COMPARISON OF BLSTM-ATTENTION
AND BLSTM-TRANSFORMER MODELS
FOR WIND SPEED PREDICTION

Zhifeng Liu, Feng Ding[#], Jianyong Lu*,
Yue Zhou*, Hetao Chu

Received on March 16, 2021

Presented by E. Bachvarova, Corresponding Member of BAS, on September 29, 2021

Abstract

Accurate estimation of wind speed is essential for many meteorological applications. A novel short-term wind speed prediction method of Bi-directional LSTM and Transformer Network (BLSTM-TRA) model is proposed by combining the Transformer model and LSTM model, and a hybrid model of Bi-directional Long Short-term Memory and Attention Network (BLSTM-ATT) is proposed based on Attention mechanism and LSTM model. The proposed BLSTM-ATT and BLSTM-TRA model are used for predicting the wind speed of seven meteorological stations in Qingdao. In combination with historical observation data, the proposed models outperform the Numerical Weather Prediction (NWP) system of European Centre for Medium-Range Weather Forecasts (ECMWF). By comparing the results of BLSTM-ATT, BLSTM-TRA and ECMWF forecast model, RMSE and MAE of BLSTM-ATT are reduced by 44.7% and 50.3% on average, respectively, as well as an average decrease of 43.0% in the RMSE, an average decrease of 47.4% in the MAE of the BLSTM-TRA model. This demonstrates that the BLSTM-ATT model and the BLSTM-TRA model are more accurate than the ECMWF model in wind speed prediction.

Key words: attention mechanism, transformer model, LSTM, wind speed forecasting

[#]Corresponding author.

This research was supported by Project of Qingdao Meteorological Bureau (No 2021qdxq06).

DOI:10.7546/CRABS.2022.01.10

1. Introduction. Wind speed forecasting is a challenging problem because of the intermittency and uncertainty of the wind data. A lot of researches have been carried out on wind speed forecasting technology. Most of them adopt the general physical method, statistical method or the combination of these two methods to forecast wind speed. However, the atmospheric wind field has strong randomness and obvious nonlinear characteristics. Emerging artificial intelligence methods have been widely employed since they can effectively capture the nonlinear characteristics of wind speed.

With the rise of machine learning in recent years, an increasing number of researchers have used machine learning methods to predict wind speed [1–8]. Attention neural networks have recently demonstrated success in a wide range of tasks ranging from question answering, machine translations, speech recognition, to image captioning [9–13]. The application of the attention mechanism in time series models is also an active research field [14–16].

Transformer is a relatively new model proposed by VASWANI et al. [17] in “Attention Is All You Need”. Its main operation lies in the use of position embedding to calculate the attention weight of the input sequence. In this model, the focus of input and output is placed on themselves, while the focus of the target is placed on the source. This model extends the conventional attention mechanism through self-attention and it has been proved to have sound prediction effect in many fields [18–24].

At present, there are few studies on the use of the attention mechanism, especially the Transformer model, to predict wind speed. In this regard, the observation data and numerical forecast products are applied in this study to carry out the research on wind forecast technology based on the deep learning algorithm of the Attention mechanism. The bidirectional LSTM neural network model based on the Attention mechanism and the bidirectional LSTM model combining Transformer are established to predict wind speed. These models are expected to reduce the error of wind speed prediction and provide technical support for fine forecast business.

2. Methodology. 2.1. Long Short-Term Memory network. LSTM is introduced as a variant of the RNN model to overcome the long-term dependence of RNN. Generally speaking, wind speed time series have long-term and short-term dependence. LSTM is able to learn long-term dependence and process time data with long-term and short-term dependence in order. Therefore, the LSTM model is applied in this study to predict wind speed.

2.2. The ensemble of deep learning time series prediction for wind speed forecasting. To achieve a better wind speed forecasting, the ensemble of deep learning time series prediction based on LSTM, Attention mechanism (or Transformer model) is developed in this paper. The bidirectional LSTM (BLSTM) neural network with attention layer proposed by ZHOU [24] is used in this study. The network can utilize the information from the past and the future because a

second layer is added to BLSTM to extend the unidirectional LSTM network, in which the connections between hidden layers flow in reverse chronological order. In the structure of the ensemble learning, predictions of BLSTM are input into the Attention layer (or Transformer module) to produce the final forecasting.

2.2.1. Bidirectional long short-term memory network with attention layer. Extracting the hidden features of wind speed time series is an effective method to improve the prediction performance. Traditional model gives the same weights to each input feature, while Attention mechanism gives different weights according to different features. Attention mechanism helps to snatch the temporal characteristics better. Therefore, in this paper the Attention layer is added to the BLSTM model (BLSTM-ATT).

2.2.2. Bidirectional long short-term memory network with Transformer model. Transformer is a more recent model. The main idea of Transformer is to remove the requirement of a recurrent neural network and compute attention weights over an input sequence with positional embeddings instead. The model uses self-attention mechanism, which has been shown to be more effective in some domains than the standard attention model. Therefore, in this paper the Transformer module is added to the BLSTM model (BLSTM-TRA) to extract wind speed features, and the feature extraction structure used is encoder part of Transformer.

3. Data. Qingdao, Shandong Province, China is located in the southeast of Shandong Peninsula, bordering on the Yellow Sea. Figure 1 lists the location distribution of the stations from which the historical data set used in this study is collected. The black points in Fig. 1b are the meteorological stations, and the red points are the ECMWF forecast grid points used in this paper. Twelve points fall in the research area.

The data used in this study include:

1. The historical hourly data is obtained from seven meteorological stations, namely QingDao (QD), HuangDao (HD), JiaoZhou (JZ), LaoShan (LS), JiMo (JM), PingDu (PD), and LaiXi (LX) (Fig. 1). The observation elements consist of time, wind direction, wind speed, precipitation, temperature, relative humidity, dew point temperature, air pressure and visibility. The time range of the historical data is from January 1, 2006 to December 31, 2018.

2. The 6-hour forecast data provided by ECMWF at a spatial resolution of $0.5^\circ \times 0.5^\circ$, including 10 meter U wind component and 10 meter V wind component, wind speed is synthesized as the forecast data of ECMWF in this paper. The time range of forecast data is from January 1, 2018 to December 31, 2018. The inverse distance weighted interpolation method is used to interpolate the forecast data of $0.5^\circ \times 0.5^\circ$ grid point.

The historical hourly observation data collected from seven stations is the input data of the models. Sliding window training method is used to train the models, and the wind speed data of the previous 48 hours sampled at an interval

Fig. 1. Location distribution of seven stations

of 1 hour is used to predict the wind speed of the next hour.

4. Evaluation of forecasting performance. To compare performance between different methods for predicting wind speed time series, there are three evaluation indexes to consider the error between the predicted value and the real value. They are defined as below.

Root mean square error (RMSE):

$$(1) \quad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \tilde{x}_i)^2}.$$

Mean absolute error (MAE):

$$(2) \quad MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \tilde{x}_i|.$$

Correlation coefficient (CC):

$$(3) \quad CC = \frac{Cov(x_i, \tilde{x}_i)}{\sqrt{Var(x_i)Var(\tilde{x}_i)}},$$

where x_i is the observed value; $\tilde{x}_i$ is the predicted value of the model; and n is the total number of statistical objects. Lower values of RMSE and MAE suggest that the model has better performance. When the value of CC is closer to 1, it means that the performance of the model is better.

5. Experiment. 5.1. Analysis of experimental results. In order to evaluate the generalization capabilities, the proposed BLSTM-ATT and BLSTM-TRA are compared with several traditional wind speed forecasting methods including, GRU and LSTM. According to the configuration characteristics of deep

Fig. 2. Four test samples (LSTM, GRU, BLSTM-TRA and BLSTM-ATT) at two stations (HD and JZ) are chosen to visualize the forecasting of wind speed in the future 200 hours

learning models, the data set is divided into training set (from January 1, 2006 to December 31, 2016), validation set (from January 1, 2017 to December 31, 2017) and test set (from January 1, 2018 to December 31, 2018), and the sampling interval is 1 hour. To test the accuracy of models, the trained and calibrated prediction model is used to test the independent test set.

Figure 2 shows the actual and forecasted wind speeds of LSTM, GRU, BLSTM-TRA and BLSTM-ATT, for 200 samples in January 1th~8th. From the time of 175–200 at HD station and 125–200 at JZ station, it is obvious that the prediction values of LSTM and GRU is still far behind the actual values, while BLSTM-TRA and BLSTM-ATT models, especially BLSTM-TRA, agree more with real results and can predict the time of high wind speed. The BLSTM-TRA and BLSTM-ATT achieve promising results in comparison with the classic LSTM by handling the uncertainties existing in the wind speed data. This shows the superiority of BLSTM-TRA and BLSTM-ATT over classic models.

5.2. Comparison with ECMWF. In order to verify the effectiveness of the models, the 6-hour ECMWF forecast data from January 1, 2018 to December 31, 2018, at a 6-hour interval in the next 24 hours from 12 UTC every day are selected, there are 1460 ($365 \text{ days} \times 4/\text{day}$) nodes compared with the observational values.

Disastrous gale process is the focus of marine meteorological forecast. Spring is a season of frequent occurrence of strong winds. Figure 3 shows wind speed of a certain month in spring 2018. The prediction effects of different models are analyzed in Fig. 3 and Table 1. Table 1 shows the RMSE, MAE and CC of the forecasting methods to determine 6-hour interval wind speed forecasts.

As shown in Table 1 and Fig. 3, BLSTM-ATT and BLSTM-TRA methods have fairly good performance for wind speed forecasts. There is a certain error between the prediction results of ECMWF numerical model and the observed values of wind speed, which may be caused by: (I) Systematic error of the model.

T a b l e 1
Statistics of wind speed errors of five models

Station	Model	RMSE(m/s)	MAE(m/s)	CC
QD	GRU	1.27	0.83	0.74
	LSTM	1.44	1.01	0.67
	BLSTM-TRA	0.97	0.71	0.85
	BLSTM-ATT	0.94	0.70	0.85
	ECMWF	1.83	1.45	0.68
HD	GRU	1.23	0.92	0.66
	LSTM	0.83	0.6	0.79
	BLSTM-TRA	0.75	0.56	0.83
	BLSTM-ATT	0.70	0.52	0.85
	ECMWF	2.49	2.13	0.61
LS	GRU	1.86	1.33	0.68
	LSTM	1.50	1.03	0.75
	BLSTM-TRA	1.23	0.88	0.82
	BLSTM-ATT	1.22	0.88	0.85
	ECMWF	1.93	1.53	0.6
JZ	GRU	1.39	1.09	0.80
	LSTM	1.20	0.84	0.82
	BLSTM-TRA	1.27	0.97	0.82
	BLSTM-ATT	1.08	0.8	0.86
	ECMWF	1.50	1.14	0.81
JM	GRU	1.07	0.77	0.75
	LSTM	1.03	0.72	0.78
	BLSTM-TRA	0.92	0.69	0.82
	BLSTM-ATT	0.91	0.68	0.82
	ECMWF	1.46	1.15	0.72
PD	GRU	1.03	0.76	0.78
	LSTM	0.91	0.63	0.78
	BLSTM-TRA	0.89	0.67	0.8
	BLSTM-ATT	0.76	0.55	0.85
	ECMWF	1.49	1.22	0.73
LX	GRU	1.20	0.89	0.75
	LSTM	1.03	0.76	0.76
	BLSTM-TRA	0.83	0.63	0.85
	BLSTM-ATT	1.03	0.73	0.77
	ECMWF	1.34	1.08	0.79
Average	GRU	1.34	0.97	0.73
	LSTM	1.13	0.80	0.76
	BLSTM-TRA	0.98	0.73	0.83
	BLSTM-ATT	0.95	0.69	0.84
	ECMWF	1.72	1.39	0.71

Fig. 3. Comparison of wind speed forecasts by BLSTM-ATT, BLSTM-TRA and ECMWF models at seven stations for a month in spring 2018. The X-axis shows the forecast data at 6-hour intervals

Due to factors such as grid distance and integral time step, the simulated values and the actual values in the numerical model cannot be completely consistent, but only represent the actual wind speed to a certain extent, and the error is inevitable. (II) The local change of environment in the observation field also has influence on the accuracy of the observational data. In this study area, JZ station has the highest terrain and is less affected by the surrounding environment. The forecast

value is smaller than the observed one, which conforms to the general law. QD, HD and LS stations are representative of the coastal character of the region, and the climatological stations at JM, PD and LX – of the meteorological conditions that prevail at the Shandong Peninsula. Numerical prediction is affected by topography and local effects, so it is difficult to forecast. The error of wind speed decreases gradually from the coastal area to the inland area, and the spatial distribution of wind speed error is large in the inland and coastal areas. So ECMWF forecasts are bigger than the observed one. Machine learning models predict future trends based on a large number of historical data, so they will be more accurate. These two approaches have significantly lower RMSE and MAE compared to ECMWF for longer horizons. BLSTM-ATT and BLSTM-TRA improve RMSE by 44.7% and 43.0% for 6-hour interval forecasting compared to ECMWF. They have basically the same performance in coastal stations (QD, HD and LS).

6. Conclusion. BLSTM-TRA and BLSTM-ATT models are constructed in this paper to predict wind speed. The methodology is applied for seven sites at the greater area of Qingdao city, with promising results. The predictive accuracy of the two trained neural networks is superior to the traditional methodologies for the study area. Its high effectiveness in extracting trend information enables attention mechanism to play an important role in model prediction.

Compared with the ECMWF model, there is a decrease of 44.7% and 50.3%, respectively, in the RMSE and MAE of BLSTM-ATT model; there is a decrease of 43.0% and 47.4%, respectively, in the RMSE and MAE of BLSTM-TRA model. Although the trained networks are site specific the methodology can be applied in other locations. In general, BLSTM-TRA and BLSTM-ATT have good performance in coastal stations, but different performance results in inland stations. More inland sites can be selected for testing in the future.

REFERENCES

- [¹] LIU H., H. Q. TIAN, C. CHEN, Y. F. LI (2010) A hybrid statistical method to predict wind speed and wind power, *Renewable Energy*, **35**(8), 1857–1861.
- [²] LIU H., H. Q. TIAN, Y. F. LI (2012) Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction, *Applied Energy*, **98**, 415–424.
- [³] ZHU Q. M., J. F. CHEN, L. ZHU, X. DUAN, Y. LIU (2018) Wind Speed Prediction with Spatio-Temporal Correlation: A Deep Learning Approach, *Energies*, **11**(4), 705.
- [⁴] LIU H., X. W. MI, Y. F. LI (2018(a)) Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network, *Energy Conversion and Management*, **156**, 498–514.

- [5] LIU H., X. W. MI, Y. F. LI (2018(b)) Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network, *Energy Conversion and Management*, **166**, 120–131.
- [6] LIU H., X. W. MI, Y. F. LI (2018(c)) Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM, *Energy Conversion and Management*, **159**, 54–64.
- [7] LIU H., X. W. MI, Y. F. LI (2019) Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional Gated Recurrent Unit network and Support Vector Regression, *Renewable Energy*, **143**, 842–854.
- [8] XU W., H. PENG, X. ZENG, F. ZHOU, X. TIAN, X. PENG (2019) A hybrid modelling method for time series forecasting based on a linear regression model and deep learning, *Applied Intelligence*, 1–14.
- [9] HERMANN K. M., T. KOCISKY, E. GREFFENSTETTE, L. ESPEHOLT, W. KAY et al. (2015) Teaching machines to read and comprehend. In: *Advances in Neural Information Processing Systems*, 1684–1692.
- [10] BAHDANAU D., K. CHO, Y. BENGIO (2014) Neural machine translation by jointly learning to align and translate, *Computer Science*, [arXiv:1409.0473](https://arxiv.org/abs/1409.0473).
- [11] CHOROWSKI J. K., D. BAHDANAU, D. SERDYUK, K. CHO, Y. BENGIO (2015) Attention-based models for speech recognition. In: *Advances in Neural Information Processing Systems*, 577–585.
- [12] XU K., J. BA, R. KIROS, A. COURVILLE, R. SALAKHUTDINOV et al. (2015) Show, attend and tell: Neural image caption generation with visual attention, *Computer Science*, 2048–2057.
- [13] GAO Y., C. FANG, Y. RUAN (2019) A novel model for the prediction of long-term building energy demand: LSTM with Attention layer, *Earth and Environmental Science*, **294**, 012033.
- [14] ZHANG X., X. LIANG, A. ZHIYULI (2019) AT-LSTM: An Attention-based LSTM Model for Financial Time Series Prediction, *Materials Science and Engineering*, **569**(5), 052037.
- [15] SHIH S. Y., F. K. SUN, H. LEE (2019) Temporal pattern attention for multivariate time series forecasting, *Machine Learning*, **108**(8), 1421–1441.
- [16] WANG X., Z. LI, J. ZHANG, H. LIU, X. CAI (2019) An LSTM-attention wind power prediction method considering multiple factors, *8th Renewable Power Generation Conference*.
- [17] VASWANI A., N. SHAZEER, N. PARMAR, J. USZKOREIT, L. JONES et al. (2017) Attention is all you need. In: *Advances in Neural Information Processing Systems*, **30**, 5998–6008.
- [18] FU S., L. ZHAO, G. P. ZANK et al. (2020) An ACE/CRIS-observation-based Galactic Cosmic Rays heavy nuclei spectra model II, *Sci. China Phys. Mech. Astron.*, **63**, 219511.
- [19] ZHOU L., Y. ZHOU, J. J. CORSO, R. SOCHER, C. XIONG (2018) End-to-end dense video captioning with masked transformer, *Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [20] CHATTERJEE J., N. DETHLEFS (2020) A Dual Transformer Model for Intelligent Decision Support for Maintenance of Wind Turbines, *International Joint Conference*

- on Neural Networks.
- [²¹] FU S., L. L. ZHAO, G. P. ZANK et al. (2020) An ACE/CRIS-observation-based Galactic Cosmic Rays heavy nuclei spectra model II, *Sci. China Phys. Mech. Astron.*, **63**(1), 100–111.
 - [²²] LI J., W. WANG, J. LU, J. YUE, A. G. BURNS et al. (2019) A modeling study of the responses of mesosphere and lower thermosphere winds to geomagnetic storms at middle latitudes. *Journal of Geophysical Research: Space Physics*, **124**(5), 3666–3680.
 - [²³] LI J., W. WANG, J. LU, T. YUAN, J. YUE et al. (2018) On the responses of mesosphere and lower thermosphere temperatures to geomagnetic storms at low and middle latitudes, *Geophysical Research Letters*, **45**, 10128–10137.
 - [²⁴] ZHOU P., W. SHI, J. TIAN (2016) Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification, *Proceedings of the 54th annual meeting of the association for computational linguistics*, 207–212.

Qingdao Meteorological Bureau
Qingdao 266003, China
e-mail: lzflyjk@163.com
fdingqd@126.com
chuwang90@sina.com

**Nanjing University of Information Science*
and Technology
Nanjing 210044, China
e-mail: yuezhou@nuist.edu.cn
jylu@nuist.edu.cn