IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 26, 2020, accepted August 30, 2020, date of publication September 7, 2020, date of current version September 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3022246

Short-Term PV Power Prediction Based on Optimized VMD and LSTM

LISHU WANG[1], YANHUI LIU[1,2], TIANSHU LI[1], XINZE XIE[1], AND CHENGMING CHANG[1]
[1]Institute of Electrical and Information, Northeast Agricultural University, Harbin 150030, China
[2]Institute of Electrical Engineering, Suihua University, Suihua 152061, China

Corresponding author: Lishu Wang (wanglishu@neau.edu.cn)

This work was supported in part by the Heilongjiang Education Department, Harbin, China, under Grant 12521038.

ABSTRACT Because of intermittence and fluctuation of photovoltaic (PV) power, it is difficult to enhance prediction accuracy. To sustain high-efficient operation of power system, this paper proposes a hybrid method to predict the short-term PV power. It consists of components separation of PV power, parameters optimization and re-construction of prediction result. Firstly, the methods based on the identifying of feature frequency and mutual information maximum are proposed to optimize the mode number and penalty factor of VMD, respectively. The optimized VMD (OVMD) is used to decompose the complicated fluctuation components of PV power into single component. Then, the improved PSO (IPSO) based on non-linear inertia weight of anti-sine function is proposed to optimize the number of hidden layer nodes, learning rate and iteration number of LSTM network. The optimized LSTM is used to predict each single component of OVMD decomposition. Thirdly, the prediction result of each single component is re-constructed to obtain the final PV prediction power. The experiment result indicates that the prediction accuracy of the proposed method (OVMD-IPSO-LSTM) outperformances the other typical methods. By the improvement of the traditional method (VMD and PSO) and the parameter optimization of LSTM, this hybrid method makes a contribution to the prediction of short-term PV power.

INDEX TERMS Power system, photovoltaic power prediction, parameter optimization, PSO, LSTM.

I. INTRODUCTION

A. PROBLEM DEFINITION

With the rapid development of science and technology, the energy demand becomes very important [1], [2]. Comparison with the other energy resources, the solar power is the one of most desired energy due to the advantages of cleanliness, good potential, extendibility and universality [3]. Many counties have developed the large scale photovoltaic (PV) plants to reduce the environmental pollution. However, the PV power system has the drawbacks of intermittence and fluctuation due to un-governable solar irradiance and the metrological factors, such as wind speed, temperature, humidity and dew-retting [4], [5]. However, the accuracy prediction of PV power can improve the reliability of power system and decrease the uncertainty of PV power on the power grid [6]. According to the difference of the predicted during time, the prediction of PV power can be divided into long term prediction (during time >1 month), medium term prediction (during time $1\sim7$ days) and short term prediction (during time <4 hours). Among the prediction mode, the short term prediction is very important for energy storage and load control. The short term prediction can make an important contribution for supervision and regulation of power market.

B. STATE OF THE ART

Several methods have been introduced to predict the PV power. In early stage, the time series method is widely used in the domain of prediction of the PV power [7]–[9]. However, due to instability of meteorological condition, the prediction accuracy is low. Afterwards, with the development of artificial intelligence (AI), the problem of the low prediction accuracy is improved effectively. The long short term memory(LSTM) neural network is the one of the outstanding prediction methods [10]. LSTM is a special case of recurrent neural network (RNN). LSTM can realize the remembering and forgetting of long-term historical status by the different gate structures. Recently, the prediction model of PV power

The associate editor coordinating the review of this manuscript and approving it for publication was Jonghoon Kim.

based on the LSTM have been developed and represented some advantages. Wang *et al.* [11] established the prediction model based on LSTM-RNN to predict a head-day PV power. Comparison with the SVM and BPNN, the LSTM-RNN has the more contribution for PV power prediction. Zhang *et al.* [12] proposed the CNN-LSTM to predict PV power. By the evaluation of the prediction performance and generation ability, the proposed model has some advantages of PV power prediction. Gao *et al.* [13] realized the prediction of the short-term power production in a large-scale photovoltaic plant. The result indicates that the LSTM is suitable to predict the PV power. However, it still exist two problems in the prediction of the PV power based on LSTM:

(1) The parameters sensitivity of LSTM. LSTM is very sensitive to hyperparameters. If the hyperparameters are not determined effectively, the prediction accuracy will be decreased heavily; (2) The intermittence and fluctuation of PV power. Because the PV power includes the components of randomness and volatility, they may disturb the prediction performance of PV power.

Thus, the parameter optimization and the separation of the disturbed information are focused on research in this paper.

At present, because the simplification and validity of PSO, PSO is considered as a promised parameter optimization method [14], [15]. However, due to the drawback of pre-mature convergence and local optimal solution, it limits the application in the prediction of PV power domain. Some researchers have improved the performance of PSO. Qin *et al.* [16] proposed an exponential center symmetric inertia weight to enhance the convergence speed and the optimal accuracy. Wang *et al.* [17] proposed a quantum PSO method with flighting and jumping operation to improve the accuracy and enhance the search ability. The other methods are also developed to improve the performance of the PSO [18], [19]. The enhancement of the global (local) search ability and convergence speed is the key mask in parameter optimization with PSO.

In addition, the decomposition-reconstruction method is also developed to separate the different information component in the PV power. The most representative method included the wavelet analysis [20] and the (expended) empirical mode decomposition (EMD, EEMD and CEEMDAN) [21]–[24]. However, the wavelet basis function need pre-definition and is non-adaptive in nature. EMD is prone to mode mixing. Although the EEMD can avoid the drawback of EMD, due to the interaction of added white noise and the signal, it may introduce the residue noise and spurious mode in IMFs. The CEEMDAN can solve the drawback of EEMD. However, the two critical parameters, which include the amplitude of added white noise and the number of ensemble trials, are difficult to determine [25], [26]. Afterwards, Dragomiretskiy *et al.* [27] proposed the variational mode decomposition (VMD). VMD has a strong theoretical foundation and can overcome the drawback of the EMD and wavelet. Recently, many scholars have applied the VMD to predict PV power. Oveis *et al.* [28] proposed prediction model based on VMD, information theoretic and multilayer perceptron neural network. Xie *et al.* [29] applied the VMD to decompose the PV power into different fluctuated components. And then deep belief network and auto-regressive moving average were used to predict the fluctuated component. However, the VMD has a drawback of parameter setting of the mode number and penalty factor by experience decision.

C. CONTRIBUTIONS

In this paper, a hybrid prediction method of PV power is proposed to enhance the prediction accuracy of the short-term PV power. Firstly, in the VMD decomposition, the mode number is optimized by identifying feature frequency of PV power, and the penalty factor is optimized by mutual information maximum. The optimized VMD (OVMD) can enhance the decomposition accuracy and avoid the information loss. Secondly, a new non-linear inertia weight function based on anti-sine is introduced to make the inertia weight change with the iteration number and particle position. The new inertia weight makes the strengthen of the global and local search ability. The accuracy and convergence performance is improved. The improved PSO is named as the IPSO. Thirdly, the OVMD is applied to decompose the PV power into single fluctuated component. Each fluctuated component is predicted with LSTM. The parameters of LSTM including the number of hidden layer nodes, learning rate and iteration number are optimized with the IPSO. Finally, the prediction result of each component is re-constructed to obtain the final prediction of PV power.

The main contribution of this paper with respect to the previous research work are as follows:

(1) The parameter optimization method based on feature frequency and mutual information maximum is proposed to optimize the mode number and penalty factor of traditional VMD. The optimized VMD can enhance the decomposed performance and avoid the information loss.
(2) A new inertia weight function of PSO is proposed to make the inertia weight change with the iteration number and particle position. The new inertia weight can make the strengthen of the global and local search ability. It enhances the optimized accuracy and convergence performance.
(3) The improved PSO is used to optimize the parameter of LSTM. The optimal LSTM can avoid the drawback of parameter setting by the experience and enhance prediction accuracy effectively.
(4) The decomposition-reconstructed method is proposed to obtain the final prediction of PV power by combination of each prediction result of each component.

The proposed method is verified on a real case study of PV power in Changzhou power network in China. The verified result indicates that the prediction accuracy and time complexity of the proposed model (OVMD-IPSO-LSTM) have a better performance comparison with the other traditional

methods. The hybrid method is more suitable for the local application in the current environment.

D. PAPER ORGANIZATION

The rest of the paper are organized as follows. Section 2 describes the related work; Section 3 presents the proposed method; Section 4 discusses the prediction result by comparison proposed model with the traditional model; Section 5 is the conclusion and future research.

II. RELATED WORK

A. VMD PRINCIPLE

In 2014, Dragomiretskiy *et al.* [27] proposed the VMD algorithm. Because of the strong theoretical background, the VMD is widely applied in the many fields such as fault diagnosis [30], [31], image processing [32] and signal processing [33], [34].

In the VMD decomposition, each IMF can be considered as the AM-FM signal $u_k(t)$. The work principle of the VMD is as follows:

(1) Do hilbert transform for the mode function $u_k(t)$, and then extract the single-side spectrum

$$\left[\delta(t) + \frac{j}{\pi t}\right] \cdot u_k(t) \tag{1}$$

where, $\delta(t)$ refers to the impulse function, j refers to the imaginary part.

(2) Transfer the single-side spectrum into the corresponding baseband based on estimation of enter frequency of each mode

$$\left[\left(\delta(t) + \frac{j}{\pi t}\right) \cdot u_k(t)\right] e^{-j\omega_k t} \tag{2}$$

(3) Estimate the bandwidth of each modal signal (IMF) by calculating the square of the norm of gradient of the modal signal

$$\begin{aligned} &\min\left\{\sum_k \left\|\partial_t\left[\left(\delta(t) + \frac{j}{\pi}\right) \cdot u_k(t)\right] e^{-j\omega_k t}\right\|_2^2\right\} \\ &s.t. \sum_k u_k = f \end{aligned} \tag{3}$$

where, u_k refers to the k_{th} IMF, ω_k refers to the corresponding center frequency, k refers to the decomposed mode number.

(4) Introduce the second penalty factor α and Lagrange multiplication operator $\theta(t)$. The expanded Lagrange is expressed as

$$\begin{aligned} &L(\{u_k\}, \{\omega_k\}, \theta) \\ &= \alpha \sum_{k=1}^{K} \left\|\partial_t\left[\left(\delta(t) + \frac{j}{\pi}\right) \cdot u_k(t)\right] e^{-j\omega_k t}\right\|_2^2 \\ &\quad + \left\|f(t) - \sum_{k=1}^{K} u_k(t)\right\|_2^2 \\ &\quad + \left\langle \theta(t), f(t) - \sum_{k=1}^{K} u_k(t)\right\rangle \end{aligned} \tag{4}$$

Finally, the mode function $u_k(t)$ and corresponding center frequency ω_k can be obtained by iteration updating.

The mode number (k) and penalty factor (α) determine if the decomposed result exists the information loss and over decomposition [35]. Thus, the two key parameters play an important role for the separation of complexed components of PV power.

B. MUTUAL INFORMATION

Mutual information, which is derived from information theory, is a good indicator of the similarity measure. Generally, the stronger correlation the two time series is, the larger mutual information value is [36]. Based on this, the mutual information can be used to optimize the parameter of VMD. If the decomposed effectiveness of the VMD is higher, the decomposed IMF and re-constructed will have the stronger correlation with original signal. The following is the principle of mutual information:

The joint entropy of random variants X and Y is defined as

$$\begin{aligned} H(X, Y) = -\sum_i p(x = a_i, y = b_i) \\ \cdot \log p(x = a_i, y = b_i) \end{aligned} \tag{5}$$

For a discrete random variable X with n elements length, the entropy of the random variant X is expressed as

$$H(X) = -\sum_{i=0}^{n} p(i) \log p(i) \tag{6}$$

where, $p(i)$ is the probability density function of discrete point i within the variant X. The mutual information is defined as

$$I(X, Y) = H(X) + H(Y) - H(X, Y) \tag{7}$$

where, $H(Y)$ is the entropy of the variant Y.

C. LONG SHORT TERM MEMORY

LSTM can avoid the drawback of gradient disappearance and realize the long term memory. It consists of the different gate structure: forget gate, input gate and output gate. These gates can memory and forget the relevant information (previous and current information) status of the time series data. For input gate, it can determine how much input will be saved into cell at current moment. For the forget gate, it indicates 'memory' level at last moment. Finally, the output gate controls the number of information outputting. The figure 1 shows the structure of LSTM.

The following is the work principle of LSTM:

$$f(t) = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{8}$$

$$i(t) = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{9}$$

$$\tilde{C}(t) = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{10}$$

$$C(t) = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{11}$$

$$O_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O) \tag{12}$$

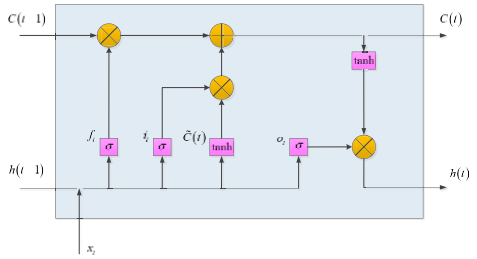

FIGURE 1. Structure of LSTM neural network.

$$h(t) = O_t * \tanh(C_t) \tag{13}$$

$$sigmoid(x) = \frac{1}{1+e^{-x}} \tag{14}$$

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{15}$$

From the equation (8) to equation (15), the σ refers to the sigmoid function and controls the information passing state. When the σ is 0, the nothing can pass; When the σ is 1, everything can pass. W_f, W_i, W_C and W_O refer to the input weight. The corresponding b_f, b_i, b_C and b_O refer to the biasing. The t and $t-1$ refer to the current and previous time status. The x and h refer to the input and output, and C refers to the cell status.

The LSTM includes some key parameters: the neuron number, iterations number in the hidden layer and learning rate [37], [38]. The neuron number is related to the fitting ability of LSTM, and the iteration number is related to the training effectiveness, and learning rate is concerned with convergence performance. However, these parameters are generally determined by experience setting. It leads that the prediction result can not be obtained effectively.

D. PARTICLE SWARM OPTIMIZATION

In this paper, the algorithm, which include particle swarm optimization (PSO) [39], ant colony optimization (ACO) [40], genetic algorithm (GA) [41] and simulated annealing (SA) [42], are compared to determine the suitable optimization algorithm. The PSO is easy and simple to apply for the prediction of PV power and has the fast convergence speed. Although the PSO may get local optimum solution, it can be avoided by tuning parameter as much as possible. The ACO has the drawback of low convergence speed. Although the ACO has the strong robustness, the performance is affected by the initial parameter easily. The convergent speed of the GA is general and keep stable. However, it has the drawback of Hamming Cliff. Likely, the SA has also the drawback of low convergent speed. If the temperature drops too fast, the global optimal solution may not be determined. Thus, the PSO is selected to consider as the fundamental optimized algorithm due to the fast efficiency and simple implementation. The following is the principle of the PSO.

It is assumed that the number of particle swarm is N. Each particle has position (v_{ij}) and velocity (x_{ij}). The particle swarm updating is expressed as follows:

$$v_{ij}(t+1) = w \cdot v_{ij}(t) + c_1 r_{1j}(t)\left(p_{ij}(t) - x_{ij}(t)\right) + c_2 r_{2j}(t)\left(p_{gj}(t) - x_{ij}(t)\right) \tag{16}$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \tag{17}$$

where, w refers to inertia weight, t indicates iteration number. The c_1 and c_2 refer to the acceleration factor. The $r_{1j}(t)$ and $r_{2j}(t)$ refer to the random number between [0,1]. The vector $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \cdots, x_{iD})$ is the potential solution. The vector $\boldsymbol{p_i} = (p_{i1}, p_{i2}, \cdots, p_{iD})$ and $\boldsymbol{p_g} = \left(p_{g1}, p_{g2}, \cdots, p_{gD}\right)$ refers to the optimal local and global position of the i-th particle.

The optimization ability of PSO is related to the inertia weight. If the inertia weight is too big, the convergence speed may be low (pre-mature convergence), and if the weight is too small, the local optimal solution may be happened.

III. PROPOSED WORK

A. OPTIMIZED VMD

In this section, the mode number and penalty factor of VMD will be optimized by the identifying of feature frequency and mutual information maximum.

It is assumed that the original signal $x(t)$ can be decomposed into a set of IMFs by VMD.

$$x(t) = \sum_{i=1}^{K} A_i(t)\cos(\Phi_i(t)) = \sum_{i=1}^{K} \mathrm{IMF}_i(t) \tag{18}$$

where, K refers to the mode number, $A_i(t)$ refers to the amplitude of each IMF, and $\Phi_i(t)$ refers to the corresponding phase.

1) OPTIMIZATION OF MODE NUMBER

Theoretically, the each IMF contains a specified frequency component. If the original signal contains the K feature frequencies, the decomposed number of IMFs should be K. However, the mode number is generally determined in experience setting. If the mode number is bigger than the theoretical one, it means that the different mode will contain the same frequency component, and if the mode number is smaller than the theoretical mode number, it indicates that there are various frequency components in the same IMF. They will lead that the signal will be not decomposed effectively.

To determine the mode number, this paper proposes the method based on the frequency bandwidth. It means that the signal is firstly done the Fast Fourier Transform (FFT), and then extract the feature frequency in overall frequency spectrum. Finally, the frequency bandwidth of each IMF can be determined by the feature frequency. The number of segmenting zone of the feature frequency is to determine mode number. This method can guarantee that the IMF only contain a kind of frequency component.

2) OPTIMIZATION OF PANALTY FACTOR

The penalty factor is needed to determine the other key parameter. It is related to decomposed performance and reconstructed performance.

A. The determination of the optimal decomposed performance If the decomposed IMF contains the major information of original data, the IMF is more similar with the original data. The similarity can be reflected by mutual information. It illustrates that the decomposition performance can be determined by the mutual information between the IMF and original data. The following is the determined process.

(1) Set the range of penalty factor $\alpha \in [\alpha_{start}, \alpha_{end}]$ with step K_{step}

(2) Decompose the original signal into K modes by VMD;

(3) Obtain the mutual information (MI) between each IMF and original signal

$$\mathrm{MI}_\alpha = \left(\mathrm{MI}_1^\alpha, \mathrm{MI}_2^\alpha, \ldots, \mathrm{MI}_K^\alpha\right) \tag{19}$$

(4) Look for the local maximum under each penalty factor

$$\mathrm{MI}_\alpha^{\max} = \left(\mathrm{MI}_{a_{start}}^{\max}, \ldots, \mathrm{MI}_{\alpha_{end}}^{\max}\right) \tag{20}$$

B. The determination of the optimal re-constructed performance

It is found through experiment that the smaller the penalty factor is, the better the re-constructed effectiveness is. It means that the IMFs can re-construct the original signal effectively. It does not happen the loss of original information.

It is assumed that the original signal $x(t)$ is decomposed into a set of $\mathrm{IMF_i}$ (i=1,…,K). Here, K denotes the IMF number. The re-construction of signal can be expressed as:

$$x'(t) = \mathrm{IMF}_1 +, \ldots, +\mathrm{IMF}_K \tag{21}$$

The re-constructed performance can be reflected by the mutual information between the re-constructed signal and original signal.

$$\beta = \mathrm{MI}\left(x'(t), x(t)\right) \tag{22}$$

Under the range of penalty factor, the β can be obtained:

$$\beta_\alpha = \left(\beta_{\alpha_{start}}, \ldots, \beta_{\alpha_{end}}\right) \tag{23}$$

The bigger the β_α is, the better re-constructed effectiveness is.

Thus, the optimized penalty factor can be reflected by the combination of the decomposed performance and the reconstructed performance.

$$\gamma_\alpha = \left(\beta_{\alpha_{start}} \cdot \mathrm{MI}_{a_{start}}^{\max}, \ldots, \beta_{\alpha_{end}} \cdot \mathrm{MI}_{a_{end}}^{\max}\right) \tag{24}$$

where, the γ_α refers to the comprehensive evaluated factor. Finally, the optimized penalty factor can be obtained:

$$\alpha_{opt} = \arg\max(\gamma_\alpha) \tag{25}$$

where, α_{opt} refers to the optimized penalty factor. The table 1 shows the algorithm of the parameter optimization of VMD.

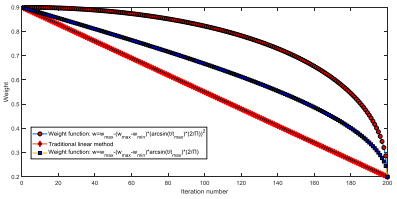

FIGURE 2. Comparison of inertia weight of traditional method with proposed method.

B. LSTM OPTIMIZATION WITH IMPROVED PSO

The traditional PSO has the drawback of pre-mature convergence and local optimal solution. The reason is that the linear inertia weight does not enhance the global search ability and convergence performance. To overcome the drawback of linear method, the non-linear weight, which is based on anti-sine function, is proposed in this paper.

$$w = w_{\max} - (w_{\max} - w_{\min}) \cdot \left(\arcsin\left(\frac{t}{t_{\max}}\right) \cdot \frac{2}{\pi}\right)^2 \tag{26}$$

where, $w_{\max}$ refers to the maximum inertia weight, $w_{\min}$ refers to the minimum inertia weight, $t_{\max}$ refers to the maximum iteration number. It indicates that a larger weight is used to enhance global search ability and make the particle traverse the all space in the early search stage, and in the late stage, a small inertia weight is applied to enhance local search performance to increase the speed of convergence.

The traditional linear inertia weight (as shown the equation (27)) and non-linear inertia weight based on the first power of anti-sine (as shown the equation (28)) are compared with the that of proposed method.

$$w = w_{\max} - (w_{\max} - w_{\min}) \cdot \frac{t}{t_{\max}} \tag{27}$$

$$w = w_{\max} - (w_{\max} - w_{\min}) \cdot \arcsin\left(\frac{t}{t_{\max}}\right) \cdot \frac{2}{\pi} \tag{28}$$

Here, $w_{\max}$ refers to 0.9, $w_{\min}$ refers to 0.2, and $t_{\max}$ refers to 200. The figure 2 plots the compared result. It is found that the inertia weight w is closed to the maximum inertia weight $w_{\max}$ in the early stage. The descent rate of the inertia weight is slower than the other methods. Thus, the IPSO has the stronger global search performance. With the increment of the iteration number, the inertia weight w decreases non-linearly. The descent rate of inertia weight is faster than the other methods. Thus, the IPSO has also the stronger local search ability. Meanwhile, the particle velocity go down faster. When the particle is nearby the global optimal solution, it can search for the optimal solution easily. The acquirement of the optimal solution can be guaranteed. In the late stage, the inertia weight w also decreases non-linearly. The decent speed of the inertia weight is faster than the other methods. The convergent speed is high. The global search ability become weaken. The particle can jump out the local minimum value. The local optimal solution can be avoided. Here, the improved PSO is sign as IPSO.

TABLE 1. Algorithm of the optimized parameters.

Input: original signal $x(t)$

Optimize mode number K

Extract feature frequency: $(f_1, f_2, \ldots f_K) = \mathrm{FFT}(x(t))$

Segment the frequency bandwidth based on feature frequency;
Determine the number of frequency bandwidth;
Obtain the optimized mode number K_{opt};

Optimize penalty factor α

Set penalty factor range: $\alpha \in [\alpha_{start}, \alpha_{end}]$

Set penalty factor step: α_{step}

For $\alpha = \alpha_{start} : \alpha_{step} : \alpha_{end}$

$\mathrm{IMF}_{\alpha,K}(t) = \mathrm{VMD}(x(t),\ \alpha,\ K)$

$\mathrm{MI}^{\max}_{\alpha,K} = \mathrm{MI}(\mathrm{IMF}_{\alpha K}(t), x(t))$

$x'_{\alpha}(t) = \mathrm{IMF}_{a,1} + , \ldots , + \mathrm{IMF}_{a,\mathrm{K}}$

$\gamma_{\alpha} = (\beta_{\alpha_{start}} \cdot \mathrm{MI}^{\max}_{a_{start}}, \ldots, \beta_{\alpha_{end}} \cdot \mathrm{MI}^{\max}_{a_{end}})$

End

Get comprehensive parameter: $\gamma_{\alpha} = (\beta_{\alpha_{start}} \cdot \mathrm{MI}^{\max}_{a_{start}}, \ldots, \beta_{\alpha_{end}} \cdot \mathrm{MI}^{\max}_{a_{end}})$

Obtain optimized parameters: $\alpha_{opt} = \arg\max(\gamma_{\alpha})$

Output: Optimized VMD decomposition: $\mathrm{IMFs}_{opt}(t) = \mathrm{VMD}(x(t), K_{opt}, \alpha_{opt})$

Thus, the IPSO is used to optimize the two layer LSTM. The optimized parameter include the first and the second hidden layers nodes number, learning rate and iteration number. The following are the optimal steps:

(1) Divide the original data into training data and testing data set;
(2) Initialize the parameters: iteration number (n), learning rate (ε), the first hidden layer node number (h_1) and the second hidden layer node number (h_2);
(3) Produce a population particle: $P_{i,0}(h_1, h_2, \varepsilon, n)$;
(4) Specify the particle $P_{i,0}(h_1, h_2, \varepsilon, n)$ as the parameter in LSTM, and train the LSTM neural network with the training sample set (y_t);
(5) Obtain the training sample output ($\hat{y}_t^j$), and determine the fitness function (fit_i);

$$fit_i = \sum_{i=1}^{I} \frac{\hat{y}_t^i - y_t^i}{y_t^i} \times 100 \tag{29}$$

(6) Calculate the fitness value of each particle $P_{i,0}(h_1, h_2, \varepsilon, n)$, and determine the individual and population extremum according to initialized fitness value, and record the best position of each particle as the historical best location;
(7) Apply the equation (16) and (17) to update the speed and position of each particle itself by the individual and global extremum, and calculate the fitness value by the updated extremum;
(8) Obtain the optimized LSTM model with IPSO;
(9) Predict the PV power with optimized LSTM. The picture 3 is the flowchart of optimized LSTM with IPSO.

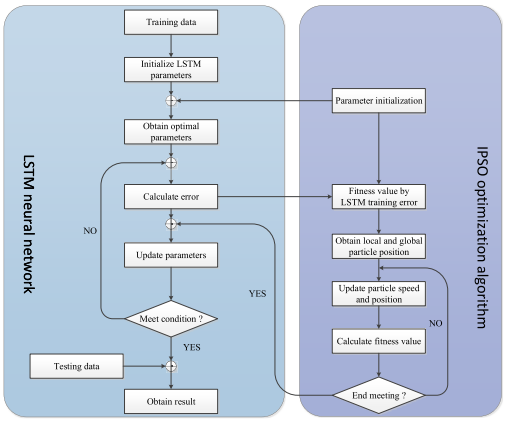

FIGURE 3. Parameter optimization of LSTM with IPSO.

C. FINAL FORECASTING MODEL

Assuming that the original PV power can be decomposed into a set of IMFs ($\mathrm{IMF}_1, \ldots, \mathrm{IMF}_N$) with OVMD. The prediction model (IPSO-LSTM) is used to predict the individual IMF, and then the prediction result is re-constructed to obtain the

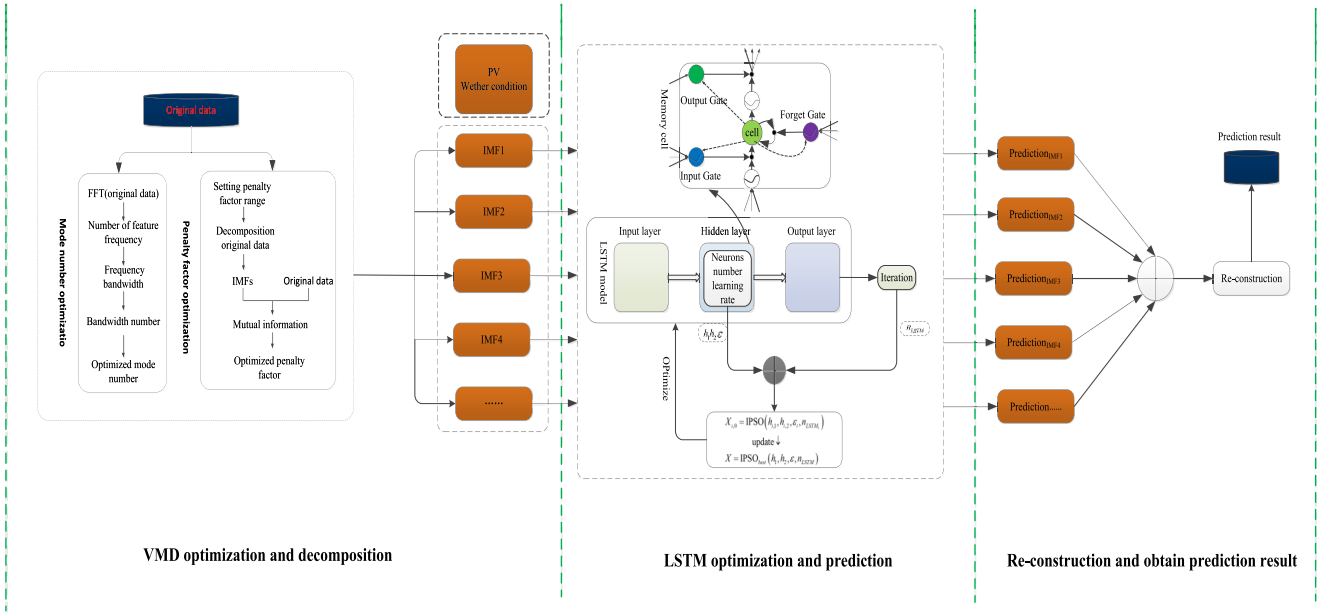

FIGURE 4. Flowchart of proposed method for prediction of PV power.

final prediction of PV power.

$$\mathrm{PV}_{final} = \mathrm{ILSTM}\left(\mathrm{IMF}_1\right), +\ldots, +\mathrm{ILSTM}\left(\mathrm{IMF}_N\right) \tag{30}$$

where, PV_{final} refers to the final prediction of PV power, $\mathrm{ILSTM}\left(\mathrm{IMF}_1\right)$ refers to the prediction of IMF1 with IPSO-LSTM, and $\mathrm{ILSTM}\left(\mathrm{IMF}_N\right)$ refers to the prediction of final IMF.

D. EVALUATION OF FORECASTING RESULT

The determination coefficient (R^2), mean absolute error (MRE), and mean absolute percentage error (MPAE) are used to evaluate the prediction accuracy (as shown the expression (31), (34) and (35)).

(1) Determination coefficient (R^2)

$$R^2 = 1 - \frac{RSS}{TSS} \tag{31}$$

$$RSS = \sum_{i=1}^{N} \left(x_i - \hat{x}_i\right)^2 \tag{32}$$

$$TSS = \sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2 \tag{33}$$

(2) Mean relative error (MRE)

$$\mathrm{MRE} = \frac{1}{N} \frac{\sum_{i=1}^{N} \left|x_i - \hat{x}_i\right|}{\bar{x}} \tag{34}$$

(3) Mean absolute percentage error (MPAE)

$$\mathrm{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left|\frac{\hat{x}_i - x_i}{x_i}\right| \times 100\% \tag{35}$$

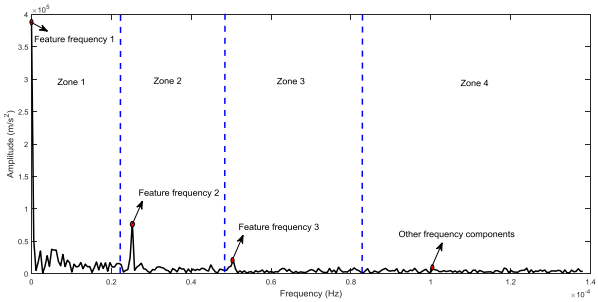

FIGURE 5. Determining of mode number by FFT.

where, x_i and $\hat{x}_i$ is the tested and predicted value, and N is the length of tested sample. The R^2 is between [0,1]. When the determination coefficient is 1, the prediction performance is the best, and vice versa. A smaller MRE and MPAE is, the better the prediction effectiveness is. The figure 4 shows the flowchart of proposed method for prediction of PV power.

IV. RESULT AND DISCUSSION

This paper uses PV power from Changzhou power network in China to verify the effectiveness of proposed method. The time period comes from the 21 May 2018 to 4 July 2018. The only period between 7a.m. and 5p.m. of every day is selected to establish the prediction model. That is because that the PV power is almost not generated except the selected period. The PV power data from the 21 May 2018 to 2 July 2018 is considered as the training set, and the rest is used to verify the prediction performance.

Firstly, the parameter of VMD is optimized with the identified feature frequency and mutual information (MI) maximum in section 3.1. The FFT is done for PV power to

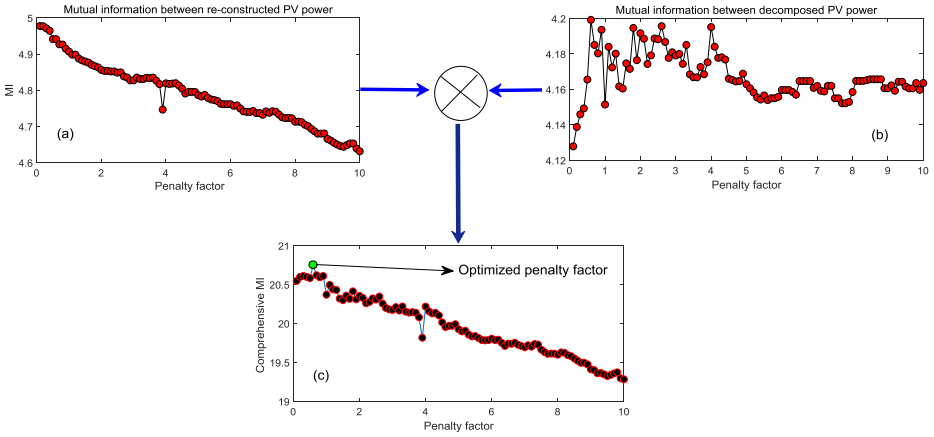

FIGURE 6. Optimization of penalty factor based on mutual information maximum.

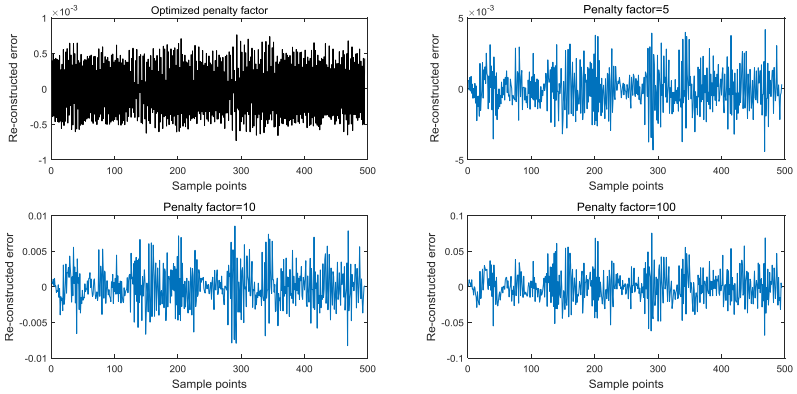

FIGURE 7. Comparison of reconstruction relative error under the different penalty factor.

determine the optimized mode number. Because this paper focuses on the short-term prediction of PV power (every 1 hour), the sample frequency in FFT is set as 1/3600. The figure 5 shows the result of FFT. It is found that the frequency spectrum main contain the four kinds of feature frequency component. Thus, the overall frequency spectrum is segmented into four zones (from zone1 to zone4). It means that the optimized mode number is 4.

Then, the optimal penalty factor (α) is determined. The search range of penalty factor is form 0.1 to 10 ($\alpha \in [0.1, 10]$) with the step 0.1. The figure 6 shows the optimized result. It is found that the MI of re-constructed performance is negatively correlated with penalty factor. For the decomposed performance, the MI changes obviously within the range of penalty factor 5, and for the penalty factor between 5 and 10, the MI fluctuates around 4.16. However, the maximum MI of comprehensive evaluated factor is concentrated on penalty factor 0.6. It indicates that the optimized penalty factor is 0.6.

Finally, the optimized mode number and penalty factor $[K, \alpha]$ is $[4, 0.6]$ for the PV power decomposition in this paper.

In this section, the penalty factor are set as 5, 10 and 100 to compare with the decomposed effectiveness of OVMD, respectively. Here, the re-constructed performance is quantified by relative error (as shown the equation (36)).

$$err = \frac{\left|x'(t) - x(t)\right|}{\max\left(x(t)\right)} \tag{36}$$

where, $x'(t)$ refers to re-constructed PV power, and $x(t)$ refers to the original PV power. The compared result is plotted in the figure 7. It is found that the relative error of proposed method is smallest, and the other one become larger with the increase of the penalty factor. It means that the MI become smaller with the increase of penalty factor. The result agrees with analysis in the figure 6(a). The re-constructed result is plotted in the figure 8. It is found that the re-constructed PV power is almost consistent with original one, and the others one can not re-construct with original PV power effectively. The figure 9 shows the decomposed IMFs under the optimized mode number and penalty factor.

In addition, the feature frequency of each IMF is also compared with that in PV power (as shown in the figure 10). It is found that each IMF only contain a kind of feature frequency, and the frequency component is all consistent with

TABLE 2. Parameters setting in LSTM and IPSO.

Order	Name	Value	Order	Name	Value
1	Maximum iteration number n for IPSO	200	5	Random number r (r_{1j} and r_{2j})	0.3
2	Started weight $w_{\max}$	0.9	6	Learning rate ε	[0.001,0.1]
3	Ended weight $w_{\min}$	0.2	7	Hidden nodes h_1, h_2	[1,200]
4	acceleration coefficients c_1, c_2	2	8	Search dimension	4
9	Iteration number range for LSTM	[100, 450]			

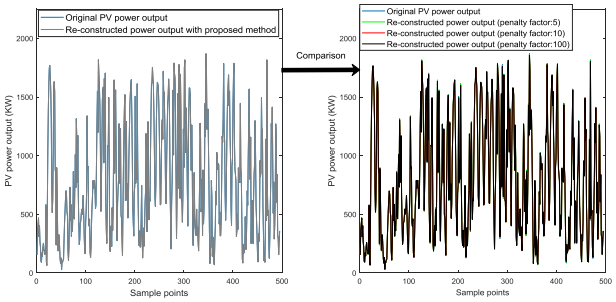

FIGURE 8. Comparison of reconstructed result under the different penalty.

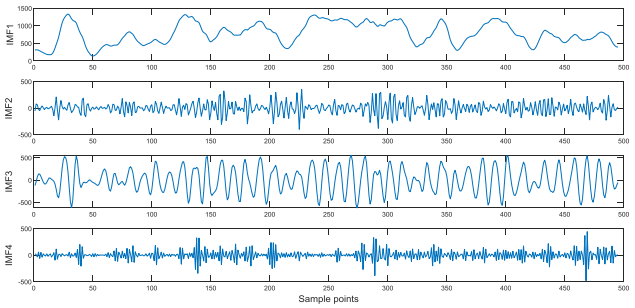

FIGURE 9. Decomposed IMFs with improved VMD for PV power.

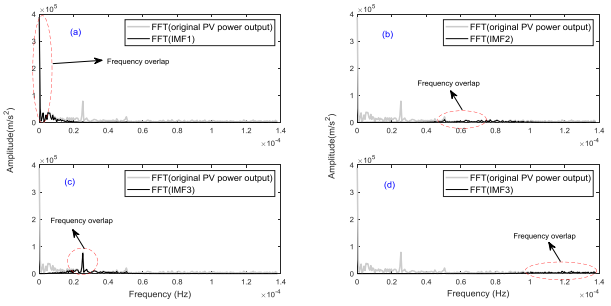

FIGURE 10. Comparison of feature frequency between the original data and each IMF.

that of original PV power (as shown the red dotted line), respectively. It indicates that the proposed method is more suitable to decompose the PV power effectively.

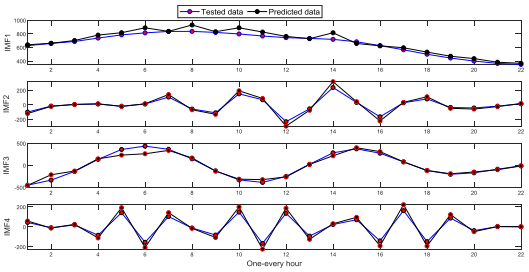

FIGURE 11. Comparison between the predicted result and original PV.

Next, the IMFs in the figure 9, which are considered as the independent component of PV power, are used to predict the PV power to establish the prediction model. The input condition includes the time series in everyday, temperature and solar radiation.

The IPSO is used to optimal the parameter of LSTM. For the determination of IPSO parameter, generally, the started and ended weight are set as the 0.9 and 0.4, respectively [43]. In this paper, the ended weight is extended to 0.2 to enhance performance of parameter optimization of LSTM. The acceleration coefficients (C_1 and C_2) are all considered as 2, which is an acceptable setting [43]. The random number belongs to [0,1]. It's value is set as 0.3 in this paper.

For the LSTM, at present, there is no specified literature to illustrate about the determination of parameter. Based on the consideration of prediction performance for PV power, this paper sets learning rate, hidden node and iteration number as [0.001,0.1], [1,200] and [100, 450], respectively.

The table 2 shows the main parameters in LSTM and IPSO. The figure 11 shows the comparison between the prediction result and original PV power for the each IMF. Then, the final prediction PV power is obtained by re-construction prediction result of IMF component.

The prediction performance of the proposed method (OVMD-IPSO-LSTM) is also compared with that of CEEMDAN-IPSO-SVM, CEEMDAN-IPSO-LSTM, IPSO-LSTM, VMD-IPSO-LSTM, PSO-LSTM, EMD-IPSO-LSTM, EMD-IPSO-SVM, IPSO-SVM and back propagation

TABLE 3. Comparison of the different forecasting method.

Order	Model index	Model name	Error index		
			R^2	MRE	MAPE(%)
1	Model 1	OVMD-IPSO-LSTM	0.9578	0.0042	8.49
2	Model 2	CEEMDAN-IPSO-SVM	0.9364	0.0053	12.07
3	Model 3	CEEMDAN- IPSO-LSTM	0.9236	0.0055	14.37
4	Model 4	IPSO-LSTM	0.9207	0.0061	17.46
5	Model 5	VMD-IPSO-LSTM	0.9167	0.0063	20.78
6	Model 6	PSO-LSTM	0.9088	0.0062	21.34
7	Model 7	EMD- IPSO-LSTM	0.9084	0.0067	21.75
8	Model 8	EMD- IPSO-SVM	0.9014	0.0065	21.85
9	Model 9	IPSO-SVM	0.8924	0.0067	21.93
10	Model 10	BP	0.8904	0.007	22.71

neural network (BP). The prediction model are named from model 1 to model 10 (as shown in the table 3). To realize the comparison fairly, the parameter setting of LSTM, IPSO, PSO are identical to the table 2. For parameter setting of CEEMDAN, it involves two important parameters: the ensemble size (EN) and the amplitude of the added white noise (AAWN). Because the determination of the two parameters are not guided theoretically, Wu *et al.* [23] suggested that the EN is set as hundreds of realizations, and the AAWN is set from 0.2 to 0.5. In this paper, the EN and AAWN are set as 200 and 0.3. EMD does not resolve the parameter setting. For the VMD, the mode number is set as optimized number 4, and the penalty factor is set as 100 (as shown the figure 7) to compare the prediction performance between the OVMD-IPSO-LSTM and VMD-IPSO-LSTM under univariate variation. Likely, for the SVM, the penalty parameter C,epsilon parameter ε, kernel function parameter γ and cross-validation are determined by the data characteristic of PV power and reference suggestion [44], [45]. Finally, the searching range of the parameters C, ε and γ are $[0, 100]$, $[0, 1]$ and $[0, 100]$, respectively, and the cross-validation is set as 5. For the BP algorithm, the training period, the learning rate and training target are set as the 200, 0.1 and 10^{-5}.

For the model 2, model 3, model 5, model 7 and model 8, firstly, the PV power is decomposed into a set of IMFs with the EMD, CEEMDAN and VMD. Each IMF is predicted with LSTM and SVM. The final predicted result are obtained by the re-construction of each IMF prediction. For the model 4, model 6 and model 9, the (I)PSO are directly used to optimize the parameter of LSTM and SVM to obtain the optimized prediction result. For the model 10, the optimized result can be obtained by repeated calculation. The compared result is plotted in the figure 12.

The MRE, MAPE and R^2 are used to evaluate the prediction effectiveness of each method (as shown in the table 3), respectively. The corresponding result is also plotted in the figure 13. It is found that the MRE and MAPE of proposed method is the smallest among the all methods, and the corresponding R^2 is the biggest. It indicates that the prediction

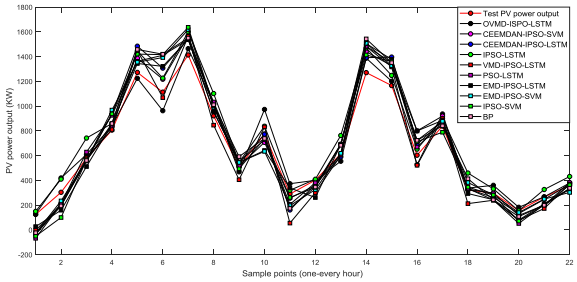

FIGURE 12. Comparison of the different forecasting method.

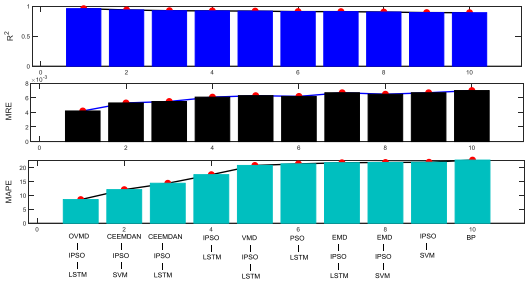

FIGURE 13. Comparison of the evaluated method for the different prediction method.

effectiveness of the proposed method outperformances the other methods.

Furtherly, the absolute percentage error (RAPE, as shown the equation (37)) is also used to evaluate the prediction accuracy.

$$\mathrm{RAPE}_i = \frac{|\mathrm{model}_i - \mathrm{model}_1|}{\mathrm{model}_1} \times 100\% \tag{37}$$

where, model_1 refers to the model index of proposed method including R^2, MRE and MAPE, and the model_i means the i_{th} model in the table 3. The table 4 shows the compared result with RAPE method. The compared result is plotted in the figure 14. It is found that for the determination coefficient (R^2), the proposed method improves the prediction accuracy between 2.2343% and 7.037%; for mean absolute error (MRE), the range of improved accuracy is from 26.1905% to 66.6667%, and for mean absolute

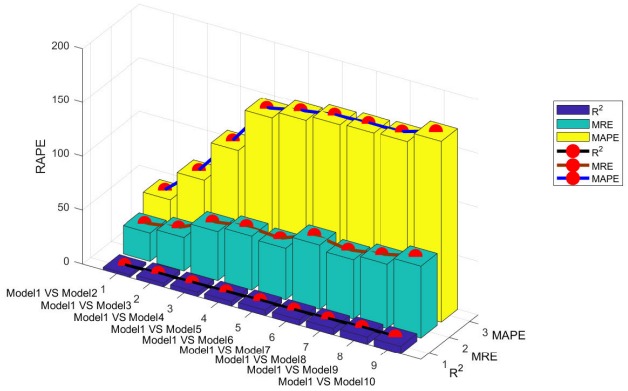

FIGURE 14. Comparison of the improved accuracy.

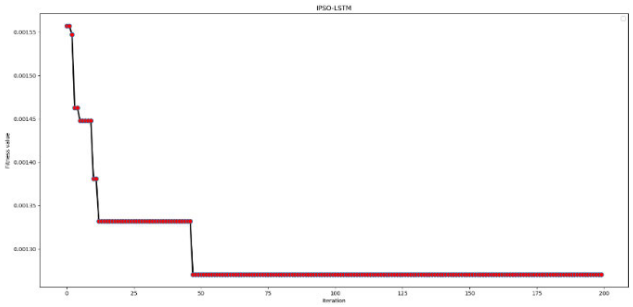

FIGURE 15. Convergent speed of the improved PSO-LSTM.

percentage error (MPAE), the range of improved accuracy is from 42.1673% to 167.4912%.

It is also found from the table 3 and table 4 that the prediction accuracy is related to the model structure. On the whole, the more complex the model structure is, the better the prediction accuracy is. The following illustrates the influence of the model structure on the prediction accuracy.

(1) The proposed method versus the model 2
The CEEMDAN and SVM have been applied for the time series prediction successfully [46], [47]. Although the IPSO can optimize the parameter of SVM, the CEEMDAN and SVM is not suitable to predict the PV power in this paper. The reason is that for the CEEMDAN, the decomposed IMF maybe contain residue noise and spurious mode due to the interaction between the signal and the added white noise [26], [48], it maybe effect the prediction of PV power. In addition, for the SVM, the hyperplane is fixed when the training is over. It leads that the model may not be updated by the current input and output. It leads that the history PV information is not retained into the model. It maybe make the prediction accuracy be low. However, the LSTM can employ the current information effectively and optimize the parameters and weight by feedback regulation continually [49]. Thus, the prediction accuracy based on LSTM maybe outperformance that based on SVM.

(2) The proposed method versus the model 3
The difference of algorithm structure concentrates on the component separation of PV power. Because the decomposed IMFs maybe contain the residue noise and spurious mode with CEEMDAN, it makes that the decomposed effectiveness is lower than the OVMD.

(3) The proposed method versus the model 4
There is no the component separation of PV power in the model 4. The prediction accuracy is low when the complex component of PV power is considered as the train data.

(4) The proposed method versus the model 5
The main reason of low prediction accuracy is that the parameters of the VMD are not optimized. It leads that the information component of original PV power is lost when the IMFs is re-constructed. Finally, the establish of the model maybe used with the no completed PV power.

(5) The proposed method versus the model 6
The structure of model 6 are consisted of the traditional PSO and LSTM. Comparison with the proposed method, the main reason of low prediction accuracy is that the model 6 is based on the prediction of the PV power and lacks the component separation of PV power. In addition, the traditional PSO may fall into local optimal value. Meanwhile, the convergent speed of the IPSO and PSO is also compared between the model 4 and model 6 (as shown the figure 15 and figure 16). It is found that the model 4 has the faster convergence speed. It needs about 48 iterations to keep the fitness be unchanged. However, the model 4 needs about the 135 iterations until the minimum fitness is obtained. It means that the traditional PSO has low convergent speed.

(6) The proposed method versus the model 7
Because the traditional EMD happens the mode mixing [50], [51], the each IMF contains the different frequency component of the PV power. It leads that the fluctuation component of the PV power may not be separated effectively. Finally, the prediction accuracy is lower than that based CEEMDAN and OVMD.

(7) The proposed method versus the model 8
The main reason of low prediction accuracy is that EMD has the drawback of the mode mixing, and the prediction model based on SVM is established by the fixed hyperplane.

(8) The proposed method versus the model 9
The model 9 lacks the decomposition of PV power effectively, and the SVM does not update the prediction model by the current input and output when the training is over. It leads that the prediction accuracy is lower than the proposed method.

(9) The proposed method versus the model 10
Because the model 10 is consisted of single algorithm. Comparison with the proposed method, it does not have the function of the component separation of PV power and the parameter optimization. Thus, it leads that the prediction accuracy is the lowest among the prediction models in this paper.

TABLE 4. Comparison of improvement accuracy.

Order	Model comparison	Relative absolute percentage error (RAPE)		
		$RAPE_R^2$	$RAPE_{MRE}$	$RAPE_{MAPE}$
1	Model 1 VS Model 2	2.2343	26.1905	42.1673
2	Model 1 VS Model 3	3.5707	30.9524	69.2580
3	Model 1 VS Model 4	3.8735	45.2381	105.6537
4	Model 1 VS Model 5	4.2911	50.000	144.7585
5	Model 1 VS Model 6	5.1159	47.6190	151.3545
6	Model 1 VS Model 7	5.1577	59.5238	156.1837
7	Model 1 VS Model 8	5.8885	54.7619	157.3616
8	Model 1 VS Model 9	6.8281	59.5238	158.3039
9	Model 1 VS Model 10	7.0370	66.6667	167.4912

TABLE 5. Time complexity and time cost (h).

Model order	IMF decomposition ($t_{1,i}$)	IMF number	Parameters optimization and prediction ($t_{2,i,j}$)	Re-constructed time ($t_{3,j}$)	Time cost
1	0.0311	4	3.1234	1.9444×10^{-7}	3.1547
2	0.0019	9	0.8751	2.2222×10^{-7}	0.8771
3	0.0019	9	7.0242	2.2222×10^{-7}	7.0262
4	—	—	0.8086	—	0.8086
5	2.9661×10^{-4}	4	3.1435	1.9444×10^{-7}	3.1438
6	—	—	0.8536	—	0.8536
7	1.3111×10^{-5}	7	5.4637	2.2222×10^{-7}	5.4637
8	1.3111×10^{-5}	7	0.6846	2.2222×10^{-7}	0.6847
9	—	—	0.0988	—	0.0988
10	—	—	—	—	2.7443×10^{-5}

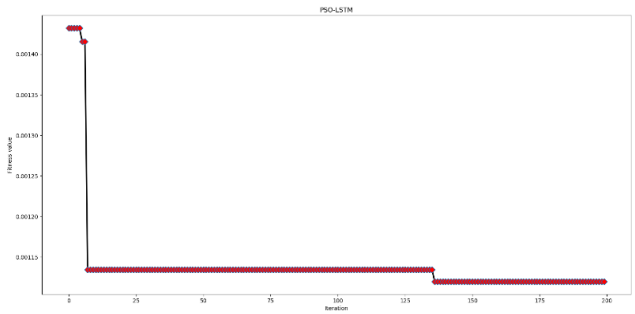

FIGURE 16. Convergent speed of the traditional PSO-LSTM.

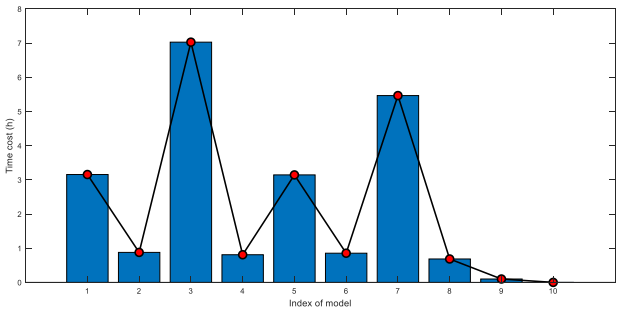

FIGURE 17. Comparison of time cost of the model.

In addition, the time complexity and time cost is also compared with the each prediction model. The evaluated method is shown in the equation (38).

$$t_i = t_{i,1} + \sum_{j=1}^{N} t_{i,2,j} + t_{i,3} \quad (i = 1, \ldots, 10) \tag{38}$$

where, the subscript i in $t_{1,i}$, $t_{2,i,j}$ and $t_{3,i}$ means the model order, and the subscript 1,2,3 means the configuration of each model. For example, the model 1 are consisted of three parts: OVMD (sign as $t_{1,1}$), IPSO and LSTM (sign as $t_{1,2,j}$), and re-construction (sign as $t_{1,3}$). The j refers to the number of IMFs. Thus, the time cost are consisted of the time of the decomposition IMF, parameter optimization and prediction, and the re-construction prediction result.

The table 5 shows the time complexity of each model and corresponding time cost. The time cost is plotted in the figure 17. Here, '−' means that the algorithm is not involved in the model. It is found that the time cost of the model 10 is the smallest. However, the prediction accuracy is the lowest. The time cost of the model 3 is the longest. It needs about 7 hours. The main reason is that the PV power is

decomposed into nine IMFs. The each IMF is all re-considered as the original data, and the IPSO is used to optimize the parameter of LSTM for each IMF under the maximum iteration. It needs the repetitive operation of nine times. Thus, the main consuming of time cost is the part of the repetitive parameter optimization and prediction. However, the proposed method decomposes the PV power into four IMFs. The time cost is relative low than that based CEEMDAN and EMD. It needs about the three hours. And the prediction accuracy is the highest. The other models such as model 4, model 6 and model 9 do not involve the decomposition-reconstruction and only use (I)PSO to optimize the parameter and predict. The time cost is moderate. In addition, the model establishment of LSTM employs the current information and optimize the parameter and weight by feedback regulation continually. The model establishment of SVM only apply for the fixed hyperplane and don't involve the feedback. Thus, on the whole, the time cost based on LSTM is longer than that based on SVM. However, the prediction accuracy based on LSTM is higher than that based on SVM. The reason why the prediction accuracy of the proposed method is. Firstly, the OVMD can decompose PV power into the IMFs effectively. The decomposed IMFs can reflect the fluctuated component, the information component of PV power is not lost. Secondly, the improved PSO can enhance convergence speed and does not fall into local optimal solution. Thirdly, the prediction performance of optimized LSTM is higher than that based on SVM. Finally, the prediction accuracy of the hybrid method outperformances the other methods. Although the time cost of proposed method is slightly longer due to the repeated optimization of LSTM parameter, the time cost is not the longest among the compared methods in this paper. Thus, comprehensive consideration of prediction accuracy and time cost, the proposed method is more suitable to predict the PV power in this paper.

Here, the LSTM in development process of program is based on Python 3.7.7 and Keras 1.0.8 deep learning tools package (open resource library). Adam optimizer [52] is used to make the loss function be minimized. The all programs are performed by the following computer configuration: Windows 10 64-bit operation system, intel core i7-8500U processor, 1.99GHz frequency, 8GB memory.

V. CONCLUSION AND FUTURE WORK

In this work, a novel hybrid method based on OVMD-IPSO-LSTM is proposed to predict the short-term PV power. Firstly, the methods based on feature frequency bandwidth and mutual information maximum are introduced to optimize the mode number and penalty factor of VMD. When the optimized VMD decompose the PV power, the fluctuated component can be separated effectively, and the information component of the PV power is not lost. The optimal VMD decompose the PV power in Changzhou power networks in China into the four different components. Secondly, the non-linear inertia weight based on anti-sine function is proposed. The inertia weight not only enhance convergence speed but also avoid the occurrence of local optimal solution. The improved PSO (IPSO) can enhance the optimized performance of traditional PSO significantly. Thirdly, the IPSO is used to optimize the hidden nodes, learning rate and iteration number of LSTM. Fourthly, the each separated component of PV power is predicted by IPSO-LSTM. Finally, the prediction result for each component is re-constructed to obtain the prediction output of PV power. The proposed model is compared with the other nine models, respectively. It is found from compared result based on actual PV power data that the OVMD-IPSO-LSTM can enhance prediction accuracy 2.2343 - 7.037% (R^2), 26.1905% - 66.6667% (MRE) and 42.1673% - 167.4912% (MAPE). Due to effective decomposition of PV power, improved performance of PSO and the effective optimization of LSTM parameter, it makes that the prediction accuracy of proposed model is higher than the other models. It indicates that the proposed method can meet the requirement of high prediction application for PV power.

However, the improvement of the prediction accuracy takes up a certain time cost. It main reflect on the repeated optimization of LSTM parameter. Thus, the future work will be developed towards the decrease of time cost under enhancement of prediction accuracy. It main focus on: (1) the other neural network such as the ring probabilistic logic neural network (RPLNN) is researched to evaluate prediction accuracy and time cost; (2) for the proposed model in this paper, the parallel program is developed to save the time cost; (3) the advanced decomposition method is developed to decompose the PV power into the different fluctuation components more effectively; (4) the prediction model is designed in embedded system to meet the practical demand more conveniently.

CONFLICTS OF INTEREST

All the authors do not have any possible conflicts of interest.

REFERENCES

[1] P. Kundur, "Sustainable electric power systems in the 21st century: Requirements, challenges and the role of new technologies," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, Jun. 2004, pp. 2297–2298.
[2] A. Moro and A. Holzer, "A framework to predict consumption sustainability levels of individuals," *Sustainability*, vol. 12, no. 4, p. 1423, Feb. 2020.
[3] W. Tao-Zhe, "Research on influence of PV power station model on power system voltage stability," *Telecommun. Power Technol.*, vol. 35, no. 6, pp. 58–63, Jun. 2018.
[4] P. Ariyaratna, K. M. Muttaqi, and D. Sutanto, "A novel control strategy to mitigate slow and fast fluctuations of the voltage profile at common coupling point of rooftop solar PV unit with an integrated hybrid energy storage system," *J. Energy Storage*, vol. 20, pp. 409–417, Dec. 2018.
[5] C.-J. Huang and P.-H. Kuo, "Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting," *IEEE Access*, vol. 7, pp. 74822–74834, 2019.
[6] U. Das, K. Tey, M. Seyedmahmoudian, M. I. Idris, S. Mekhilef, B. Horan, and A. Stojcevski, "SVR-based model to forecast PV power generation under different weather conditions," *Energies*, vol. 10, no. 7, p. 876, Jun. 2017.
[7] D. Xu, L. Kang, L. Chang, and B. Cao, "Optimal sizing of standalone hybrid wind/PV power systems using genetic algorithms," in *Proc. Can. Conf. Electr. Comput. Eng.*, 2005, pp. 1722–1725.
[8] J. G. De Gooijer and R. J. Hyndman, "25 years of time series forecasting," *Int. J. Forecasting*, vol. 22, no. 3, pp. 443–473, Jan. 2006.
[9] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Sol. Energy*, vol. 83, no. 10, pp. 1772–1783, Oct. 2009.

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[11] F. Wang, Z. Xuan, Z. Zhen, K. Li, T. Wang, and M. Shi, "A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework," *Energy Convers. Manage.*, vol. 212, May 2020, Art. no. 112766.
[12] H. Zang, L. Liu, L. Sun, L. Cheng, Z. Wei, and G. Sun, "Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations," *Renew. Energy*, vol. 160, pp. 26–41, Nov. 2020.
[13] M. Gao, J. Li, F. Hong, and D. Long, "Short-term forecasting of power production in a large-scale photovoltaic plant based on LSTM," *Appl. Sci.*, vol. 9, no. 15, p. 3192, Aug. 2019.
[14] P. Wang, J. Zhao, Y. Gao, M. A. Sotelo, and Z. Li, "Lane work-schedule of toll station based on queuing theory and PSO-LSTM model," *IEEE Access*, vol. 8, pp. 84434–84443, 2020.
[15] X. Song, Y. Liu, L. Xue, J. Wang, J. Zhang, J. Wang, L. Jiang, and Z. Cheng, "Time-series well performance prediction based on long short-term memory (LSTM) neural network model," *J. Petroleum Sci. Eng.*, vol. 186, Mar. 2020, Art. no. 106682.
[16] C. Qin and X. Gu, "Improved PSO algorithm based on exponential center symmetric inertia weight function and its application in infrared image enhancement," *Symmetry*, vol. 12, no. 2, p. 248, Feb. 2020.
[17] Y. Wang and X. Chen, "Hybrid quantum particle swarm optimization algorithm and its application," *Sci. China Inf. Sci.*, vol. 63, no. 5, May 2020, Art. no. 159201.
[18] T.-Y. Chen and T.-M. Chi, "On the improvements of the particle swarm optimization algorithm," *Adv. Eng. Softw.*, vol. 41, no. 2, pp. 229–239, 2010.
[19] Y. Chen, L. Li, J. Xiao, Y. Yang, J. Liang, and T. Li, "Particle swarm optimizer with crossover operation," *Eng. Appl. Artif. Intell.*, vol. 70, pp. 159–169, Apr. 2018.
[20] P. Zhang, H. Takano, and J. Murata, "Daily solar radiation prediction based on wavelet analysis," in *Proc. SICE Annu. Conf.*, 2011, pp. 712–717.
[21] F.-F. Li, S.-Y. Wang, and J.-H. Wei, "Long term rolling prediction model for solar radiation combining empirical mode decomposition (EMD) and artificial neural network (ANN) techniques," *J. Renew. Sustain. Energy*, vol. 10, no. 1, Jan. 2018, Art. no. 013704.
[22] N. Nayak and A. K. Pani, "Short term PV power forecasting using empirical mode decomposition based orthogonal extreme learning machine technique," *Indian J. Public Health Res. Develop.*, vol. 9, no. 11, pp. 2170–2182, 2018.
[23] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.
[24] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 4144–4147.
[25] X. Xue, J. Zhou, Y. Xu, W. Zhu, and C. Li, "An adaptively fast ensemble empirical mode decomposition method and its applications to rolling element bearing fault diagnosis," *Mech. Syst. Signal Process.*, vols. 62–63, pp. 444–459, Oct. 2015.
[26] L. Zhan, F. Ma, J. Zhang, C. Li, Z. Li, and T. Wang, "Fault feature extraction and diagnosis of rolling bearings based on enhanced complementary empirical mode decomposition with adaptive noise and statistical time-domain features," *Sensors*, vol. 19, no. 18, p. 4047, Sep. 2019.
[27] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
[28] O. Abedinia, D. Raisz, and N. Amjady, "Effective prediction model for hungarian small-scale solar power output," *IET Renew. Power Gener.*, vol. 11, no. 13, pp. 1648–1658, Nov. 2017.
[29] T. Xie, G. Zhang, H. Liu, F. Liu, and P. Du, "A hybrid forecasting method for solar output power based on variational mode decomposition, deep belief networks and auto-regressive moving average," *Appl. Sci.*, vol. 8, no. 10, p. 1901, Oct. 2018.
[30] C. Wang, H. Li, G. Huang, and J. Ou, "Early fault diagnosis for planetary gearbox based on adaptive parameter optimized VMD and singular kurtosis difference spectrum," *IEEE Access*, vol. 7, pp. 31501–31516, 2019.
[31] J. Ding, D. Xiao, and X. Li, "Gear fault diagnosis based on genetic mutation particle swarm optimization VMD and probabilistic neural network algorithm," *IEEE Access*, vol. 8, pp. 18456–18474, 2020.
[32] D. K. Agrawal, B. S. Kirar, and R. B. Pachori, "Automated glaucoma detection using quasi-bivariate variational mode decomposition from fundus images," *IET Image Process.*, vol. 13, no. 13, pp. 2401–2408, Nov. 2019.
[33] H. Hu, L. Zhang, H. Yan, Y. Bai, and P. Wang, "Denoising and baseline drift removal method of MEMS hydrophone signal based on VMD and wavelet threshold processing," *IEEE Access*, vol. 7, pp. 59913–59922, 2019.
[34] W. Liu, S. Cao, Z. Jin, Z. Wang, and Y. Chen, "A novel hydrocarbon detection approach via high-resolution frequency-dependent AVO inversion based on variational mode decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2007–2024, Apr. 2018.
[35] Z. Li, J. Chen, and Y. Zi, "Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive," *Mech. Syst. Signal Process.*, vol. 85, pp. 512–529, Feb. 2017.
[36] C. Li, L. Zhan, and L. Shen, "Friction signal denoising using complete ensemble EMD with adaptive noise and mutual information," *Entropy*, vol. 17, no. 12, pp. 5965–5979, Aug. 2015.
[37] D. Geng, H. Zhang, and H. Wu, "Short-term wind speed prediction based on principal component analysis and LSTM," *Appl. Sci.*, vol. 10, no. 13, p. 4416, Jun. 2020.
[38] C. Yu, X. Qi, H. Ma, X. He, C. Wang, and Y. Zhao, "LLR: Learning learning rates by LSTM for training neural networks," *Neurocomputing*, vol. 394, pp. 41–50, Jun. 2020.
[39] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw. (ICNN)*, 1995, pp. 1942–1948.
[40] M. Dorigo, G. D. Caro, and L. M. Gambardella, "Ant algorithms for discrete optimization," *Artif. Life*, vol. 5, no. 2, pp. 137–172, Apr. 1999.
[41] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Mach. Learn.*, vol. 3, no. 2, pp. 95–99, 1988.
[42] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.
[43] A. R. Jordehi and J. Jasni, "Parameter selection in particle swarm optimisation: A survey," *J. Exp. Theor. Artif. Intell.*, vol. 25, no. 4, pp. 527–542, Dec. 2013.
[44] C.-L. Huang and J.-F. Dun, "A distributed PSO-SVM hybrid system with feature selection and parameter optimization," *Appl. Soft Comput.*, vol. 8, no. 4, pp. 1381–1391, Sep. 2008.
[45] L. Zhan and C. Li, "A hybrid PSO-SVM-based method for predicting the friction coefficient between aircraft tire and coating," *Meas. Sci. Technol.*, vol. 28, no. 2, Feb. 2017, Art. no. 025004.
[46] H. Lu, X. Ma, K. Huang, and M. Azimi, "Carbon trading volume and price forecasting in China using multiple machine learning models," *J. Cleaner Prod.*, vol. 249, Mar. 2020, Art. no. 119386.
[47] S. Dai, D. Niu, and Y. Li, "Daily peak load forecasting based on complete ensemble empirical mode decomposition with adaptive noise and support vector machine optimized by modified grey wolf optimization algorithm," *Energies*, vol. 11, no. 1, p. 163, Jan. 2018.
[48] P. Fuentealba, A. Illanes, and F. Ortmeier, "Independent analysis of decelerations and resting periods through CEEMDAN and spectral-based feature extraction improves cardiotocographic assessment," *Appl. Sci.*, vol. 9, no. 24, p. 5421, Dec. 2019.
[49] Y. Tian, K. Zhang, J. Li, X. Lin, and B. Yang, "LSTM-based traffic flow prediction with missing data," *Neurocomputing*, vol. 318, pp. 297–305, Nov. 2018.
[50] W.-C. Shen, Y.-H. Chen, and A.-Y. Wu, "Low-complexity sinusoidal-assisted EMD (SAEMD) algorithms for solving mode-mixing problems in HHT," *Digit. Signal Process.*, vol. 24, pp. 170–186, Jan. 2014.
[51] Y. Gao, G. Ge, Z. Sheng, and E. Sang, "Analysis and solution to the mode mixing phenomenon in EMD," in *Proc. Congr. Image Signal Process.* Washington, DC, USA: IEEE Computer Society, 2008, pp. 223–227.
[52] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980