Training Set of Support Vector Regression Extracted by Empirical Mode Decomposition

Han Zhong-he
Department of Power Engineering
North China Electric Power University
Baoding, China, 071003
e-mail: zhonghehan@yahoo.com.cn

Zhu Xiao-xun
Department of Power Engineering
North China Electric Power University
Baoding, China, 071003
e-mail:zhuxiaoxun@163.com

Abstract—**Support vector regression (SVR) is a common learning method for machines which is developed these years. Comparing with the other regression models, this method has the advantages of structural risk minimization and strong generalization ability. It is widely used in the prediction field and acquires good effects. The training characters of SVR model are very important to SVR. To solve the problem, this paper puts forward a method of SVR training by the characters which extracted by empirical mode decomposition (EMD). This method firstly uses the EMD to decompose the signal to get the intrinsic mode function (IMF), and then uses the components data of each time spot as features to train SVR. Meanwhile, the forecasting model is obtained. This method is used to forecast the wind speed. The experiment shows that the method improves the calculating precision greatly, increases the number of effective forecasting points, and has the self-adoptive characteristic.**

Keywords-Support Vector Regression; Empirical Mode Decomposition; Feature Extraction; Wind Speed Forecast

I. INTRODUCTION

Currently, the method which is used in prediction mainly include support vector regression (SVR) [1-2], neural network [3-5], wavelets neural network [6], chaos [7] and so on. Comparing with neural network and other forecasting methods, SVR is used widely in predication field, with the advantages of structural risk minimization and strong generalization ability, which acquires good effects.

In the training of SVR, features play an important role in prediction accuracy. Presently, the feature use to train the SVR is generally associated factors. For example, Document [8] adopt the data of wind direction, temperature, air pressure, etc to predict wind speed; document [9] adopt the data of wind speed of prediction spot in the same time of several days to train SVR model and predict wind speed; However, in practice, there has some difficult in extraction feature and choose it has much subjectivity, then the information will be not enough or superfluous in this condition. Thus it will not only cause the precision of the forecasting result lower, but also very little time for the valid predicting.

To solve this problem, this paper puts forward a method which based on empirical mode decomposition (EMD) [10-12] to input character to train SVR .This method adopts EMD to decompose time series, and intrinsic mode function (IMF) is get, by means of IMFs of each time spot to train SVR. So it can enlarge the data information on the basis of not adding other data. With the experiments proved, the method improves a lot in the prediction accuracy and increases valid predictors of time.

II. SUPPORT VECTOR REGRESSION

A. Support Vector Regression

In the 1990s, it is Vapnik that put forwards the Support Vector Machine (SVM) theory [13-14] for the first time. SVM is machinery technique that is based on structural risk minimization, which has the quality of generation and accuracy. This method is aimed for classifications at first, later people apply it to regression, which acquire better effects. Regression is similar to classification. For the given train samples:

$$T = \left\{ \left(x_1, y_1\right), \left(x_2, y_2\right), \dots, \left(x_n, y_n\right) \right\}$$

To restrict function $f(x)$ to be linear function:

$$y = f(x) = (\omega \cdot x) + b \tag{1}$$

Construct and solve the optimization:

$$\begin{aligned} \min_{\alpha^{(*)} \in R^{2n}} \frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(x_i \cdot x_j) \\ + \varepsilon \sum_{i=1}^{n} (\alpha_i^* + \alpha_i) - \sum_{i=1}^{n} y_i(\alpha_i^* - \alpha_i) \end{aligned} \tag{2}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) = 0 \tag{3}$$

$$0 \le \alpha_i^*, \alpha_i \le C, \quad i = 1, 2, ..., n \tag{4}$$

Gain the optimum value:

$$\alpha^{(*)} = (\alpha_1, \alpha_1^*, ..., \alpha_n, \alpha_n^*)^T$$

By means of optimum value, it will get ω and b :

978-1-4244-6255-1/11/$26.00 ©2011 IEEE

$$\omega = \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) x_i \quad (5)$$

$$b = y_i - \sum_{j=1}^{n} (\alpha_j^* - \alpha_j)(x_j \cdot x_i) + \varepsilon \quad (6)$$

Finally, construct the linear regression function. In which, C is error costs, $\alpha_i^{(*)}$ is Lagrange multiplier, ε is insensitive loss function.

However, there are more nonlinear situations, so the linear conditions must be expanded [15]. The key is the kernel function $K(x_i, x_j)$, with the mapping of reduced dimensionality space into higher dimensional space, optimize the question to be:

$$\min_{\alpha^{(*)} \in R^{2n}} \frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(x_i \cdot x_j) + \varepsilon \sum_{i=1}^{n} (\alpha_i^* + \alpha_i) - \sum_{i=1}^{n} y_i (\alpha_i^* - \alpha_i) \quad (7)$$

Thus, the nonlinear question transforms into higher dimensional linear question, and effectively avoids “dimension disaster”. At present, the widely used kernel functions include RBF kernel function, polynomial kernel function, Fourier Kernel function etc [16]. Because RBF kernel function can be easily acquired and parameter can select simple, so the model in this paper adopts RBF kernel function.

B. The selection of training set

The basic data of SVR is the training set. It is vital for SVR to choose an appropriate training set. A proper training set input contains feature neither too much nor too little. If too much, it will include the feature which unrelated to the SVR model. If too little, it will lack of the important feature that response to the model.

At the present time, it is difficult and subjective to extract the feature, so the information will be not enough or superfluous. Thus it will cause the precision of the prediction low not only, but also make the valid forecasting time less. So we have to extract the useful feature to train the machine under the condition of the existing time sequence.

III. THE SVR TRAINING METHOD BASED ON EMD FEATURE EXTRACTING

This method deal with time series based on EMD so as to acquire intrinsic mode function (IMF) of time sequence, and then uses the IMF of each time spot as features to train SVR, which gets the prediction model of this sequence.

A. Empirical mode decomposition

The method of EMD is to decompose the signal into several relatively stable and unrelated IMF.

An IMF should satisfy two conditions: first, the number of the extreme points and the zero-crossing must equal or up to a difference; second, the mean of the upper envelope line defined by the maximum points and the lower envelope line defined by the minimum points is zero. In other words, the upper envelope line and the lower envelope line symmetry in time axis.

EMD method first to find out the local extreme point, then get the upper and lower envelope line by the interpolation methods, finally calculates the mean of the upper and lower envelope line and label as m1, calculates h_{11} through formula (8);

$$h_{11} = x(t) - m_1 \quad (8)$$

Repeat the progress above till h_{11} satisfy two conditions of IMF, then we consider h_{11} as the first IMF and label as $c_1 = h_{1,k}$, continue to calculates in this way, we will finally obtain all of the IMF c_n, thus the signal can express as:

$$x(t) = \sum_{i=1}^{n} c_i \quad (9)$$

These IMFS have some good properties, such as self-adaptive, completeness, approximate orthogonality, and the modulation components and so on [17].

B. Training method of SVR based on EMD

Aiming at the problem of the SVR training, we raise the method of using EMD to construct input feature in this paper. The detailed steps are as follows:

First, for a given set of time series *X*={x(1), x(2),…, x(n)}, we use EMD to decompose the time series into several IMFS

$$\{c_1, c_2, \ldots, c_m\}.$$

And then use the IMFS which are gained by EMD to construct training set:

$$\min_{\alpha^{(*)} \in R^{2n}} \frac{1}{2} \sum_{i,j=1}^{n-1} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(dx_i \cdot dx_j) + \varepsilon \sum_{i=1}^{n-1} (\alpha_i^* + \alpha_i) - \sum_{i=1}^{n-1} y_i (\alpha_i^* - \alpha_i) \quad (10)$$

In which $y_i = x(i+1)$,

By finding the optimal solution of formular (8), then the regression function can be gained.

$$f(dx) = \sum_{i=1}^{n-1} (\alpha_i - \alpha_i^*) K(dx_i, dx) + b \quad (11)$$

In brief, this method establishes a mapping relation *f:* $R^m \to R$ between IMF values of the point before forecasting point and forecasting point.

IV. RESEARCH ON APPLICATION

A. Research on wind speed forecast

The experiment in this paper is wind speed prediction, the interval of the sample data is 10 minutes, and take 120 points in total (20 hours). The first 100 points is used as training set and the 20 points behind as the forecasting sample points. Figure 1 is the wind wave diagram.

Use the feature extracted by EMD to construct SVR model and compare to the SVR training methods constructed by temperature, angle and pressure.

We define the average error and relative error as the standard measure of forecast accuracy, see formula (12) (13). The too large prediction error will cause forecast meaningless, so we define the point that prediction error is less than 10 percents as valid forecast point.

$$ME = \frac{\sum_{i=1}^{N} |x(i) - x'(i)|}{N} \tag{12}$$

$$RE = \frac{|x(i) - x'(i)|}{x(i)} \tag{13}$$

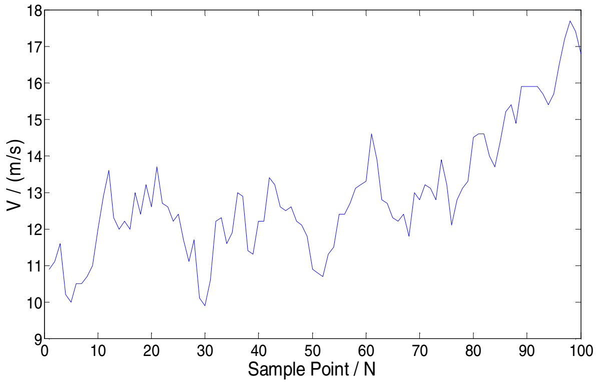

Figure 1. *Original time signal series*

B. Experimental procedure

First, decompose the time series of wind speed, and get five IMFs c1～c5, the amplitude is zero mean amplitude, show in Figure .2.

Second, use the IMF value of each time point as the feature to train SVR, finally gain the forecast modal of the wind speed. Figure 3 is comparison of the precision of the two different models.

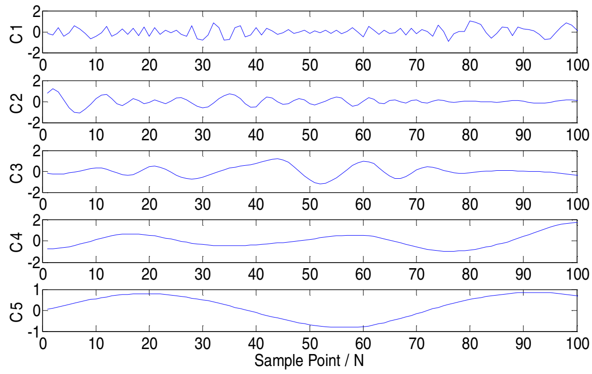

Figure 2. IMFs of time signal series

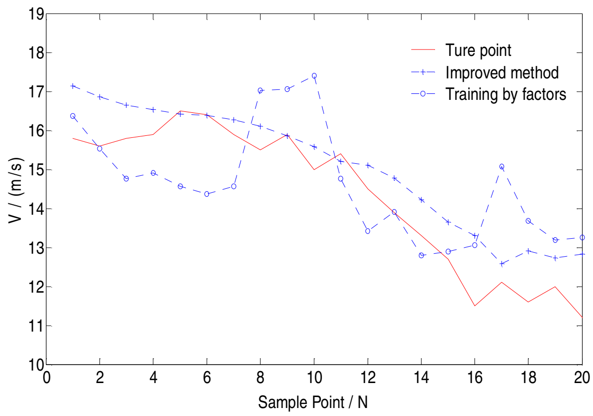

Figure 3. Comparison of prediction results

C. Analysis of prediction result

According to the figure3, the prediction values of SVR model trained by EMD are close to the true value, and receive more satisfactory predicted results. But the prediction values of SVR model trained by factors are too different to the true value and the prediction error is big. The more accurate prediction can be found through the contrast between the two.

The table 1 shows the comparison of the two forecasting results based on the total absolute error and average error, relative error and the valid prediction points.

TABLE I. PREDICTION RESULT COMPARISON

	training by EMD	**training by factors**
total absolute error	15.3053	25.3102
Average error	0.7653	1.2655
relative error(%)	4.64	7.67
valid prediction point(n)	17	13

According to table 1, it shows that the first method is better than the second method in total absolute error and relative error. The number of effective forecasting points of the first is 17 (85 percents of the total), however the number of the second is 13 (65 percents of the total).

V. CONCLUSION

Aiming at the problem of SVR feature extraction, the SVR training method based on EMD feature extracting is put forward in this paper. First the time series is decomposed into several mode functions by EMD, then train SVR by the IMF of each point. This method has some advantages:

- The adaptive characteristic of the feature extraction;
- The prediction error is greatly reduced;
- The valid prediction length greatly increased.

Overall, this method is an effective，adaptive and accurate forecasting methods.

REFERENCES

[1] Zhao Dengfu, Wang Meng, Zhang Jiangshe, A Support Vector Machine approach for short Term load forecasting[J], Proceedings of the CSEE, 2002, 22(4): 26-30. (in Chinese)

[2] LI Yuancheng, FANG Tingjian, YU Erkeng, Study of Support Vector Machines for Short-Term load forecasting[J], Proceedings of the CSEE, 2003, 23(6): 55-59. (in Chinese)

[3] Hippert H S, Pedreira C E, Souza R C, Neural networks for short-term load forecasting: A review and evaluation[J], IEEE Trans on Power System, 2001, 16(1): 44—55.

[4] Saini L M, Soni M K, Artificial neural network-based peak load forecasting using conjugate gradient methods [J], IEEE Trans on Power System, 2002, 17(3): 907—912.

[5] Feng Zhipeng, Song Xigeng, Xue Dongxin, General regression neural network based prediction of time series[J], Journal of Vibration Measurement & Diagnosis, 2003, 23(2), 105-109. (in Chinese)

[6] Alexander Y, Ronald L, Recognition of Bearing Failures using Wavelets and Neural Networks, TFTS97, Coventry(UK), August 27-29, 1997.

[7] Zhao Xiaomei, Research on Chaotic Time Series Forecasting and Chaotic optimization[D], Doctoral thesis, Hangzhou: Zhejiang University, 2002. (in Chinese)

[8] Peng Huaiwu , Yang Xiaofeng , Liu Fangrui, Short-Term Wind Speed Forecasting of Wind Farm Based on SVM Method[J], Power System and Clean Energy, 2009, 25(7), 48-52. (in Chinese)

[9] Zhu Zhiyong, Lin Mugang, Zhang Shenji, Short-Term load forecasting based on wavelet transform and support vector machine[J], Microcomputer Applications, 2005, 26(4), 440-442. (in Chinese)

[10] Huang N E, Shen Z, The empirical mode decomposition and the hilbert spectrum for non-linear non—stationary time series analysis [C] Proc. R. Soc, London, l998.

[11] Huang N E, A new view of non-linear waves: the hilbert spectrum [J]. Annual Review of Fluid Mechanics, 1999, 31(5): 417—457.

[12] Xiang Ling, Zhu Yongli, Tang Guiji, Application of Hilbert-Huang Transform Method in Vibration Faults Diagnosis for Rotor System[J], Proceedings of the CSEE, 2007, 27(35), 84-89. (in Chinese)

[13] Vapnik V, The nature of statistical learning theory[M], New York: Springer-Verlag, 1995.

[14] Vapnik V, An overview of statistical learning theory [J], IEEE Trans on Neural Network, 1999, 10(5), 988-999.

[15] Deng Naiyang, Tian Yingjie, New Method for Data Mining-Support Vector Machine [M], Beijing: Science Press, 2004. (in Chinese)

[16] Fang Ruiming, Theory and Application Analysis of Support Vector Machine [M], Beijing, China Electric Power Press, 2007. (in Chinese)

[17] Hu Aijun, Research on the application of Hilbert-Huang Transform in vibration signal analysis of rotating machinery[D], Doctoral thesis, Baoding: North of China Electric Power University, 2008. (in Chinese)