Short-Term Wind Power Forecasting Using Support Vector Regression and Harmony Search Algorithm

Yann-Chang Huang^b, Chao-Ming Huang^{a*}, and Kun-Yuan Huang^b

^bDepartment of Electrical Engineering, Cheng Shiu University, No.840, Chengcing Rd., Kaohsiung 833, Taiwan, R.O.C.

^aDepartment of Electrical Engineering, Kun Shan University, No.195, Kunda Rd., Tainan 710, Taiwan, R.O.C.

*Corresponding Author: h7440@ms21.hinet.net

Abstract

This study proposes a hybrid method for short-term wind power forecasting using support vector regression (SVR) and harmony search algorithm (HSA). The proposed method comprises data classification, training and forecasting stages. A k-means clustering algorithm is used in the data classification stage to classify the historical wind power generation data into various classes. In the training stage, five training models are established, according to various wind speed forecasts of the Taiwan Central Weather Bureau (TCWB). Each training model is constructed using a SVR, for which the control parameters of each SVR are optimized using a HSA. In the forecasting stage, an adequate forecasting model is selected according to the wind speed forecasts of the TCWB every three hours. Testing on a practical wind power generation system, the results show that the proposed method provides better forecasting accuracy than the ANN method over one-year testing data.

Keywords: wind power forecast, k-means clustering, support vector regression, harmony search algorithm.

1. Introduction

Wind power generation is a clean and renewable energy that generates low pollution electricity and provides diversity in electricity supply. There is a great relation between wind power output and wind speed. Due to the irregularity of the wind speed, wind power generation has an intermittent nature that brings a great challenge to the energy dispatchers.

To allow efficient planning for distributed generation (DG), the forecast of wind power generation is necessary for a grid operator. Accurate short-term wind power forecast improves the wind power penetration level, increases system reliability, reduces operating costs and allows for efficient load management strategies. However, because it is affected by amount of wind speed, wind power generation is highly uncertain and difficult to predict accurately. To make effective use of wind power generation, accurate forecasting of wind power output is essential for a DG system.

Since the wind power output is greatly affected by amount of wind speed, many studies focus on predicting wind speed, according to the historical meteorology information. The forecasted wind speeds are then converted to wind power outputs using a power curve model. These techniques include adaptive wavelet neural network (AWNN) method [1], wind vector prediction method [2], smoothing technique [3] and adaptive local learning technique [4]. Other studies use historical data and associated meteorology information to directly forecast wind power output. These studies use radial base function neural network (RBFNN) and fuzzy methods [5], ridgelet neural network method [6] and wavelet transformation and SVR techniques [7].

Most existing studies, as mentioned above, improve the forecasting accuracy for wind power output. However, the ANN based methods [1][5-6] require more complicated structure to capture the non-linear characteristic between input and output datasets, though it has good approximation for a wide range of nonlinear functions. The fuzzy method [5] uses rule-based knowledge to design an efficient

inference engine, but it is difficult to draw conclusions from a large amount of linguistic terms. Through a multi-resolution technique, the wavelet-based method [7] decomposes the original time series data into a stationary series in different frequency bands of subsets. The structure of this method is complicated and requires long execution time. The SVR method [7] has been proven to be an effective method for regression prediction. However, the parameters of SVR are often tuned by the operators' experiences.

To achieve more accurate forecasting, this paper proposes a hybrid method which combines SVR and HSA to predict the wind power output for a 3-hour ahead in steps of 15 minutes. First, a k-means clustering algorithm is used to classify the historical wind power generation data into various classes. Five HSA-based SVR models are then established to train the collected data. Since the HSA provides an efficient scheme to optimize the SVR parameters, the selected forecasting model obtained from the wind speed forecasts every three hours by the TCWB is then used to produce accurate forecasts for the wind power generation.

The remainder of this paper is organized as follows. In Section II, the characteristics of wind power generation are briefly reviewed. Section III describes the proposed approach to produce future 3 hours wind power forecast in steps of 15 minutes. Section IV presents the forecasting results for ANN and proposed methods. Finally, conclusions are given in Section V.

2. Characteristics of wind Power Generation

The output of wind power generation is affected by meteorological conditions, such as the wind speed and air density, as follows [8]:

$$P_{wind} = \frac{1}{2} \rho A v^3 C_p \quad (1)$$

where P_{wind} is the power output of a wind turbine, A is the rotor area of a wind turbine (m^2), ρ is the density of air (kg/m^3), v is the wind speed (m/s) and C_p is the utilization coefficient of wind energy. The conversion efficiency of a wind turbine is about 20~40%.

As shown in (1), power output of a wind turbine is proportional to the cube of the wind speed. There are three types of method to control the turbine generation: pitch control, tip control and stall control. Fig. 1 shows the wind power curve for a 2 kW wind turbine generator. There are some control devices to limit the wind power generation

according the different wind speeds. The speed to start the turbine generation is called the cut-in speed, while a cut-out speed is set to prevent the turbine suffering from damage. The speed between 12 m/s and cut-out speed is called the rated speed, which indicates the output capability of a wind turbine. Table 1 shows the wind speed as well as wind grade classification provided by TCWB. Consideration of the rated output of a wind turbine, the wind grades from 2 to 6 are used to develop the training models.

Since the power output of a wind power system is intrinsically unstable and has an intermittent nature, providing an accurate forecast of wind power output is important for a DG system. This paper presents a hybrid method to make more accurate forecast for PV power output. Details of the proposed scheme are given in the next section.

Fig. 1 Wind power curve for a 2kW wind turbine generator

Table 1 Wind speed classification

Grade	Verbal grade	Range of wind speed (m/s)
0	windless	0.0-0.2
1	soft wind	0.3-1.5
2	slight wind	1.6-3.3
3	breeze wind	3.4-5.4
4	moderate wind	5.5-7.9
5	cool breeze	8.0-10.7
6	strong wind	10.8-13.8
7	powerful wind	13.9-17.1
8	gale wind	17.2-20.7
9	violent wind	20.8-24.4
10	very strong wind	24.5-28.4
11	storm wind	28.5-32.6
12	typhoon wind	32.7-36.9
more than 12	no definition	more than 37.0

3. The Proposed Approach

3.1 Data classification

The collected historical wind power generation patterns are classified using a k-means clustering algorithm.

K-means clustering was first proposed by Lloyd in 1982 [9]. It is a method of vector quantization and is used for cluster analysis in data mining. Given a set of observations ($X_1, X_2, \dots, X_n$), the k-means clustering partitions n observations into k clusters, as follows:

$$\arg \min \sum_{i=1}^k \sum_{j=1}^n w_{ji} \|X_j - R_i\|^2 \quad (2)$$

where X_j is the j th observation, R_i is the i th clustering center, w_{ji} is a synaptic weight and $\|\bullet\|$ is the Euclidean distance. R_i and w_{ji} are respectively expressed as follows:

$$R_i = \frac{\sum_{j=1}^n w_{ji} X_j}{\sum_{j=1}^n w_{ji}} \quad (3)$$

$$w_{ji} = \begin{cases} 1, & \text{if } \|X_j - R_i\| \leq \|X_j - R_m\|, \forall m \neq i \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

As shown in (3), the k-means clustering partitions n observations into k clusters ($k \leq n$), in order to minimize the within-cluster sum of the squares. This algorithm assigns using the least sum of the squares of the Euclidean distance to achieve the near global optimum.

3.2 The creation of training models

As described in the data classification section, a k-means clustering algorithm provides information on similar days and their associated wind speed types. A similar day is defined as having the same wind speed type. In this stage, the SVR evolved using a HSA is used to train the collected input-output datasets.

Since the TCWB provides the verbal wind speed forecasts for the next 3 hours, in terms such as the slight wind, breeze wind, moderate wind, cool breeze and strong wind, the training stage consists of five different models.

Figure 2 shows the structure of the training stage. The inputs for each training model contain the actual wind power output for future 3 hours in steps of 15 minutes on the latest similar day, a prediction of the wind speed for future 3 hours, a prediction of the wind direction for future 3 hours and time zone for future 3 hours. The outputs are the wind power forecasts for future 3 hours in steps of 15 minutes. The five training models have the same training structure. In this paper, a HSA-based SVR is used to train the input/output datasets. The SVR and the HSA-based SVR models are hereby described.

(1) SVR

SVR was first introduced by Vapnik in 1995 [10]. It is a

technique for data classification and regression analysis. SVR can be roughly divided into linearly separable SVR, linearly inseparable SVR and nonlinear SVR. In this paper, a nonlinear SVR is used to find the best hyperplane from n -dimensional spaces to capture the nonlinear mapping between input and output data. Fig. 3 shows the hyperplane of a SVR. The SVR used to divide the data into a high dimension space is described as follows [11].

$$\begin{aligned} \text{minimize} \quad & f(u, \psi) = \frac{1}{2} u^T u + \sigma \sum_{k=1}^n \psi_k \\ \text{subject to} \quad & \begin{cases} Q_k(u^T H(x_k) + h) \geq 1 - \psi_k \\ \psi_k \geq 0 \end{cases} \end{aligned} \quad (5)$$

where u is an unit normal vector to the hyperplane, h is the distance from the origin to the hyperplane, n is the number of training data, ψ_k is the k th slack variable, σ is a weight for penalty function, x_k is the k th input data set and $H(x_k)$ is a nonlinear mapping function. The term $\sum \psi_k$ is a penalty function.

The Lagrange equation for (5) can be expressed as $F_{Lagrange}(u, h, \psi, \lambda, \gamma)$

$$= f(u, \psi) - \sum_{k=1}^n \lambda_k (Q_k(u^T H(x_k) + h) - 1 + \psi_k) - \sum_{k=1}^n \gamma_k \psi_k \quad (6)$$

where λ_k and γ_k are the Lagrange multipliers.

Fig. 2 Structure of training stage.

Fig. 3 The hyperplane of a SVR.

Minimization of (6) by partially differentiating u , h and ψ , respectively, yields

$$u = \sum_{k=1}^n \lambda_k Q_k H(x_k) \quad (7)$$

$$\sum_{k=1}^n \lambda_k Q_k = 0 \quad (8)$$

$$\sigma - \lambda_k - \gamma_k = 0 \quad (9)$$

Substituting (7)-(9) into (6) forms the dual optimization problem as follows:

$$\text{maximum } F_{dual}(\lambda) = -\frac{1}{2} \sum_{k,l=1}^n Q_k Q_l H(x_k)^T H(x_l) \lambda_k \lambda_l + \sum_{k=1}^n \lambda_k \quad (10)$$

$$\text{subject to } \begin{cases} \sum_{k=1}^n \lambda_k Q_k = 0 \\ \sigma \geq \lambda_k \geq 0, \quad k = 1, 2, \dots, n \end{cases}$$

The term $H(x_k)^T H(x_l)$ in (10) is defined as the kernel function, $K(x_k, x_l)$, and must satisfy the condition as follows.

$$\int K(x_k, x_l) g(x_k) g(x_l) dx_k dx_l \geq 0 \quad (11)$$

where $g(x)$ is an integrable function. The kernel function used in this paper is the radial basis function as follows:

$$K(x_k, x_l) = e^{-\left(\frac{\|x_k - x_l\|^2}{\varepsilon^2}\right)} \quad (12)$$

where ε is the dilation parameter.

The parameters, σ in (10) and ε in (12), determine the mapping properties of SVR. In this paper, the HSA algorithm is used to automatically tune these two parameters.

(2) HSA for tuning the parameters of SVR

A HSA [12] operates similarly to the process of musical improvisation, where musicians tune the pitches of their instruments to attain better harmony. It is a population-based random search algorithm and has been applied to various fields. A HSA used to tune the parameters of SVR is described as follows.

1) Initialization

In this step, an initial harmony matrix (HM) is randomly generated as follows:

$$HM = \begin{bmatrix} z_1^1 & z_1^2 & \dots & z_1^{S-1} & z_1^S \\ z_2^1 & z_2^2 & \dots & z_2^{S-1} & z_2^S \end{bmatrix}^T \quad (13)$$

where z_l^j is the j th weight of penalty function, and $j=1, 2, \dots, S$, S is the size of the harmony matrix, z_2^j is the dilation parameter of the j th kernel function.

As shown in (13), each column vector ($[z_l^j, z_2^j]$) represents a feasible solution which is randomly generated as follows:

$$z_h^j = z_{\min}^j + rand \times (z_{\max}^j - z_{\min}^j), \quad (14)$$

where $h=1, 2$ and $j=1, 2, \dots, S$, $z_{\min}^j$ and $z_{\max}^j$ are the lower and upper bounds of z^j and $rand$ is a uniform distribution value between 0 and 1.

2) Determination of the best hyperplane

For each vector solution obtained from (14), the hyperplane is evaluated using (10). The solution with the maximum value in HM is regarded as the optimal solution for the SVR.

3) Creation of a new HM

A new harmony matrix, $Z'=[z_1', z_2']$, is generated using the memory consideration, pitch adjustment and random selection processes. In the memory consideration, the decision variables in Z' are updated by the difference ($z^1 - z^S$). Then the harmony memory considering rate ($HMCR$) is used to improve the diversity of the solution vector, as follows:

$$z_i' = \begin{cases} z_i^1 \in [z_i^1, z_i^2, \dots, z_i^S], & \text{if } rand \leq HMCR \\ z_i' \in Z_i, & \text{otherwise} \end{cases} \quad (15)$$

where $HMCR \in [0,1]$ is set by users. If $HMCR$ is set at 0.85, the HSA has 85% probability to choose decision values from historically stored values and 15% probability to choose decision values from the entire range.

For a new harmony matrix obtained using (15), the pitch adjustment is checked to improve the vector solution, as follows:

$$z_i' = \begin{cases} z_i' \pm rand \times B_w & \text{if } rand \leq R_{par} \\ z_i' & \text{otherwise} \end{cases} \quad (16)$$

where $R_{par} \in [0,1]$ is the pitch adjusting rate and B_w is the bandwidth of the variations.

4) Parameter update

In the parameter adjustment process, R_{par} is tuned to improve the performance of the HSA, as follows:

$$R_{par}(t) = R_{par}^{\min} + \frac{R_{par}^{\max} - R_{par}^{\min}}{I_p} \times t \quad (17)$$

where $R_{par}^{\min}$ and $R_{par}^{\max}$ are the lower and upper bounds of R_{par} , t is the number of iteration and I_p is the number of improvisations so far in the optimization process. The improvisation means that the best hyperplane of the offspring harmony matrix is better than that of the parent harmony matrix.

5) Selection of the best vector solution

After adjusting the parameters, the size of *HM* is kept at 2*S*. *S* sets of the better row vector in the harmony matrix are then selected.

Processes from 3) to 5) are repeated, until the maximum number of iterations is reached. The best vector solution in *HM* is regarded as the optimal hyperplane for the SVR.

4. Numerical Results

4.1 Data description and evaluation criteria

The proposed method was used to forecast the power output for a 2 kW wind turbine system. The forecasting period is 3-hour ahead in steps of 15 minutes. To verify the superiority of the proposed approach, the traditional ANN method was also tested using the same database. The practical wind power generation data, wind speed and wind direction data are sampled every 5 seconds, which are collected from January to December in year 2013. The 15-minutely wind power output data are then obtained by averaging the data collected over 15 minutes.

To verify the proposed approach, the data for the last day of each month was used for testing and that for the other days are for training. Note that the traditional ANN and proposed methods have the same number of training models and same input/output variables of each model, while the parameters remained to be determined independently.

To evaluate the forecasting performance of each method, the criterion of root mean square error (RMSE) which measures the average magnitude of the errors was used as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (P_f - P_t)^2}{N}} \quad (18)$$

where P_f is the forecasting value, P_t is the actual value and N is the number of forecasting points.

Besides *RMSE*, the mean relative error (MRE) was also used as an evaluating index for trend estimation, as follows:

$$MRE = \frac{1}{N} \sum_{i=1}^N \frac{|P_f - P_t|}{P_{total}} \times 100\% \quad (19)$$

where P_{total} is the capacity of the wind power system.

4.2 Forecasting results

Table 2 shows the parameters of ANN and proposed methods. Figure 4 shows the forecasting results on Mar. 31 in year 2013. Both the ANN and proposed methods give a

large error from 3rd to 5th forecasting periods. For *RMSE* value, the ANN method obtained is 127.0492 W, while the proposed method gives a value of 113.4993 W. For *MRE* index, the ANN and proposed methods obtained are 4.8304% and 4.48674%, respectively.

Figure 5 shows the forecasting results on Jun. 30 in year 2013. The ANN and proposed methods also give a large error from 3rd to 5th forecasting periods. But the obtained *RMSE* and *MRE* are 10.0654 W and 0.3534% for the ANN method, while the proposed method gives values of 7.9569 W and 0.2970%, respectively. The forecasting errors are very low because the wind power levels are very low on that day.

Figure 6 shows the forecasting results on Sep. 30 in year 2013. Since the wind speed is changed frequently, both the ANN and proposed methods produce larger prediction error. Fig. 7 shows the forecasting results on Dec. 31 in year 2013. The ANN method has better forecasts in the 5th forecasting period but worse in the 6th forecasting period than the proposed method.

Table 3 summarizes the forecasting results of *RMSE* and *MRE* for each method. For the average *RMSE*, the ANN gives a value of 68.0519 W while the proposed method gives a value of 58.0490 W. The average *MRE* is 2.0542% for the proposed method and the ANN gives a value of 2.3587%. The proposed method produces a better forecasting result than the ANN method in terms of *RMSE* and *MRE* criteria.

Conclusions

A hybrid approach that combines SVR and HSA for the short-term wind power forecasting is proposed in this paper. The proposed approach comprises data classification, training and forecasting stages. The data classification stage uses a k-means clustering algorithm to classify the wind power output patterns into different wind-speed types. In the training stage, five training models, representing different wind-speed types, are created using a HSA-based SVR. The forecasting stage uses verbal wind speed forecasts of the TCWB for the next 3 hours to choose an adequate model for accurate forecasting. For one year testing data, the proposed approach achieves better forecasting accuracy than the ANN method, in terms of MRE and RMSE criteria. Furthermore, the proposed method is also feasible to deal with more forecasting models for more accurate forecasts.

Acknowledgment

Financial supports from the Ministry of Science and Technology, Taiwan, R.O.C. under the Grant No. 104-2221-E-230-003 and 103-2221-E-168-015 are acknowledged.

Table 2 Parameters of ANN and proposed methods

Method	Parameters	Value or Range
ANN	no. of input neurons	15
	no. of output neurons	12
	no. of neurons for the first hidden layer	7
	no. of neurons for the second hidden layer	5
	learning rate	0.05
	momentum constant	0.85
	max. iteration	100
Proposed HSA- Based SVR	no. of input neurons	15
	no. of output neurons	12
	penalty weight (c)	[0 10]
	dilation parameter (σ)	[0 1]
	HMCR	0.8
	PAR	0.85
	BW	0.6
	population size	30
	max. iteration	100

	RMSE(W)	MRE(%)
ANN	127.0492	4.8304
proposed	113.4993	4.4867

Fig. 4 Forecasting results on Mar. 31 in year 2013.

References

- (1) K. Bhaskar and S. N. Singh, "AWNN-assisted wind power forecasting using feed-forward neural network," *IEEE Trans. on Sustainable Energy*, vol. 3, no. 2, pp. 306-315, 2012.

	RMSE(W)	MRE(%)
ANN	10.0654	0.3534
proposed	7.9569	0.2970

Fig. 5 Forecasting results on Jun. 30 in year 2013.

	RMSE(W)	MRE(%)
ANN	148.3715	5.4710
proposed	120.8687	4.6398

Fig. 6 Forecasting results on Sep. 30 in year 2013.

- (2) M. Khalid and A. V. Savkin, "A method for short-term wind power prediction with multiple observation points," *IEEE Trans. on Power Systems*, vol. 27, no. 2, pp. 579-586, May 2012.
- (3) K. Bhaskar and S. N. Singh, "AWNN-assisted wind power forecasting using feed-forward neural network," *IEEE Trans. on Sustainable Energy*, vol. 3, no. 2, pp. 306-315, 2012.

	RMSE(W)	MRE(%)
ANN	90.9904	3.1112
proposed	67.8051	2.3957

Fig. 7 Forecasting results on Dec. 31 in year 2013.

Table 3 Forecasting results for different methods

Date	ANN method		Proposed method	
	RMSE(W)	MRE(%)	RMSE(W)	MRE(%)
1/31	11.5676	0.4241	10.4030	0.3600
2/28	71.9993	2.6636	64.8425	2.4576
3/31	127.0492	4.8304	113.4993	4.4867
4/40	100.0855	2.9210	77.4300	2.0392
5/31	18.7735	0.6490	16.6526	0.6038
6/30	10.0654	0.3534	7.9569	0.2970
7/31	56.7245	1.5183	55.6267	1.2981
8/31	49.9713	1.4298	36.7434	1.2724
9/30	148.3715	5.4710	120.8687	4.6398
10/31	118.1248	4.4127	113.4002	4.3446
11/30	12.8997	0.5195	11.3594	0.4572
12/31	90.9904	3.1112	67.8051	2.3957
Max.	148.3715	5.4710	120.8687	4.6398
Min.	10.0654	0.3534	7.9569	0.2970
Ave.	68.0519	2.3587	58.0490	2.0543

- (4) M. Khalid and A. V. Savkin, "A method for short-term wind power prediction with multiple observation points," *IEEE Trans. on Power Systems*, vol. 27, no. 2, pp. 579-586, May 2012.
- (5) M. G. Lobo and I. Sánchez, "Regional wind power forecasting based on smoothing techniques with application to the Spanish peninsular system," *IEEE Trans. on Power Systems*, vol. 27, no. 4, pp. 1990-1997, Nov. 2012.
- (6) A. Vaccaro, G. Bontempi, S. B. Taieb, and D. Villacci, "Adaptive local learning techniques for multiple-step-ahead wind speed forecasting," *Electric Power Systems Research*, vol. 83, pp. 129-135, 2012.

- (7) G. Sideratos and N. D. Hatziaargyriou, "An advanced statistical method for wind power forecasting," *IEEE Trans. on Power Systems*, vol. 22, no. 1, pp. 258-265, Feb. 2007.
- (8) N. Amjadi, F. Keynia, and H. Zareipour, "Short-term wind power forecasting using ridgelet neural network," *Electric Power Systems Research*, vol. 81, pp. 2099-2107, 2011.
- (9) Y. Liu, J. Shi, and Y. Yang, "Short-term wind-power prediction based on wavelet transform - support vector machine and characteristics analysis," *IEEE Trans. on Industry Applications*, vol. 48, pp. 1136-1141, 2011.
- (10) M. A. Mohandes, T. O. Halawani, S. Rehman, and A. A. Hussain, "Support vector machines for wind speed prediction," *Renew. Energy*, vol. 29, no. 6, pp. 939-947, May 2004.
- (11) S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, 1982, vol. 28, no.2, pp.129-137.
- (12) C. Cortes and V. Vapnik, "Support Vector Networks," *Machine Learning*, 20, pp. 273-297, 1995.
- (13) H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in Neural Information Processing Systems*, vol. 9, The MIT Press, 1997, pp. 155.
- (14) Z. W. Geem, J. H. Kim, et al., "A new heuristic optimization algorithm: harmony search," *Simulation*, vol. 76, no. 2, pp. 60-68, 2001.