**PAPER • OPEN ACCESS**

# Wind Power Prediction Based On Improved Genetic Algorithm and Support Vector Machine

View the article online for updates and enhancements.

# Wind Power Prediction Based On Improved Genetic Algorithm and Support Vector Machine

**Li Zhang[1], Kui Wang[1, *], Wenli Lin[1], Tianxiang Geng[2], Zhen Lei[3], Zheng Wang[4]**

[1]CHSCOM Electric Technology Co.Ltd, Nanjing 210019, China
[2]State Grid Ningxia Electric Power Co. Ltd. Yinchuan 750000, China
[3]State Grid Jiangsu Electric Power Co. Ltd. Nanjing 210000, China
[4]China Electric Power Research Institute Co. Ltd. Beijing 100192, China

*Corresponding author e-mail: 813585818@qq.com

**Abstract**. With the continuous improvement of wind power generation technology, the installed capacity of wind farms and the scale of wind power interconnection are increasing, which brings huge economic benefits to our country, at the same time, alleviates energy security and brings environmental benefits. However, as an intermittent power supply, the fluctuation and uncertainty of wind power increase the difficulty of security dispatch of power grid and increase the burden of system reserve capacity. Based on this, aiming at single-point wind power prediction and support vector machine method, this paper studies the improved support vector machine wind power prediction based on genetic algorithm. In the prediction model of support vector machine, the improper setting of parameters will lead to the phenomenon of "under-learning" or "over-learning", which directly affects the prediction accuracy. In this paper, genetic algorithm is used to optimize the parameters. The example shows that the GA-SVM model has a better prediction effect than the model with default parameters.

## 1. Introduction

Based on economic and social benefits, wind power generation has attracted worldwide attention and become the fastest growing renewable energy in the world. Many countries have taken the development of wind power generation as an important measure to improve energy structure, protect the ecological environment and reduce environmental pollution, and incorporated it into the national energy development plan [1-2]. It is expected that the proportion of wind power in the whole energy structure will gradually increase and become the most important clean energy in the 21st century [3]. However, due to the randomness, volatility and uncontrollability of wind, this uncertainty will affect the security and stability of the power grid, increase the difficulty of dispatching, and increase the backup burden of the entire power network [4]. Therefore, accurate prediction of wind power is becoming more and more important.

Wind power prediction plays an important role in wind farm operation and power grid dispatching. Based on power characteristic curve of wind farm, an estimation method for wind power prediction error distribution is proposed in paper [5]. In order to make system operate economically and reliably, a multi-objective unit commitment optimization model was proposed based on comprehensive consideration of negative peak load regulation capability and operation cost of conventional generators in paper [6]. Characteristics analysis of wind power forecasting error can provide more accurate

reference for optimal dispatch and stable operation of power system in paper [7]. This paper proposed a numerical error characteristics analyzing approach which stratified the errors into different layers according to the probability density of the ultra-short-term wind power forecasting errors. A method of probabilistic wind power forecasting with the consideration of ramp characteristics is proposed in paper [8]. To control harmful wind power events, this paper proposed a combined model of different meteorological variables to realize wind power prediction with high accuracy in paper [9]. In paper [10], an intelligent multi-objective optimized prediction interval (PI) model for wind power is proposed with wavelet neural network as its basic prediction model.

Wind power prediction is an interdisciplinary application and also a mutual fundamental discipline. Global state-of-art of prediction error is even below 3%, whereas in China more efforts should be made to top out. According to present trend, benefits of accuracy improvement from accumulative data slow down, and incredible breakthrough of numeric weather prediction (NWP) is unexpected. For wind power prediction in near future, it is important to improve prediction skills in every step considering NWP limits, select and combine proper methods to reduce the final 2%-3% in error learning curve[11].
Aiming at wind power single point value prediction and support vector machine method, this paper studies wind power prediction based on improved support vector machine based on genetic algorithm. The parameters are optimized by genetic algorithm. Under the condition of guaranteeing accuracy, 5 fold cross validation is adopted, and the mean square error of the model under different parameters is selected as the result. Criteria for evaluating the quality of models. Through cross validation, the minimum mean square error prediction model is determined, and the optimal parameters are found. The example shows that the GA-SVM model has a better prediction effect than the model with default parameters.

## 2. SVM nonlinear regression

Support Vector Machine Regression (SVM) is a learning method based on VC dimension theory and structural risk minimization principle. It seeks the best compromise between model complexity and learning ability according to limited sample information in order to obtain the best generalization ability.

For input samples $X_i \in R^d$, the corresponding output is $Y_i \in R$ , a training set containing N samples, which can be expressed as:

$$s = \{(X_1, Y_1), (X_2, Y_2)...(X_N, Y_N)\} \tag{1}$$

The basic idea of SVM regression is to obtain the parameters of the model by optimizing the training sample set, and to establish the corresponding relationship between input and output, so that the estimated value of output can be obtained for any given new input sample.
The linear regression function can be written as follows:

$$f(X) = w^T gX + b \tag{2}$$

Insensitive loss function $\varepsilon$ is introduced to ensure the existence of the global optimal solution and to satisfy the generalization ability. Under the given precision, and assuming that all training samples can be fitted without error under this precision, the following constraints are formed:

$$s.t. \begin{cases} Y_i - w^T gX - b \le \varepsilon \\ w^T gX + b - Y_i \le \varepsilon \end{cases} \quad i = 1,...,N \tag{3}$$

According to the SVM regression method, the coefficients $w$ are determined by minimizing the objective function. According to the concepts of interval and geometric interval in statistical theory,

the coefficients are equivalent to minimization $\frac{1}{2}\|w\|^2$. At the same time, penalty parameters $C$ are introduced to balance the minimum training error, and the regression problem is equivalent to the quadratic convex optimization problem about the $w$ and $b$. Further relaxation variables $\xi_i \geq 0, \xi_i^* \geq 0$ are introduced to deal with other allowable fitting errors in the optimization problem, and the objective function is further transformed into the following optimization problems:

$$\min : \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

$$s.t.\begin{cases} Y_i - w^T gX - b \leq \varepsilon + \xi_i \\ w^T gX + b - Y_i \leq \varepsilon + \xi_i^* \quad i = 1,\ldots,l \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{4}$$

In formula (4), $\frac{1}{2}\|w\|^2$ is empirical risk, $C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$ is. confidence risk, which satisfies the principle of structural risk minimization.

Then, the Lagrange function is introduced to the dual problem, and the KKT (Karush-Kuhn-Tucker) condition is used to calculate the dual problem. Finally, the principle of structural risk minimization is obtained.

When the relationship between input and output is complex, it is impossible to describe the corresponding relationship between input and output through a simple linear model, which requires SVM non-linear regression. Its basic idea is to map $X \in R^d$ to a high-dimensional space by a pre-determined non-linear mapping, in which the non-linear regression operation in low-dimensional space is transformed into linear regression operation in high-dimensional space.

The nonlinear function model can be written as follows:

$$f(X) = \phi(w)^T g\phi(X) - b \tag{5}$$

The model optimization function is updated to:

$$\min : \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

$$s.t.\begin{cases} Y_i - \phi(w)^T g\phi(X_i) - b \leq \varepsilon + \xi_i \\ \phi(w)^T g\phi(X_i) + b - Y_i \leq \varepsilon + \xi_i^* \quad i = 1,\ldots,N \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{6}$$

However, there is no systematic method to find the appropriate non-linear function; even if this $\phi(g)$ is found, the dimension of the mapped high-dimensional feature space is often very large and the direct inner product is very complex. In this case, by introducing the kernel function, the inner product of high-dimensional space can be obtained directly after input in low-dimensional space.

In other words:

$$f(X) = \phi(w)^T g\phi(X) + b \tag{7}$$

$$g(X) = K(w, X) + b \tag{8}$$

Statistical learning theory points out that the kernel function as SVM must satisfy the Mercer condition. The common kernel functions are:

(1) Linear kernel function:

$$K(x_i, x_j) = (x_i g x_j) \tag{9}$$

The simplest kernel function is equivalent to direct computation in low dimensional space.

(2) Polynomial kernel function:

$$K(x_i, x_j) = [(x_i g x_j) + 1]^d \quad , \quad d \in R \tag{10}$$

Polynomial kernel function has poor local performance, and far away samples will also affect the regression curve. The parameter refers to the dimension of the polynomial kernel function. If it is too large, the computational complexity will be greatly increased, and at the same time, it will lead to "over-fitting".

(3) RBF kernel function:

$$K(x_i, x_j) = e^{-\gamma(x - x_i)^2} \tag{11}$$

The local performance of RBF kernel function is excellent, but the selection of parameters needs special attention.

Among them, the RBF kernel function is the most widely used kernel function. Compared with linear kernels, it can map a sample to a higher dimensional space. It can be said that linear kernels are only a special case of RBF kernels. Compared with polynomial kernels, RBF kernels need to determine relatively few parameters, the number of parameters directly affects the complexity of the function; in addition, when the order of polynomial kernels is relatively high, the calculation will become very difficult, and the program runs very time-consuming.

In this paper, RBF kernel function is selected as the kernel function used in support vector machine modeling, and genetic algorithm is used to optimize its parameters.

## 3. Support vector machine improved by genetic algorithm

### 3.1. Optimize target parameters

This paper optimizes the most widely used RBF kernel function parameters $\gamma$ and error penalty factor C.

In order to further improve the prediction effect of support vector machine and reduce the prediction error, genetic algorithm is used to optimize these two parameters.

In this paper, under the condition of ensuring accuracy and considering the running time of the program, the average mean square error of the model under different parameters is selected as the evaluation criterion of the model. Through cross validation, the minimum mean square error prediction model is determined, and the optimal parameters are found.

Based on the theoretical basis of support vector machine (SVM), this paper uses genetic algorithm (GA) and cross validation as evaluation criteria to optimize the key parameters C and $\gamma$ in SVM, and establishes a GA-SVM prediction model, which is validated by several sets of data in an example.

### 3.2. Selection of input variables for models

Linear correlation analysis is used to analyze the real-time data of a wind farm to determine the input parameters of the model.

Linear correlation coefficient $r$ :

$$r = \frac{S_{xy}^2}{S_x S_y}$$
$$= \frac{\sum(x-\bar{x})(y-\bar{y})/n}{\sqrt{\sum(x-\bar{x})^2/n}\,\mathrm{g}\sqrt{\sum(y-\bar{y})^2/n}} \qquad (12)$$
$$= \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\,\mathrm{g}\sqrt{n\sum y^2 - (\sum y)^2}}$$

The value range of $r$ is [-1,1]. $r>0$ is positive correlation, $r<0$ is negative correlation, $|r|=0$ indicating that there is no linear relationship, $|r|=1$ indicating a complete linear relationship.

$|r|\leq 0.3$ : there is no linear correlation.

$0.3\leq|r|\leq 0.5$ : low linearity correlation;

$0.5<|r|\leq 0.8$ : significant linear correlation.

$|r|>0.8$ : highly linear correlation.

Based on the above correlation analysis, for the independent variable input of SVM, four independent variables are selected as NWP predicting wind speed and tracing back three historical power points.

Mean relative error (MAPE), mean square error (MSE), root mean square error (RMSE) were used as the evaluation criteria of model prediction effect.

$$e_{MAPE} = \frac{1}{N}\sum_{i=1}^{N}|\frac{A_i - P_i}{P}|\times 100\%$$
$$e_{MSE} = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(A_i - P_i)^2} \qquad (13)$$
$$e_{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(A_i - P_i)^2}$$

In formula, $A_i$ is the true value of the $i$ prediction point. $P_i$ is the prediction value of the $i$ prediction point, and $N$ is the number of the prediction points.

## 4. Case study

In this case, the 10-day data of a wind farm is used as training set. The sampling period is 5 minutes, totaling 2880 time periods. The 24 hours after selecting training set samples, the sampling period is 5 minutes, totaling 288 time periods. The input of the model is NWP forecasting wind speed and retrospective 3-point historical power, with a total of 4 independent variables; the output of the model is the wind speed of the forecasting point. This section predicts a single point prediction method.

Data normalization preprocessing:

$$\hat{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \qquad (14)$$

The essence of normalization is to transform the dimension into dimensionless without changing the physical meaning of the data, and map the uncertain range to the definite range, so that the data can be controlled and the complexity of the model can be reduced effectively.

For training set and test set, the default parameters and optimized parameters are used to model and predict respectively. The predicted results are shown in Figure 1.
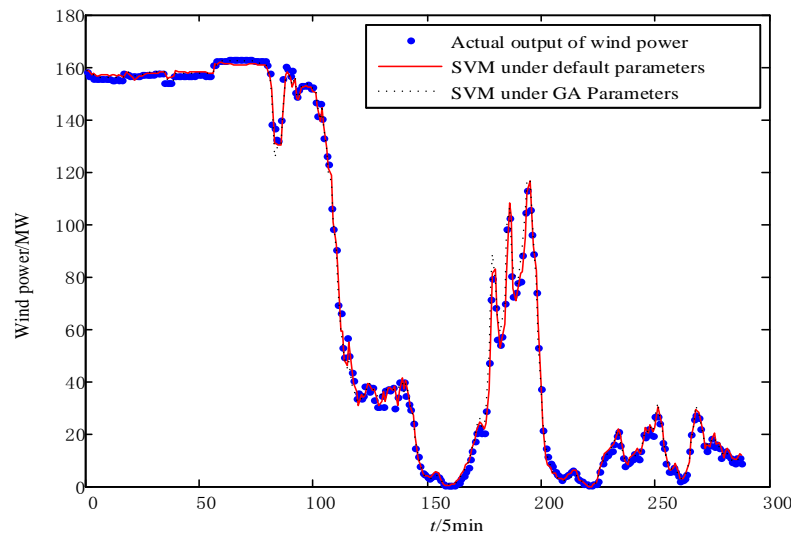


**Figure 1.** SVM Prediction Results under Default Parameters and GA Optimized Parameters

The experimental results show that the relative error, mean square error and root mean square error are all reduced and the prediction accuracy is improved after using genetic algorithm to optimize the parameters of kernel function and penalty parameter.

From the above analysis results, we can see that for the sample selected in this section, because the sample data itself is relatively ideal, the SVM method has higher prediction accuracy and better prediction effect. However, in practical production applications, bad data may occur due to data loss, data errors and other reasons in the process of data acquisition, communication and storage of wind farms, which may have a certain impact on the prediction effect of wind power. In order to verify the prediction effect of GA-SVM when bad data appear. In this section, we do some research on the above examples by adding artificial disturbance.

Using the processed training set data to model and using the same test set sample input, the prediction results are obtained, as shown in Figure 2:
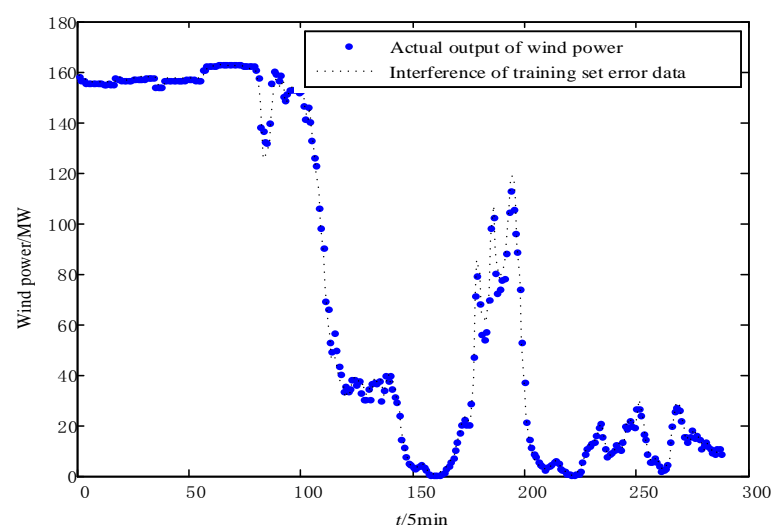


**Figure 2.** Prediction effect of SVM with a large number of error data in training set

Through the observation and analysis of Figure 2, it can be seen that the prediction effect of SVM is still good under the condition that there are a lot of wrong data in the training set. This is due to the combined effect of relaxation variables and penalty functions in the process of support vector machine modeling. A few support vectors which play a decisive role in the prediction effect are selected to automatically eliminate the influence of those obviously unreasonable training set samples. Therefore, when a large number of error data appear in the training set samples, the support vector machine can also show a more accurate prediction effect, which shows that the SVM method has good anti-jamming performance.

**5. Conclusion**

In this paper, based on the single point value prediction of wind power, the following conclusions are obtained from the research of support vector machine method:

In the prediction model of support vector machine, the parameter setting plays an important role. The improper parameter setting often leads to the phenomenon of "under-learning" or "over-learning", which directly affects the prediction accuracy. In this paper, the genetic algorithm is used to optimize the parameters. Compared with the model with default parameters, the prediction effect of GA-SVM model has been significantly improved.

**Acknowledgments**

**References**

[1]    Kusiak A, Zheng H, Song Z. Short-Term Prediction of Wind Farm Power: A Data Mining Approach [J]. IEEE Transactions on Energy Conversion, 2009, 24 (1): 125-136.

[2]    Dong L, Wang L, Khahro S F, et al. Wind power day-ahead prediction with cluster analysis of NWP [J]. Renewable & Sustainable Energy Reviews, 2016, 60: 1206-1212.

[3]    Wan C, Wang J, Lin J, et al. Nonparametric Prediction Intervals of Wind Power via Linear Programming [J]. IEEE Transactions on Power Systems, 2017, 33 (1): 1074-1076.

[4]    Wang H Z, Li G Q, Wang G B, et al. Deep learning based ensemble approach for probabilistic wind power forecasting [J]. Applied Energy, 2017, 188:56-70.

[5]    WU Xiaogang, SUN Rongfu, QIAO Ying, et al. Estimation of Error Distribution for Wind Power Prediction Based on Power Curves of Wind Farms [J]. Power System Technology, 2017, 41 (6): 1801-1807.

[6]    XIANG Hongji, DAI Chaohua, MING Jie, et al.. Research on Multi-Objective Optimization of Unit Commitment Considering Negative Peak Load Regulation Ability in Valley Load Period and Wind Power Prediction Interval [J]. Power System Technology, 2017, 41 (6): 1912-1918.

[7]    YE Lin, REN Cheng, ZHAO Yongning, et al. Stratification Analysis Approach of Numerical Characteristics for Ultra-short-term Wind Power Forecasting Error [J]. Proceedings of the CSEE, 2016, 36 (3): 692-700.

[8]    GAN Di, KE Deping, SUN Yuanzhang, et al. Short-term probabilistic wind power forecast considering ramp characteristics [J]. Electric Power Automation Equipment, 2016, 36 (4): 145-150.

[9]    OUYANG Tinghui, ZHA Xiaoming, QIN Liang, et al. Short-term wind power prediction based on kernel function switching [J]. Electric Power Automation Equipment, 2016, 36(9):80-86.

[10]   CHEN Jie, SHEN Yanxia, LU Xin, et al. An Intelligent Multi-Objective Optimized Method for Wind Power Prediction Intervals [J]. Power System Technology, 2016, 40 (8): 2281-2287.

[11]   QIAO Ying, LU Zongxiang, MIN Yong. Research & Application of Raising Wind Power Prediction Accuracy [J]. Power System Technology, 2017, 41 (10): 3261-3268.