Photovoltaic Power Forecasting Based on SVM Optimized by Improved ABC

Lin Liu¹,², Yanyun Lu¹, Shuqin Li¹
¹School of Management, Hefei University of Technology, Hefei 230009, China
²Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei 230009, China
*Corresponding author’s e-mail: 1623957852@qq.com

Abstract. In order to improve the accuracy of photovoltaic power output prediction, a photovoltaic power prediction method based on similar days and improved artificial bee colony support vector machine is proposed. Firstly, through calculating the Euclidean distance of history day and measured day meteorological factors to determine similar days. Secondly, select historical data of photovoltaic power output, temperature, humidity and daily radiation on the slope of similar days and temperature, humidity and daily radiation on the slope of test date as input variables of support vector machine. And we adopt the improved artificial bees colony to optimize kernel function parameters and the penalty factor of support vector machine. Finally get the output in each period of photovoltaic power prediction. The experimental results showed that the proposed method can effectively improve the prediction accuracy of photovoltaic power.

1. Introduction
Due to the limitation of resource reserves and the environmental pollution caused by the use of fossil energy such as oil and natural gas, the development and utilization of renewable energy has become an inevitable trend of world energy transformation. In recent years, Photovoltaic (PV) energy with its advantages of clean, pollution-free has been rapid development, is becoming the main direction of energy transformation. Therefore, it is an effective way to solve this problem to establish an accurate prediction model of photovoltaic power generation and improve the prediction accuracy of short-term photovoltaic power generation. Meanwhile, it is of great significance to power grid scheduling and energy management.

For PV power prediction, domestic and foreign scholars have done more research in recent years. In [1], A hybrid short-term photovoltaic power prediction method based on Variational Mode Decomposition, Deep Belief Network and Auto-Regressive Moving Average model is proposed. In [2], A photovoltaic power generation prediction model based on multiple reservoirs echo state network is presented. In [3], the Gradient Boost Decision Tree is presented to forecast the power output of PV generation. In [4], the Extreme Learning Machines is proposed to provide accurate 24 h-ahead photovoltaic power production predictions. In [5], A photovoltaic power prediction model is built based on empirical mode decomposition and support vector machine (SVM) optimized with an Artificial Bee Colony (ABC). In [6], A ensemble Empirical mode decomposition and variable-weight combination forecasting is proposed to forecast Photovoltaic power generation. In [7], A two-stage model is proposed to quantify the relevant prediction interval of Photovoltaic power generation. In [8],...
A new method for photovoltaic power generation prediction based on genetic algorithm, particle swarm optimization algorithm and adaptive neuro-fuzzy inference systems is proposed. In [9], a photovoltaic power generation model combined with environmental factors and support vector machine optimized by genetic algorithm is proposed.

According to the above researches at home and abroad, it can be concluded that although there are some methods based on support vector machine optimized by intelligent algorithm for PV power prediction, there is still room for improvement in the speed and accuracy of network optimization. In this paper, firstly, select the prediction days and find the similar days. Then put forward an improved ABC which add increasing weighting function and global boot factor to the position updating formula of the artificial colony algorithm. Finally, forecast the power output of PV generation base on the SVM optimized by improved artificial bees colony.

2. Modeling principle

2.1. Improved Artificial Bee Colony algorithm

ABC algorithm is a novel swarm intelligence optimization algorithm proposed by Karaboga in 2005. Suppose the solution space of the problem is D dimensional, and the number of leading bees and following bees are NP, which are equals to the number of food sources. The basic steps of the improved Artificial Bee Colony algorithm are as follows.

Step1: According to Eq. (1), the initial values of NP feasible solutions corresponding to the location of food source are generated randomly.

\[ x_{id} = x_{id}^{\min} + r(x_{id}^{\max} - x_{id}^{\min}) \]  

(1)

where \( x_{id}^{\min} \) and \( x_{id}^{\max} \) are the lower and upper bounds of the d-th dimension; \( d=1, 2, ..., D \), is the dimension of optimization solution parameters; \( i=1, 2, ..., NP \).

Step2: According to the original position update Eq. (2), a new position update Eq.(3) is proposed.

\[ x_{id} = x_{id} + \phi_{id}(x_{id} - x_{id}) \]  

(2)

\[ x_{id} = x_{id} + \phi_{id}(x_{id} - x_{id}) + \theta(x_{best,d} - x_{id}) \]  

(3)

where \( i=1, 2, ..., NP; d=1, 2, ..., D; x_{id} \) is the candidate solution, \( x_{id} \) is the current solution, \( \phi_{id} \) is a random number on \([-1,1] \); in Eq. (3), input increasing weight function, let \( W_t = W_{\text{max}} + (W_{\text{max}} - W_{\text{min}}) \times t/t_{\text{max}} \), \( t \) is the current cyclic iteration number, \( t_{\text{max}} \) is the maximum number of iterations; \( \theta \) is a random number on \([0, 1] \); \( (x_{best,d} - x_{id}) \) is the global boot factor.

Step3: According to Eq.(4), following bees chose the location of food source, that means chose the food source according to the probability of the amount of nectar.

\[ P_{i} = \frac{\text{fit}_i}{\sum_{j=1}^{NP} \text{fit}_j} \]  

(4)

where \( \text{fit}_i \) is the fitness value of candidate solution \( x_i \). The higher fitness value, the greater chance of food source is selected.

Step4: If employed bees and onlooker bees search the whole space, the fitness of a food source in a given step has not been improved, then discards the food source location, and correspond to the food source of leading bee change to scout bee, according to Eq.(1), scout bees search for a new candidate solution to replace the original solution, after a finite number of iterations, to obtain the optimal solution.

2.2. Support vector machines

Support vector machines(SVM) were proposed by Vapnik and were used in fields such as classification, nonlinear regression and prediction, et, al. It is a machine learning method based on statistical learning theory.
Suppose given the training sample set \( \{(x_i, y_i), i = 1, 2, 3, \ldots, n\} \), where \( n \) is the number of training samples, \( x_i \in \mathbb{R}^N \), \( y_i \in \{-1, 1\} \). In order to avoid overfitting, a tolerance value of \( \varepsilon > 0 \) is added. Insensitivity loss function is shown in Eq. (5):
\[
[y - f(x)]_+ = \begin{cases} 0, & |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon, & |y - f(x)| > \varepsilon 
\end{cases}
\]
where \( f(x) \) is the regression estimation function and \( y \) is the target value. Assume that the mapping function of the sample data set is a non-linear function \( \phi(x) \), and Eq. (6) represents the prediction model after \( f(x) \) mapping:
\[
f(x) = w^T \phi(x) + b
\]
where \( w \) is the weight vector; \( b \) is the offset vector. According to the minimization principle of structural risk, the optimal problem to be solved is shown in Eq. (7):
\[
\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \\
s.t. \begin{cases} y_i - w^T \phi(x_i) - b \leq \varepsilon + \xi_i \\ w^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0 \\ \xi_i^* \geq 0, i = 1, 2, \ldots, n \end{cases}
\]
where \( C \) is the error penalty variable, \( \xi_i \) and \( \xi_i^* \) are the relaxation variables.

The Lagrangian function is introduced, and establish the Lagrangian equation. By solving the partial derivative of the parameters \( w, b, \xi_i \) and \( \xi_i^* \), according to the dual principle, the linear regression function is obtained as follows:
\[
f(x) = w^T \Phi(x) + b = \sum_{i=1}^{n} (\alpha_i^* - \alpha_i^*) K(x_i, x_j) + b
\]
where \( K(x_i, x_j) \) is the kernel function. The commonly used kernel functions are Linear kernel function, Polynomial kernel function, Gaussian kernel function and Sigmoid function.

Among these kernel functions, Gaussian kernel function is the most commonly used, which has the characteristics of less set parameters, lower computational complexity and better performance. Gaussian kernel function is adopted as the kernel function of SVM, and the improved process of optimizing SVM parameters by ABC is shown in Figure1.

3. Photovoltaic power forecasting based on SVM optimized by improved ABC

In this paper, monitoring data of a photovoltaic intelligent cloud service platform in Hefei are used to define the power output period as 8.00~17.00. And power data are collected every 10 minutes to obtain 55 times one day. The data used in modeling include ambient temperature, ambient humidity, daily radiation on the slope, date data and output power value of photovoltaic power generation system.
3.1. The sample data normalization
Since the data has different dimensions and its value range varies greatly, it can be unified into a reference range so as to improve the prediction efficiency and accuracy. The power data of the original photovoltaic power generation system are normalized as follows:

\[ X_i^* = \frac{X_i - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \]  

where \( X_i \) represents power value at the \( i \)th time point \((1 \leq i \leq 55)\); \( X_{\text{min}} \) and \( X_{\text{max}} \) represents the minimum and maximum values of photovoltaic power generation; \( X_i^* \) represents the normalized sequence value.

3.2. The similar days
For the same weather type, in the set \( D \) containing \( n \) days of historical data, the Euclidean distance \( d_{ik} \) (equation 10) is used to calculate the set \( \{d_{i1}, d_{i2}, \ldots, d_{i5}, \ldots, d_{in}\} \) on the predicted day, and then the Euclidean distance set is sorted from small to large. The data of 5 days which are most similar to the predicted day \( i \) are selected for subsequent training.

\[ d_{ik} = \left[ \sum_{j=1}^{5} (X_{ij} - X_{kj})^2 \right]^{1/2} \quad i, \ k \in [1, n], \ k \neq i \]  

where \( X_{i1}, X_{i2}, \ldots, X_{i5} \) represent the daily highest temperature, the daily lowest temperature, the daily maximum humidity, the daily minimum humidity and the daily radiation on the slope on the predicted day \( i \); \( X_{k1}, X_{k2}, \ldots, X_{k5} \) represent the daily highest temperature, the daily lowest temperature, the daily maximum humidity, the daily minimum humidity and the daily radiation on the slope in the data of day \( k \) in sample \( D \).

3.3. Flow of photovoltaic power prediction based on the SVM optimized by improved ABC
Due to the good performance of gaussian kernel function, this paper chooses gaussian kernel function as kernel function of SVM prediction model. In order to improve the prediction performance of SVM model, the improved artificial bee colony algorithm proposed in this paper was used to optimize SVM and find the optimal kernel function parameter \( \sigma \) and penalty factor \( C \). Similar day and improved artificial bee colony algorithm optimization SVM PV power prediction flow chart is shown in Figure 2.
The sample data are normalized
Select similar days
Select the training samples
The data of similar days about daily maximum and minimum temperature,
daily maximum and minimum humidity, average daily Radiation of inclined
The pv power data of similar days
The data of prediction days about daily maximum and minimum temperature,
daily maximum and minimum humidity, average daily Radiation of inclined
SVM model
Get the best parameters $\sigma$ and $C$ optimized by ABC
Output the value of prediction days of pv power generation

Figure 2. Flow chart of photovoltaic power prediction

3.4. Comparison of the prediction errors
In this paper, Mean Absolute Percentage Error (MAPE) and Root mean square Error (RMSE) are used as evaluation indexes of prediction results.

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Y_i - Y_i^*}{Y_i} \right| \times 100\%$$ \hspace{1cm} (11)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (Y_i - Y_i^*)^2}{N}}$$ \hspace{1cm} (12)

where $Y_i$ represents the actual PV power value; $Y_i^*$ represents the predicted value of PV power generation; $i$ represents the time variable; $N$ represents the number of predicted samples.

4. Experiment and results
All the codes involved in the experiment were written on Matlab R2016a software. The sample data of the experiment are the photovoltaic power generation data and meteorological data provided by a company in hefei from January to May 2018. Firstly, select March 21, 2018, April 25, 2018 and May 25, 2018 as the prediction days. The weather types are sunny, cloudy and rainy. According to the historical meteorological data of the prediction days, the similar days sample are calculated by Euclidean distance. Then, data of prediction days of the maximum and minimum daily temperature, maximum and minimum daily humidity, and average daily radiation on the slope are inputed to predict the photovoltaic power generation. Finally, in order to verify the validity and accuracy of the prediction model, forecasting the photovoltaic power generation of the prediction days by using SVM model, SVM model optimized by original artificial bees colony and the method in this paper. Error analysis is showed in table 1.

- The MAPE fluctuation range of the PV power prediction by using the method proposed in this paper is 6.60%~16.33%, and the average value is 11.21%. The MAPE fluctuation range of SVM optimized by original ABC is 8.94%~18.13%, and the mean value is 13.61%. The MAPE fluctuation range of SVM model is 10.15%~23.46%, and the mean value is 16.18%. The RMSE range of the method in this paper is 5.98~8.69, and the mean value is 7.40. The RMSE range of the SVM model optimized by ABC is 7.63~9.98, and the mean value is 8.81. The RMSE range of the SVM model is 8.86~10.96, and the mean value is 10.05. The MAPE and RMSE of the last two methods are both larger than the method proposed in this paper, the method proposed in this paper effectively improves the prediction accuracy of photovoltaic power generation.

- Due to high air humidity in the morning, and often accompanied by fog and haze, leading to
unstable photovoltaic power generation. Moreover, due to the influence of automobile exhaust emission, dust and other factors, there is a big error between the predicted value and the actual value of photovoltaic power generation. If the water vapor content and dust in the air can be collected accurately, the accuracy of prediction can be improved better.

Although the method proposed in this paper improves the prediction accuracy effectively, the error between the predicted and actual values of photovoltaic power is also big. The one reason for the error is the initial randomness of intelligence algorithm. In addition, the PV power data monitored by the PV power station in different locations in Hefei is also the reason for the error.

| Prediction methods                      | Prediction days | MAPE (%) | RMSE |
|----------------------------------------|-----------------|----------|------|
| SVM                                    | Sunny           | 23.46    | 10.96|
|                                        | Cloudy          | 14.93    | 10.33|
|                                        | Rainy           | 10.15    | 8.86 |
|                                        | Average         | 16.18    | 10.05|
| SVM model optimized by original ABC    | Sunny           | 18.13    | 8.81 |
|                                        | Cloudy          | 13.75    | 9.98 |
|                                        | Rainy           | 8.94     | 7.63 |
|                                        | Average         | 13.61    | 8.81 |
| The method in this paper               | Sunny           | 16.33    | 8.69 |
|                                        | Cloudy          | 10.70    | 7.54 |
|                                        | Rainy           | 6.60     | 5.98 |
|                                        | Average         | 11.21    | 7.40 |

5. Conclusion
In this paper, firstly, the similar days are obtained by the Euclidean distance, selecting the temperature, humidity and average daily radiation on the slope data of the prediction days and similar days as the sample data, and SVM prediction model is bulited. Then, using an improved artificial colony algorithm which add increasing weighting function and global leading factor to the position updating formula of the artificial colony algorithm to optimize the parameters of the SVM. The experimental results show that the method proposed in this paper can effectively improve the prediction accuracy of photovoltaic power generation which is beneficial for photovoltaic enterprises to better guarantee the operation of PV power stations, and for electric power departments to make scheduling plans.

Acknowledgments
Supported by the Decision Science and Technology Research Institute, Hefei University of Technology, Hefei, China is greatly appreciated. This research was supported in part by the National Natural Science Foundation under the Grant Nos.: 71521001 and Anhui Province Natural Science Foundation under the Grant No.: 1908085MG223.

References
[1] Xie T, Zhang G, Liu HC, et al. (2018) A Hybrid Forecasting Method for Solar Output Power Based on Variational Mode Decomposition, Deep Belief Networks and Auto-Regressive Moving Average. Applied Sciences-Basel, 8: 1-24.
[2] Yao XS, Wang ZS, Zhang HG. (2019) A novel photovoltaic power forecasting model based on echo state network[J]. Neurocomputing, 325:182-189.
[3] Wang JD, Li P, Ran R, et al. (2018) A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree. Applied Sciences-Basel, 8: 1-14.
[4] Al-Dahidi S, Ayadi O, Adeb J, et al. (2018) Extreme Learning Machines for Solar Photovoltaic Power Predictions. Energies, 11: 1-18.
[5] Gao XM, Yang SF, Pan SB. (2017) Optimal Parameter Selection for Support Vector Machine Based on Artificial Bee Colony Algorithm: A Case Study of Grid-Connected PV System
Power Prediction. Computational Intelligence and Neuroscience, 8: 1-14.

[6] Wang H, Sun JB, Wang WJ. (2018) Photovoltaic Power Forecasting Based on EEMD and a Variable-Weight Combination Forecasting Model. Sustainability, 10: 1-11.

[7] Liu LY, Zhao Y, Chang DL, et al. Prediction of short-term PV power output and uncertainty analysis. Applied Energy, 228: 700-711.

[8] Semero YK, Zhang JH, Zheng DH. (2018) PV Power Forecasting Using an Integrated GA-PSO-ANFIS Approach and Gaussian Process Regression Based Feature Selection Strategy. Csee Journal of Power and Energy Systems, 4: 210-218.

[9] Wang JD, Ran R, Song ZL, et al. (2017) Short-Term Photovoltaic Power Generation Forecasting Based on Environmental Factors and GA-SVM. Journal of Electrical Engineering & Technology, 12: 64-71.