of the brain and the artificial neural network (ANN) algorithm for the process of information processing. The nonlinear approximation ability of ANN makes it the most widely used
artificial intelligence method in wind power prediction. Although the theory and method of wind power prediction technology have been very mature, but the single prediction method has its limitations, it is difficult to meet the requirements of the prediction accuracy of the power in the power grid. Therefore, many scholars combine the characteristics of two or more methods to improve the prediction accuracy of the wind power, that is, to set up a combined prediction model of wind power. In fact, most of the prediction models proposed in recent years are combined forecasting methods.

Recently, the deep learning algorithm has made great achievements and has gained wide attention from academia and industry. In this article, we propose a new wind power forecasting method, which combines WPD and GRU to achieve short-term wind power prediction. WPD can decompose the high-frequency and low-frequency components of complex signals into different frequency bands, thus improving the accuracy and efficiency of subsequent prediction. Both GRU and LSTM are extended for the recurrent neural network (RNN) model, but compared to the LSTM model, the GRU model reduces the gate control unit from three to two, and the model is more simple and has a higher efficiency. Experiments show that GRU can avoid gradient vanishing and gradient explosion when using BPTT for network training. And this model uses a new activation function SELU instead of the traditional sigmoid function to squash the hidden states to calculate the output of the model.

2. Related work

2.1. Wavelet packet decomposition

Wavelet packet decomposition is generated and developed on the basis of wavelet decomposition. However, the wavelet decomposition is a mathematical method that decomposes the original signal into multiple sub-sequences, and each time only decomposes the low frequency part. The wavelet packet decomposition not only decomposes the low frequency part, but also decomposes the high frequency part, and can adaptively select the corresponding frequency band to match the signal spectrum according to the signal characteristics and analysis requirements. Therefore, for complex gradient signals, wavelet packet decomposition can help to grasp the details of the signal and improve the resolution of the signal in the time domain, so it has a wider application value.

Figure 1 shows the schematic diagram of the three level wavelet decomposition. Figure 2 gives a schematic diagram of the three level wavelet packet decomposition, where $S$ is the input signal. For wind power forecasting, $S$ is the original wind power.

As shown in Figure 2, the WPD is essentially obtained by further decomposition of the high frequency part obtained from the wavelet decomposition. The decomposition result is the final mapping of the signal $S$ to $2^i$ (as the number of decomposed layers) in the wavelet packet space.

![Figure 1. Process of wavelet decomposition.](image1)

![Figure 2. Schematic diagram of wavelet packet decomposition with three layers.](image2)

2.2. Gated recurrent unit

The GRU is a larger variant of the long short term memory unit (LSTM), and LSTM is a kind of recurrent neural network. LSTM contains three gate functions, input gate, forget gate, and output gate, respectively, and GRU model is a simplified version of the LSTM model, containing only two gate functions, reset gate and update. Reset gate determines how the previous information is combined with the current input, and update gate decides how much previous information is retained. Practice has
proved that GRU is very effective for long-term dependence. The GRU model is shown below Figure 3.

![GRU structure](image)

Figure 3. GRU structure.

At time $t$, the calculation formula is as follows.

1. $r_t = \sigma_{\text{sig}} (W_r x_t + U_r h_{t-1})$  
2. $z_t = \sigma_{\text{sig}} (W_z x_t + U_z h_{t-1})$  
3. $\tilde{h} = \phi_{\text{tanh}} (W_h x_t + r_t \circ U_h h_{t-1})$  
4. $h_t = (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t$

In the above formula, $\tilde{h}_t$ represents the candidate value of the current hidden node, and $h_t$ represents the activation value of the current hidden node output. $r_t$ indicates reset gate and $z_t$ represents update gate. $\circ$ means the element-wise multiplier. $\sigma_{\text{sig}}$ and $\phi_{\text{tanh}}$ are activation functions to activate control gates and candidate states respectively. The expressions for the sigmoid and tanh functions are:

5. $\text{sig}(x) = \left(1 + e^x\right)^{-1}$
6. $\text{tanh}(x) = 2 \times \text{sig}(x) - 1$

In this paper, in order to improve the prediction accuracy, we use an improved GRU model to predict the wind power. We introduce SELU as the activation function of the output $h_t$. The SELU formula is:

7. $\text{selu}(x) = \begin{cases} 
    x, & x > 0 \\
    \alpha e^x - \alpha, & x \leq 0 
\end{cases}$

Where $\lambda = 1.0507009873554804934193349852946$, $\alpha = 1.6732632423543772848170429916717$.

The GRU update formula for introducing the SELU activation function is as follows:

8. $r_t = \sigma_{\text{sig}} (W_r x_t + U_r \phi_{\text{SELU}}(h_{t-1}))$
9. $z_t = \sigma_{\text{sig}} (W_z x_t + U_z \phi_{\text{SELU}}(h_{t-1}))$
10. $\tilde{h} = \phi_{\text{tanh}} (W_h x_t + r_t \circ U_h \phi_{\text{SELU}}(h_{t-1}))$
11. $h_t = (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t$
12. $\text{Output}_t = \phi_{\text{SELU}}(h_t)$

In the above formula, $h_t$ calculates the output value of the model through the SELU activation function. Meanwhile, the SELU squashed state $h_{t-1}$ also contributes to the calculation of $r_t$, $z_t$ and $\tilde{h}_t$.

3. Models and forecasting process

3.1. Data pre-processing

3.1.1. Abnormal data processing. Due to misoperation and other reasons, data collection errors, data loss, etc. may be caused. The wrong data will affect the accuracy of power prediction. So we need to
modify and fill up the abnormal data. In this paper, we use the mean filling method to deal with outliers. The mean filling method is to average data as padding data for the same period of time.

3.1.2. Data standardization. Since GRU uses sigmoid and tanh function as activation function, the original data needs to be normalized and converted into data in the \([0,1]\) interval. The normalized formula is:

\[
y = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}
\]

\(x_{\max}\) is the maximum value of the original power, \(x_{\min}\) is the minimum value of the original power, \(x_i\) is the actual power value, and \(y\) is the normalized power value.

3.2. Wavelet packet decomposition processing

WPD of the historical power data after preliminary processing is carried out. The optimal wavelet packet decomposition space corresponding to different frequency bands and power data in corresponding frequency bands are obtained. The smaller part of the correlation coefficient can be considered as a random wave part of the power, similar to the signal noise, and the power data that neglects the rest of the random wave part will be used as the training sample and predicting data for GRU neural network.

3.3. GRU model construction

GRU neural networks with different structures are constructed for different band power. According to the different frequency bands of different optimal wavelet packets, the GRU neural networks with different structures are constructed respectively, and the structure of each GRU neural network and the number of neurons in each layer are determined.

3.4. GRU model training

The BPTT algorithm is used to train the GRU neural network. First, use the optimal wavelet packets to decompose the power data of different frequency bands in the space, and then, use the BPTT algorithm to train the GRU neural network corresponding to the different frequency bands. When the target prediction error reaches the threshold, the training is stopped. Because of the characteristics of GRU neural network, BPTT algorithm can be used to avoid the gradient vanishing problem.

3.5. Wind power prediction

The WPD-GRU model is used to predict the wind speed of each band. The predictive power data of different frequency bands in the space are decomposed by each optimal wavelet packet as input of GRU neural network, and the power prediction in the corresponding frequency band is carried out respectively. Finally, the predicted power is obtained by the results of the power prediction within each frequency band.

4. Prediction results and comparative analysis

This paper use the actual operating data of wind farms disclosed by Belgian power operator Elia, the data has a time resolution of 15 min, that is 96 data samples per day. The sample data are pre-processed, and the time resolution of the data is converted to one hour, and the average power per hour is the power at that time. After trial and error test several times, considering the speed and accuracy of the calculation, we use the three level WPD. In order to verify the effectiveness of this method, we compare the experimental results with the simple GRU model, the GRU-SELU model, the WPD-GRU-SELU model and the WD-GRU-SELU model. The results show that the prediction accuracy of the proposed model is obviously higher than that of the other models.
4.1. Forecast evaluation index
In order to evaluate the prediction results of each model comprehensively and effectively, the prediction performance of each model was evaluated by absolute mean error (MAE), average relative percentage error (MAPE).

\[
MAE = \frac{1}{N} \sum_{t=1}^{N} |X(t) - \hat{X}(t)|
\]

(14)

\[
MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{X(t) - \hat{X}(t)}{X(t)} \right| \times 100\%
\]

(15)

In the formula, \(N\) is the time point of the test, \(N=24\), \(X(t)\) is the wind power prediction value at \(t\) time, \(\hat{X}(t)\) is the actual wind power at \(t\) time.

4.2. Comparison of the prediction results of models
The above several models are used to predict the wind field data published by Elia. The time resolution of the prediction is 1h. The prediction results of each model in one day are shown in Figure 4, and the prediction error analysis of all models is shown in Table 1.

![Figure 4. Curves of predicted value and real value of wind power during 0:00-24:00 of the typical day.](image)

| model          | MAPE(%) | MAE(m/s) |
|----------------|---------|----------|
| GRU            | 11.32   | 1.0675   |
| GRU-SELU       | 9.58    | 0.9542   |
| WD-GRU-SELU    | 7.21    | 0.7458   |
| WPD-GRU-SELU   | 5.35    | 0.5124   |

It is known from table 1 that: compared with WPD-GRU-SELU, WD-GRU-SELU, GRU-SELU, WPD-GRU and GRU, the former obviously has a higher prediction accuracy. It shows that the optimized neural network can effectively improve the prediction ability. Compared with the two decomposition prediction results, the prediction accuracy after WPD is better than the prediction after WD, indicating that WPD is more conducive to improving the prediction accuracy. WPD-GRU-SELU is the best in all hybrid models, especially in the time period of large fluctuation of wind power. The advantage is more obvious, which shows that the model has a higher prediction accuracy.
5. Conclusion
A short-term wind power forecasting method based on WPD and GRU neural network is proposed in this paper. GRU neural network can mine long-term information of time series, and its characteristics are suitable for short-term wind power forecasting. In order to further improve the accuracy of prediction, WPD is used to decompose the wind power time series. The decomposed sub-sequences have relatively simple wave characteristics and changing rules. In order to verify the validity of the model, we use the actual operating data of the Belgian electric power operator Elia to verify the actual operating data of the wind farm, and compare it with WPD-GRU-SELU, GRU-SELU, WPD-GRU and GRU. The final experimental results show that the model is compared with other prediction methods in predicting the short-term wind power performance. This method can significantly improve the accuracy of short-term wind power prediction.

References
[1] Soder L, Hofmann L, Orths A, et al. 2007 Experience from wind integration in some high penetration areas[J]. IEEE Transactions on Energy Conversion. 22 4-12
[2] Ren Y and Suganthan PN 2015 A comparative study of empirical mode de-composition-based short-term wind speed forecasting methods. Sustain Energy IEEE Trans. 6 236–44
[3] Han L, Romero CE and Yao Z 2015 Wind power forecasting based on principle component phase space reconstruction. Renew Energy. 81 737–44
[4] Wang J, Zhang W, Li Y, et al. 2014 Forecasting wind speed using empirical mode de-composition and Elman neural network. Appl Soft Comput. 23 452–9
[5] Tascikaraoglu A and Uzunoglu M 2014 A review of combined approaches for prediction of short-term wind speed and power. Renew Sustain Energy Rev. 34 243–54
[6] Giorgi MGD, Ficarella A and Tarantino M 2011 Error analysis of short term wind power prediction models. Appl Energy. 88 1298–311
[7] Guo ZH, Wu J, Lu HY, et al. 2011 A case study on a hybrid wind speed forecasting method using BP neural network. Knowl-Based Syst. 24 1048–56
[8] Cho K, van Merrienboer B, et al. 2014 Learning phrase representations using RNN encoder-decoder for statistical machine translation. EMNLP. pp 1724–1734
[9] Sutskever I, Vinyals O and Le Q 2014 Sequence to sequence learning with neural networks. Advances in neural information processing systems. 3104–3112
[10] Chang Li, Yang Qingzhou and Huang Kaibin 2013 A wind speed forecast method based on wavelet transform and ann. Acta Energiae Solaris Sinica. 34 604-609
[11] Han Xiaojuan, Chen Yuyan, Zhang Hao, et al. 2013 Application of hybrid energy storage technology based on wavelet packet decomposition in smoothing the fluctuation of wind power. Proceeding of the CSEE. 33 8-13
[12] CHO K, VAN MERRIENBOER B et al. 2014 Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. arXiv preprint arXiv:1406.1078.
[13] Miranda MS and Dunn RW 2006 One-hour-ahead wind speed prediction using a bayesian methodology. IEEE power engineering society general meeting
[14] Gunter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter 2017 Self-normalizing neural networks. CoRR, abs/1706.02515