Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-6, Issue-8, August 2024
DOI: 10.53469/jrse.2024.06(08).04

Ultra-Short-Term Wind Power Forecasting Based on VMD-GRU-Transformer

Wei Liu*, Xinfu Liu

School of Mechanical and Automotive Engineering, Qingdao University of Technology, Qingdao 266520, Shandong, China
**Correspondence Author, 2065920268@qq.com*

Abstract: ***Accurate wind power prediction is essential for the stable operation of power systems. Aiming at the problem of insufficient accuracy of ultra-short-term wind power prediction, a combined prediction model based on VMD-GRU-Transformer is proposed. Variational Mode Decomposition (VMD) is used to split the wind power data into different intrinsic mode functions (IMFs) to weaken the non-stationarity of the original series. The combined GRU-Transformer network structure is designed to utilize gated recurrent unit (GRU) instead of the original word embedding and positional coding links, and feature fusion is performed on the input data to fill in the gaps in Transformer where the relevant information is not fully considered. Relying on the self-attention mechanism in Transformer to capture the time dependence of sequence data for prediction. Finally, a case analysis is performed with a public dataset, and the results show that the proposed model has higher prediction accuracy compared to other existing models.***

Keywords: Wind power forecasting, Variational Mode Decomposition, GRU, Transformer**.**

1. Introduction

As a new type of clean energy, wind power has significant environmental and sustainable advantages. The rapid advancement of wind power projects is highly important in mitigating global warming and achieving the transition of energy infrastructure. However, due to the rapid growth of the wind power sector, as well as the randomness and volatility characteristics of wind energy, the centralized grid connection of large-scale wind power is bound to impact the secure and steady operation of local power system [1]. Therefore, accurate and reliable ultra-short-term wind power prediction is particularly important to enhance the capacity of the receiving end of the grid to utilize wind power and promote safe and economic operation of the grid [2].

Presently, wind power prediction approaches that are widely employed can be categorized into three primary groups: physical models, statistical models, and machine learning models [3]. Physical and statistical models, although mature in their methodological development, have insufficient coping ability in the face of complex environmental changes in wind power prediction, resulting in unsatisfactory prediction results. Machine learning methods, in contrast to the first two models, depend on data-driven approaches to discover the inherent relationships between data. These methods have found extensive application in the domain of wind power prediction. In addition, in the face of the strong non-linearity and non-stationarity characteristics of wind power, the algorithms can be combined with each other to form an integrated model or deep learning structure to improve the overall performance.

Gao *et al.* [4] propose a CNN-GRU model, which effectively improves prediction accuracy by extracting global and local features from the input multichannel signals via Convolutional Neural Network (CNN) and combining these features with those obtained in the convolution process. However, the method lacks the processing of randomness and volatility signals in power data and has insufficient coping ability and generalization ability when facing complex data. Zhang *et al.* [5] used VMD to decompose the raw power into multiple modal components, and then optimized the hyperparameters of the TCN-BIGRU model through sparrow search algorithm (SSA), and applied the optimized model to power prediction, which effectively suppressed the volatility of wind power. However, the model accumulates predictions for each component, which increases the computational burden of the model significantly. Zhang *et al.* [6] propose a CGAN-CNN-LSTM model that utilizes Conditional Generative Adversarial Networks (CGAN) to complete the missing parts of the dataset. A combined CNN-LSTM model is constructed for feature extraction, after which the attention mechanism is applied to assign weights to the features to accelerate model convergence. The model has good prediction and generalization ability. However, the CNN model is not flexible enough to handle time series data, and the feature relationships are not captured well enough.

In summary, a VMD-GRU-Transformer method for ultra-short-term forecasting of wind power is proposed. VMD decomposes the power series data into several smooth components, which capture the non-linearity and non-stationarity in the data characterized by changes at different frequencies. Combining the information on wind speed and historical wind power compose the input features of the model. The combined GRU-Transformer prediction model is constructed, utilizing GRU instead of the original word embedding and positional coding links to fuse the features of the input data, which significantly improves the ultra-short-term prediction accuracy of wind power. The public dataset is used to analyze the arithmetic cases and compared with the other five existing models for validation.

2. Variational Mode Decomposition (VMD)

VMD is a fully non-recursive adaptive modal decomposition and signal processing technique designed to decompose complex original signals into multiple smoother IMFs, effectively addressing challenges such as signal feature extraction difficulties [7]. Compared to other existing modal decomposition techniques, VMD is more suited for dealing with strong non-linearity and non-stationarity in wind power, reducing modal aliasing and enhancing prediction accuracy. The decomposition process of the VMD model involves the following steps:

(1) Construct the constrained variational problem as follows:

$$F = \min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{\mathrm{j}}{\pi t} \right) * u_k(t) \right] \mathrm{e}^{-\mathrm{j}\omega_k t} \right\|_2^2 \right\} \qquad (1)$$
$$\text{s.t.} \sum_{k=1}^{K} u_k(t) = f(t)$$

where K is the number of modal decompositions set in advance; $f(t)$ is the original signal sequence; $\delta(t)$ is the Dirac distribution; * is the convolution operator; $\{u_k(t)\}$ is the IMF component; $\{\omega_k\}$ is the center frequency of each IMF component;

(2) Lagrange Multiplier λ(t) and penalty factor α are introduced to turn the constrained variational problem into an unconstrained variational problem, and the extended Lagrangian function is expressed as:

$$L(\{u_k(t)\},\{\omega_k\},\lambda) =$$
$$\alpha \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{\mathrm{j}}{\pi t} \right) * u_k(t) \right] \mathrm{e}^{-\mathrm{j}\omega_k t} \right\|_2^2 + \qquad (2)$$
$$\left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle$$

(3) The IMF component {uk(t)} and center frequency {ωk} are updated using the alternating direction multiplier method, yielding the final decomposition result.

3. Forecasting Model Combing GRU and Transformer

3.1 Gated Recurrent Unit

GRU is an improvement of recurrent neural network (RNN), which can effectively solve the problem of gradient vanishing. GRU controls the updating and extraction of the input information and its relative position through the structure of the memory gate [8]. Relative to the Long Short-Term Memory (LSTM) network structure is more concise, has fewer parameters to be optimized, and has a faster convergence speed. The model structure is shown in Figure 1.

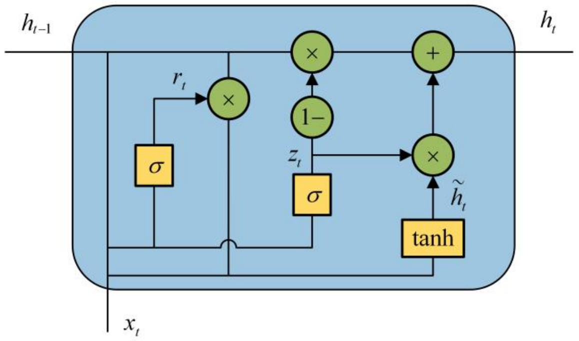

Figure 1: The structure of GRU model

The z_t and r_t denote the update gate and reset gate, respectively. The update gate quantifies the amount of pertinent information from both the previous and current time steps that should be transmitted to the next step, while the reset gate regulates the extent to which prior information should be disregarded. The GRU network uses the following formula for forward propagation:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \qquad (3)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \qquad (4)$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t \odot h_{t-1}, x_t]) \qquad (5)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \qquad (6)$$

where h_t and are the hidden states of the unit to be updated and the next node at moment t, respectively; x_t is the input at moment t; $\boldsymbol{W}$ is weight matrice; σ represents the sigmoid activation function; and ⊙ is the Hadamard product.

3.2 GRU-Transformer Neural Network

The GRU-Transformer combinatorial model consists of an input layer, N Transformer coding layers, and an output layer. Each layer of the encoder in Transformer uses residual connections to improve the model fitting ability and convergence speed. The input layer of the traditional Transformer consists of word embedding and positional coding links, which are designed to better handle natural language-like problems. In this paper, GRU is used instead of the input layer of the traditional Transformer model to feature process the input vectors and improve the loss of temporal positional information of Transformer to better target the wind power prediction problem. The Transformer coding layer is used to compute the self-attentive expression of the input feature vectors, and the output layer is responsible for outputting the prediction results of the model. The model structure is shown in Figure 2.

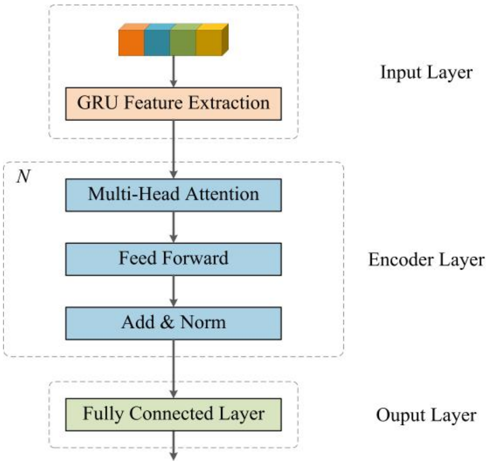

Figure 2: The structure of GRU-Transformer model

Transformer neural network has been gaining attention in dealing with time series forecasting research because of their long-term dependency and interaction with time series data [9]. Transformer is a model based on the attention mechanism. The coding component of Transformer consists of a multilayer encoder, and each encoder layer consists of a multi-head self-attention layer and a feed-forward Network layer. The formula for the self-attention mechanism is

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{softmax}\left(\frac{\boldsymbol{QK}}{\sqrt{n}}\right)\boldsymbol{V} \qquad (7)$$

$$\boldsymbol{Q} = \boldsymbol{HW}_Q, \boldsymbol{K} = \boldsymbol{HW}_K, \boldsymbol{V} = \boldsymbol{HW}_V \qquad (8)$$

where: n is the dimension of the $\boldsymbol{Q}$ and $\boldsymbol{K}$ matrices; $\boldsymbol{Q}$ is the query matrix; $\boldsymbol{K}$ is the key-value matrix; $\boldsymbol{V}$ is the value matrix; $\boldsymbol{H}$ is the input matrix; and $\boldsymbol{W}$ is the weight matrices.

The output of the Multi-head Attention mechanism layer is obtained by concatenating several Attention outputs and then applying a linear transformation. The mathematical expression for the Multi-head Attention mechanism is:

$$\text{MultiHead} = \text{Concat}(\boldsymbol{h}_1, \cdots, \boldsymbol{h}_h)\boldsymbol{W}_o \tag{9}$$

$$\boldsymbol{h}_h = \text{Attention}(\boldsymbol{QW}_{Qh}, \boldsymbol{KW}_{Kh}, \boldsymbol{VW}_{Vh}) \tag{10}$$

where W_o is the matrix of linear transformation coefficients; Concat denotes multiple matrix splicing, $\boldsymbol{W}_{Qh}$, $\boldsymbol{W}_{Kh}$, and $\boldsymbol{W}_{Vh}$ are the parameter matrices for performing the linear transformation; the subscript h is the number of heads of the multi-head attention mechanism.

The Feed Forward neural network comprises two fully connected layers. The activation function of the first layer is Rectified Linear Unit (ReLU), while the second layer employs a linear activation function with the given expression.

$$F(\boldsymbol{H}) = \max(0, \boldsymbol{HW}_1 + \boldsymbol{b}_1)\boldsymbol{W}_2 + \boldsymbol{b}_2 \tag{11}$$

where $\boldsymbol{W}$ is weight matrice, and $\boldsymbol{b}$ is bias matrice.

The output layer consists of fully connected layers and is computed as

$$Y = \boldsymbol{W}^o \boldsymbol{X}_D + \boldsymbol{b}^o \tag{12}$$

where: $\boldsymbol{Y}$ is the output result of the output layer; $\boldsymbol{X_D}$ is the output result of the coding layer result; $\boldsymbol{W}^o$, b^o are the full connection layer parameter matrix and bias.

3.3 Evaluation Metrics

This research chooses the following three indicators to precisely evaluate the effectiveness of the wind power forecast model: Root Mean Square Error (E_{RMSE}), Mean Absolute Error (E_{MAE}), and Coefficient of Determination (R^2). They are calculated as follows:

$$R^2 = 1 - \frac{\sum_{t=1}^{\rho}(\hat{y}_t - y_t)^2}{\sum_{t=1}^{\rho}(\overline{y}_t - y_t)^2} \tag{13}$$

$$E_{\text{RMSE}} = \sqrt{\frac{\sum_{t=1}^{\rho}\left(y_t - \hat{y}_t\right)^2}{\rho}} \tag{14}$$

$$E_{\text{MAE}} = \sum_{t=1}^{\rho}\frac{\left|y_t - \hat{y}_t\right|}{\rho} \tag{15}$$

where y_t and be the actual and predicted values at time t, represents the average value of y_t, and ρ denotes the length of the predicted data.

4. Case Studies

The National Grid Renewable Energy Generation Forecasting Competition data provided by the literature [10] is used as the research object to validate the model and method proposed in this paper. The output capacity of the wind farm is 99 MW, and the turbine models are GW1500/85 and H93 L-2.0 mw. The dataset used was gathered between January 16, 2019, and March 16, 2020, with data collected every 15 minutes. The goal is to predict wind power for the next 15 minutes in advance. 80% of the dataset is allocated for training, while the remaining 20% is divided equally between validation and testing, with a 1:1 ratio for each.

4.1 VMD Decomposition of Wind Power Series

The wind power data information is shown in Figure 3. According to its changes, it can be seen that the power of wind power has a cyclic cycle characteristic, and with the advancement of the time dimension, the data information has a certain degree of evolutionary characteristics. Simultaneously, as a result of the fluctuating and unpredictable wind speed, the data includes certain elements of randomness.

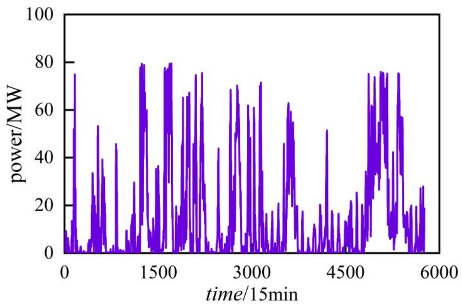

Figure 3: Wind power data information

In order to accurately analyze the variation of wind power and to weaken the negative effects of nonlinear and nonsmooth features, the sequence is modally decomposed using VMD to improve the prediction accuracy of wind power. The modal number K=9 is chosen and the decomposition results are shown in Figure 4.

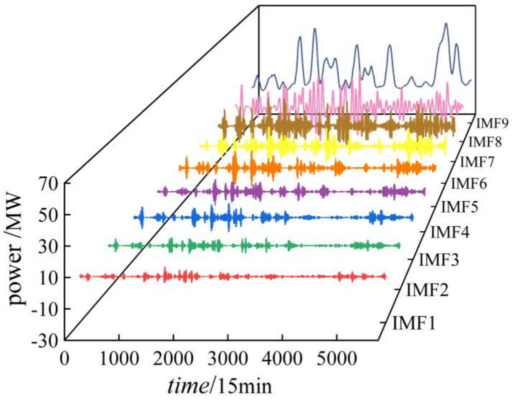

Figure 4: Wind power VMD decomposition results

4.2 Construction of Input Features

The dataset includes meteorological and calendar information such as wind speed, wind direction, and day. When selecting input characteristics for wind farms, it is important to avoid using data that has little correlation with the output target. Alternatively, it will increase the computational burden of the model and affect the training efficiency of the model and the prediction accuracy. Hence, this article uses the Pearson correlation coefficient to examine the link between various parameters. The correlation heat map is shown in Figure 5.

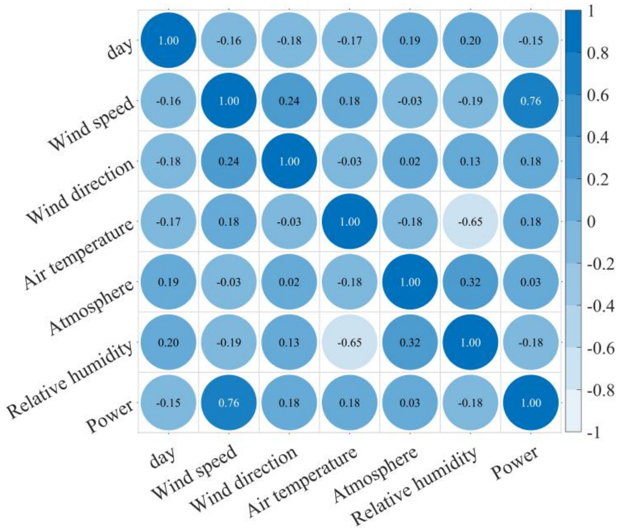

Figure 5: Heat map of Pearson correlation coefficient for wind power

Table 1: Input and output features for the forecasting model

Input feature	*Output feature*
Wind power IMFs in the previous three hours	Wind power to be forecast at the moment
Wind speeds at different heights in the previous three hours	

Considering the high correlation between data with correlation coefficients of 0.5 and above in absolute value, wind speed and power were selected as input features. Among them, the wind speed contains data related to the heights of 10 meters, 30 meters, 50 meters, and wheel hubs. The wind speed at different heights and the VMD decomposition result of wind power together form the input features of the model. The corresponding parameter settings are shown in Table 1

4.3 Comparison of Different Forecasting Models

To validate the accuracy of the proposed prediction model, we conducted wing power forecasting and comparative analysis using existing individual models such as BiLSTM, GRU, and TCN, as well as existing hybrid models like CNN-BiLSTM, in comparison to the proposed VMD-GRU-Transformer hybrid model. Figure 6 shows the prediction results of the different models and their local scaling plots. The green curve represents the forecast made by the VMD-GRU-Transformer model, while the red curve represents the actual measured wind power value. The VMD-GRU-Transformer model has superior accuracy in predicting test results compared to other existing models.

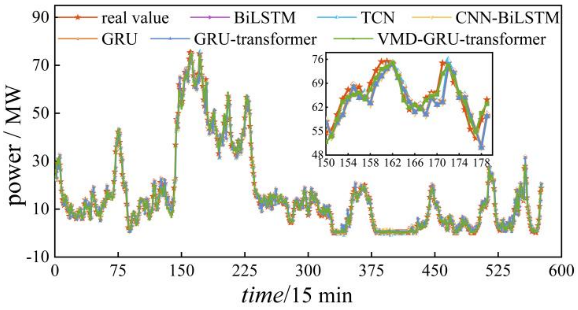

Figure 6: Wind power forecasting results of different models

The evaluation metrics of the prediction results of different models on the test set are shown in Table 2. The R^2, E_{MAE} and E_{RMSE} metrics of the prediction results of the combined VMD-GRU-Transformer model are 0.9971, 0.7005 MW, and 0.9391 MW, respectively, which are better than the prediction results of any one of the existing single models such as BiLSTM, GRU and TCN. Moreover, compared with the combined GRU-TRransformer and CNN-BiLSTM models, the R^2 metrics of the proposed model are improved by 1.7552% and 1.7345%, the E_{MAE} metrics are reduced by 0.9920 MW and 0.9928 MW, and the E_{RMSE} metrics are decreased by 1.5146 MW and 1.5019 MW, and the prediction accuracies are significantly improved. To summarize, the VMD-GRU-Transformer model integrates the strengths of the VMD and GRU-Transformer models, with more accurate prediction results and better applicability in wind power prediction.

Table 2: Metrics of different forecasting models for wind power

Prediction Model	R^2	E_{MAE}/MW	E_{RMSE}/MW
GRU	0.9796	1.7150	2.4724
BiLSTM	0.9799	1.7052	2.4575
GRU-TRransformer	0.9799	1.6925	2.4537
TCN	0.9800	1.6948	2.4484
CNN-BiLSM	0.9801	1.6933	2.4410
VMD-GRU-Transformer	0.9971	0.7005	0.9391

5. Conclusion

Upon analysis of the acquired findings, the following conclusions can be proposed:

(1) Considering the non-linear and non-stationary characteristics of wind power, and combined with wind speed and other meteorological factors, an ultra-short-term prediction model of wind power based on the combination of VMD-GRU-Transformer is proposed.

(2) Construct a mathematical model of VMD decomposition of wind power, and use the variational modal decomposition to decompose the wind power sequence into different IMFs, which effectively extracts the characteristics of sequence changes and weakens the complexity of the power sequence.

(3) Design the network structure of the combined GRU and Transformer wind power prediction model, and perform feature fusion of the input sequences through GRU as the input embedding layer to fill the gap of Transformer that does not fully consider the relevant information. The Transformer utilizes the self-attention mechanism to capture the temporal dependencies in sequence data, hence enhancing prediction accuracy.

(4) The evaluation metrics for the prediction results of the combined VMD-GRU-Transformer model are superior to those of existing single models like BiLSTM, GRU, and TCN, as well as combined models like CNN-BiLSTM. In the future, the VMD-GRU-Transformer model can be practically applied to wind power prediction by combining various application scenarios based on site requirements. Our goal is to enhance the precision of ultra-short-term wind power forecasting in wind farms.

References

[1] Wang Y, Zou R, Liu F, et al. A review of wind speed and wind power forecasting with deep neural networks[J]. Applied Energy, 2021, 304: 117766. https://doi.org/10.1016/j.apenergy.2021.117766.

[2] Zha W, Liu J, Li Y, et al. Ultra-short-term power forecast method for the wind farm based on feature selection and temporal convolution network[J]. ISA transactions, 2022, 129: 405-414. https://doi.org/10.1016/j.isatra.2022.01.024.

[3] Xiao Y, Zou C, Chi H, et al. Boosted GRU model for short-term forecasting of wind power with feature-weighted principal component analysis[J]. Energy, 2023, 267: 126503. https://doi.org/10.1016/j.energy.2022.126503.

[4] Gao J, Ye X, Lei X, et al. A multichannel-based cnn and gru method for short-term wind power prediction[J]. Electronics, 2023, 12(21): 4479. https://doi.org/10.3390/electronics12214479.

[5] Zhang Y, Zhang L, Sun D, et al. Short-term wind power forecasting based on VMD and a hybrid SSA-TCN-BiGRU network [J]. Applied Sciences, 2023, 13(17): 9888. https://doi.org/10.3390/app13179888.

[6] Zhang J, Zhao Z, Yan J, et al. Ultra-short-term wind power forecasting based on CGAN-CNN-LSTM model supported by lidar[J]. Sensors, 2023, 23(9): 4369. https://doi.org/10.3390/s23094369.

[7] Dragomiretskiy K, Zosso D. Variational mode decomposition [J]. IEEE transactions on signal processing, 2013, 62(3): 531-544. https://doi.org/10.1109/ TSP.2013.2288675.

[8] Zhao Z, Yun S, Jia L, et al. Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features[J]. Engineering Applications of Artificial Intelligence, 2023, 121: 105982. https://doi.org/10.1016/j.engappai.2023.105982.

[9] Sun S, Liu Y, Li Q, et al. Short-term multi-step wind power forecasting based on spatio-temporal correlations and transformer neural networks[J]. Energy Conversion and Management, 2023, 283: 116916. https://doi.org/10.1016/j.enconman.2023.116916.

[10] Chen Y, Xu J. Solar and wind power data from the Chinese state grid renewable energy generation forecasting competition[J]. Scientific Data, 2022, 9(1): 577. https://doi.org/10.1038/s41597-022-01696-6