Short-term wind speeds prediction of SVM based on simulated annealing algorithm with Gauss perturbation

Short-term wind speeds prediction of SVM based on simulated annealing algorithm with Gauss perturbation

Yan Chen[1,*], Rui Chen[1], Chunyan Ma[1], Peiran Tan[2]

[1]College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, Shanxi Province, China;

[2]State Grid Shanxi Metering Center, Taiyuan 030032, Shanxi Province, China;

[*]Correspondence: chenyanlxq@163.com; Tel: +86-135-9318-2148.

Abstract. Wind speed prediction is an efficient means to reduce the downside effects of large-scale wind power generation on the grid, but the behaviour of wind speeds is nonlinear and non-stationary, which yields adverse challenge for its prediction. This work proposes a method of prediction for short-term wind speed, which makes Simulated Annealing Fruit fly Optimization Algorithm based on Gaussian Disturbance (GDSAFOA) to optimize the Support Vector Machine (SVM). In the method, Grey Relational Analysis (GRA) is used to select the factors which influence wind speeds prediction. A time series of wind speeds is decomposed by the Ensemble Empirical Mode Decomposition (EEMD). The wind speeds predication is the linear combination of the SVM and the dynamic neural network model based on the nonlinear autoregressive models with exogenous inputs (NARX). This method is applied for the model with wind speeds data measured from a wind farm in China's Shanxi Province, where results exhibit that the proposed method is feasible and competitive.

1. Introduction

The high precision prediction of winds speed in wind farm is an effective measure to ensure power system to operate in the way of safe, economics, reliability [1]. Currently, commonly used prediction methods are: spatial correlation method [2], neural network method [3-5], time series method [6], wavelet algorithm [7], empirical mode decomposition EMD (Mode Decomposition Empirical) algorithm [8], etc. As the various methods have different advantages and disadvantages, single method of prediction is difficult to meet the requirements of power grid operation.

In this paper, the Ensemble Empirical Mode Decomposition (EEMD) algorithm is used to decompose a time series of wind speeds into a stationary time series. Compared with the wavelet and EMD algorithms, the EEMD algorithm that adds small amplitude white noise is not only stronger in adaption but also has the advantage of anti-modal aliasing. In order to make use of the advantages of different algorithms in prediction, we propose support vector machine (SVM) [9] and the nonlinear auto-regressive exogenous models (NARX) for linear combination forecasting in this paper. In this combinatorial model, the reasonable penalty parameters that provide to SVM have a great impact on the prediction. Consequently, this paper proposes to use the Simulated Annealing Fruit fly Optimization Algorithm based on Gaussian Disturbance (GDSAFOA) to optimize the SVM. The algorithm not only introduces the inferior solution with certain probability, but also makes the Gaussian variation to the optimal solution, which can make up the shortcoming that the Fruit Fly Optimization Algorithm (FOA) easily is trapped in the local minimum.

2. Grey Relational Analysis

The basic idea of grey correlation method is to transform discrete observations into continuous curves, and to judge the degree of correlation by comparing the changing trends of geometric shapes of each curve. Firstly, the reference sequence and comparison sequence are determined. The reference sequence is X_0, X_0=[x(1),x(2),x(3),…,x(m)]. They represent the information characteristics of the sequence. The sequence of comparison was X_1, X_2, ⋯, X_n. Dimensionless processing of reference sequence and comparison sequence is carried out:

$$x_i'(k) = x_i(k) / x_i(1) \qquad i = 0,1,2,\mathrm{L}\ ,n; k = 1,2,\mathrm{L}\ ,m \tag{1}$$

The degree of grey correlation is the difference of geometric shapes between curves. The difference between curves is taken as the measure of the degree of correlation:

$$\xi_{0i}(k) = \frac{\min\limits_i \min\limits_k \left|x_0'(k) - x_i'(k)\right| + \rho \max\limits_i \max\limits_k \left|x_0'(k) - x_i'(k)\right|}{\left|x_0'(k) - x_i'(\mathrm{k})\right| + \rho \max\limits_i \max\limits_k \left|x_0'(k) - x_i'(k)\right|} \tag{2}$$

ξ_{0i} is the correlation coefficient between X_0 and X_i at K point, where i=0, 1, 2,⋯, n; k=1, 2,⋯, m; usually ρ=0.5. By averaging the discrete correlation number, the correlation degree of each curve can be obtained:

$$r_{0i} = \frac{1}{m}\sum_{k=1}^{m} \xi_{0i}(k), i = 1,2,\mathrm{L}\ ,n \tag{3}$$

3. Fruit Fly Optimization Algorithm

FOA is a new evolutionary algorithm based on foraging characteristics of fruit flies. Compared with other swarm intelligence algorithms, FOA has some prominent advantages, such as fast operation speed, few parameters, and excellent global search performance. However, the development of FOA is also restricted by the lack of local search ability. In view of the FOA defects of the need to improve, it is proposed to introduce the inferior solution to the FOA with a certain probability in this paper, and the optimal value of the FOA is Gaussian variation to help the FOA to jump out of the local extreme smoothly.

Drosophila foraging behavior is divided into two steps, first search for food sources by the keen sense, flying in the vicinity of food sources, then find the best position in the companion, and fly to the location.

Based on the feeding characteristics of fruit flies, FOA steps are as follows:

(1) Setting the initial position of the fruit fly is (X_{axis}, Y_{axis}).

(2) Drosophila with olfactory organs find food, its position is (X_i, Y_i) at this time.

$$\begin{cases} X_i = X_{axis} + R \\ Y_i = Y_{axis} + R \end{cases} \tag{4}$$

Where R is a real number.

(3) Computing the value of S_i that is defined as the taste concentration decision, the value of S_i is equal to $Dist_i$ that is defined as the reciprocal of the distance between the fruit fly and the origin.

$$\begin{cases} Dist_i = \sqrt{X_i^2 + Y_i^2} \\ S_i = \dfrac{1}{Dist_i} \end{cases} \tag{5}$$

(4) Taking the value of S_i into taste concentration decision function, we can obtain the $smell_i$ that is defined as the current taste concentration.

$$smell_i = function(S_i) \tag{6}$$

Where function is the fitness function.

(5) Computing the optimum concentration and the corresponding location of the fruit fly at this time, and record. Afterwards, the drosophilae fly to the location by virtue of the visual organ.

$$\begin{cases} [bestsmell, bestindex] = \min(smell_i) \\ smellbest = bestsmell \\ X_{axis} = X(bestindex) \\ Y_{axis} = Y(bestindex) \end{cases} \tag{7}$$

(6) If the current is superior to the parent, best *bestsmell* and position of the fruit fly is updated. When the stop condition is satisfied, the FOA is finished, otherwise it will turn to step (2).

4. Simulated Annealing Fruit Fly Optimization Algorithm Based on Gaussian Disturbance

Simulated Annealing Algorithm (SAA) is a heuristic combinatorial optimization algorithm, which is very suitable for complementary with FOA. The Metropolis criterion not only allows the SAA to accept the high quality solution, but also to accept the inferior solution with a certain probability, and the probability decreases with the temperature decreasing. In this paper, the SAA is introduced into the FOA, and Gaussian perturbation is carried out in the neighbourhood of the optimal solution to further improve the algorithm optimization ability.

The algorithm steps are as follows:

(1) Setting fruit flies the initial position is (X_{axis}, Y_{axis}), and SAA initial temperature is T.

(2) Computing the current concentration of taste and fruit flies location (X_i, Y_i).

(3) Compute he best value of the concentration of taste $smellbest_a$ and the corresponding fruit flies position (X_a, Y_a).

(4) Compute $smellbest_b$ that is defined as the optimum taste concentration of the new fruit fly population and the corresponding fruit fly population location (X_b, Y_b).

(5) If the variation $\Delta E < 0$ that caused by the change of position of concentration happens, then record the position (X_b, Y_b) where best concentration of fruit fly locates; If $exp(-\Delta E/T) > \alpha$, α is [0,1], then (X_b, Y_b) is also recorded as the best concentration of fruit flies. In which, ΔE means:

$$\Delta E = smellbest_b - smellbest_a \tag{8}$$

(6) Generate a new set of positions (X_c, Y_c) where make a Gaussian disturbance, according equation (6). The corresponding taste concentration is $smell_c$. If $semll_c < smellbest$, then $smellbest = semll_c$, (X_{axis}, Y_{axis}) = (X_c, Y_c). The operation of annealing temperature is performed.

$$(X_c, Y_c) = (X_{axis}, Y_{axis}) + a \oplus \varepsilon \tag{9}$$

In which, ε is the matrix with the same order of (X_c, Y_c); $\varepsilon_{ij} \sim N(0,1)$; a is a constant; $\oplus$ is the point-to-point multiplication.

(7) Iteration is Repeated until termination conditions is met.

5. EEMD Algorithm Principles

The non-stationary wind power signal can be decomposed into several different scales of Intrinsic Mode Functions (IMF) by EEMD algorithm, and each IMF should meet the following two conditions:

(1) The number of extreme points and zero points of the signal is at most one difference.

(2) The mean value tends to zero.

The steps of the algorithm are as follows:

(1) First, determine all extreme points of $x(t)$ that is defined as the wind power signal, and use the cubic spline function to fit the upper and lower envelopes, then calculate the mean value m_1, and compute $h_1(t)$ that is defined as the difference between the original signal and the mean m_1.

(2) If $h_1(t)$ satisfies the IMF condition, $h_1(t)$ is regarded as the first IMF component. If not, it will be regarded as the original signal. And Step (1) is repeated until the difference $h_{1n}(t)$ appears, which meets the conditions of IMF, then regard it as an IMF. Denoted by $c_1(t) = h_{1n}(t)$.

(3) According to the step (1), $c_1(t)$ is separated from the original signal, and $r_1(t)$ is obtained:

$$r_1(t) = x(t) - c_1(t) \tag{10}$$

(4) The remaining signal $r_1(t)$ serves as a new original signal, which repeats steps of (1), (2), (3) to get the rest of the N - 1 of the IMF component and $r_n(t)$, when the remaining component $r_n(t)$ is monotone function, decomposition is terminated, and the original signal can be expressed as:

$$x(t) = \sum_{n=1}^{N} c_n(t) + r_N(t) \tag{11}$$

Where, $c_n(t)$ is IMF component; $r_n(t)$ is residual component.

The decomposition processes of EEMD are as follows:

(1) Add white Gaussian noise to the original signal;

(2) According to the EMD decomposition algorithm, $x(t)$ is decomposed to get a series of IMF signals;

(3) Repeat step (1) and step (2) for n times, when different white noise is added. Then calculate the mean of the IMF that acquired for the process of decomposition in each time.

6. SVM

Support Vector Machine (SVM) is a machine learning algorithm based on small sample method, which maps the input sample space to the high-dimensional feature space by nonlinear kernel function. The non-linear relationship between input and output is obtained through the optimal classification surface. The regression function of SVM is:

$$f(x) = \sum_{i=1}^{n} (a_i - a_i^*) k(x_i, x) + b \tag{12}$$

Where, n is for training sample; a_i and a_i^* are as the Lagrange multiplier, a_i and a_i^* are parameters corresponding vector x_i that is defined as support vector; b is constant; $K(x_i, x)$ for the kernel function, the expression is:

$$K(x_i, x) = \exp \frac{\|x_i - x\|}{2\delta^2} \tag{13}$$

Where δ is the parameter of the kernel function.

7. NARX

NARX network has been more and more widely used in nonlinear prediction system, especially when dealing with time series, the unique characteristic of dynamic closed loop feedback can be made use of. NARX network mainly contains four components, which namely are input layer, hidden layer and output layer, the time delay. Because of feedback layer and the delay layer, so NARX network has a memory function, and can feedback to the input of a state of the past period of time. The NARX network is defined as:

$$\begin{aligned} y(t) = f[&y(t-1), y(t-2), \mathrm{L}\ , y(t-n_y), \\ &u(t-1), u(t-2) \mathrm{L}\ , u(t-n_u)] \end{aligned} \tag{14}$$

Where, $y(t - n_y)$ is the output feedback value of the network at some time; $u(t - n_u)$ is the input value of the network at a certain time.

The ith hidden layer node h_i of NARX is:

$$h_i = g(\sum_{r=0}^{R} w_{ir} x_r + \sum_{l=0}^{L} w_{il} y_l + b_i) \tag{15}$$

Where, g is the hidden layer activation function; R is the input delay; L is the output delay; w_{ir} is the weights between the ith hidden layer node and the rth delay point of the input point; w_{il} is the weight between the ith hidden layer node and the lth delay point of the output point; x_r is the rth delay point of input signal; y_l is the lth delay point of the output feedback signal.

The output point of the jth output layer of NARX is defined as:

$$O_j = \sum_{j=1}^{J} w_{ji} h_i + \beta_j \tag{16}$$

Where, w_{ji} is the weight of between the jth output point and the ith hidden layer point; β_j is the threshold of the jth output node.

8. The Flow of Prediction Algorithm

The flow of prediction algorithm is shown in figure 1.

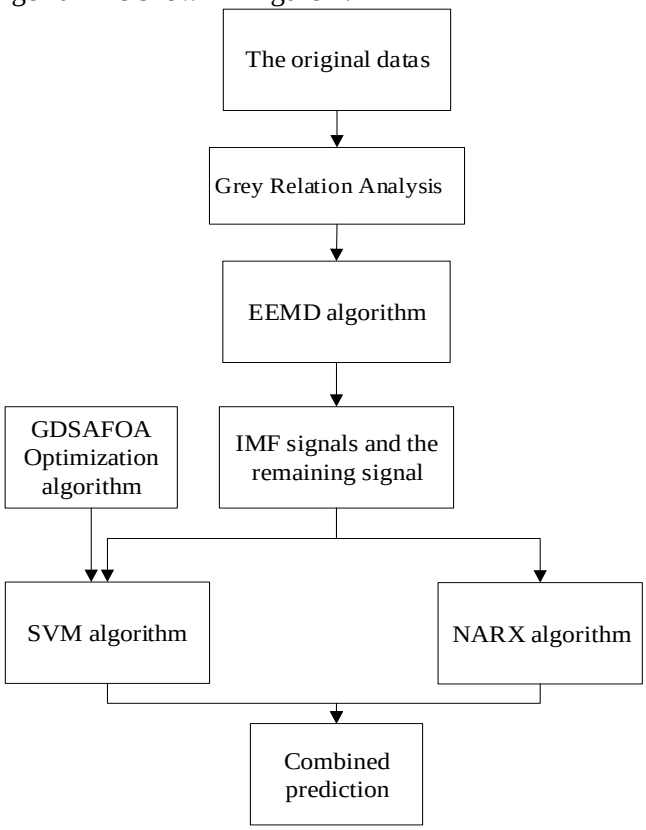


Figure 1. Flow chats of prediction algorithm

The EEMD algorithm is used to decompose the continuous time series and is obtained to stable IMF signals and a residual signal $R(t)$. Then, the combination of NARX and SVM is optimized by GDSAFOA, which are used to predict the wind speeds and are assigned different weight value. In this paper, Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) are used to evaluate the prediction result.

$$\mathrm{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{v_i - v_i'}{v_i} \right| \tag{17}$$

$$\mathrm{RMSE} = \left(\frac{1}{N} \sum_{i=1}^{N} (v_i - v_i')^2 \right)^{1/2} \tag{18}$$

Where, N is the number of points of prediction; v_i is the real value; v_i' is the value of prediction.

9. Example analysis

Dates of wind speeds are shown in figure 2. Dates are used to take the experiment of simulation, which were selected from June 1, 2015 to July 30, 2015, and a total of 5760 in a wind farm of Shanxi Province. The sampling period of wind power is 10 minutes.

Variation of wind speeds is influenced by many factors. In this paper, some main factors are selected based on GRA, including wind speeds, wind direction, humidity, temperature and pressure, and their value of r_{0i} are all greater 0.5 shown on table 1.

Table 1. The Relevance Between wind speeds and Meteorological Data

Factors of effecting wind speeds prediction	The value of r_{0i}
Previous Wind speeds	0.9563
Wind direction	0.6741
Humidity	0.5223
Temperature	0.6866
Pressure	0.5084

It can be seen from figure 2 that the sample of wind speeds has obvious mutation. So such training samples of wind speeds may bring interference to the result of prediction.

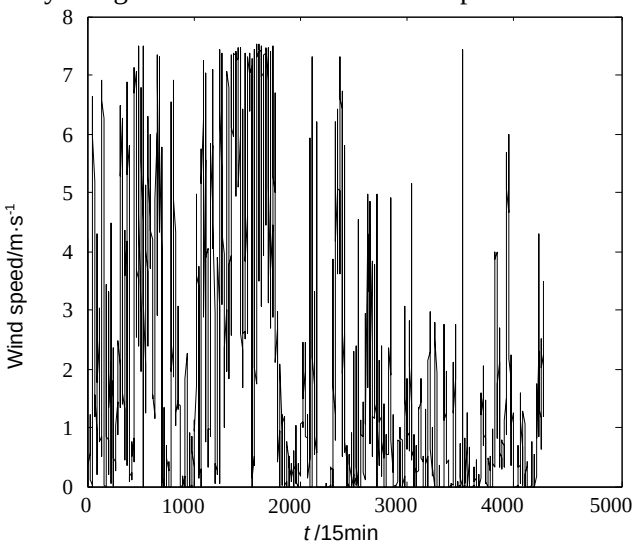


Figure 2. Dates of wind speeds

In order to prevent this situation, this paper adopts the EEMD algorithm that is used to decompose the curve of wind speeds into 6 IMF signals and a residual signal $R(t)$, as shown in figure 3. Although the IMF1 and IMF2 signals are more volatile, they have been isolated from other IMF signals, which can reduce the impact of the nonlinear characteristics of the signal to the predicted results; Compared with the IMF1 and IMF2 signals, IMF3~IMF6 and the residual signal $R(t)$ are relatively more gentle, and more easy to predict. A satisfied outcome of forecast will be acquired after sum the prediction value of the IMF and the $R(t)$.

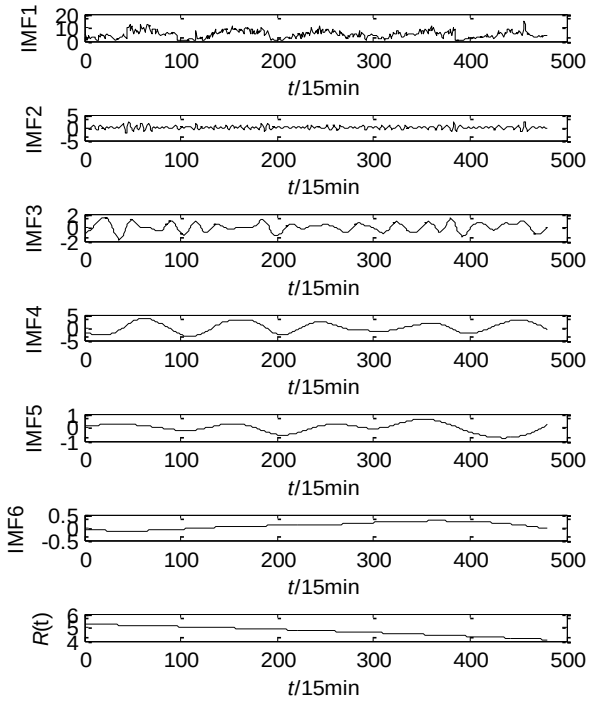
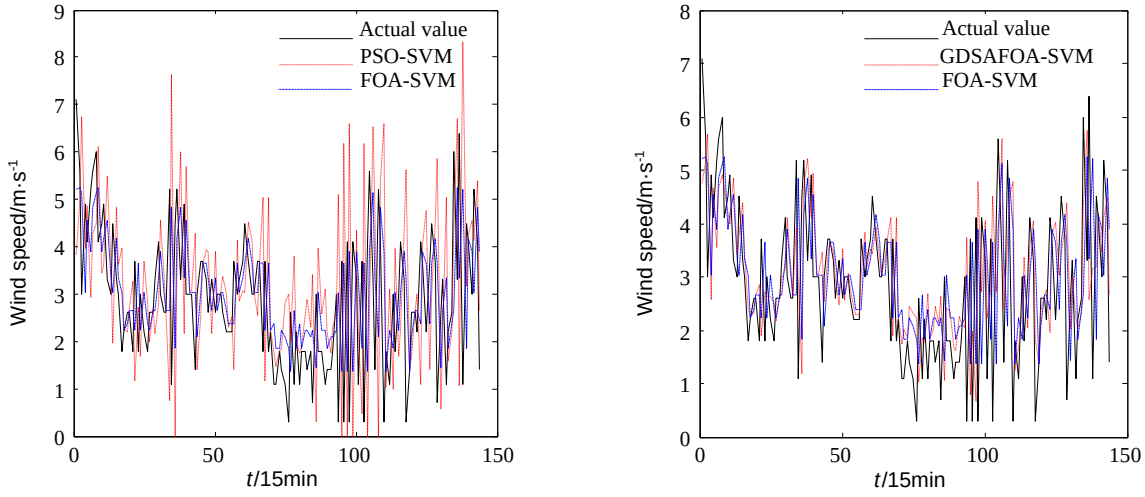


Figure 3. EEMD decomposition

Figure 4 is forecasting result of wind speed: (a) is the forecasting result comparison of PSO-SVM, FOA-SVM and actual value, (b) is the forecasting result comparison of FOA-SVM, GDSAFOA-SVM and actual value.



(a) (b)

Figure 4. Forecasting result of wind speed

Table 2 is the error result of PSO-SVM, FOA-SVM and actual value. Table 3 is the error result of FOA-SVM, GDSAFOA -SVM and actual value. It is obviously that the prediction curve which is calculated by PSO-SVM deviates the actual curve is much farther than the curve which is calculated by GDSAFOA-SVM. Especially, the degree of amplitude deviation is improved by the GDSAFOA-SVM algorithm in some extreme value point. It can be seen from the Table 2 and Table 3 that the MAPE and RMSE of SVM is respectively 11.37 and 0.2118, which are decreased nearly doubled than PSO-SVM. Furthermore, the MAPE and RMSE of GDSAFOA-SVM is 5.67 and 0.0532 that is smaller.

Table 2. Errors of different models

Mode	MAPE/%	RMSE(m/s)
FOA-SVM	11.37	0.2118
PSO-SVM	20.25	0.3443

Table 3. Errors of different models

Mode	MAPE/%	RMSE(m/s)
FOA-SVM	11.37	0.2118
GDSAFOA-SVM	5.67	0.0532

The hidden layer node of NARX is 10 and the input delay is 6. The result of NARX and the combined prediction model of NARX and SVM are shown in the figure 5 and figure 6.

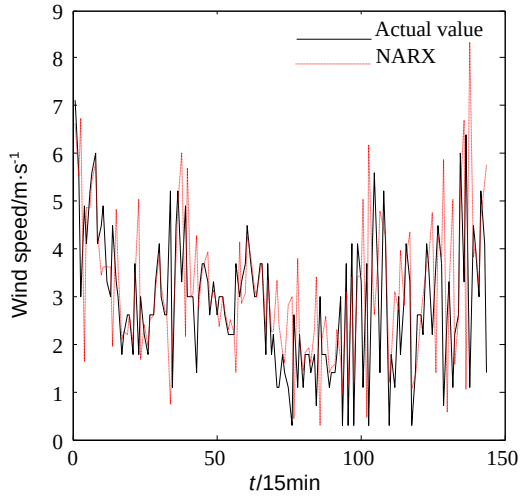


Figure 5. Forecasting result of wind speed

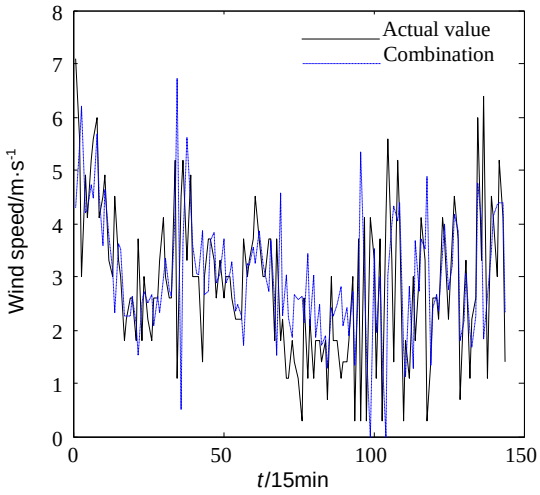


Figure 6. Forecasting result of wind speed

The index value of prediction accuracy can be known from Table 4, it can be seen that MAPE and RMSE of NRAX network are increased compared with SVM. Although we can see that the GDSAFOA-SVM is more suitable for prediction than the NARX network in this paper, GDSAFOA-SVM doesn't has the closed-loop dynamic characteristic which NARX has. NARX and SVM are combined by the linear way and the corresponding weights are allocated in the way of minimizing the sum of squares of errors. It can be seen from Table 4 that the error of the combined model is smaller than the errors of single model SVM or NARX. The NARX describes the nonlinear system of prediction from a dynamic angle. On the contrary, the SVM represents the system from a static point of view.

It is verified by experiment that the dynamic closed-loop feedback characteristics of NARX and SVM can be well complemented. When they can be assigned the appropriate weight, they can maximize the advantages of their algorithms. The combination of prediction error is less than any single prediction error.

Table 4. Errors of different models

Mode	MAPE/%	RMSE(m/s)
NARX	6.66	0.0732
GDSAFOA-SVM	5.67	0.0532
Combined	3.32	0.0360

10. Conclusions

The parameters of SVM, which are optimized by GDSAFOA, can jump out of the local extremum quickly. the GDSAFOA algorithm is easy to promote because of the simple structure and the rapid

convergence. Compared with the unmodified FOA, GDSAFOA algorithm with poor quality solution at a certain probability is more suitable for optimizing SVM.

SVM that is complemented by the dynamic closed-loop feedback characteristics of NARX make an admirable performance, Since the NARX and SVM describe the prediction system from dynamic angle and static point of view respectively. When different methods can be assigned the appropriate weight, they can maximize their advantages. According to the consequence, we can see that the combination of prediction error is less than any single prediction error.

Funding
This research is supported by the Natural Science Foundation of Shanxi Province (201701D121127) and by the Key Research and Development Program of Shanxi Province (201803D221028-2).

Acknowledgments
Thanks to Wentao Wu, Xiangnan Hou, Zaihe Shen and Yanzhao Hao for their assistance during the research.

References
[1] Feng S L, Wang W S, Liu C and Dain H Z 2011 Short term wind speed prediction based on physical principle *Acta Energ. Sol. Sin.* **32** 611-15
[2] Chen N Y, Qian Z, Meng X F and Meng K F 2013 Multi-step ahead wind speed forecasting model based on spatial correlation and support vector machine *Transactions of China Electrotechnical Society* **28** 15-21
[3] Shi H T, Yang J L, Ding M S and Wang J M 2011 A short-term wind power prediction method based on wavelet decomposition and BP neural network *Automation of Electric Power Systems* **35** 44-7
[4] Wang J J, Zhang W J, Wang J Z, Han T T and Kong L B 2014 A novel hybrid approach for wind speed prediction *Inf. Sci.* **273** 304-18
[5] Gnana Sheela K and Deepa S N 2013 A review on neural network models for wind speed prediction *Wind Engineering* **37** 111-24
[6] Liu Mingfeng and Xiu Chunbo 2013 Hybrid prediction method for wind speed series based on ARMA and neural network *J. Cent. South Univ.* **44** 16-20
[7] Xiao Q, Li W H, Li Z G, Liu J L and Liu H Q 2014 Wind speed and power prediction based on improved wavelet-BP neural network *Power System Protection and Control* **42** 80-6
[8] Wang J J, Zhang W J, Li Y N, Wang J Z and Dang Z L 2014 Forecasting wind speed using empirical mode decomposition and Elman neural network *Appl. Soft. Comput*. **23** 452-9
[9] Liu D, Niu D X, Wang H and Fan L L 2014 Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm *Renew. Energy* **62** 592-7