International Journal of Electric Power and Energy Studies
ISSN: 3006-2004 (Print), ISSN: 3006-0826 (Online) | Volume 2, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijepes.v2n1.05
Journal homepage: https://wepub.org/index.php/IJEPES/index

Short-term Wind Power Forecasts based on VMD-KPCA and CFSBOA-BiLSTM

Rui Wang[1, a], Yingying Dong[1, b], Jing Lu[2, c]

[1] School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China

[2] School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China

[a]wangrui@hpu.edu.cn, [b]1244557330@qq.com, [c]lujing@hpu.edu.cn

ABSTRACT

Accurate prediction of wind power is of great significance to reduce the impact of large-scale wind power connection and improve the security and stability of power grid. A short-term wind power forecasting method based on VMD-KPCA and CFSBOA-BiLSTM is proposed. Firstly, the key environmental factors restricting the volatility of wind power are fully considered, and the variational mode decomposition method is used to decompose the wind power and meteorological factors. Secondly, the kernel principal component analysis method is used to reduce the dimension and construct the sub-sequence characteristic data set. Finally, the multi-strategy improved butterfly algorithm is used to optimize the BiLSTM network hyperparameters and establish the CFSBOA-BiLSTM prediction model. The simulation analysis uses the measured data of a wind farm in northwest China. The experimental results show that the combined forecasting model can fully mine the hidden information of the data and improve the accuracy of short-term wind power forecasting. Compared with the BiLSTM model, RMSE, MAE and MAPE decreased by 60.37%, 66.6% and 67.59%, respectively.

KEYWORDS

Short-term Wind Power Forecasting; Variational Modal Decomposition; Kernel Principal Component Analysis; Butterfly Optimization Algorithm; Bi-directional Long Short-term Memory.

INTRODUCTION

In the current global economic landscape, the issue of depletion of global resources is escalating, particularly in the energy sector which is transitioning towards a greener and low-carbon structure. Wind energy, due to its widespread availability, ease of access, and environmental friendliness, has emerged as a focal point for research in transforming power generation in China [1]. Accurately predicting wind power is crucial for enhancing wind power utilization, ensuring power system reliability, and optimizing power grid operations [2]. Existing research on wind power forecasting is extensive, with traditional methods categorized into physical and statistical approaches [3]. Physical methods involve establishing a model based on meteorological parameters from Numerical Weather Prediction (NWP) to forecast wind power trends [4], while statistical methods predict power output by analyzing historical data [5]. However, with the increasing integration of wind power into the grid, traditional forecasting methods struggle to handle complex wind power data. The introduction of a combined model utilizing data decomposition techniques and neural network optimization algorithms offers a promising avenue to enhance power forecasting accuracy.

In terms of data decomposition, [6] proposed a wind power prediction method based on empirical mode decomposition (EMD) and improved extreme value learning machine. [7] used integrated empirical mode decomposition (EEMD) to transform irregular wind power time series into relatively easy to analyze subseries, and the resulting smooth subseries were predicted using the least absolute shrinkage selection operator-quantile regression neural network (LASSO-QRNN) model. A hybrid model combining variational modal decomposition (VMD), sparrow search algorithm (SSA), and time convolutional network-based bi-directional gated recurrent unit (TCN-BiGRU) was proposed in [8]. These methods all involve decomposing wind power time series into components with different frequencies, predicting these components, and combining the predicted values to obtain the final result. The study suggests that VMD provides more comprehensive decomposition and leads to higher prediction accuracy. However, it is important to note that wind power data are influenced by meteorological factors like wind speed, air pressure, and temperature. Subsequences generated by different decomposition techniques will inherently reflect these factors. Neglecting meteorological factors in the analysis of wind power data will likely result in reduced prediction accuracy.

In terms of neural network optimization algorithms, [9] used genetic algorithm to optimize the hyperparameters of RBF neural network to predict the wind power of wind farms, and the GA-RBF model is beneficial to improve the accuracy of prediction. [10] established an optimized LSTM model based on the modified bald eagle search (MBES) algorithm to solve the problem that the selection of LSTM hyperparameters may affect the prediction results, which improves the convergence speed of the model, but the improvement of accuracy is limited, and the LSTM only considers the forward transfer of information, and does not have the ability of two-way learning.

A short-term wind power combination forecasting model is proposed in this study, utilizing VMD-KPCA and CSFBOA-BiLSTM techniques. The model combines variational mode decomposition (VMD), kernel principal component analysis (KPCA), multi-strategy improved butterfly optimization algorithm (CSFBOA), and bi-directional long short-term memory (BiLSTM). Initially, the VMD method decomposes the original wind power data and meteorological factors individually, extracting detailed components of historical meteorological features while addressing non-stationarity in the power series. Subsequently, kernel principal component analysis optimizes features and reduces input variable dimensions. Finally, the CFSBOS-BiLSTM model is established, with the CSFBOA enhancing the learning and generalization capabilities of the BiLSTM network. Simulation results based on data from a wind farm in northwest China demonstrate the effectiveness of the proposed forecasting method.

1. METHOD THEORY

1.1. VMD Decomposition

The wind power series is susceptible to meteorological factors and exhibits high-frequency characteristics, making it necessary to use decomposition techniques to divide it into multiple stationary sub-sequences. VMD [11] can decompose the original signal $f(t)$ into K discrete modal components, where each modal component's finite frequency band surrounds its central frequency, and the optimal solution is obtained through iteration.The specific process is as follows:

1) Construct the variational constraint formula:

$$\begin{cases} \min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial(t) \left[\left(\delta(t) + \frac{\mathrm{j}}{\pi t} \right) * u_k(t) \right] \mathrm{e}^{-\mathrm{j}\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} \sum_k u_k = f(t) \end{cases} \tag{1}$$

In the formula, $\{u_k\}$ and $\{\omega_k\}$ are the modal set and the central frequency set, respectively; $\partial(t)$ is the partial derivative operator of t; $*$ is the convolution operator; $*$ is the Dirac function.

2) The Lagrangian operator $\lambda(t)$ inverse and the quadratic penalty factor α are utilized to create the augmented Lagrange function, which converts the variational constrained problem into an unconstrained problem for finding the optimal solution.

$$\begin{aligned} L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_k \| \left(\partial(t) \left[\left(\delta(t) + \frac{\mathrm{j}}{\pi t} \right) * u_k(t) \right] \mathrm{e}^{-\mathrm{j}\omega_k t} \|_2^2 \\ + \left\| f(t) - \sum_k u_k \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k \right\rangle n \end{aligned} \tag{2}$$

In the formula, $\alpha > 0$; $\langle \cdot, \cdot \rangle$ is the inner product operator of the function.

3) The saddle point of the augmented Lagrange function is determined through the alternating direction multiplier method, where the update of variables u_k and ω_k is alternated until the optimal solution of the variational model is achieved.

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{3}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k^{n+1}(\omega)^2 \right| \mathrm{d}\omega}{\int_0^\infty \left| \hat{u}_k^{n+1}(\omega)^2 \right| \mathrm{d}\omega} \tag{4}$$

In the formula, n is the number of iterations, and $\hat{f}(\omega)$, $\hat{u}_i(\omega)$, $\hat{\lambda}(\omega)$, $\hat{u}_k^{\hat{n}+1}(\omega)$ are the Fourier transformations of $f(t)$, $u_i(t)$, $\lambda(t)$, $u_k^{\hat{n}+1}(\omega)$ respectively.

1.2. KPCA Dimensionality Reduction

To address the challenge of handling complex data in conventional machine learning prediction models, KPCA [12] is employed for dimensionality reduction of input variables. The principle of KPCA involves mapping linearly inseparable data to a high-dimensional feature space, followed by the use of a kernel function to compute principal components. Feature extraction is then achieved based on the cumulative contribution rate. The kernel function represents the nonlinear mapping φ:

$$\varphi : \mathbf{x} \rightarrow \varphi(\mathbf{x}) \tag{5}$$

Assuming that the data sample of matrix X is $\mathbf{x}_\mathrm{i} \in \mathbf{R}^{\mathrm{m} \times \mathrm{n}}$, then after mapping to the high dimensional feature space, the covariance matrix C can be expressed as:

$$\mathbf{C} = \frac{1}{m}\sum_{i=1}^{m}\varphi(\mathbf{x}_i)\varphi^{T}(\mathbf{x}_i) \tag{6}$$

The eigenvalue λ_i and eigenvector $\mathbf{P_i}$ of matrix C are:

$$\mathbf{CP_i} = \lambda_i\mathbf{P_i} \tag{7}$$

$$\mathbf{P_i} = \sum_{i=1}^{m}\alpha_{\mathbf{i}}\varphi(\mathbf{x}_i) \tag{8}$$

In the formula, α_i is a linear coefficient matrix.

Calculate the projection of the vector $\varphi(x)$ on the feature vector $\mathbf{P_i}$ to extract the nonlinear principal components:

$$\varphi(\mathbf{x})\mathbf{P_i} = \sum_{i=1}^{m}\alpha_{\mathbf{i}}\varphi(\mathbf{x})\varphi(\mathbf{x}_i). \tag{9}$$

The contribution rate of data matrix X and variance α after dimensionality reduction is:

$$\mathbf{X} = \mathbf{X}\times\mathbf{P_i}\,(i = 1,\cdots,m) \tag{10}$$

$$\alpha = \frac{\lambda_i}{\sum_{i=1}^{m}\lambda_i}(i = 1,\cdots,m) \tag{11}$$

1.3. BiLSTM

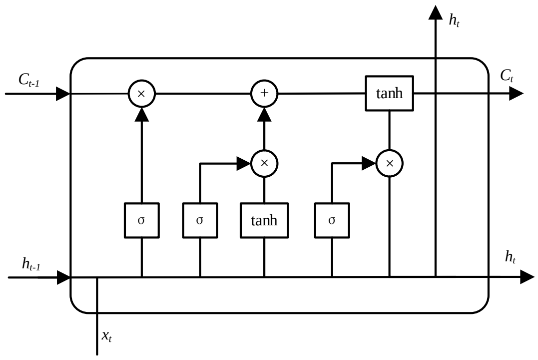

Fig. 1 Schematic diagram of LSTM network structure

LSTM [13] is a type of recurrent neural network that builds upon the traditional RNN [14] model by incorporating mechanisms like the forget gate, input gate, output gate, and memory unit to address the issue of vanishing gradients. These 'gates' help regulate the flow of sequential information, enabling the model to learn complex temporal relationships between sequences and enhance its capability to process long sequences of data. The standard architecture of LSTM is illustrated in fig. 1.

The BiLSTM [15] neural network consists of forward and reverse LSTM neural networks. In comparison to traditional LSTM networks, BiLSTM can simultaneously capture time series characteristics from past and future data, enhancing overall feature utilization. Typically, BiLSTM outperforms LSTM in prediction accuracy. The structure of BiLSTM is illustrated in Fig. 2.

Fig. 2 shows that the forward LSTM layer encodes sequence features in the positive direction of the time series, capturing dependencies from past moments to the current moment (recorded as $\overleftarrow{h}$). On the other hand, the backward LSTM layer encodes sequence features in the reverse direction of the time series, capturing dependencies from the current moment to future moments (recorded as $\vec{h}$), and ultimately outputs the calculation result y.

$$h_t = \alpha\overleftarrow{h} + \beta\vec{h} \tag{12}$$

$$y = \sigma\left(h_t\right) \tag{13}$$

Where α and β are constant, and the sum of them is 1.

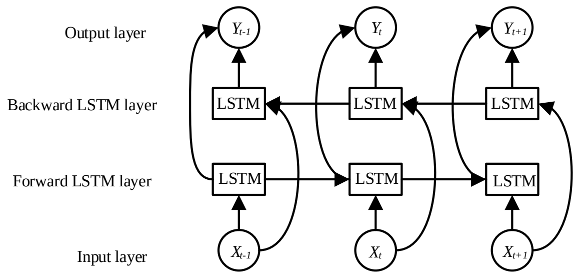

Fig. 2 Schematic diagram of BiLSTM network structure

1.4. CFSBOA

The super-parameter of BiLSTM plays a crucial role in its prediction performance, and the efficient optimization capability of CFSBOA enables quick and accurate determination of these super-parameters. Butterfly Optimization Algorithm (BOA) [16] is a novel natural meta-heuristic algorithm inspired by butterfly foraging behavior. The algorithm generates butterfly positions randomly using formula (14) and facilitates the transition of search modes during the iterative optimization process.

$$\begin{cases} X = lb + r \times (ub - lb) \\ x_i^{t+1} = x_i^t + \left(r^2 \times x_j^t - x_k^t\right) \times f_i \quad , \quad r \geqslant P \\ x_i^{t+1} = x_i^{\,t} + \left(r^2 \times g^* - x_i^{\,t}\right) \times f_i \quad , \quad r < P \end{cases} \tag{14}$$

In the formula, ub and lb represent the upper and lower bounds of the search space, r is a random number within the range [0,1], f stands for fragrance concentration, x denotes the solution vector, and g^* signifies the optimal solution of the current iteration.

While BOA outperforms other similar algorithms, it still suffers from premature convergence, limited global search capability, and inadequate local optimization. To address these issues, reference [17] proposes three strategic enhancements to the standard BOA:

1) Incorporating the Hénon mapping initialization mechanism. The Hénon map is a two-dimensional discrete chaotic map known for generating uniformly distributed sequences. Initializing the population with the Hénon chaotic map helps reduce search stagnation, enhance population diversity, and prevent local optima traps. The mathematical model of the Hénon map is expressed as follows:

$$\begin{cases} x(k) = 1 - \gamma(x(k-1))^2 + y(k-1) \\ y(k) = \varphi x(k-1) \\ z_i^k = \dfrac{x_i^k - \varphi_i}{\gamma_i - \varphi_i}, i = 1, 2, \cdots, n \end{cases} \tag{15}$$

$$x_i^t = lb_i + z_i^k \left(ub_i - lb_i\right) \tag{16}$$

In the form, $\gamma = 1.4$; $\varphi = 0.3$;x(0)=0;y(0)=0; k is the number of chaotic iterations. n chaotic variables z_i^k are generated by equation (15), and the individual search space variable x_i^t is obtained by inverse mapping.

2) Introduce a feedback sharing mechanism into the search formula, incorporating the sharing coefficient and feedback factor. Positive or negative feedback is generated by calculating and comparing the fitness value of each butterfly's objective function. Updates are then made to a random individual position or the optimal position. The specific formula is as follows:

$$\alpha = \left(F_{\max} - F_{\min}\right) r + F_{\min} \tag{17}$$

$$\theta = \frac{1}{1 + \exp(-\lambda)} \tag{18}$$

$$x_i^t = \begin{cases} x_i^t + \left(g^* - x_i^t\right)\alpha + \left(x_h^t - x_i^t\right)\alpha + \theta, & F\left(x_h^t\right) < F\left(x_i^t\right) \\ x_i^t + \left(g^* - x_i^t\right)\alpha + \theta, & F\left(x_h^t\right) \geqslant F\left(x_i^t\right) \end{cases} \tag{19}$$

Where, α is the sharing coefficient; Fmax and Fmin were the fitness ranges of population. θ is the feedback factor, λ is the number of current iteration entities; x_h^t is the location of random butterflies; F is the calculation fitness.

3) The group synergistic effect location updating mechanism is introduced. Using the golden sine algorithm[19] as a medium, the golden section number was used to improve the distance formula, and cooperative individual correction factor $|\sin\psi_1|$ and population coordination factor $\psi_2 \sin\psi_1$

were introduced to guide individual flight adjustment and population search direction at the individual and population levels respectively. The specific formulas are as follows:

$$\begin{cases} l_1 = -\pi + 2\pi(1-\tau) \\ l_2 = -\pi + 2\pi\tau \\ \tau = \dfrac{\sqrt{5}-1}{2} \end{cases} \tag{20}$$

Local search formula updated:

$$x_i^{t+1} = x_i^t \left|\sin\psi_1\right| + \left|l_1 x_j^t - l_2 x_i^t\right| f_i \psi_2 \sin\psi_1 \tag{21}$$

Global search formula updated:

$$x_i^{t+1} = x_i^t \left|\sin\psi_1\right| + \left|l_1 g^* - l_2 x_i^t\right| f_i \psi_2 \sin\psi_1 \tag{22}$$

Where, τ is the golden ratio; l1 and l2 are distance renewal coefficients with the advantage of golden section number. The random number of $\psi_1 \in [0,\pi]$, $\psi_2 \in [0, 2\pi]$.

2. SHORT-TERM WIND POWER PREDICTION MODEL

2.1. Prediction Model Building

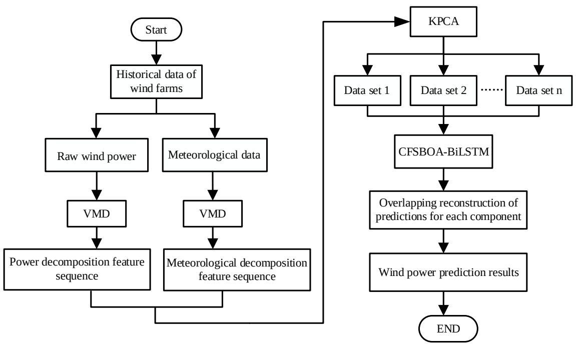

Fig. 3 Flow chart of VMD-KPCA-CFSBOA-BiLSTM model

This study presents a short-term wind power prediction model based on VMD-KPCA and CFSBOA-BiLSTM, leveraging the benefits of VMD-KPCA meteorological feature mining and the high precision of CFSBOA optimization. The specific process is as follows:

Step 1: Screen the key meteorological factors needed by the prediction model by calculating Pearson correlation coefficient.

Step 2: The original wind power and meteorological factors are decomposed by VMD to obtain a series of IMF components.

Step 3: KPCA is used to optimize the feature sequence dimension, and the principal component is selected according to the cumulative contribution rate to construct the input feature data set.

Step 4: Construct the CFSBOA-BiLSTM prediction model to predict the feature data set one by one.

Step 5: Stack the predicted values and output the final prediction results.

2.2. CFSBOA-BiLSTM

In this paper, the multi-strategy improved CFSBOA algorithm is used to optimize the hyperparameters of the BiLSTM model. The specific steps are as follows:

Step 1: Fill the missing value of the data set and normalize the data set.

Step 2: Set the initial parameters. Among them, the CFSBOA parameter includes the population size and the number of iterations, and the BiLSTM parameter includes the super-parameters of the neural network.

Step 3: Hénon mapping initializes the butterfly population.

Step 4: Calculate the individual fitness of the butterfly, compare the fitness of the random butterfly and the current butterfly, and update the position according to the feedback sharing machine system and group synergy effect.

Step 5: Determine whether the iteration termination condition is satisfied, if so, end and output the combination of the optimal parameters, otherwise jump to step 3.

Step 6: Establish the CFSBOA-BiLSTM regression prediction model according to the optimal parameters.

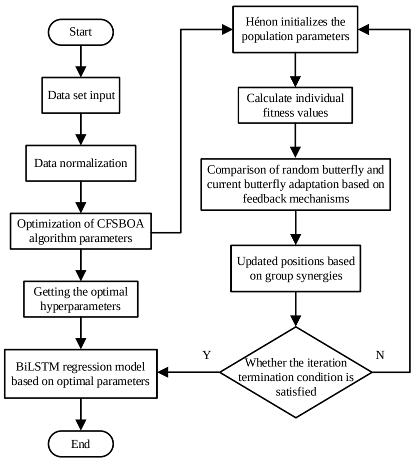

Fig. 4 Flow diagram of CFSBOA optimization BiLSTM

3. ANALYSIS OF EXPERIMENTAL EXAMPLES

To evaluate the proposed method, data from a wind farm in northwest China was utilized for research. The dataset includes wind direction, wind speed, temperature, humidity, air pressure, and actual power readings at different altitudes (10m, 30m, 50m, and 70m). The data was sampled at 15-minute intervals over a period of 31 days from July 1 to July 31, 2020, resulting in 2976 data points. The first 30 days' data (2880 points) were allocated for training, while the remaining 96 points from the last day were used for testing. A single-step prediction approach with a time step of 7 was employed, where historical multi-dimensional data from t-6 to t time steps were used to predict the power value at time t+1. All computations and simulations were conducted on a computer platform equipped with an R7-6800H 3.2GHz processor, core graphics card, and 16GB memory.

In order to quantify the prediction error, the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and prediction time are selected as prediction evaluation indexes. The formula is as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{c=1}^{N}\left(y_c - y_c^{'}\right)^2} \tag{23}$$

$$MAE = \frac{1}{N}\sum_{c=1}^{N}\left|y_c - y_c^{'}\right| \tag{24}$$

$$MAPE = \frac{1}{N}\sum_{c=1}^{N}\left|\frac{y_c - y_c^{'}}{y_c}\right| \times 100\% \tag{25}$$

Where, y is the actual wind speed; y' is the predicted wind speed; N is the predictor.

3.1. Person Correlation Analysis

The selection of input features will directly affect the prediction performance of the algorithm. Meteorological factors have varying influences on wind power, making the selection of meteorological features beneficial for reducing training time and data redundancy. In this study, Pearson correlation analysis is utilized for feature selection. Initially, the Pearson correlation coefficient [20] between meteorological factors and wind power is calculated using equation (26), and correlation analysis is conducted between various influencing factors and wind power to identify key environmental factors necessary for the machine learning model.

$$r_{x_1 y} = \frac{\sum_{i=1}^{n}\left(x_{1,i} - \overline{x}_1\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_{1,i} - \overline{x}_1\right)^2\left(y_i - \overline{y}\right)^2}} \tag{26}$$

In the formula, $x_{1,i}$ is the i-th value of eigenvalue x_1; y_i is the i-th power value; $\overline{x_1}$ and $\overline{y}$ are the mean values of eigenvalue x_1 and power value y respectively; n is the sequence length; $r_{x_1 y}$ is the correlation coefficient between characteristic x_1 and y; the range of values is $[-1,1]$.

According to the division of correlation degree in reference [21], there is a strong positive correlation between wind speed and wind direction at different heights, a moderate correlation between air pressure and wind power, and a weak negative correlation between humidity and temperature, of which the temperature correlation is the smallest. Therefore, nine meteorological factors except temperature and humidity and the actual generating power are selected to construct the initial data set.

Table 1. Correlation coefficient between wind power and environmental factors

variate	Correlation coefficient	variate	Correlation coefficient
Wind speed of 10m	0.68	wind direction of 10m	0.61
Wind speed of 30m	0.71	wind direction of 30m	0.63
Wind speed of 50m	0.75	wind direction of 50m	0.71
Wind speed of 70m	0.83	wind direction of 70m	0.73
temperature	-0.28	humidity	-0.39
atmospheric pressure	-0.57		

3.2. VMD Decomposition Result

When the original wind power and the original environmental sequence data are decomposed by VMD, it is necessary to select the appropriate number of modal decomposition K according to the characteristics of the signal. If the K value is too small, the noise left by the general assembly will be decomposed insufficiently, resulting in modal aliasing. Therefore, the K value is determined by the center frequency method, and the penalty parameter $\alpha = 2000$ is set, the initial center frequency is $\omega = 0$, and the tolerance of the convergence criterion is $\tau = 1 - e^{-7}$. Taking the original wind power series as an example, the decomposition result is shown in fig. 5.

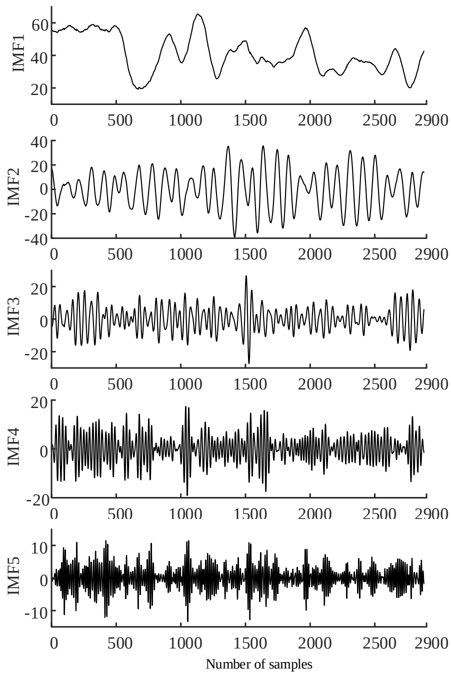

Fig. 5 Wind power decomposition

After the 11 meteorological factors selected above are decomposed by VMD, 79-dimensional IMF components can be obtained, which are combined with the 5-dimensional IMF components from the original wind power decomposition to construct five feature sequence sets.

3.3. Input Dataset Feature Dimensionality Reduction

In order to accurately mine the nonlinear features of multi-dimensional feature sequences, the above five feature sequence sets are reduced by KPCA method respectively. The Gaussian kernel function suitable for nonlinear feature extraction is selected, the principal components are arranged in descending order according to the variance contribution rate, and the cumulative variance contribution rate is calculated. The dimensionality reduction results of five feature sequence sets are given in tables 2 to 6 respectively.

Table 2. Feature sequence set 1

principal component	variance contribution	Cumulative variance contribution
X_1	0.6313	0.6313
X_2	0.1317	0.7630
X_3	0.0741	0.8371
X_4	0.0495	0.8866
X_5	0.0287	0.9153

Table 3. Feature sequence set 2

principal component	variance contribution	Cumulative variance contribution
X_1	0.6326	0.6326
X_2	0.1320	0.7646
X_3	0.0743	0.8389
X_4	0.0496	0.8885
X_5	0.0287	0.9171

Table 4. Feature sequence set 3

principal component	variance contribution	Cumulative variance contribution
X_1	0.6339	0.6339
X_2	0.1323	0.7662
X_3	0.0745	0.8407
X_4	0.0497	0.8904
X_5	0.0288	0.9192

Table 5. Feature sequence set 4

principal component	variance contribution	Cumulative variance contribution
X_1	0.6342	0.6342
X_2	0.1324	0.7666
X_3	0.0745	0.8411
X_4	0.0497	0.8908
X_5	0.0288	0.9196

Table 6. Feature sequence set 5

principal component	variance contribution	Cumulative variance contribution
X_1	0.6343	0.6343
X_2	0.1324	0.7666
X_3	0.0745	0.8412
X_4	0.0497	0.8909
X_5	0.0288	0.9197

It can be seen from the table that the cumulative variance contribution rate has reached more than 90% when calculated to X_5, which basically contains the information of all features, so the first five principal components are selected for each feature set to construct the input samples of the wind power rate prediction model.

3.4. Optimization Performance Analysis of CFSBOA

In order to verify the optimization performance of CFSBOA algorithm, this algorithm is used to optimize BiLSTM parameters with particle swarm optimization algorithm PSO, whale algorithm WOA and basic butterfly algorithm BOA at the same time, and the optimization results are compared. The parameters of the algorithm are as follows: the number of iterations $N \in (10,100)$, the number of neurons $\text{units} \in [50,150]$, and the learning rate $\text{lr} \in [0.001,0.01]$. The mean square deviation of the verification set is taken as the fitness function, and the fitness value of each optimization algorithm is shown in fig. 6.

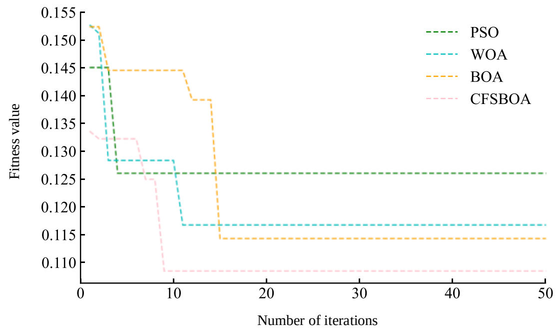

Fig.6 Fitness values of different optimization algorithms

Within the set maximum number of iterations, the learning ability of the algorithm is characterized by the fitness value and the speed of jumping out of the local optimal. As can be seen from figure 6, the convergence speed of PSO is the fastest, but the premature phenomenon occurs at the beginning of the iteration of the algorithm, and it is easy to fall into local optimization; although BOA continues to explore better parameters in the later stage, the convergence speed is slow; CFSBOA and WOA achieve local optimization in the 9th generation and 11th generation respectively, and the optimization effect is relatively good. Compared with WOA, CFSBOA, the convergence is faster and the fitness value is smaller. Therefore, the optimization performance of CFSBOA is the best among the four optimization algorithms.

3.5. Validity Analysis of Predictive Model

In order to verify the effectiveness of the model, a variety of single models and combination models are built to compare with the model in this paper. Among them, the single model selects BP, ELM, SVR, LSTM and BiLSTM, and the combined model selects EEMD-BiLSTM, VMD-BiLSTM, VMD-KPCA-BiLSTM (model one), VMD-KPCA-BOA-BiLSTM (model two) and this article model.

Among them, the number of neurons in BP network is 64, the activation function is tanh, the number of iterations is 80, the number of neurons is 64, the activation function is tanh; SVR, the radial basis kernel function is RBF, the number of neurons is 64, the learning rate is e=0.01, and the number of generations is 80. The number of neurons is 64, the learning rate is 0.01, and the number of generations is 80. The parameter setting of the combined method is the same as that of the single method, in which the population number of BOA and CFSBOA algorithm is set to 20, the maximum number of iterations is 50, and the number of hidden layer neural units, iterations and learning rate are selected as the hyperparameters to be optimized. The prediction results of the single model are shown in fig. 7, and the error indicators are shown in Table 7.

As can be seen from fig. 7, a single model can predict the fluctuation law of wind power, but there is a certain lag between the overall fluctuation curve and the real fluctuation curve. Combined with Table 7, it can be seen that the RMSE, MAE and MAPE of LSTM prediction model are 4.3688, 3.1409 and 0.0720 respectively, which is significantly lower than that of BP, SVR and ELM, and the prediction time increases, indicating that LSTM has the ability to deal with complex data, and the regression prediction effect of wind power real value is better. Compared with LSTM, the three error indicators of BiLSTM model are increased by 7.28%, 3.83% and 0.14% respectively, and the prediction time is the longest, indicating that the running time of deep learning is longer. BiLSTM can effectively mine the forward and reverse timing rules of power series, has better learning ability, and is more suitable for dealing with wind power data with high volatility.

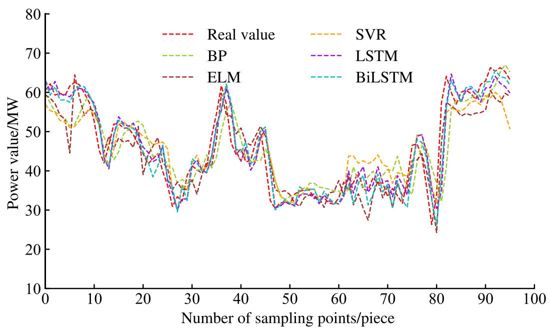

Fig.7 Results of single model prediction

Table 7. Errors of a single model

Forecast model	RMSE	MAE	MAPE	Forecast time/s
BP	6.3848	4.6720	0.1071	45
SVR	4.9204	4.2224	0.0992	2
ELM	5.7278	4.5816	0.1002	3
LSTM	4.3688	3.1409	0.0720	48
BiLSTM	4.0722	3.0251	0.0719	51

Fig. 8 shows the prediction results of the combined model, and the error is shown in Table 8. Compared with the results of a single model, it can be seen that after adding the decomposition strategy, the peak offset of the combination method is obviously improved, and the lag effect is weakened as a whole, this is because the decomposition algorithm can decompose the original data into multiple stationary subsequences. reduce the random fluctuation of the original data. Among them, compared with the EEMD-BiLSTM combination model, the RMSE value of the VMD-BiLSTM combined model is reduced by 15.04%, the mae value is reduced by 13.77%, and the MAPE value is reduced by 13.63%, which proves that VMD has a better decomposition effect than EEMD and is more conducive to reducing the error of wind power prediction.

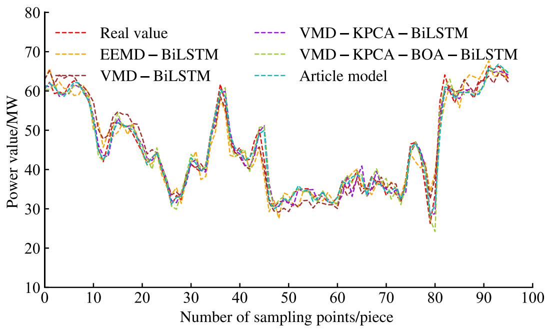

Fig. 8 Forecast results of combined mode

Table 8. Errors of combinatorial models

Forecast model	RMSE	MAE	MAPE	Forecast time/min
BiLSTM	4.0722	3.0251	0.0719	0.85
EEMD-BiLSTM	3.0004	2.4644	0.0565	6.15
VMD-BiLSTM	2.5490	2.1251	0.0488	6.3
Model one	2.1988	1.7012	0.0406	8.9
Model two	2.1784	1.5263	0.0366	243.45
Article model	1.6137	1.0101	0.0233	267.7

With the addition of KPCA dimensionality reduction algorithm, the RMSE value of the model 1 is 13.74% lower than that of the VMD-BiLSTM combination model, the score mae value is 19.95% lower, the map value is 16.8% lower, the error is significantly improved, and the prediction time of the model is increased. Although the processing of meteorological factors increases the prediction time of the model, the KPCA dimensionality reduction method can screen nonlinear features, enhance the logical correlation between predicted values and input values, and greatly improve the accuracy of training prediction. Comparing the prediction results of model 1, model 2 and this model, it is proved that using intelligent algorithm to optimize the super-parameters of BiLSTM can significantly improve the accuracy of the model and avoid the uncertainty of manual parameter adjustment, but the running time will be significantly increased. Among them, the three error indexes of the prediction results of this model are increased by 25.92% and 36.34% respectively compared with model 2, and the prediction time is increased by 578min compared with model 2, which proves that the prediction time of CFSBOA optimization algorithm is slightly longer than that of BOA optimization algorithm, but the prediction performance of the algorithm is significantly improved. the improvement of multi-strategy can make the butterfly individuals in the population converge to the optimal value quickly and accurately. Improve the optimization accuracy and generalization energy of the optimization algorithm, so that the model has the best prediction effect.

3.6. Other Experimental Analysis

The impact of season on wind turbine output in northwest China is significant. To validate the proposed method, data from a wind farm in northwest China for the month of January was selected for analysis. The data was collected at 15-minute intervals, with 2880 data points from the first 30 days used as the training set and 96 data points from the last day used for testing. The predicted results can be seen in fig. 9, with error indicators presented in Table 9.

As can be seen from fig. 9, at the 20-40 sample points with large wind power disturbance, the model in this paper is more consistent with the actual value, while the predicted values of other models are offset to a large extent. It can be seen that in the winter when the wind power output is strong, this method still has better fitting effect than other methods, and the proposed combination model can have better adaptability to seasonal changes. It can also be seen from Table 9 that the values of RMSE, MAE and MAPE of this method are 1.9188, 1.526 and 0.0403 respectively, which are kept within 2, and have the highest accuracy in several models, which further proves that this method has good accuracy and applicability.

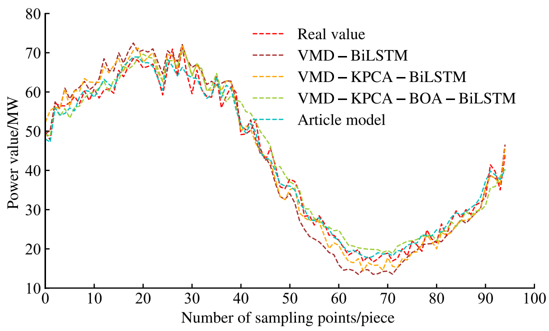

Fig. 9 Forecast results of combined model

Table 9. Errors of combinatorial models

Forecast model	RMSE	MAE	MAPE	Forecast time/min
VMD-BiLSTM	3.7263	3.2815	0.1101	7.4
Model one	2.7905	2.396	0.0675	9.75
Model two	2.6974	2.1863	0.0575	288.3
Article model	1.9188	1.526	0.0403	305.15

4. CONCLUSION THEORY

Wind power exhibits strong random fluctuations influenced by environmental factors. This paper introduces a combined forecasting approach using VMD-KPCA and CFSBOA-BiLSTM for short-term wind power prediction. The method takes into account the key environmental factors that impact wind power and enhances the accuracy of the forecasting model. Experimental results demonstrate that the proposed algorithm outperforms other algorithms in terms of accuracy and generalization for short-term wind power prediction across various meteorological conditions. In conclusion, the study highlights the effectiveness of the proposed method in improving short-term wind power forecasting.The main conclusions are as follows:

(1) By applying Variational Mode Decomposition (VMD) to analyze wind power data, stationary subsequences are extracted to mitigate the volatility of the original power output. Similarly, the key meteorological factors influencing wind power are decomposed to obtain stationary detail components at varying frequency scales. This approach helps address the issue of mismatched fluctuation trends between the environmental data series and the decomposed wind power subseries.

(2) The kernel principal component analysis method is employed to decrease the dimensionality of a multi-dimensional feature dataset. This method filters out key influencing factors of wind power by removing variables with high collinearity, reducing the input parameters of the model while preserving essential feature information for comprehensive feature data analysis.

(3) The multi-strategy improved CFSBOA algorithm is utilized to optimize the hyperparameters of a BiLSTM neural network, enhancing the model's learning and generalization capabilities for complex data features. The characteristic parameters, after being processed by KPCA dimensionality reduction, serve as input for the combined forecasting model of VMD-KPCA and CFSBOA-BiLSTM to predict wind power.

ACKNOWLEDGEMENTS

Funding: National Natural Science Foundation of China (62273133); Science and Technology Tackling Project of Henan Province (222102210120).

REFERENCES

[1] Zhang Y, Sun H, Guo Y. Wind power prediction based on PSO-SVR and grey combination model[J]. IEEE Access, 2019, 7: 136254-136267.

[2] Dong X, Sun Y, Li Y, et al. Spatio-temporal convolutional network based power forecasting of multiple wind farms[J]. Journal of Modern Power Systems and Clean Energy, 2021, 10(2): 388-398.

[3] Foley A M, Leahy P G, Marvuglia A, et al. Current methods and advances in forecasting of wind power generation[J]. Renewable energy, 2012, 37(1): 1-8.

[4] Gu B, Zhang T, Meng H, et al. Short-term forecasting and uncertainty analysis of wind power based on long short-term memory, cloud model and non-parametric kernel density estimation[J]. Renewable Energy. 2021, 164: 687-708.

[5] An G, Jiang Z, Cao X, et al. Short-term wind power prediction based on particle swarm optimization-extreme learning machine model combined with AdaBoost algorithm[J]. IEEE access, 2021, 9: 94040-94052.

[6] Tian Z, Ren Y, Wang G. Short-term wind power prediction based on empirical mode decomposition and improved extreme learning machine[J]. Journal of Electrical Engineering & Technology, 2018, 13(5): 1841-1851.

[7] He Y, Wang Y. Short-term wind power prediction based on EEMD–LASSO–QRNN model[J]. Applied Soft Computing, 2021, 105: 107288.

[8] 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.

[9] Guo P, Qi Z, Huang W. Short-term wind power prediction based on genetic algorithm to optimize RBF neural network[C]//2016 Chinese Control and Decision Conference (CCDC). IEEE, 2016: 1220-1223.

[10] Tuerxun W, Xu C, Guo H, et al. A wind power forecasting model using LSTM optimized by the modified bald eagle search algorithm[J]. Energies, 2022, 15(6): 2031.

[11] Zhang Y, Zhao Y, Kong C, et al. A new prediction method based on VMD-PRBF-ARMA-E model considering wind speed characteristic[J]. Energy Conversion and Management, 2020, 203: 112254.

[12] Wang G, Yang J, Qian Y, et al. KPCA-CCA-Based Quality-Related Fault Detection and Diagnosis Method for Nonlinear Process Monitoring[J]. IEEE Transactions on Industrial Informatics, 2022.

[13] Ding G, Qin L. Study on the prediction of stock price based on the associated network model of LSTM[J]. International Journal of Machine Learning and Cybernetics, 2020, 11: 1307-1317.

[14] Wang Y, Zhao Y, Addepalli S. Practical options for adopting recurrent neural network and its variants on remaining useful life prediction[J]. Chinese Journal of Mechanical Engineering, 2021, 34(3): 45-64.

[15] Cai R, Qin B, Chen Y, et al. Sentiment analysis about investors and consumers in energy market based on BERT-BiLSTM[J]. IEEE access, 2020, 8: 171408-171415.

[16] Arora S, Singh S. Butterfly optimization algorithm: a novel approach for global optimization[J]. SOFT COMPUTING, 2019, 23(3): 715-734.

[17] LI S Y, HE Q, DU N S. Butterfly optimization algorithm for chaotic feedback sharing and group synergy[J]. Journal of Frontiers of Computer Science & Technology, 2022, 16(7): 1661.

[18] Shah D, Shah T, Jamal S S. Digital audio signals encryption by Mobius transformation and Hénon map[J]. Multimedia systems, 2020, 26: 235-245.

[19] Tanyildizi E, Demir G. Golden sine algorithm: A novel math-inspired algorithm[J]. Advances in Electrical and Computer Engineering, 2017, 17(2): 71-78.

[20] Jebli I, Belouadha F Z, Kabbaj M I, et al. Prediction of solar energy guided by pearson correlation using machine learning[J]. Energy, 2021, 224: 120109.

[21] Liu Y J, Fan Y F, Bai X Y, et al. Short-term prediction of wind power based on feature crossover mechanism and error compensation [J]. Journal of Electrotechnics. 2022: 1-12.