Ultra Short Term Power Forecasting of Wind Generation Based on Improved LSTM

Haitao Zhang
Lincang Power Supply Company
Yunnan Power Grid Co., Ltd.
Lincang, China
198066318@qq.com

Qihu Zhu
Lincang Power Supply Company
Yunnan Power Grid Co., Ltd.
Lincang, China
153976127@qq.com

Wenjuan Li
Lincang Power Supply Company
Yunnan Power Grid Co., Ltd.
Lincang, China
liwj@163.com

Xuefeng Li
Lincang Power Supply Company
Yunnan Power Grid Co., Ltd.
Lincang, China
xuefengli@sina.com.cn

Changqing Xie
Lincang Power Supply Company
Yunnan Power Grid Co., Ltd.
Lincang, China
xiecq@sina.com.cn

Chunyong Xiang
Lincang Power Supply Company
Yunnan Power Grid Co., Ltd.
Lincang, China
cyxiang@163.com

Jinghao Fan
College of Information Sciences and Engineering
Hohai University
Changzhou, China
1045618442@qq.com

2023 3rd Power System and Green Energy Conference (PSGEC) | 979-8-3503-4009-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/PSGEC58411.2023.10255886

***Abstract*—In response to the characteristics of intermittent, high variability, and strong randomness in wind power output, A multi-layer stacked bidirectional LSTM (MBLSTM) with Temporal Pattern Attention(TPA) mechanism is proposed as a forecasting method for wind power. Firstly, clean up the dataset using a combination of DBSCAN algorithm and linear threshold method; Secondly, based on the TPA mechanism, the complex correlation between features is extracted from the hidden row vectors obtained from the MBLSTM network, and the forecasting precision is enhanced by reasonable allocation of weights at different time steps. Finally, experimental simulation data proves that the proposed approach demonstrates significant efficacy in enhancing the precision of ultra short-term power forecasting.**

Keywords—wind power prediction; TPA; Improved LSTM; anomaly detection; DBSCAN; linear regression

I. INTRODUCTION

The progress made in wind power technology presents a practical means of tackling worldwide energy crises and mitigating the adverse impacts of greenhouse effects. However, wind power systems exhibit strong intermittency and volatility[1], posing new challenges to grid planning and stable operation. Precise wind power prediction is essential for optimizing operational efficiency of wind energy generation systems, enhancing the safety and reliability of power system operation, and facilitating improved coordination with alternative power generation devices while simultaneously reducing carbon dioxide emissions[2]. A wide array of physical modeling methods can be employed for wind power forecasting, including downscaling models and wind-to-power conversion models[3]. These modeling processes are complex, computationally expensive, and challenging. Statistical approaches predominantly investigate the statistical distribution attributes of historical dataset. The observed wind power displays a non-linear association with historical wind power. The primary methodologies encompass ARMA model and artificial neural network model.

With the advancement of technology, the field of wind power forecasting has seen widespread utilization of AI and deep learning techniques[4]. Incomplete or insufficient grid operational data caused by grid regulation and equipment failures can lead to poor performance in wind power forecasting[5]. In reference [6], the isolation forest algorithm was employed for anomaly detection in wind power datasets. While this method is simple, it tends to misclassify many normal values as anomalies. In reference [7], The K-means algorithm was utilized to detect anomalies in wind turbine data. However, the determination of the optimal number and the precise values of cluster centers heavily influence outcomes.

To address the aforementioned challenges, This study introduces a prediction method based on TPA mechanism and MBLSTM network. First, the combination of DBSCAN and linear regression algorithm is introduced to detect Outlier, and the abnormal data is reconstructed by KNN interpolation method; Secondly, the MBLSTM network is used to extract the forward and reverse characteristics of wind power data, and the TPA mechanism is combined to further extract the complex links between time series data, thus enhancing the accuracy of predictions. Ultimately, experiments were conducted to validate the efficacy and precision of TPA-MBLSTM model outlined in this research paper.

II. DETECTION AND HANDLING OF ANOMALOUS VALUES IN WIND POWER DATA

Existing research indicates that wind speed and power exhibit an approximate cubic relationship. However, the presence of numerous outliers in the actual dataset results in a discrete nature of the wind speed-power characteristics. Accurate identification and removal of these abnormal data points can significantly enhance the precision and dependability of wind power prediction.

This research was supported by the Science and Technology Project of Yunnan Power Grid Co., Ltd. （YNKJXM20220039）

A. Data Anomaly Detection

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) offers broader applicability compared to methods like K-means, which are limited to convex sample sets. DBSCAN does not require manual determination of the number of clusters and is capable of clustering data of arbitrary shapes. It can effectively filter out noise and outliers, making it a versatile and robust approach for data clustering and anomaly detection.

Assuming the sample set is $D=(x_1,x_2,\ldots,x_n)$, the definition of density-based clustering using DBSCAN can be described as follows:

Neighborhood: The region centered around a selected target object with a radius ε is considered the neighborhood of that target object.

Core Point: A target point is referred to as a core point if the number of samples within its neighborhood exceeds $minPts$.

Directly Density-Reachable：In a sample set D, point m is considered directly density reachable from point n if it falls within the vicinity of a core point.

Density-Reachable：In sample set D, for any given points x_i and x_j, if there exists points $p_1,p_2,\ldots,p_t$ such that $p_1=x_i$, $p_t=x_j$, and point p_{t+1} is directly density reachable from point p_t, then x_j is considered density reachable from x_i.

Density-Connectivity：In dataset D, if there exists a point o such that both point m and point n are density reachable from, then points m and n are considered density connected.

DBSCAN clustering is a process of forming a set of dense connected components by extending the density-reachable relationship. By utilizing two parameters, namely the neighborhood radius ε and the minimum number of samples $minPts$, the sample set can be divided into three categories: core points, boundary points, and outliers. Within a neighborhood radius and when the number of sample points exceeds $minPts$, these points are defined as core points. Within the neighborhood radius but with a lower count of sample points than $minPts$, these points are defined as boundary points, while the remaining sample points are considered outliers.

Because of the approximately cubic correlation between wind velocity and power. this study employs a linear regression algorithm for supplementary detection of anomalous data in order to accurately identify such data. Firstly, a linear regression model is constructed to fit the power and wind speed data. Subsequently, the trained linear regression model is utilized to predict wind speed values based on the input power values. The deviation between the predicted values and the true values is computed, and a threshold is set. If the actual deviation exceeds the threshold, the data point is classified as anomalous.

B. Reconstruction of Anomalous Data

Handling anomalous data typically involves deleting some anomalous and missing data when the dataset is sufficiently large. However, This methodology may compromise the data integrity and diminish the temporal coherence of the original dataset, thus affecting the accuracy of subsequent modeling. By reconstructing the deleted anomalous and missing data, the integrity and temporal continuity of the data can be preserved. K-nearest neighbors (KNN) is a method that utilizes the average or median values of neighboring samples to impute the missing values in anomalous data. The specific process be outlined as follows:

First, calculate the distance d_i between the sample and the target point using Equation (1)：

$$d_i=\sqrt{\sum_{i=1}^{n}(\mathrm{c}_x-\mathrm{c}_i)^2} \tag{1}$$

$$\omega_i=\frac{1}{d_i} \tag{2}$$

c_x represents the target sample, c_i represents the neighboring sample, ω_i represents the sample weight.

By calculating the distances between different samples, the distance matrix D can be obtained. Then, the weight matrix is computed, and finally, the estimated value g of the target sample is calculated using Equation 3.

$$g=\sum_{i=1}^{n}\omega_i\cdot x_i \tag{3}$$

Where x_i represents the corresponding numerical values of neighboring samples.

III. TPA-MBLSTM Prediction Model

A. Multi-layer Bidirectional LSTM Model（MBLSTM）

Due to the limitations of the unidirectional nature of LSTM neural networks, they can only store and process information from the past and current time steps, which can result in cumulative errors in prediction. Generally, the use of bidirectional LSTM neural networks (BiLSTM) overcomes this limitation by simultaneously utilizing information from both past and future time steps. BiLSTM networks consist of both forward and backward LSTM units, which enhances memory training. Deep LSTM neural networks further capture intricate data features and ensure data accuracy.

The BiLSTM networks are utilized to capture the temporal information of the input sequence from past and future time steps for forward and backward training. The mathematical model of the BiLSTM neural network is as in equations (4).

$$\begin{cases} \vec{Q}_t = LSTM(m_t, \vec{Q}_{t-1}) \\ \overleftarrow{Q}_t = LSTM(m_t, \overleftarrow{Q}_{t-1}) \\ Q_t = [\vec{Q}_t, \overleftarrow{Q}_t] \end{cases} \tag{4}$$

Where m_t represents the input at time step t, $\vec{Q}_{t-1}$ represents the output of the forward LSTM hidden layer at time step $t-1$, $\overleftarrow{Q}_{t-1}$ represents the output of the backward LSTM hidden layer at time step $t-1$, and Q_t represents the outputs of hidden layer.

The multi-layer stacked bidirectional LSTM is a process of continuously expanding the depth of the BiLSTM network, wherein the neural network undergoes multiple training iterations to improve forecasting precision. The structure of MBLSTM is depicted in Fig 1.

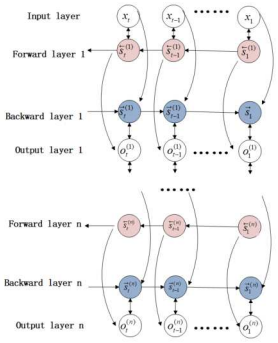

Fig. 1. MBLSTM network structure

B. Temporal Pattern Attention（TPA）

Traditional attention mechanisms primarily concentrate on extracting information pertinent to the present time step, which works well for data with a single time series. In wind power forecasting, each time step consists of multiple variables., including wind power time series and weather factors. Due to the diverse and complex nonlinear relationships among these variables, traditional attention mechanisms, which average information across multiple time steps, struggle to effectively capture the importance of a specific time step. The TPA mechanism utilizes the frequency domain information of time series to extract intrinsic temporal characteristics from the data. It can effectively uncover the inherent relationships between variables and the wind power time series. The architecture of TPA is depicted in Fig 2. h represents the output information of the previous-level neural network hidden layer, where each information represents the temporal state of the previous-level network. Then, multiple convolutional layer filters are used to extract features.

Assuming the hidden layer information generated by the MBLSTM network is $H = \{h_{t-\omega}, h_{t-\omega+1}, \ldots, h_{t-1}\}$, h_t represents the hidden layer information at time step t, ω denotes the length of the time sliding window used to select a certain length of time data, and T represents the maximum length of weights extracted by the filters, which is typically equal to the sliding window length. The original hidden layer information H is convolved to produce an $n*k$ dimensional matrix $H_{i,j}^C$.

$$H_{i,j}^C = \sum_{l=1}^{\omega} H_{i,(t-\omega-1+l)} \times C_{j,T-\omega+l} \tag{5}$$

C represents the filter, $H_{i,j}^C$ denotes the i row vector passing through the j length feature extracted by filter C_j. In $H_{i,(t-\omega-1+l)}$, i represents the row vector, t represents time, ω denotes the filter length, and T denotes the maximum size of weights derived from the filter., typically taken as ω.

The weight α_i is calculated by computing the hidden layer function h_t with H_i^C, representing the strength of influence of each row in H_i^C on the hidden layer function h_t. The weight coefficient W_α for $\mathrm{m}\times\mathrm{k}$ is denoted as in equations (6) and (7) as follows:

$$f\left(H_i^C, h_t\right) = \left(H_i^C\right)^T W_\alpha h_t \tag{6}$$

$$\alpha_i = sigmoid\left(f\left(H_i^C, h_t\right)\right) \tag{7}$$

Where H_i^C represents the vector in the i row of H^C.

The weighted summation of H_i^C is performed, where v_t represents the cumulative impact of all rows in H_i^C on h_t, i.e., the temporal influence.

$$v_t = \sum_{i=1}^{m} \alpha_i H_i^C \tag{8}$$

m represents dimensionality of the input variables.

Finally, considering the overall influence of the temporal impact v_t, the final predicted value $h_t^{'}$ is obtained as follows:

$$h_t^{'} = W_h h_t + W_v v_t \tag{9}$$

Where W_v and W_{h} are matrix parameter weights.

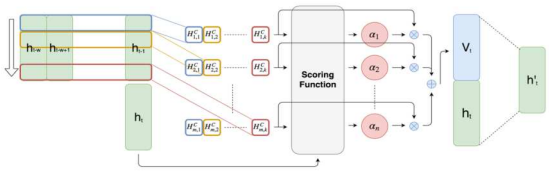

Fig. 2. Schematic diagram of attention mechanism in time mode

C. TPA-MBLSTM Prediction Model

This study introduces the TPA-MBLSTM prediction model, which utilizes a MBLSTM network to capture long and short-term temporal features from the data. The TPA mechanism is employed to extract crucial information and improve prediction performance. The workflow of the TPA-MBLSTM framework is depicted in Fig 3.

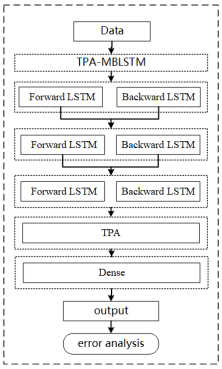

Fig. 3. Forecast flowchart

IV. Example Analysis

A. Data Description

The dataset utilized in this research was collected from a wind farm situated in western China. The dataset contains operational data from 2018 to 2019, with a sampling frequency of 15 minutes per data point, resulting in 96 data points per day. The dataset comprises various meteorological parameters, including air density, temperature, wind speed, wind direction, humidity, and atmospheric pressure.

Six variables, namely air density, temperature, wind speed, wind direction, humidity, and atmospheric pressure, were selected for calculating the Pearson correlation coefficient and significance level (p-value). The findings are displayed in Table I.

In Table I, it can be inferred that wind speed is highly correlated with power generation, followed by air density and wind direction. Atmospheric pressure, temperature, and humidity show weaker correlations with power generation. Therefore, in this study, wind speed, the sine and cosine of wind direction, and air density are selected as influential factors for wind power generation for further investigation.

TABLE I. CORRELATION COEFFICIENT

Meteorological parameters	The Pearson correlation coefficient	P-Value
Wind Speed	0.71	0.00
Wind direction	0.21	3.11e-215
Air density	0.02	3.9 e-264
Temperature	-0.13	0.00
Humidity	-0.16	0.00
Atmospheric Pressure	-0.04	1.48 e-38

B. Detection and Handling of Data Outliers

In this study, a combination of the DBSCAN clustering algorithm and linear fitting method was used to perform data cleaning. Firstly, the scattered points in the wind speed-power scatter plot were detected using the DBSCAN density-based clustering method for outlier detection. DBSCAN has two important parameters, namely the neighborhood radius ε and the minimum number of samples $minPts$, which are generally set manually. In this study, ε was set to 0.2 and $minPts$ was set to 600. The experimental outcomes are depicted in Fig 4, where a few sample groups were identified as outliers after clustering. However, due to the high density of some outlier data points and their connection to the normal wind speed-power curve, it was difficult to differentiate them by manually adjusting ε and $minPts$. Therefore, a linear regression algorithm was applied to complementarily identify the outliers. By applying a linear model to the wind speed and power data, and evaluating the deviation between the predicted values and the true values, points with errors larger than a threshold were considered as outliers. The experimental results are depicted in Fig 5. This combined approach accurately identifies a significant quantity of outliers in the dataset that exhibit a discrete pattern in the wind speed-power characteristics.

To guarantee the temporal consistency of data after removing outliers, KNN algorithm is employed to fill in the missing data by selecting neighboring samples and using the mean or median of their values.

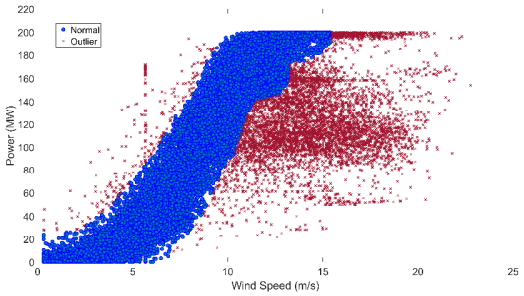

Fig. 4. DBSCAN density clustering

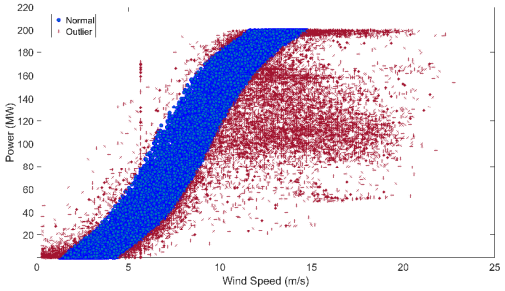

Fig. 5. linear regression

To validate the effectiveness of the outlier detection and processing methods, the original dataset and the dataset after outlier handling are compared in terms of prediction using the LSTM algorithm. Prior to the experiments, the experimental data are divided, with the data from 2018 being allocated as the training set , while that from 2019 is assigned as the testing set. The prediction errors obtained from the experiments are showcased in Table II. It can be observed that, after undergoing data cleaning through the combination of DBSCAN clustering algorithm and linear fitting method, both the mean absolute error (MAE) and root mean square error (RMSE) have decreased, indicating improved prediction performance compared to the original dataset without outlier preprocessing.

TABLE II. C OMPARISON OF PREDICTION ERROR BETWEEN ORIGINAL DATA SET AND PROCESSED DATA SET

Dataset	MAE/MW	RMSE/MW

Raw dataset	4.573	9.590
DBSCAN processed dataset	3.803	7.412
DBSCAN and linear regression processed dataset	3.711	7.126

C. *Results and Analysis of Predictive Models*

To validate the precision of the algorithm proposed in this study LSTM, BiLSTM, and TPA-LSTM models were selected for experimental comparison. The test set consists of two representative daily datasets, one from summer and one from winter, selected from the original dataset. Fig 6-7 display the predicted outcomes.

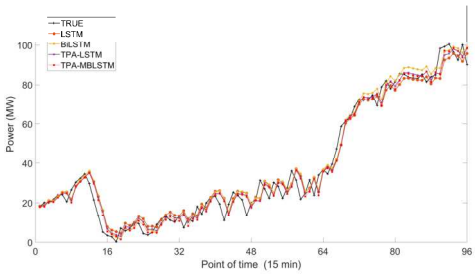

Fig. 6. Comparative results (Summer)

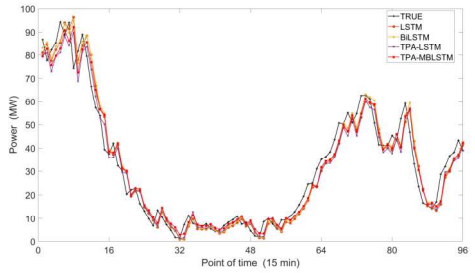

Fig. 7. Comparative results (Winter)

TABLE III. PREDICTION ERROR COMPARISON

Test set	Error/MW	LSTM	BiLSTM	TPA-LSTM	TPA-MBLSTM
Summer	RMSE	5.167	5.046	4.856	4.796
	MAE	4.020	3.991	3.827	3.759
Winter	RMSE	5.905	5.738	5.707	5.571
	MAE	4.586	4.419	4.399	4.380

From Fig 6-7, the TPA-MBLSTM model has higher accuracy and predictive capability for different climates and wind turbine operation modes. From Table III, From the results presented, the method described in this article achieves RMSE and MAE values of 4.796 MW, 5.571 MW, 3.759 MW, and 4.380 MW on two representative days during summer and winter, respectively. Especially during typical summer days, compared to LSTM, BiLSTM, and TPA-LSTM, The TPA-MBLSTM model reduced RMSE by 7.18%, 4.95%, and 1.24%, respectively, and MAE by 6.49%, 5.81%, and 1.78%, respectively. Other data are shown in Table III.

Based on the prediction process of the model in Fig 6-7, it becomes apparent that a single neural network has a significant prediction accuracy error when facing severe changes in wind speed. By mining the temporal correlation of wind speed and the coupling between wind speed and power through the TPA mechanism, and utilizing the nonlinear features between layers and the deep hidden association learning ability of MBLSTM network, we can quickly capture the relevant features of drastic changes in meteorological factors. In addition, from Fig 6-7, it can be seen that all models can effectively predict wind power during periods of mild power fluctuations. The traditional LSTM model cannot adapt to rapid changes in power under severe power fluctuations, and its predicted results are smaller than the actual power, resulting in significant deviations.

Furthermore, from Table III, compared to LSTM model, the RMSE index of the TPA-LSTM model in typical days decreased by 6.02% and 3.35%, respectively, while the MAE index decreased by 4.80% and 4.08%, respectively. This is due to the fact that the TPA mechanism conducts feature extraction on the hidden state matrix, identifying significant information from various time steps and assigning weights to the temporal pattern matrix. Therefore, it can perform well in predicting various time periods. The MBLSTM model repeatedly trains input data, increases model's Data dependency and system robustness. At the same time, the TPA mechanism can capture more important historical information in the time series.

V. CONCLUSION

In this study, an approach for ultra-short-term wind power prediction is proposed, which integrates MBLSTM neural network with TPA mechanism. This method comprehensively investigates the intrinsic correlation between multiple features and wind power, and utilizes the TPA mechanism and MBLSTM neural network to establish a coupling relationship between wind power and wind speed, as well as historical wind power data, in order to improve prediction accuracy; In addition, for the abnormal training data, DBSCAN and linear fitting methods are used to accurately identify the Outlier of training data. The efficacy and validity of the proposed method have been demonstrated through experimental validation.

REFERENCES

[1] Han L, Jing H, Zhang R, et al. Wind power forecast based on improved Long Short Term Memory network. Energy 2019;189:116300.

[2] Wu Z, Luo G, Yang Z, et al. A comprehensive review on deep learning approaches in wind forecasting applications[J]. CAAI Transactions on Intelligence Technology, 2022, 7(2): 129-143.

[3] Feng Shuanglei, Wang Weisheng, Liu Chun, et al. Study on the physical approach to wind power prediction. Proceedings of the CSEE, 2010,(2):1-6.

[4] YANG Hai-min, PAN Zhi-song and BAI Wei. Review of Time Series Prediction Methods[J].Computer Science,2019,46(1):21-28.

[5] Yang Mao, Yang Qiongqiong. Review of modeling of wind speed-power characteristic curve for wind turbine[J]. Electric Power Automation Equip-ment,2018,38(2):34-43.

[6] Kisvari A, Lin Z, Liu X. Wind power forecasting–A data-driven method along with gated recurrent neural network[J]. Renewable Energy, 2021, 163: 1895-1909.

[7] Kusiak A, Verma A. Monitoring wind farms with performance curves[J]. IEEE transactions on sustainable energy, 2012, 4(1): 192-199.