The Real-time Forecast of Photovoltaic Output Based on Empirical Mode Decomposition

Cui Du [a], Zongqi Liu [b]

School of electrical and electronic engineering, North China Electric Power University, Beijing 102206, China

[a]dc891022@sina.com, [b]lzq@ncepu.edu.cn

Abstract. The randomness and volatility of photovoltaic (PV) output brings difficulties to do real-time forecast. This paper proposed a new method called EMD-SVM-GA to forecast PV output real-time. Three steps were needed to carry out the method. Firstly, decompose historical PV output data sequence into several Intrinsic Mode Functions (IMF) and a Residual Component based on Empirical Mode Decomposition (EMD) theory. Secondly, construct Support Vector Machine (SVM) models for every component and search best parameters by Genetic Algorithm. Then superpose all forecast results to obtain the final data. Examples show that EMD-SVM-GA method has better performance and higher accuracy.

1. Introduction

Studies indicate that PV output is influenced by meteorological factors, like solar irradiance, temperature, wind speed, cloud, shadow and so on[1-2], showing obvious randomness and volatility. When large-scaled PV station access to the grid, the randomness and volatility can bring huge impact to the grid and increase difficulties of operation. So it is important to do PV output forecasting.

There are less papers discussing real-time forecasting in PV output. Some great achievements are introduced now. For example, Reference [4] proposed a forecasting model by combining numerical weather forecast with ground-based cloud images. Reference [5] proposed a self-adapting very-short-term forecast model. Based on the established SVM classifier, the power curve type of the sequential 15 minutes is determined through historical PV output data of the previous 30mimutes.

References show that SVM is suitable for forecasting PV real-time output. For higher accuracy and better performance, we need to make some improvement in the single SVM model.

So, this paper proposed a new method of PV output real-time forecasting called EMD-SVM-GA method. Three steps were needed to carry out it. Firstly, decompose PV output data sequence into several Intrinsic Mode Functions (IMF) and a Residual Component based on Empirical Mode Decomposition (EMD) theory. Secondly, construct Support Vector Machine (SVM) models for every component and search best parameters by Genetic Algorithm. Then superpose all forecast results to obtain the final data. Examples show that EMD-SVM-GA method has better performance and higher accuracy.

2. Modeling Theory

2.1 EMD Theory

Doctor Huang put up EMD theory in 1980s [6]. It is a new signal separated method, which making components separated from the original non-linear, non-stationary signal. The several different characteristic scaled components are called Intrinsic Mode Functions (IMFs), and the last signal showing the changing trend is called Residual Component. IMF can be regarded as free of complex harmonic vibration component characterization system inherent vibration mode. Compared to the traditional sequential decomposition method, wavelet decomposition, EMD avoids empirically selecting wavelet basis and decomposition level, having better adaptive features. So it has great advantage to be applied in the PV output sequence decomposition.

2.2 Support Vector Machine.

SVMs are statistics learning tools introduced by Vapnik in 1995[7], these are usually used in classification and regression problems. The approach allows defining complex surfaces in spaces of large dimensions with very concise representation. SVM can treat problems of regression (linear or nonlinear) and have good learning ability of small sample data. SVM model can be described as follows:

Given a training data $\{(x_i, y_i)\}(i = 1, 2, ..., n)$, where $x_i \in R^n$ is the input vector of dimension n; $y_i \in R^n$ is associated to the desired output value of x_i.

The SVR algorithm determines f as a linear function of the form:

$$f(x) = (w \cdot \varphi(x)) + b \qquad (8)$$

Where $\varphi(x)$ is a nonlinearly mapping from the input space x; w is a vector of weights, and b a constant.

Finally, the model can be written as:

$$f(x) = \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*)k(x_i, x) + b \qquad (9)$$

Where the positive constants α_i, α_i^* are the Lagrange multipliers. The bias parameter b can be calculated with the conditions of Karush-Kuhn-Tucker (KKT).The function k is called kernel function; polynomial kernels are the most used for SVM models, sigmoidal and radial basis function (RBF) defined as follows:

$$K(x_i, x) = \exp\left(-\|x_i - x\|^2 / 2p^2\right) \qquad (10)$$

Thus *SVM* model needs to select parameters C and p.

3. Sample Study

In this paper, generation data from Oct1 to Oct3 which were got from a 60kW PV plant in southwest China were used as samples. The method called EMD-SVM-GA was used to do real-time forecasting for PV output testing samples. For the PV plant didn't generate before 8o'clock and after 18o'clock, so we choose generation data from 8:15 to 17:45 in the three days to do test.

Decomposed those 120 data by EMD, four IMFs IMF$_1$~IMF$_4$ and a Residual Component r were obtained. These components were shown in Fig.1.Set the first 85 data as the training samples and the last 35 data as the testing data. Establish 5 SVM models for every component. After that, add all the results together and get the final forecasting generation. The forecasting data and real data were compared in Fig.2.
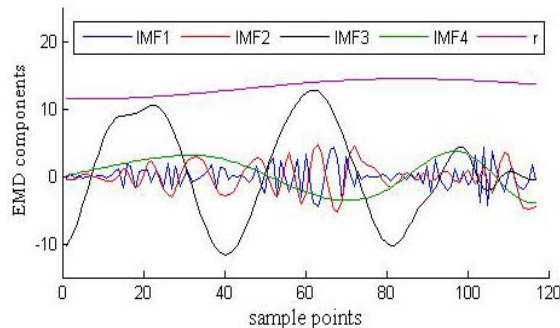

Fig. 1 The signal components of PV output data decomposed by EMD method
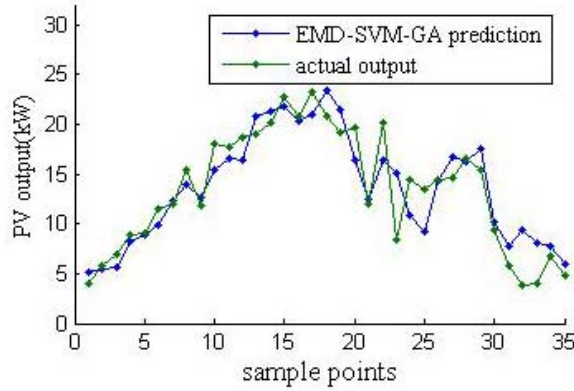
Fig. 2 The comparison of EMD-SVM-GA method and the real data

As shown here, EMD-SVM-GA model has good performance to do real-time forecasting. Prediction curve are close to the actual generation curve, even around the corner. All the samples' RMSE is 2.38% after calculation, satisfying real-time operation requirements.

To further explore the model performance and observe different steps' effect, establish similar models that some steps were replaced.

To study the impact that SVM style brings, establish EMD-LSSVM-GA model that replace SVM for LSSVM. LSSVM stands for Least Squares Support Vector Machine. It changes equality constraints in standard SVM to inequality constraints, so it has weaker simulation difficulty. Forecast by the same samples above, we can get the results in Fig.3.
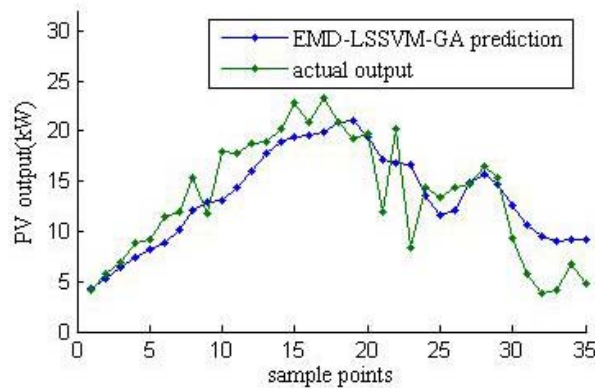

Fig.3 The comparison of EMD-LSSVM-GA method and the real data

The RMSE that predicted by EMD-LSSVM-GA is 4.49%, higher than EMD-SVM-GA method. Figure 4 shows that the predicting curve is relatively flat, although there is no large error among all points, but it also cannot response change. For PV output real-time forecasting, it is not a good model.

Theoretically, EMD plays a role that smooth fluctuating series, so that samples are more likely to be learning. To verify the actual contribution of EMD made in the prediction, design a new test that use output data without decomposition and use SVM and LSSVM mentioned above to predict. SVM-GA model and LSSVM-GA model were built and their results are shown in Fig.4.
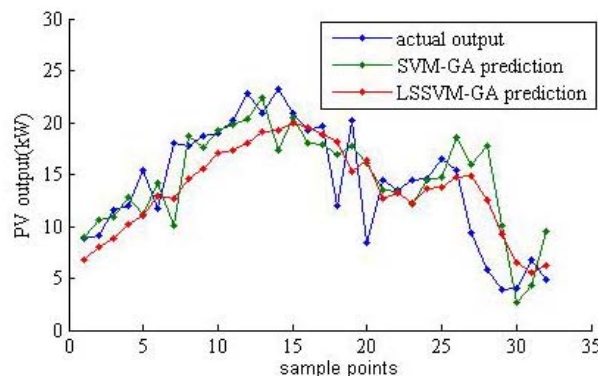

Fig. 4 The comparison of SVM-GA method, LSSVM-GA method and the real data

The calculated RMSE of SVM-GA model is 3.34%, and LSSVM-GA model is 5.05%, both higher than EMD-SVM-GA. SVM-GA has some large errors at inflection points, leading to misjudgment of decision-makers. LSSVM forecasting curve fits actual data curve weakly, and does not reflect curve fluctuation. So these two models' applicability is weaker than EMD-SVM-GA model.

We can get RMSEs which are shown in Table1.

Table 1 The RMSE of four samples

Model Name	EMD-SVM-GA	EMD-LSSVM-GA	SVM-GA	LSSVM-GA
RMSE	2.38%	4.49%	3.34%	5.05%

4. Summary

The randomness and volatility of photovoltaic (PV) output bring difficulties of increasing real-time forecast accuracy. This paper proposed a new method of PV output real-time forecasting called EMD-SVM-GA method. Three steps were needed to carry out the method. Firstly, decompose PV output data sequence into several components called Intrinsic Mode Functions (IMF) based on Empirical Mode Decomposition (EMD) theory. Secondly, construct Support Vector Machine (SVM) models for every IMF and search best parameters by Genetic Algorithm. Then superpose all forecast results to obtain the final data. Examples show that EMD-SVM-GA method has better performance and higher accuracy.

References

[1]. Ren Hang, Ye Lin, "Simulation of the impact of environmental factors on small-scale solar power generation system operation," Electrical Technology, vol. 10, 2010, pp. 158-165.
[2]. Zhang Xueli, Liu Qihui, Ma Huimeng, Li Bei, "Analysis of impact factors of PV plant generation,"Grid and Clean Energy, vol.28 (5), 2012, pp.76-81.
[3]. National Energy Administration, "Interim measures for wind farm power forecast management," Solar, vol.14, pp.6-7.
[4]. Zhu Xiang, Ju Rongrong, Cheng Xu,"A very short-term prediction model for photovoltaic power based on numerical weather prediction and ground-based cloud images," Automation of Electric Power System, vol.39,2015,pp.4-10.
[5]. Gao Yang, Zhang Biling, Mao Jingli, "Machine learning-based adaptive very-short-term forecast model for photovoltaic power, "Power System Techonology, vol.39, Feb.2015, pp. 307-311.
[6]. Huang N E, Shen Z, Long S R, Wu M C, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proceedings of the Royal Society of London A, vol.454,1998,pp. 903-995.
[7]. Vapnik V N, "The Nature of Statistical Learning Theoty," Springer-Verlag, 1995.