

Research Article

Support Vector Regression Method for Wind Speed Prediction Incorporating Probability Prior Knowledge

Jiqiang Chen,^{1,2} Xiaoping Xue,¹ Minghu Ha,² Daren Yu,³ and Litao Ma²

¹ Department of Mathematics, Harbin Institute of Technology, Harbin 150001, China

² School of Science, Hebei University of Engineering, Handan 056038, China

³ School of Energy Science and Engineering, Harbin Institute of Technology, Harbin 150001, China

Correspondence should be addressed to Xiaoping Xue; xiaopingxue@263.net

Received 3 December 2013; Accepted 20 January 2014; Published 4 March 2014

Academic Editor: Huaiqin Wu

Copyright © 2014 Jiqiang Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Prior knowledge, such as wind speed probability distribution based on historical data and the wind speed fluctuation between the maximal value and the minimal value in a certain period of time, provides much more information about the wind speed, so it is necessary to incorporate it into the wind speed prediction. First, a method of estimating wind speed probability distribution based on historical data is proposed based on Bernoulli's law of large numbers. Second, in order to describe the wind speed fluctuation between the maximal value and the minimal value in a certain period of time, the probability distribution estimated by the proposed method is incorporated into the training data and the testing data. Third, a support vector regression model for wind speed prediction is proposed based on standard support vector regression. At last, experiments predicting the wind speed in a certain wind farm show that the proposed method is feasible and effective and the model's running time and prediction errors can meet the needs of wind speed prediction.

1. Introduction

Wind power is a clean, renewable energy that will play an increasingly important role in the future electricity supply [1]. Unfortunately, due to the stochastic and nonstationary nature of wind, the wind power is variable and uncontrollable. It is difficult to maintain the balance between the supply and the demand of electricity, which is required by the electricity system [2]. Wind speed prediction is a key point in the management of wind farms because it is directly related to the power produced by each of the farm's turbines, so it is usually the base of wind power forecasts, and it is necessary to increase the accuracy of the wind speed prediction for the effective use of wind energy.

At present, there are mainly two kinds of wind speed prediction methods. One is based on the physical model, and the other is based on historical data. The prediction methods based on physical model often use the numerical weather prediction (NWP) data for wind speed prediction [3, 4]. Wind speed prediction methods based on NWP do

not focus on the speed of a farm's turbines but on the speed of a region. Thus, it needs to solve the problem of how the wind speed of a region is mapped to the wind speed of a certain wind generator. Wind speed prediction methods based on historical data predict the wind speed by using correlations among the initial data. In 2008, Louka et al. [5] improved wind speed forecasts for wind power prediction using Kalman filtering. In 2012, Cao et al. [6] presented a comparative analysis of the wind speed prediction accuracy of univariate and multivariate ARIMA models with their recurrent neural network counterparts. In 2013, Woods et al. [7] developed a method to produce synthetic time series of wind power at several locations based on a measured time series of wind speed from a reference site, and so on.

In the 1990s, Vapnik et al. [8, 9] proposed support vector machines (SVMs), including support vector classifications (SVCs) and support vector regressions (SVRs). SVMs focus on the statistical learning problems for small size samples by solving a convex quadratic optimization problem and can solve the local minimization problem which cannot

be avoided by the neural network algorithm. SVMs use a kernel function to map the data in original space to a high dimensional feature space and then solve the nonlinear decision problem in high dimensional space. Thus, SVMs can successfully solve the problem of dimension disaster and have good generalization ability. However, the standard SVMs focus on historical data and cannot incorporate prior knowledge into learning process, which may causes the generalization ability of the standard SVMs to decrease. Therefore, in 2009, Guan et al. [10] proposed a modified method that incorporated prior knowledge into cancer classification based on gene expression data to improve accuracy. In 2011, Zhang et al. [11] proposed a fully Bayesian methodology for generalized kernel mixed models, which are extensions of generalized linear mixed models in the feature space induced by a reproducing kernel. In 2012, Liu and Xue [12] focused on designing a new class of kernels to incorporate the prior information into the training process of support vector regressions. Currently, SVMs have received extensive attention and are attracting more and more scholars to study from different views [13–22].

In 2011, Zhou et al. [23] presented a systematic study on fine tuning of LS-SVM model parameters for one-step ahead wind speed prediction, and Ortiz-García et al. [24] proposed an improvement to an existing wind speed prediction system using banks of regression support vector machines for a final regression step in the prediction system.

However, for the problem of wind speed prediction in practice, there is much prior knowledge. For example, the wind speed has a certain probability distribution in a season or in a day, and the probability distribution can be estimated with historical wind speed data. As the probability distribution can provide much more information about the wind speed, it is necessary to incorporate it into the wind speed prediction. Also, in a wind farm, the output wind speed v at a fixed time t is the mean value \bar{v} of many measured values v_k ($k = 0, 1, \dots, l$) during a certain period of time Δt . Assume that $v_{\max} = \max_k \{v_k\}$ and $v_{\min} = \min_k \{v_k\}$, then the larger the $v_{\max} - v_{\min}$ is, the more the fluctuation of wind speed during the period of time Δt is. Conversely, the smaller the $v_{\max} - v_{\min}$ is, the less the fluctuation of wind speed during the period of time Δt is. Nevertheless, the mean value \bar{v} does not provide this prior knowledge at all. Therefore, in order to decrease the wind speed prediction errors, it is necessary to find a way to incorporate this prior knowledge in the wind speed prediction. However, the present methods for wind speed prediction often used the historical wind speed data directly to predict the wind speed, instead of dredging information from the data, and the prediction errors are difficult to decrease. Therefore, in order to decrease the prediction errors of the wind speed at a fixed time, the probability distribution of historical wind speed data is estimated and incorporated into the training data and testing data to provide the information about the wind speed fluctuation. Then a support vector regression model for wind speed prediction is proposed combined with the standard SVR.

This paper is structured as follows. Section 2 is the preliminaries. Section 3 is the method of estimating the probability

distribution of the historical wind speed data. Section 4 is to incorporate the prior knowledge about the wind speed fluctuation into the training data and testing data and then establish the ε -support vector regression method for wind speed prediction incorporating probability prior knowledge (PPK- ε -SVR). Section 5 includes two experiments with the historical wind speed data coming from a wind farm in Gansu province, and Section 6 draws the conclusions.

2. Preliminaries

In this section, we briefly review some relevant knowledge of probability theory and the standard support vector regression often used in applications.

Let (Ω, \mathcal{F}) be a measurable space, P a function defined on \mathcal{F} , and $0 \leq P(A) \leq 1$ for any $A \in \mathcal{F}$. We call $P(A)$ the probability of event A occurring, and $P(A)$ indicates the possibility of event A occurring. Assume that $X(\omega)$ is a random variable on Ω , $F(x) = P\{X(\omega) < x\}$ ($x \in R$) is the distribution function, $E(X)$ is the expectation, and $D(X)$ is the variance of random variable X , respectively.

Definition 1 (see [25]). Let $F(x)$ be the distribution function of random variable $X(\omega)$. If there exists a nonnegative and integrable function $f(x)$ satisfying

$$F(x) = P\{X < x\} = \int_{-\infty}^x f(t) dt, \quad \forall x \in R, \quad (1)$$

then we call $f(x)$ the probability density of continuous random variable $X(\omega)$.

Definition 2 (see [25]). Let $\{X_n\}$ be a sequence of random variables. $\{X_n\}$ is said to be independent if for any finite random variables X_1, X_2, \dots, X_k ($k \leq n$) are independent.

Definition 3 (see [25]). Let the sequence of random variables $\{X_n\}$ be independent. $\{X_n\}$ is said to be independent and identically distributed if they have the same distribution function $F(x)$.

Definition 4 (see [25]). Let $\{X_n\}$ be a sequence of random variables and let a be a constant. If, for any $\varepsilon > 0$, we have

$$\lim_{n \rightarrow +\infty} P\{|X_n - a| \geq \varepsilon\} = 0 \quad (2)$$

or

$$\lim_{n \rightarrow +\infty} P\{|X_n - a| < \varepsilon\} = 1, \quad (3)$$

then the sequence $\{X_n\}$ is said to converge in probability to a . Denoted by $X_n \xrightarrow{P} a$.

Theorem 5 (Bernoulli's law of large numbers [25]). Let n_A be the number of event A occurring in n independent duplicate experiments, and let p be the probability of event A occurring. Then, for any $\varepsilon > 0$, we have

$$\lim_{n \rightarrow +\infty} P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1; \quad (4)$$

that is,

$$\frac{n_A}{n} \xrightarrow{P} p. \quad (5)$$

Remark 6. Bernoulli's law of large numbers proves the frequency stability in theory. In other words, the frequency n_A/n converges in probability to p . Then for a sufficiently large n , the frequency n_A/n almost equals the probability p . Hence, if the number of experiments is very large, the frequency n_A/n can be treated as the probability p in practice.

2.1. Regression Problem. Suppose that

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \quad (6)$$

is a given training set, where $x_i \in R^n$, $y_i \in R$, $i = 1, 2, \dots, l$. The regression problem is to find a real valued function $y = g(x)$ on R^n according to data (6) to predict the output y for any given input x .

Remark 7. For the above regression problem, if the real valued function $y = g(x)$ is linear, then we call it a linear regression problem. If the real valued function $y = g(x)$ is nonlinear, then we call it a nonlinear regression problem.

2.2. ε -Support Vector Regression. Support vector regression (SVR) differs from conventional regression in that it maps the training data (6) into a high dimensional reproducing kernel Hilbert space. In ε -SVR, the goal is to find a function $y = f(x)$ which has at most ε deviation from the actually obtained targets y_i for all the training data and at the same time as flat as possible. And ε -SVR uses an ε -insensitive loss function

$$|\xi_\varepsilon| = \begin{cases} 0, & \text{if } |\xi| < \varepsilon \\ |\xi| - \varepsilon, & \text{otherwise} \end{cases} \quad (7)$$

to solve the optimal problem

$$\begin{aligned} \min_{\omega, b, \xi, \xi^*} & \quad \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s.t.} & \quad y_i - f(x_i) \leq \varepsilon + \xi_i, \\ & \quad f(x_i) - y_i \leq \varepsilon + \xi_i^*, \\ & \quad \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, l, \end{aligned} \quad (8)$$

where $f(x_i) = \omega^T \phi(x_i) + b$ and data (6) are mapped to a higher dimensional characteristic space by the function $\phi(x)$ and the constant $C > 0$ determines the tradeoff between the flatness of $f(x)$ and the amount up to which deviations larger than ε

are tolerated. Similar to support vector classification, ω may be a huge vector variable, and then we solve the dual problem

$$\begin{aligned} \min_{\alpha^{(*)} \in R^{2l}} & \quad \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) K(x_i, x_j) \\ & \quad + \varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\ \text{s.t.} & \quad \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, \\ & \quad 0 \leq \alpha_i^{(*)} \leq C, \quad i = 1, \dots, l, \end{aligned} \quad (9)$$

where $K(x_i, x_j) = \phi(x_i)^T \cdot \phi(x_j)$ is a kernel function.

As a result, SVR has a sparse representation of solutions and hence is relatively fast in training and testing. SVR is the most common application form of SVM and has been popular for regression and function estimation problems in the past decades.

3. Method of Estimating the Probability Distribution of Historical Wind Speed Data

In practical problems, there are a lot of historical wind speed data. In order to dredge more information from them, in this part, we will give a method to estimate the probability distribution of historical wind speed data.

3.1. Method. Assume that $f(x)$ in Figure 1 is the probability density of independent and identically distributed random variables X_k ($k = 1, 2, \dots, l$) and the number and the frequency of X_k falling into interval $[x_{i-1}, x_i]$ ($i = 1, 2, \dots, n$, $n \leq l$) are u_i and u_i/l , respectively. By Definition 1, we can obtain that the probability p_i (namely, the area of trapezoid with curved edge $ABCD$) of X_k falling into interval $[x_{i-1}, x_i]$ is

$$p_i = \int_{x_{i-1}}^{x_i} f(x) dx. \quad (10)$$

By Theorem 5, we have

$$p_i \approx \frac{u_i}{l}, \quad i = 1, 2, \dots, n. \quad (11)$$

With the infinitesimal method, the area of trapezoid with curved edge $ABCD$ can be approximately substituted by the area of trapezoid $ABCD$, namely,

$$\int_{x_{i-1}}^{x_i} f(x) dx \approx \frac{1}{2} [f(x_{i-1}) + f(x_i)] (x_i - x_{i-1}). \quad (12)$$

By formulas (10), (11), and (12) we have

$$f(x_{i-1}) + f(x_i) \approx \frac{2p_i}{x_i - x_{i-1}} \approx \frac{2u_i}{l(x_i - x_{i-1})}. \quad (13)$$

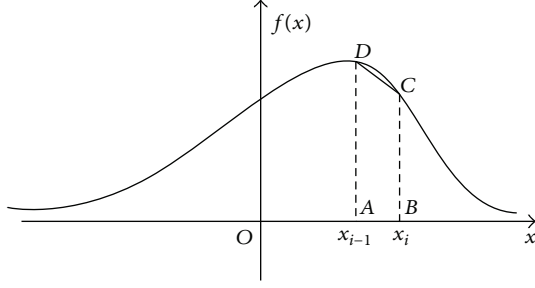


FIGURE 1: The probability density function.

Similarly, for $f(x_{i-1})$, $f(x_i)$ and $f(x_{i+1})$, we have

$$\begin{aligned} f(x_i) + f(x_{i+1}) &\approx \frac{2u_{i+1}}{l(x_{i+1} - x_i)}, \\ f(x_{i-1}) + f(x_{i+1}) &\approx \frac{2(u_{i+1} + u_i)}{l(x_{i+1} - x_{i-1})}. \end{aligned} \quad (14)$$

From formulas (13) and (14) we can obtain that

$$f(x_i) \approx \frac{1}{l} \left(\frac{u_{i+1}}{x_{i+1} - x_i} - \frac{u_{i+1} + u_i}{x_{i+1} - x_{i-1}} + \frac{u_i}{x_i - x_{i-1}} \right), \quad (15)$$

$$i = 1, 2, \dots, n.$$

3.2. Algorithm. Suppose that $v(t)$ is the wind speed at a fixed point A. Then $v(t)$ can be seen as a random variable defined on time set T . The l samples data $v(t_k)$ ($k = 1, 2, \dots, l$) can be seen as a sequence of sample data coming from l independent random variables X_k ($k = 1, 2, \dots, l$). Supposing that the probability density function of wind speed $v(t)$ is $f(x)$, then the steps of estimating the probability density $f(x)$ are as follows.

Step 1. Denote $x_0 = \min_k\{v(t_k)\}$, $x_n = \max_k\{v(t_k)\}$. Insert $n - 1$ points x_j ($j = 1, 2, \dots, n - 1$) between x_0 and x_n equidistantly, and interval $[x_0, x_n]$ is divided into n intervals $[x_{i-1}, x_i]$ ($i = 1, 2, \dots, n$) with the same length.

Step 2. Count the number u_i ($i = 1, 2, \dots, n$) of samples $v(t_k)$ ($k = 1, 2, \dots, l$) falling into interval $[x_{i-1}, x_i]$, and we have $\sum_{i=1}^n u_i = l$. By formula (11), the probability p_i of samples $v(t_k)$ falling into interval $[x_{i-1}, x_i]$ is $p_i \approx u_i/n$ ($i = 1, 2, \dots, n$).

Step 3. Calculate $f(x_i)$ ($i = 1, 2, \dots, n$) with formula (15).

Step 4. Connecting the points $(x_i, f(x_i))$ in turn with smooth curves, the probability density $f(x)$ of wind speed sample data $v(t_k)$ ($k = 1, 2, \dots, n$) can be obtained.

3.3. Experiments. In order to verify the effectiveness of the above method, in this subsection, sample data coming from normal distribution and exponential distribution are used to make experiments, respectively.

Experiment 1. Suppose that X is a standard normal random variable; then its probability density is

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right], \quad -\infty < x < +\infty, \quad (16)$$

and the graph of $f(x)$ is shown in Figure (10).

Suppose that the sample data

$$x_1, x_2, \dots, x_{10000} \quad (17)$$

are independent and identically distributed coming from the probability density $f(x)$ and $x_i \in [-3, 3]$ ($i = 1, 2, \dots, 10000$). Insert 99 points in interval $[-3, 3]$ equidistantly, and interval $[-3, 3]$ is divided into $n = 100$ small intervals with the same length. With the above proposed method, we can obtain function $g(x)$ shown in Figure 2 which is the estimation of the standard normal density $f(x)$.

Experiment 2. Supposing that X is a random variable following an exponential distribution, then the probability density is

$$f(x) = \begin{cases} \frac{1}{2} \exp\left[-\frac{x}{2}\right], & x > 0, \\ 0, & x \leq 0, \end{cases} \quad (18)$$

and the graph of $f(x)$ is shown in Figure 3. With the similar steps in the above experiment, suppose that the sample data $x_1, x_2, \dots, x_{10000}$ are independent and identically distributed coming from the probability density $f(x)$ and $x_i \in [0, 10]$ ($i = 1, 2, \dots, 10000$). Insert 99 points in interval $[0, 10]$ equidistantly, and interval $[0, 10]$ is divided into $n = 100$ small intervals with the same length. With the proposed method above, we can obtain function $g(x)$ which is the estimation of exponential density $f(x)$ (see Figure 3).

3.4. Results Analysis. From Figure 2 we can see that the estimation density $g(x)$ obtained with the sample data is not as smooth as the standard normal density $f(x)$, but the rough shapes of the two functions are almost the same. And from Figure 3 the similar conclusions can be drawn. That is to say, the method of estimating the probability density based on Bernoulli's law of large numbers and infinitesimal method is effective, and it lays a foundation for establishing the PPK- ϵ -SVR.

4. Support Vector Regression Method Incorporating Probability Prior Knowledge

In this section, we aim to predict the wind speed $v(t)$ at a fixed point A.

4.1. Incorporating the Prior Knowledge about the Wind Speed Fluctuation into the Training Data and Testing Data. Supposed that $f(x)$ is the probability density estimated with the above method, t_0 is the initial time, and v_0 is the wind speed at t_0 , denoted by (t_0, v_0) . In practice, the wind speed at a fixed point A is often measured many times t_{0k} ($k = 1, 2, \dots, l$) for

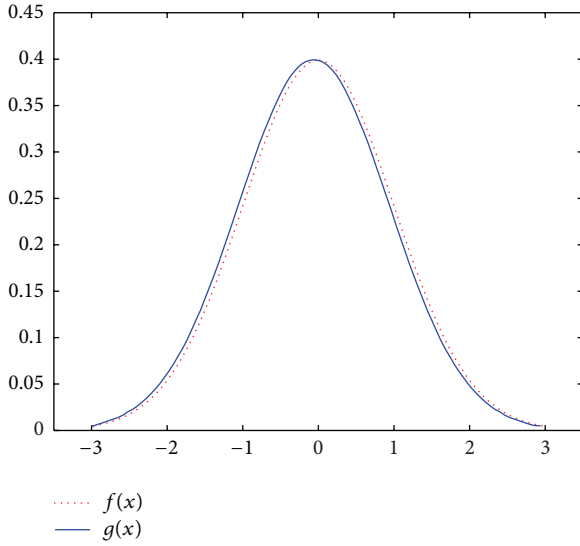


FIGURE 2: Graphs of standard normal densities.

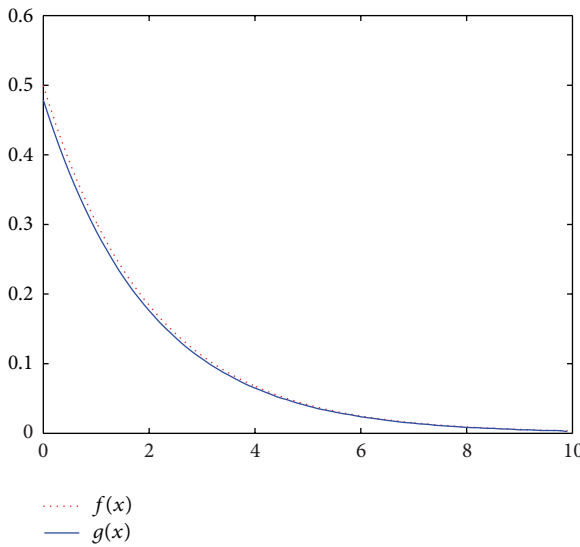


FIGURE 3: Graphs of exponential densities.

every certain period of time Δt (where $t_0 < t_{01} < \dots < t_{0l} \leq t_0 + \Delta t$), and the mean value $\bar{v} = (1/l) \sum_{k=1}^l v_{0k}$ is output as the predicted wind speed at the fixed point A . In other words, the wind speed v_1 at $t_1 = t_0 + \Delta t$ is the mean value \bar{v} of the measured values v_{0k} from t_0 to t_1 , denoted by (t_1, v_1) , where $v_1 = \bar{v}$. Of course, the mean value \bar{v} can represent the wind speed v_1 at t_1 in a sense, but in some cases the mean value \bar{v} is quite different from v_1 . For example, if the measured wind speed v_{0k} ($k = 1, 2, \dots, l$) is the same value during a certain period of time Δt , then “the mean value \bar{v} is the wind speed v_1 ” holds with a high probability. Conversely, if the measured wind speed v_{0k} fluctuates wildly during a certain period of time Δt , then “the mean value \bar{v} is the wind speed v_1 ” holds with a very low probability.

Hence, in order to incorporate this prior knowledge into the wind speed prediction, the training datum (t_1, v_1) is converted into $(t_1, 1 - P_1, v_1)$, where

$$P_1 = P\{v_{\min} \leq v_1 \leq v_{\max}\} = \int_{v_{\min}}^{v_{\max}} f(x) dx, \quad (19)$$

$$v_{\min} = \min_k \{v_{0k}\}, \quad v_{\max} = \max_k \{v_{0k}\}. \quad (20)$$

In fact, from formulas (19) and (20) we can see that the larger the $v_{\max} - v_{\min}$ is, the larger the P_1 is and, furthermore, the smaller the $1 - P_1$ is. That is to say, the possibility of “wind speed at t_1 is v_1 ” is very small. On the other hand, the large $v_{\max} - v_{\min}$ illustrates that the wind speed from t_0 to t_1 fluctuates wildly and “the mean value \bar{v} is the wind speed v_1 ” holds with a low probability (namely, the possibility of “wind speed at t_1 is v_1 ” is very small), which is in accord with the information provided by $1 - P_1$. Thus, the probability $1 - P_1$ provides the fluctuation about the wind speed during a certain period of time Δt . Therefore, datum $(t_1, 1 - P_1, v_1)$ contains the prior knowledge provided by the historical data.

Let t_i ($i = 1, 2, \dots, l$) be the l times at a fixed point A , and let

$$T = \{(t_1, 1 - P_1, v_1), (t_2, 1 - P_2, v_2), \dots, (t_l, 1 - P_l, v_l)\} \quad (21)$$

be a training set. Denote $x_i = (t_i, 1 - P_i)$; then training set (21) can be rewritten by

$$T = \{(x_1, v_1), (x_2, v_2), \dots, (x_l, v_l)\}, \quad (22)$$

where $x_i \in R \times [0, 1]$, $v_i \in [0, +\infty)$, $i = 1, 2, \dots, l$. The problem of wind speed prediction is to find a real valued function $v = g(x)$ on $R \times [0, 1]$ according to training set (22) to predict the wind speed v for any given input x .

4.2. ε -Support Vector Regression Method for Wind Speed Prediction Incorporating Probability Prior Knowledge. For the above problem of wind speed prediction, ε -support vector regression method incorporating probability prior knowledge (PPK- ε -SVR) can be constructed as follows.

Step 1. Obtain the training set

$$T = \{(x_1, v_1), (x_2, v_2), \dots, (x_l, v_l)\}, \quad (23)$$

where $x_i \in R \times [0, 1]$, $v_i \in [0, +\infty)$, $i = 1, 2, \dots, l$.

Step 2. Select a proper kernel function $K(x, x')$, $\varepsilon > 0$ and punishment parameter $C > 0$.

Step 3. Constructing and solving the convex quadratic programming problems:

$$\begin{aligned} \min_{\alpha^{(*)} \in \mathbb{R}^{2l}} \quad & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) K(x_i, x_j) \\ & + \varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l v_i (\alpha_i^* - \alpha_i), \end{aligned} \quad (24)$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, \\ & 0 \leq \alpha_i^{(*)} \leq C, \quad i = 1, \dots, l, \end{aligned}$$

we can obtain the optimal solution

$$\bar{\alpha}^{(*)} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \dots, \bar{\alpha}_l, \bar{\alpha}_l^*)^T. \quad (25)$$

Step 4. Choose the component $\bar{\alpha}_j$ or $\bar{\alpha}_k^*$ of $\bar{\alpha}^{(*)}$ in interval $(0, C)$. If $\bar{\alpha}_j$ is chosen, then

$$\bar{b} = v_j - \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x_i, x_j) + \varepsilon. \quad (26)$$

If $\bar{\alpha}_k^*$ is chosen, then

$$\bar{b} = v_k - \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x_i, x_k) - \varepsilon. \quad (27)$$

Step 5. Construct the decision function with \bar{b}

$$v = g(x) = \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x_i, x) + \bar{b}. \quad (28)$$

Remark 8. Solving regression problems with support vector regression, the kernel function can be selected according to prior knowledge, such as the characteristics of the problem, or the training set. More details about selecting kernel function with prior knowledge will be investigated in another paper.

5. Experiments

In this part, we take a wind farm in Gansu province as an example. For a fixed point A , in order to predict the wind speed $v(t)$ (m/s) at t (Hours), we recorded the wind speed from November 2006 to April 2008 and found that the wind speed had a periodicity with the change of seasons, months, or days. Therefore, the probability distribution P of wind speed in the period of the previous year (month or day) can be incorporated into the wind speed prediction of the corresponding period in this year (month or day). In this wind farm, the wind speed v is output every ten minutes, so there are 144 sets of data in a day and more than four thousand sets of data in a month. As SVR focuses on the statistical learning problems for small size samples and the wind speed had a periodicity with the change of days, the experiment is aimed at the short-term wind speed prediction, and we choose 144 sets of data to carry out the experiment. Here, without loss of generality, we take the wind speed prediction on 1 April 2008 as an example.

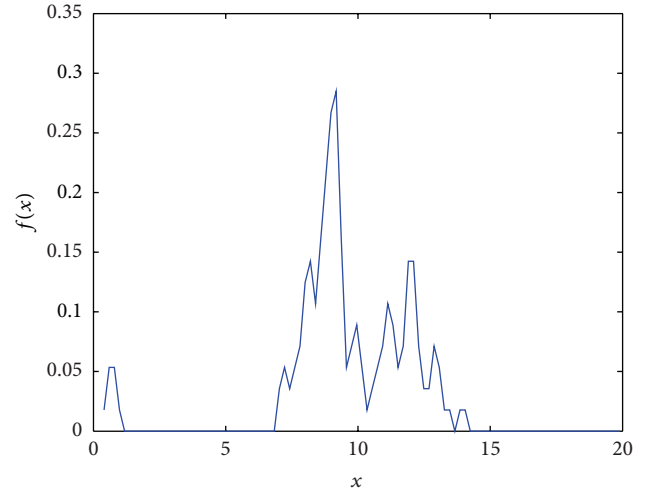


FIGURE 4: Probability density of wind speed data on 1 April 2008.

5.1. Estimation of the Probability Density of Historical Wind Speed Data. Supposed (t_i, v_i) ($i = 1, 2, \dots, 144$) indicates that the wind speed at time t_i is v_i . Wind speed was measured ten times $v(t_{ik})$ ($k = 1, 2, \dots, 10$) during every ten minutes, and the mean value $\bar{v}_i = (1/10) \sum_{k=1}^{10} v(t_{ik})$ is output as the wind speed v_{i+1} at time $t_i + 1/6$. By the proposed method in Section 3, the probability density $f(x)$ of the wind speed data on 1 April 2008 was estimated and the graph of $f(x)$ is shown in Figure 4.

5.2. Incorporating the Prior Knowledge about the Wind Speed Fluctuation into the Training Data and Testing Data. Denote that $v_i^{\max} = \max_k \{v(t_{ik})\}$ and $v_i^{\min} = \min_k \{v(t_{ik})\}$. By the estimated probability density $f(x)$ of the wind speed data on 1 April 2008 and formula (19), the probability $P_i = P\{v_i^{\min} \leq v_i \leq v_i^{\max}\}$ can be calculated. The wind speed sample data on 1 April 2008 are

$$(t_1, v_1), (t_2, v_2), \dots, (t_{144}, v_{144}), \quad (29)$$

and data (29) are converted into

$$(x_1, v_1), (x_2, v_2), \dots, (x_{144}, v_{144}) \quad (30)$$

by the method in Section 4.1, where $x_i = (t_i, 1 - P_i)$ ($i = 1, \dots, 144$). And the data $(x_1, v_1), (x_2, v_2), \dots, (x_{143}, v_{143})$ in data (30) is the training data to train a model to predict the wind speed v_{144} for the given x_{144} .

Similarly, if we want to predict the wind speed v'_1 for the given x'_1 on 2 April 2008, we can estimate the probability density of wind speed (t_i, v_i) ($i = 2, 3, \dots, 144$) on 1 April 2008 and use the data (x_i, v_i) ($i = 2, 3, \dots, 144$) as the training data.

5.3. ε -Support Vector Regression Method for Wind Speed Prediction Incorporating Probability Prior Knowledge. In the experiment, a grid search method based on 5-fold cross-validation is chosen to determine model parameters, where $\sigma \in \{2^{-10}, 2^{-9}, \dots, 2^1\}$, $C \in \{2^{-6}, 2^{-5}, \dots, 2^6\}$, and the kernel

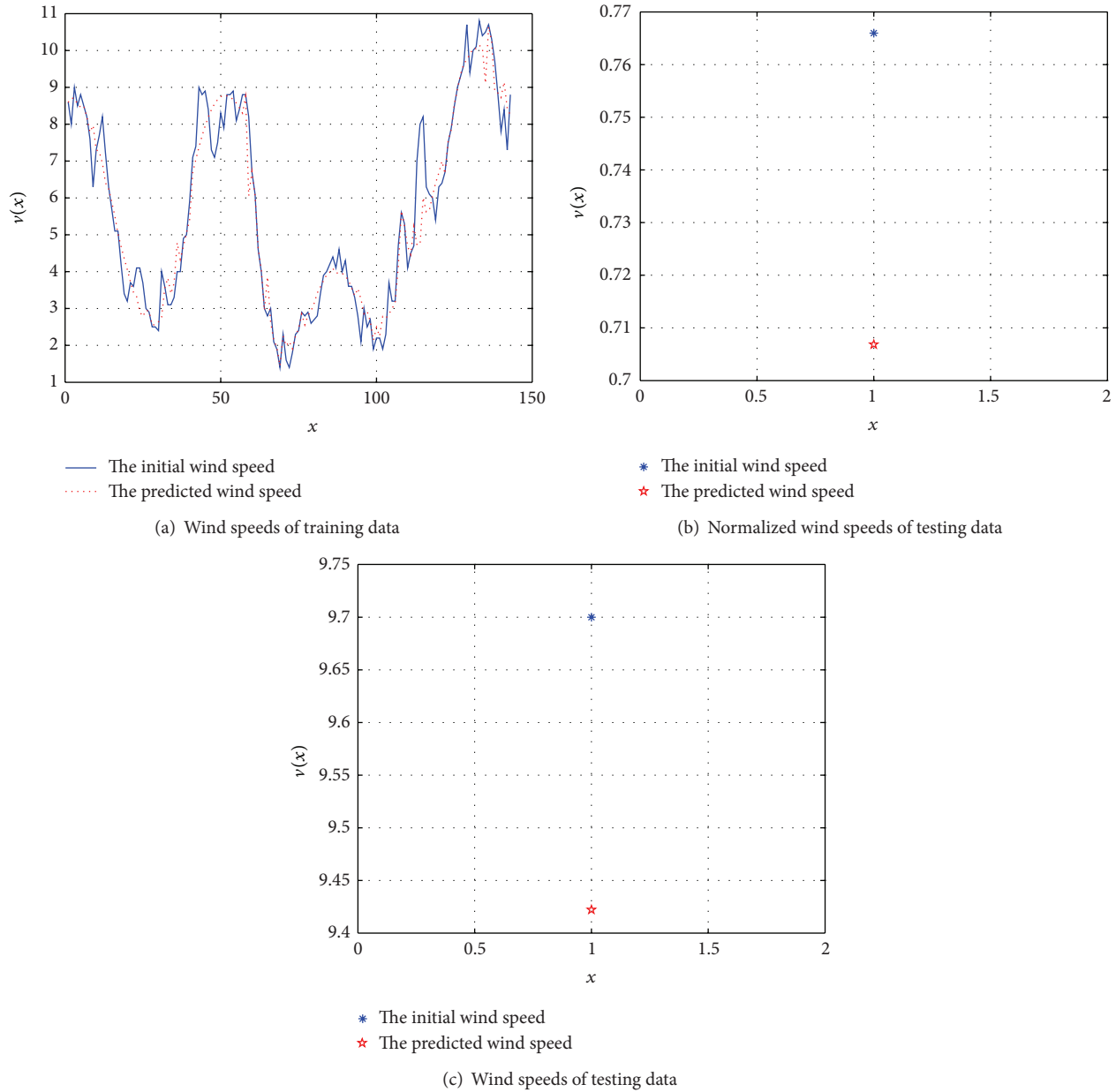


FIGURE 5: Initial wind speeds and predicted wind speeds with PPK- ϵ -SVR.

function is the radial basis function (RBF). In order to predict the wind speed v_{144} for the given x_{144} with training data $(x_1, v_1), (x_2, v_2), \dots, (x_{143}, v_{143})$ in data (30), we make experiment with PPK- ϵ -SVR and standard ϵ -SVR, respectively; the results are shown in Tables 1 and 2, respectively. The wind speeds of training data, normalized wind speeds of testing data, and wind speeds of testing data with PPK- ϵ -SVR are shown in Figures 5(a), 5(b), and 5(c), respectively. The wind speeds of training data, normalized wind speeds of testing data, and wind speeds of testing data with standard ϵ -SVR are shown in Figures 6(a), 6(b), and 6(c), respectively.

Similar to the steps of predicting wind speed v_{144} for the given x_{144} , we make experiment 50 times to predict the wind speed v_j for the given x_j ($j = 1, 2, \dots, 50$) (namely,

the former 50 wind speeds monitored on 2 April 2008), the average mean squared errors are shown in Table 3, and the numbers after \pm are the standard deviations.

5.4. Result Analysis. From Tables 1 and 2, we can see that the mean squared errors of training data and testing data with PPK- ϵ -SVR are all smaller than the corresponding ones with standard ϵ -SVR. Comparing Figure 5(a) with Figure 6(a) and Figure 5(c) with Figure 6(c), we can find that the predicted wind speeds of training data and testing data with PPK- ϵ -SVR are more close to the initial wind speeds than those with standard ϵ -SVR. This illustrates that the prediction error of PPK- ϵ -SVR is smaller than that of standard ϵ -SVR in predicting the wind speed v_{144} for the given x_{144} .

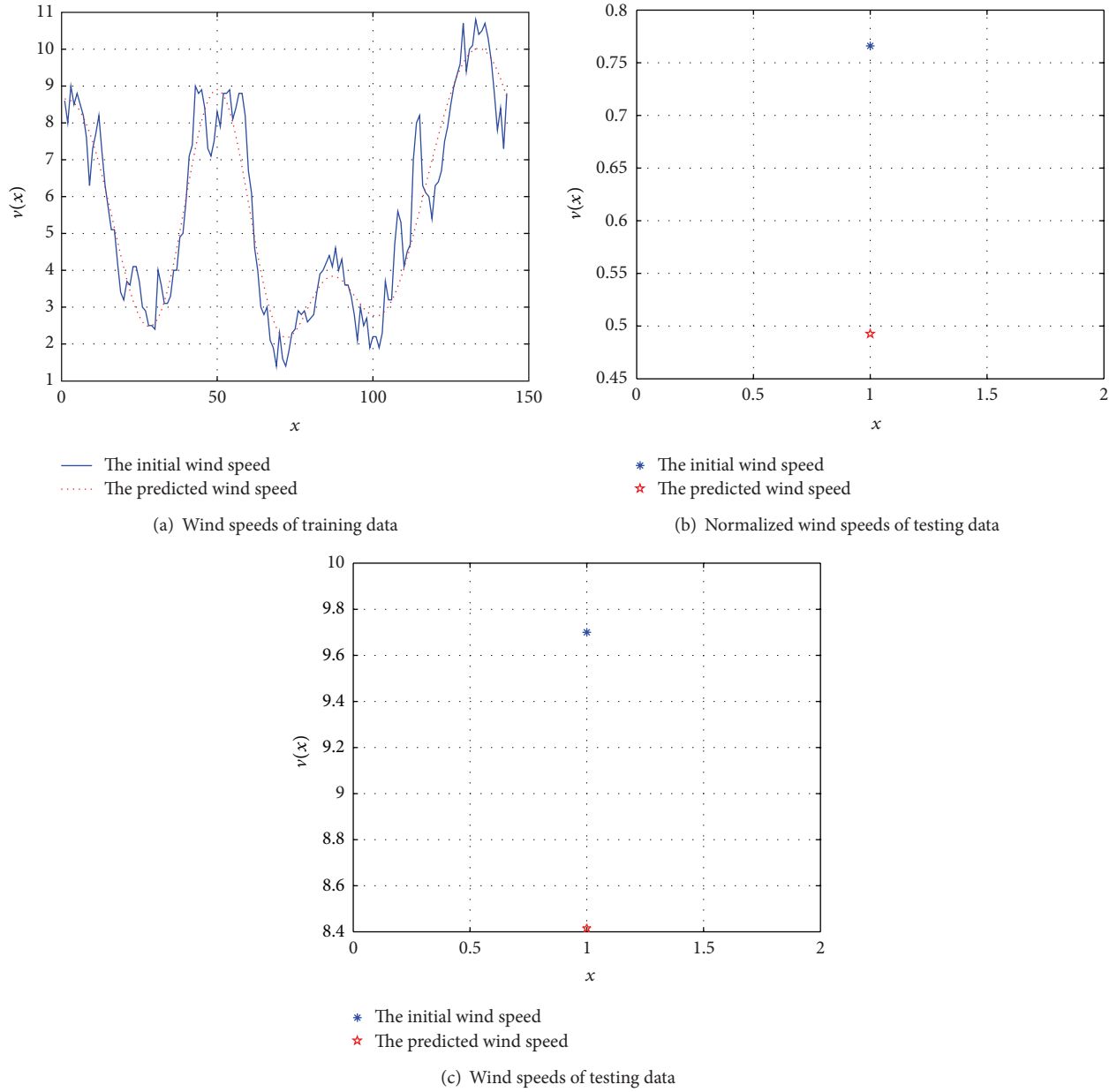


FIGURE 6: Initial wind speeds and predicted wind speeds with standard ϵ -SVR.

TABLE 1: Experiment results with PPK- ϵ -SVR.

Data	Number in the data set	ϵ	σ	C	Mean squared errors
Training data	143	0.01	0.25	8.00	0.0217
Testing data	1				0.0035

TABLE 2: Experiment results with standard ϵ -SVR.

Data	Number in the data set	ϵ	σ	C	Mean squared errors
Training data	143	0.01	0.25	4.00	0.0307
Testing data	1				0.0747

TABLE 3: Average mean squared errors.

Methods	Experiment times	Average mean squared errors
PPK- ε -SVR	50	0.0304 \pm 0.0367
Standard ε -SVR	50	0.0402 \pm 0.0408

From Figure 5(b) we can obtain that the difference between normalized initial wind speed and normalized predicted wind speed is $0.766 - 0.707 = 0.059$ which is equal to the square root of mean squared error 0.0035 in Table 1, which shows the effectiveness of the PPK- ε -SVR.

From Table 3, we can see that the average mean squared error (namely, the average prediction error) of PPK- ε -SVR is smaller than that of standard ε -SVR, which illustrates that PPK- ε -SVR method is more accurate than standard ε -SVR. What is more, the standard deviation of PPK- ε -SVR is also smaller than that of standard ε -SVR, which illustrates that the PPK- ε -SVR is more stable than the standard ε -SVR. And also the running time of PPK- ε -SVR is less than one minute, which shows that the model's running time can meet the needs of wind speed prediction in application.

6. Conclusions

In this paper, a method of estimating the probability density of historical wind speed data is proposed, and the estimated probability density is used to describe the wind speed fluctuation between the maximal value and the minimal value in a certain period of time. Then the prior knowledge provided by historical wind speed data is incorporated into the training data and the testing data. Then, based on standard ε -SVR, a kind of support vector regression for wind speed prediction incorporating probability prior knowledge is proposed. The comparing experiments show that the proposed PPK- ε -SVR is feasible and effective and the model's running time can meet the needs of wind speed prediction in application. And, how to incorporate prior knowledge into selecting the kernel function to decrease the prediction error further is our future study.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the 973 Program (no. 2012CB215201), the National Natural Science Foundation of China (no. 61073121), the Natural Science Foundation of Hebei Province of China (nos. F2012402037, A2012201033), and the Natural Science Foundation of Hebei Education Department (no. Q2012046).

References

- [1] T. H. M. El-Fouly, E. F. El-Saadany, and M. M. A. Salama, "One day ahead prediction of wind speed and direction," *IEEE Transactions on Energy Conversion*, vol. 23, no. 1, pp. 191–201, 2008.
- [2] G. Zhang, H. X. Li, and M. Gan, "Design a wind speed prediction model using probabilistic fuzzy system," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 4, pp. 819–827, 2012.
- [3] F. Cassola and M. Burlando, "Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output," *Applied Energy*, vol. 99, pp. 154–166, 2012.
- [4] S. J. Mitchell, N. Lanquaye-Opoku, H. Modzelewski et al., "Comparison of wind speeds obtained using numerical weather prediction models and topographic exposure indices for predicting windthrow in mountainous terrain," *Forest Ecology and Management*, vol. 254, no. 2, pp. 193–204, 2008.
- [5] P. Louka, G. Galanis, N. Siebert et al., "Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 96, no. 12, pp. 2348–2362, 2008.
- [6] Q. Cao, B. T. Ewing, and M. A. Thompson, "Forecasting wind speed with recurrent neural networks," *European Journal of Operational Research*, vol. 221, no. 1, pp. 148–154, 2012.
- [7] M. J. Woods, C. J. Russell, and R. J. Davy, "Simulation of wind power at several locations using a measured time-series of wind speed," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 219–226, 2013.
- [8] V. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," in *Neural Information Processing Systems*, M. Mozer, M. Jordan, and T. Petsche, Eds., vol. 9, MIT Press, Cambridge, Mass, USA, 1997.
- [9] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in Neural Information Processing Systems*, M. Mozer, M. Jordan, and T. Petsche, Eds., vol. 9, pp. 155–161, MIT Press, Cambridge, MA, 1997.
- [10] P. Guan, D. Huang, and M. He, "Lung cancer gene expression database analysis incorporating prior knowledge with support vector machine-based classification method," *Journal of Experimental and Clinical Cancer Research*, vol. 28, no. 1, article 103, 2009.
- [11] Z. Zhang, G. Dai, and M. I. Jordan, "Bayesian generalized kernel mixed models," *Journal of Machine Learning Research*, vol. 12, pp. 111–139, 2011.
- [12] F. Q. Liu and X. P. Xue, "Constructing kernels by fuzzy rules for support vector regressions," *International Journal of Innovative Computing, Information and Control*, vol. 8, no. 7, pp. 4811–4822, 2012.
- [13] M. H. Ha, C. Wang, Z. M. Zhang, and D. Z. Tian, *Uncertainty Statistical Learning Theory*, Science Press, Beijing, China, 2010, (Chinese).

- [14] O. Chapelle, V. Sindhwani, and S. S. Keerthi, "Optimization techniques for semi-supervised support vector machines," *Journal of Machine Learning Research*, vol. 9, pp. 203–233, 2008.
- [15] N. Y. Deng and Y. J. Tian, *Support Vector Machine Theory, Algorithms and Generalization*, Science Press, Beijing, China, 2009, (Chinese).
- [16] Y. C. Tsang, Y. Q. Zhang, N. V. Chawla, and S. Krasser, "SVMs modeling for highly imbalanced classification," *IEEE Transactions on Systems, Man and Cybernetics B*, vol. 39, no. 1, pp. 281–288, 2009.
- [17] M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Transactions on Neural Networks*, vol. 21, no. 3, pp. 515–519, 2010.
- [18] M. Karasuyama and I. Takeuchi, "Multiple incremental decremental learning of support vector machines," *IEEE Transactions on Neural Networks*, vol. 21, no. 7, pp. 1048–1059, 2010.
- [19] A. Ji, J. Pang, and H. Qiu, "Support vector machine for classification based on fuzzy training data," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3495–3498, 2010.
- [20] A. Korth, E. Echer, Q.-G. Zong, F. L. Guarnieri, M. Fraenz, and C. G. Mouikis, "The response of the polar cusp to a high-speed solar wind stream studied by a multispacecraft wavelet analysis," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 73, no. 1, pp. 52–60, 2011.
- [21] O. L. Mangasarian and E. W. Wild, "Nonlinear knowledge in kernel approximation," *IEEE Transactions on Neural Networks*, vol. 18, no. 1, pp. 300–306, 2007.
- [22] M. El Mokadem, V. Courtecuisse, C. Saudemont, B. Robyns, and J. Deuse, "Fuzzy logic supervisor-based primary frequency control experiments of a variable-speed wind generator," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 407–417, 2009.
- [23] J. Zhou, J. Shi, and G. Li, "Fine tuning support vector machines for short-term wind speed forecasting," *Energy Conversion and Management*, vol. 52, no. 4, pp. 1990–1998, 2011.
- [24] E. G. Ortiz-García, S. Salcedo-Sanz, Á. M. Pérez-Bellido, J. Gascón-Moreno, J. A. Portilla-Figueras, and L. Prieto, "Short-term wind speed prediction in wind farms based on banks of support vector machines," *Wind Energy*, vol. 14, no. 2, pp. 193–207, 2011.
- [25] M. Fisz, *Probability Theory and Mathematical Statistics*, Wiley, New York, NY, USA, 3rd edition, 1963.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

