IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 4 May 2023, accepted 19 June 2023, date of publication 27 June 2023, date of current version 17 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3290020

RESEARCH ARTICLE

Models LSSVR and PLSSVR With Heteroscedastic Gaussian Noise Characteristics and Its Application for Short-Term Wind-Speed Forecasting

TING ZHOU [1], GE FENG [2], AND SHIGUANG ZHANG [1]
[1]School of Information Engineering, Shandong Management University, Jinan 250357, China
[2]College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China

Corresponding authors: Ting Zhou (zt20230622@163.com) and Ge Feng (ffxxnyy18@163.com)

This work was supported in part by the Natural Science Foundation of Shandong Province under Grant ZR2022MF242, and in part by the Key Research and Development Plan of Shandong Province under Grant 2019GGX101056.

ABSTRACT Proximal least squares support vector regression is a new regression machine designed by using regularization principle technology and least squares support vector regression. In this paper, we use the above models framework to build a new regression model, called the Proximal least squares support vector regression model with Heteroscedastic Gaussian noise (PLSSVR-HGN). Based on the Heteroscedastic noise characteristics in the application field, the least square method is introduced and the regularization terms are added respectively. PLSSVR-HGN is a regression model with equality constraints based on Heteroscedasticity, which not only improves training speed and generalization ability, but also effectively improves prediction accuracy. In order to solve the parameter selection of models LSSVR-HGN and PLSSVR-HGN, the Particle swarm optimization algorithm with fast convergence speed and good robustness is selected to optimize its parameters. In order to verify the forecasting performance of LSSVR-HGN and PLSSVR-HGN, it is compared with the classical regression models on the UCI data-set and wind-speed data-set. Experimental results indicate that the proposed models not only inherit most of the merits of the original LSSVR, but also has more stable and reliable generalization performance and more accurate prediction results. These applications demonstrate the correctness and effectiveness of the proposed models.

INDEX TERMS Heteroscedastic Gaussian noise, proximal least squares support vector, regression model, short-term wind-speed forecasting.

I. INTRODUCTION

In recent decades, the rapid consumption of coal and other nonrenewable fossil fuels makes us urgently need to develop new energy. Wind energy is a renewable clean energy has been among the most rapidly growing global energy resources in recent years [1]. However, due to the intermittence and instability of wind energy, the power transmitted by wind turbines after grid connection is unstable, which has a great impact on our power system. How to effectively predict wind-speed has become a big problem in the real society [2].

Least square support vector regression (LSSVR) is a method of linear regression (LR) that implements a sum-of-squares error function together with regularization thus controlling the bias variance trade-off [3], [4]. Its purpose is to discover the linear structures hidden in raw data [5], [6]. At the same time, nonlinear mappings can be estimated by kernel LSSVR, which is an extended LR with kernel-techniques. In recent years, LSSVR as a data-rich nonlinear forecasting tool has been increasingly welcomed [7], which is applicable in many different contexts, such as machine learning [8], [9], especially wind speed/power forecasting [10], [11].

LSSVR uses equality constraints instead of inequality constraints to solve linear equations instead of classical quadratic programming problems, thus reducing the computational complexity. However, LSSVR lacks sparsity, and all samples in the data set contribute to the forecasting of new samples,

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenliang Zhang.

and the size of the contribution is determined by the size of the corresponding Lagrange multiplier [12]. Therefore, on this basis, Suykens et al [13] proposed the weighted least squares support vector machines (WLSSVM), adding weights before each error variable, assigning different weights according to the error variables, so that the less important samples or outliers have less impact on the training.

Fung and Mangasarian [14] proposed a new learning method called Proximal support vector machine (PSVM). PSVM can separate the two parallel hyperplanes as far as possible, maximize the plane spacing, and change the inequality constraint into equality constraint, which can not only improve the training speed, but also the training accuracy is higher than or equal to the traditional support vector machine. Wang et al [15] proposed a robust Proximal support vector regression based on maximum correntropy criterion (PSVR-MCC), trying to suppress the negative effects of outliers and enhance the robustness of PSVR. They improved SVR by moving the two parallel hyperplanes out for a certain distance, so that the sample points are concentrated near the two parallel hyperplanes, which can separate the two parallel hyperplanes as far as possible, maximize the plane spacing, and change the inequality constraint into equality constraint, which can not only improve the training speed, but also the training accuracy is higher than or equal to the traditional support vector machine. However, PSVR is sensitive to noise and outliers, which leads to poor generalization performance and robustness [16], [17].

In order to apply model LSSVR to better predict the wind-speed, we analyze that the wind-speed satisfies the Gaussian distribution of 0 mean and heteroscedasticity, whose variance varies with the average wind speed. That is, it becomes a task of Heteroscedasticity. Literatures [18], [19] studied the problem of Heteroscedasticity (which means that the errors at different sample points are not completely equal). Several common cases of Heteroscedasticity are: (1) the dependent variable has measurement error, and the size of the error changes with the value of the dependent variable or independent variable in the model; (2) The unit of analysis is aggregation units with different scales, and the value of dependent variable is determined by the individual value of these aggregation units; (3) The social phenomenon reflected by the dependent variable itself contains a certain trend of difference; (4) It originates from the interaction effect between the independent variables contained in the model and a missing independent variable. It is pointed out that in the case of Heteroscedasticity, in order to obtain the best linear unbiased estimation, the generalized least squares method and the weighted least squares method can be used for parameter estimation and other models.

The generalization performance of LSSVR is high, and the training effect of PSVR is efficient. Therefore, we can use the principle of hyperplane of LSSVR and the advantages of PSVR, on the basis of both, we propose least squares support vector regression model (LSSVR-HGN) and Proximal least squares support vector regression model with Heteroscedast-ic Gaussian noise (PLSSVR-HGN). Models LSSVR-HGN and PLSSVR-HGN transform inequality constraints into simpler equality constraints, which not only improves the training speed, but also effectively enhances the training efficiency and accuracy. We apply these proposed models to the short-term wind-speed forecasting. The following are the contributions of this article:

(1) Discover that the wind operation law meets a Gaussian distribution with zero mean Heteroscedasticity by investigating the properties of noise models in real wind-speed forecasting; Derive optimal empirical risk loss function of Heteroscedastic Gaussian noise characteristic by employing the Bayesian principle and maximizing posterior probability method; (2) Establish regression models LSSVR-HGN and PLSSVR-HGN by combining the framework structure of PLSSVR and the Gaussian noise characteristic of Heteroscedasticity; (3) Experimental results show that PLSSVR-HGN not only maintains the advantages of LSSVR in simple parameter setting and capability of rapid convergence, but also makes up for the disadvantages of being sensitive to noise and outliers and poor generalization performance, thus it can be easily extended to large data treatment.

The rest of this paper is organized as follows: In the second section, derive optimal empirical risk loss function of Heteroscedastic Gaussian noise characteristics, briefly introduce models LSSVR and PLSSVR. In Section III, we focus on exploring the noise model properties in wind-speed forecasting, and present models LSSVR-HGN and PLSSVR-HGN in detail. In Section IV, gives Algorithm design of LSSVR-HGN and PLSSVR-HGN. To verify the correctness of the established model, a significant number of Experiments on different data-sets including the UCI data-set and real wind-speed data-are conducted in Section V. The final section is the Conclusion of the article.

II. MATERIALS AND METHODS

The classical least squares support vector regression machine based on Gaussian noise characteristics (LSSVR) [12], [13] and the support vector regression machine based on Gaussian noise characteristics (SVR) [20] assume that the noise characteristics follow a Gaussian distribution with a mean of 0 and the same variance σ^2. In contrast, according to the continuous method [19], statistical analysis of the acquired wind-speed data-set reveals that the variance varies with the mean wind-speed, the prediction error does not obey the Gaussian distribution with 0 mean and homoscedastic variance σ^2, but obeys the Gaussian distribution with 0 mean and heteroscedastic variance $\sigma_i^2 (i = 1, 2, \cdots, N)$. At this time, the application of models LSSVR and SVR to predict wind-speed cannot achieve the expected effect.

A. OPTIMAL EMPIRICAL RISK LOSS FUNCTION OF HETEROSCEDASTIC GAUSSIAN NOISE CHARACTERISTICS

In this section, the optimal empirical risk loss function of the Heteroscedastic Gaussian noise signature will be derived by utilizing the Bayesian principle and maximizing a posteriori probability. Set the given data-set:

$$D_N = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\} \tag{1}$$

Suppose the N training data-sets with Heteroskedastic Gaussian noise property distribution are$\{(x_i, y_i)\}_{i=1}^{N}$ $(i = 1, 2, \cdots, N)$ and the association between the measured value y_i. Where $x_i = (x_{i1}, x_{i2}, \cdots, x_{iL}) \in R^L$ and $y_i \in R$, R represents the set of real numbers, R^L represents L dimensional Euclidean space, N indicates the number of samples. The predicted valuation $f(x_i)$ are as follows:

$$y_i = f(x_i) + \xi_i (i = 1, 2, \cdots, N) \tag{2}$$

In general, the noise density $P(\xi_i) = P(y_i - f(x_i))$ $(i = 1, 2, \cdots, N)$ is unknown in engineering technology applications. Assume that $\xi_i (i = 1, 2, \cdots, N)$ is a random noise variable known to be independently and identically distributed (i.i.d.) with zero mean and standard deviation $\sigma_i (i = 1, 2, \cdots, N)$. It is necessary to obtain unknown regression function $f(x)$ from training samples $D_f \subseteq D_N$. Using Bayes' principle, the optimal empirical risk loss function in the maximum likelihood sense [18], [21], [22] is stated as follows:

$$l(\xi_i) = -\log P(\xi_i) (i = 1, 2, \cdots, N) \tag{3}$$

where $P(\xi_i)$ is the probability density function (PDF) of the error variable ξ_i, $l(\xi_i)$ signifies the loss value resulting from comparisons between the predicted value $f(x_i)$ and y_i received at the sample point (x_i, y_i) for prediction, and $l(\xi_i)(i = 1, 2, \cdots, N)$ denotes the loss function.

Assuming that the noise in equation (2) is Gaussian noise with zero mean and Heteroskedastic variance $\sigma_i^2 (i = 1, 2, \cdots, N)$, the PDF of ξ_i is $P(\xi_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\{-\frac{\xi_i^2}{2\sigma_i^2}\}$. According to equation (3), the corresponding optimal loss function with Heteroskedastic Gaussian noise can be expressed as:

$$l(\xi_i) = \frac{1}{2\sigma_i^2} \xi_i^2 (i = 1, 2, \cdots, N) \tag{4}$$

If the noise in equation (2) is Gaussian noise, its mean value is zero and its covariance is σ^2, then the empirical risk loss function will be represented as $l(\xi_i) = \frac{1}{2}\xi_i^2 (i = 1, \cdots, N)$. That is, it is a square loss function.

B. REGRESSION MODEL LSSVR

Due to computational complexity, the support vector regression (SVR) model cannot predict the accumulation rate, and may generate a large amount of computational consumption. To overcome this problem, in the case of classic SVR models, the least squares support vector regression machine based on Gaussian noise characteristics (LSSVR) model was proposed, which can solve linear and nonlinear equations rather than quadratic equations. As an important extension model of SVR, LSSVR has strong fitting ability and strong generalization ability by using square error as the objective function, which greatly reduces the computational burden and improves computational efficiency. The least square method has become the most widely used method of data processing in many fields. The multiple nonlinear regression model is $f(x_i) = \omega^T \phi(x_i) + b + \varepsilon$, the parameter vector $\omega \in R^N$ is determined by the least squares regression model. The original problem of model LSSVR is:

$$\begin{aligned} g_{LSSVR} &= \frac{1}{2}\omega^T \cdot \omega + C \sum_{i=1}^{N} (y_i - \omega^T \cdot \phi(x_i) - b)^2 \\ s.t.\ y_i &= \omega^T \cdot \phi(x_i) + b + \xi_i (i = 1, 2, \cdots, N) \end{aligned} \tag{5}$$

where $\phi(\cdot)$ is the kernel function used to convert the input space into a high-dimensional space. When the noise loss follows the Gaussian distribution, the prediction results can meet the actual requirements using the least squares support vector regression machine.

The decision function of the linear regression model is $f(x) = \omega^T \cdot x + b$, where $x_i \in R^L$, the parameter vector $\omega = (\omega_1, \cdots, \omega_L) \in R^L$ and parameter $b \in R$ determine the structure of the regression model. The kernel function $K(\cdot, \cdot)$ is constructed by kernel technique, and the linear regression model is extended to the kernel regression model LSSVR. The nonlinear decision function $f(x_i) = \omega^T \cdot \phi(x_i) + b$ of model LSSVR. Where $K(x_i, x_j) = (\phi(x_i) \cdot \phi(x_j))$, $\phi : R^L \rightarrow H$, H is a Hilbert space and $(\phi(x_i) \cdot \phi(x_j))$ is the inner product of the space H, superscript T indicates transpose of vector. Common kernel functions are [3], [4], and [22]:

(1) Polynomial kernel function:
$K(x_i, x_j) = (\gamma(x_i \cdot x_j) + c)^d$, d is a positive integer;

(2) Gauss radial basis kernel function:

$$K(x_i, x_j) = \exp(-\left\|x_i - x_j\right\|^2) \Big/ \sigma^2;$$

(3) Sigmoid kernel function:
$K(x_i, x_j) = \tanh(\gamma(x_i \cdot x_j) + c)$,

tanh represents hyperbolic tangent function.

C. REGRESSION MODEL PLSSVR

The support vector classification model is based on two parallel hyperplanes with maximum interval between two classes of samples to determine the optimal solution. For more complex problems, the interval between two parallel hyperplanes is too small, which makes it difficult to obtain the optimal solution. Fung and Mangasarian [14] proposed Proximal support vector machine (PSVM). Two parallel hyperplanes can be moved out for a certain distance, so that the sample points are concentrated near the two parallel hyperplanes, and then the two parallel hyperplanes can be separated as far as possible, $(\omega^T \cdot x) + b = \pm 1$ is no longer the boundary plane, but becomes the "Proximal" plane (As shown in the Figure 1). For direction ω and relative position b, the distance between the boundary planes is maximized, and the distance between the two hyperplanes is changed to $2 \Big/ \left\| \begin{bmatrix} \omega \\ b \end{bmatrix} \right\|$.

And analytical solutions can be obtained, greatly improving the training speed.

In Literature [23], PSVM is extended to Proximal least squares support vector regression (PLSSVR). Model PLSSVR has added a bias term b to the optimization goal, thus transforming the corresponding optimization problem into a strict convex quadratic programming. And the analytical solution can be obtained, greatly improving the training speed. The original problem of model PLSSVR is as follows:

$$\min_{\omega,b}\{P_{PLSSVR} = \frac{1}{2}\|\omega\|^2 + \frac{1}{2}b^2 + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2\}$$
$$s.t. \quad y_i - \omega^T \cdot \phi(\mathrm{x}) - b = \xi_i (i = 1, 2, \cdots, N) \tag{6}$$

where ξ_i is the relaxation variable, C is the parameter, and $\phi(x_i)$ is the mapping from low dimensional space x_i to high dimensional space.

Fig.1 shows the geometric interpretation of PSVM, and obtains the partition plane by maximizing the interval. The plane $\omega^T x + b = \pm 1$ is no longer the boundary of the distribution of two kinds of points, but becomes the clustering center of the two kinds of points. The two planes are separated as far as possible because of $\|\omega\|^2 + b^2$ in the optimization problem (6), so that the training set can be separated better.

Lagrange function is constructed

$$L(\omega, b, \xi_i, \alpha_i) = \frac{1}{2}\|\omega\|^2 + \frac{1}{2}b^2 + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2 + \sum_{i=1}^{N}\alpha_i(y_i - \omega^T \cdot \phi(x) - b - \xi_i) \tag{7}$$

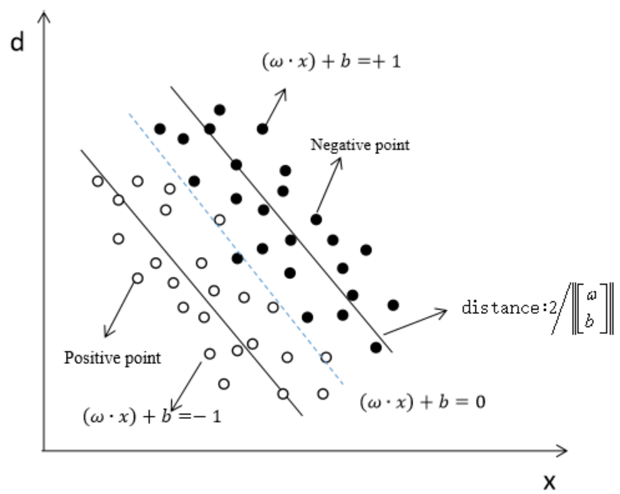


FIGURE 1. **Proximal support vector classifier machine.**

where $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_N)^T$ is the Lagrange multiplier vector. By using the optimization principle, the dual problem of the model PLSSVR can be obtained as follows:

$$\max_{\alpha}\{D_{PLSSVR} = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j K(x_i, x) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j - \frac{1}{2C}\sum_{i=1}^{N}\alpha_i^2 + \sum_{i=1}^{N}\alpha_i y_i\} \tag{8}$$

The vectors ω and $\xi_i (i = 1, \cdots, N)$ are eliminated. For the sake of brevity, the following linear equations are obtained:

$$(K + \frac{1}{C}I_N + \mathrm{ee}^T)\alpha = y$$
$$b = e^T\alpha$$

where I_N is the identity matrix $N \times N$ of $e = (1, 1, \cdots, 1)^T$ and $K = (K_{ij})_{N\times N}$ is the kernel function. The decision function of PLSSVR about b and $\alpha_i (i = 1, 2, \cdots, N)$ can be expressed as follows:

$$f(x) = \omega^T \cdot \phi(x) + b = \sum_{i=1}^{N}\alpha_i K(x_i, x) + b.$$

III. MODELS LSSVR AND PLSSVR WITH HETEROSCED-ASTIC GAUSSIAN NOISE CHARACTERIS-TICS

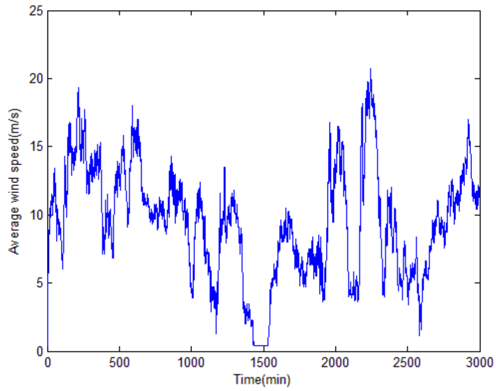
A. UNCERTAINTY OF WIND

To investigate the behavior of the noise model in the actual wind-speed forecasting, the wind-speed data from Heilongjiang province is collected, which has a sampling interval of 5 seconds. After statistical analysis and processing, the average wind-speed and variance of every 10 minutes were finally obtained. It was observed that the current forecasted wind-speed is a wind-speed in an average sense, while the real wind-speed consists of two parts, namely, hourly average wind-speed and instantaneous random fluctuations. Assuming that the time sequence of the actual instantaneous wind-speed data of the wind farm is $\{v(t)\}$, and the time sequence of the wind-speed on the hourly scale is $\{\bar{v}(t)\}$. Then according to the composition of the real-time instantaneous wind-speed, the instantaneous random fluctuation part of the wind-speed, the turbulence residual can be expressed as follows: $e(t) = \{v(t)\} - \{\bar{v}(t)\}$, and the variance of the random fluctuation of the wind-speed can be expressed [19] as follows:

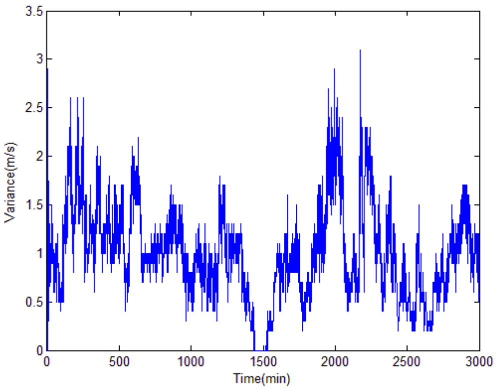
$$Var(t) = \frac{1}{N-1}\sum_{t=1}^{N}e^2(t) = \frac{1}{N-1}\sum_{t=1}^{N}[v(t) - \bar{v}(t)]^2 \tag{9}$$

According to Equation (9), when calculating the variance of wind-speed $Var(t)$, the variance of wind-speed is practically the equal in the time of $t = 1, 2, \cdots, N$ by default. However, it is investigated that the variance of wind-speed varies at different moments. As shown in the Figure 2:

By observing the two images, it can be seen that both the average wind-speed and the wind-speed variance varies with time, and the trends of both are somewhat similar. Therefore, it is reasonable to assume that there is a connection between



(a)



(b)

FIGURE 2. (a) represents the variation curve of the average wind-speed every 10 minutes; (b) Indicates the variation graph of the wind-speed variance.

wind-speed variance and wind-speed. To further investigate the association between the two, the following experiments were conducted, setting the average wind-speed as the x-axis and the wind-speed variance as the y-axis. Figure 3 illustrates the outcomes of the experiment.

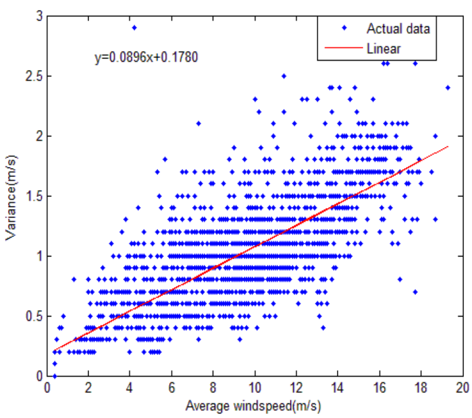


FIGURE 3. Modulation effect of wind amplitude on its variance.

From Figure 3, it can be concluded that there is a linear correlation between the two and that the wind-speed variance y varies with the average wind-speed x. The relationship expression is y = 0.0896*x + 0.1780, which implies that the variance of the wind-speed is distinct at various times and varies with the average wind-speed, which is a Heteroskedastic task.

B. LEAST SQUARES SUPPORT VECTOR REGRESSION MODEL OF HETEROSCEDASTIC GAUSSIAN NOISE CHARACTERISTICS

Generally, classical regression models (Least squares support vector regression, support vector regression, etc) assume that the noise distribution in the data-set follows the homoscedastic Gaussian distribution. Relevant research shows that in the practical application field, some noise distributions do not obey the homoscedastic Gaussian distribution, but obey the Heteroscedastic Gaussian distribution. Especially in wind-speed/wind power prediction, the noise distribution changes with seasons and regions, and can be represented by Heteroscedastic Gaussian distribution, so as to fit the unknown noise characteristics in uncertain data. At this time, the classical regression model is not the optimal prediction model in line with the actual situation. Aiming at the above problems, according to the actual distribution of noise characteristics of wind-speed data, this paper uses Heteroscedastic Gaussian noise to fit the unknown noise characteristics, and proposes a least squares support vector regression model with Heteroscedastic Gaussian noise characteristics (Abbreviated as LSSVR-HGN). The original problem is:

$$\begin{aligned} &\min_{\omega,b,\xi_i} \{g_{P_{LSSVR-HGN}} = \frac{1}{2}\omega^T \cdot \omega + \frac{C}{N}(\sum_{i=1}^{N} \frac{1}{2\sigma_i^2}\xi_i^2)\} \\ &s.t. \quad \xi_i = y_i - \omega^T \cdot \phi(x_i) - b \\ &\qquad\quad \xi_i \geq 0, i = 1, 2, \cdots, N \end{aligned} \tag{10}$$

where $C > 0$ is the penalty factor, ξ_i is the relaxation variable, and $\sigma_i(i = 1, 2, \cdots, N)$ is the Heteroscedasticity.

Proposition 1 The solution about ω of the original problem (10) of model LSSVR-HGN exists and be unique.

Proof: This proposition can be proved by referring to Theorem 1 [21].

Theorem 1 The dual problem of the original problem (10) of model LSSVR-HGN is:

$$\begin{aligned} \max_{\alpha\alpha^*}\{g_{D_{LSSVR-HGN}} &= -\frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i\alpha_j K(x_i, x_j)) \\ &+ \sum_{i=1}^{N}(\alpha_i y_i - \frac{N\sigma_i^2}{C}\alpha_i^2)\} \\ s.t. \quad &\sum_{i=1}^{N}\alpha_i = 0 \end{aligned} \tag{11}$$

where $C > 0$ is the penalty factor, ξ_i is the relaxation variable, and $\sigma_i(i = 1, 2, \cdots, N)$ is the Heteroscedasticity.

Proof: Introducing the Lagrange functional $L(\omega, b, \alpha, \xi)$, get

$$L(\omega, b, \alpha, \xi) = \frac{1}{2}\omega^T \cdot \omega + \frac{C}{N} \cdot (\sum_{i=1}^{N} \frac{1}{2\sigma_i^2}\xi_i^2) + \sum_{i=1}^{N} \alpha_i(y_i - \omega^T \cdot \phi(x_i) - b - \xi_i)$$

In order to find the minimum value of $L(\omega, b, \alpha, \xi)$, the partial derivatives of ω, b, ξ are calculated respectively. By KKT conditions:

$$\nabla_\omega L = 0, \nabla_b L = 0, \nabla_\xi L = 0$$

Get

$$\omega = \sum_{i=1}^{N} \alpha_i \cdot \phi(x_i), \sum_{i=1}^{N} \alpha_i = 0, \frac{C}{N\sigma_i^2} \cdot \xi_i - \alpha_i = 0.$$

Substitute the above extreme conditions into $L(\omega, b, \alpha, \xi)$, and seek the maximum of α, we can obtain the dual problem (11) of the original problem (10) of the model LSSVR-HGN.

And there is

$$b = \frac{1}{N}\sum_{i=1}^{N}(y_i - \sum_{j=1}^{N}\alpha_j K(x_i, x_j) - \frac{N\sigma_i^2}{C}\alpha_i),$$

the decision function of model LSSVR-HGN is

$$f_{LSSVR-HGN}(x) = \omega^T \cdot \phi(x) + b = \sum_{i=1}^{N}\alpha_i K(x_i, x) + b \quad (12)$$

where $\omega \in R^L$ is the parameter vector, $\phi : R^L \to H$ (H is the Hilbert space) is the kernel transformation, and $(\phi(x_i) \cdot \phi(x_j))$ represents the inner product in the space H, $K(x_i, x_j) = (\phi(x_i) \cdot \phi(x_j))$ is the kernel function.

C. PROXIMAL LEAST SQUARES SUPPORT VECTOR REGRESSION MODEL OF HETEROSCEDASTIC GAUSSIAN NOISE CHARACTERISTICS

In the practical application field, the noise distribution is unknown, or most of the noise distribution does not obey the homoscedastic Gaussian noise distribution. In this section, the Heteroscedastic Gaussian noise distribution is used to fit the unknown noise characteristics. The original problem of Proximal least squares support vector regression model with Heteroscedastic Gaussian noise characteristics (Abbreviated as PLSSVR-HGN) is:

$$\begin{aligned} \min_{\omega,b,\xi_i} \{g_{P_{LSSVR-HGN}} &= \frac{1}{2}\omega^T \cdot \omega + \frac{b^2}{2} + \frac{C}{N}(\sum_{i=1}^{N}\frac{1}{2\sigma_i^2}\xi_i^2)\} \\ s.t. \quad & \xi_i = y_i - \omega^T \cdot \phi(x_i) - b \\ & \xi_i \geq 0, i = 1, 2, \cdots, N \end{aligned} \quad (13)$$

where $C > 0$ is the penalty factor, ξ_i is the relaxation variable, and $\sigma_i (i = 1, 2, \cdots, N)$ is the Heteroscedasticity.

Proposition 1: The solution about ω of the original problem (13) of the model PLSSVR-HGN exists and be unique.

Proof: This proposition can be proved by referring to Proposition 1.

Theorem 1: The dual problem of the original problem (13) of the model PLSSVR-HGN is:

$$\max_{\alpha\alpha^*}\{g_{D_{PLSSVR-HGN}} = -\frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j[K(x_i, x_j) - 1] + \sum_{i=1}^{N}(\alpha_i y_i - \frac{N\sigma_i^2}{C}\alpha_i^2)\} \quad (14)$$

where $C > 0$ is the penalty factor, ξ_i is the relaxation variable, and $\sigma_i (i = 1, 2, \cdots, N)$ is the heteroscedasticity.

Proof: Introducing the Lagrange functional $L(\omega, b, \alpha, \xi)$, we can get:

$$L(\omega, b, \alpha, \xi) = \frac{1}{2}\omega^T \cdot \omega + \frac{b^2}{2} + \frac{C}{N}(\sum_{i=1}^{N}\frac{1}{2\sigma_i^2}\xi_i^2) + \sum_{i=1}^{N}\alpha_i(y_i - \omega^T \cdot \phi(x_i) - b - \xi_i)$$

In order to find the minimum value of $L(\omega, b, \alpha, \xi)$, the partial derivatives of ω, b, ξ are calculated respectively. By KKT conditions:

$$\nabla_\omega L = 0, \nabla_b L = 0, \nabla_\xi L = 0.$$

Get:

$$\omega = \sum_{i=1}^{N}\alpha_i\phi(x_i), \sum_{i=1}^{N}\alpha_i = b, \frac{C}{N\sigma_i^2}\xi_i - \alpha_i = 0.$$

Substitute the above extreme conditions into $L(\omega, b, \alpha, \xi)$, and seek the maximum of α, we can obtain the dual problem (14) of the original problem (13) of the model PLSSVR-HGN.

And there is $b = \sum_{i=1}^{N}\alpha_i$, the decision function of model PLSSVR-HGN is:

$$f_{PLSSVR-HGN}(x) = \omega^T \cdot \phi(x) + b = \sum_{i=1}^{N}\alpha_i[K(x_i, x) + 1] \quad (15)$$

where $\omega \in R^L$ is the parameter vector. $\phi : R^L \to H$ (H is the Hilbert space) is the kernel transformation, and $(\phi(x_i) \cdot \phi(x_j))$ represents the inner product in the space H, $K(x_i, x_j) = (\phi(x_i) \cdot \phi(x_j))$ is the kernel function.

IV. ALGORITHM DESIGN OF MODELS LSSVR-HGN AND PLSSVR-HGN

In this section, parameter C and kernel function $K(\cdot, \cdot)$ are not fixed, which will affect the final optimization results. Therefore, the Particle swarm optimization (PSO) and Grid search methods [15], [24], [25] are introduced. The PSO algorithm can find the parameter C and the kernel function $K(\cdot, \cdot)$ more conveniently and quickly. However, the optimal parameters and the optimal calculation accuracy found by PSO algorithm are not fixed each time, and will fall into the local optimal solution. Therefore, in order to find the optimal calculation accuracy more accurately, after determining the approximate range of parameters, the Grid search method is used to find the optimal solution. The standard PSO algorithm update iteration formula is as follows:

$$\begin{aligned}\mu_{ij}(t+1) &= \mu_{ij}(t) + \eta_{ij}(t+1)\\ \eta_{ij}(t+1) &= g\cdot\mu_{ij}(t) + \alpha\cdot\sigma(t)(\lambda_{ij}(t)-\mu_{ij}(t))\\ &\quad + \alpha^*\cdot\sigma^*(t)(\lambda^*_{ij}(t)-\mu_{ij}(t)) \qquad (16)\end{aligned}$$

where $\mu(t)$ is the position vector, $\eta(t)$ is velocity vector, g is the particle inertia,$\mu(t)$ is the positive acceleration coefficient, $\sigma^*(t)$ is the random value, $\lambda_{ij}(t)$ is the best position found by the particle in dimension i, $\lambda^*_{ij}(t)$ is the best position found by any particle in dimension j.

The specific algorithm designs of models LSSVR-HGN and PLSSVR-HGN are as follows:

A. ALGORITHM DESIGN OF THE REGRESSION MODEL LSSVR-HGN

The algorithm design of least squares support vector regression model with Heteroscedastic Gaussian noise characteristics (LSSVR-HGN) is:

Step 1. Let a given data-set $D_l = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$ with noise characteristics, where $x_i \in R^L$, $y_i \in R$ $(i = 1, 2, \cdots, N)$, R^L represents the L dimensional Euclidean space, R represents the real number set, and N represents the number of samples;

Step 2. By using the Particle swarm optimization (PSO) algorithm to determine the optimal parameter C and select the appropriate kernel function $K(\cdot,\cdot)$; Set t in PSO algorithm to 1;

Step 3. Construct and solve the optimization problem (11) of model LSSVR-HGN, the optimal solution $\alpha = (\alpha_1, \alpha_2, \cdots \alpha_N)$ at different time t are obtained and updated continuo- usly, where $\alpha_1, \alpha_2, \cdots, \alpha_N$ are Lagrange multipliers;

Step 4. If the time $t \le m$, skip to Step 2 and Step 3; If so $t > m$, go to Step 5;

Step 5. Construct the decision function

$$f_{LSSVR-HGN}(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x) + b,$$

where

$$b = \frac{1}{N}\sum_{i=1}^{N}(y_i - \sum_{j=1}^{N}\alpha_j K(x_i, x_j) - \frac{N\sigma_i^2}{C}\alpha_i).$$

B. ALGORITHM DESIGN OF THE REGRESSION MODEL PLSSVR-HGN

The algorithm design of Proximal least squares support vector regression model with Heteroscedastic Gaussian noise characteristics (PLSSVR-HGN) is:

Step 1. Let a given data set $D_l = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$ with noise characteristics, where $x_i \in R^L$, $y_i \in R$ $(i = 1, 2, \cdots, N)$, R^L represents the L dimensional Euclidean space, R represents the real number set, and N represents the number of samples;

Step 2. By using the PSO algorithm to determine the optimal parameter C and select the appropriate kernel function $K(\cdot,\cdot)$; Set t in PSO algorithm to 1;

Step 3. Construct and solve the optimization problem (14) of model PLSSVR-HGN, the optimal solution $\alpha = (\alpha_1, \alpha_2, \cdots \alpha_N)$ at different time t are obtained and updated continu- ously, where$\alpha_1, \alpha_2, \cdots, \alpha_N$ are Lagrange multipliers;

Step 4. If the time $t \le m$, skip to Step 2 and Step 3; If so $t > m$, go to Step 5;

Step 5. Construct the decision function

$$f_{PLSSVR-HGN}(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x) + b$$

where $b = \sum_{i=1}^{N} \alpha_i$.

V. EXPERIMENTS AND DISCUSSION

In this section, we mainly conducted three experiments. The first experiment mainly used UCI data-set to test and compare the prediction results of models SVR, LSSVR, LSSVR-HGN, PLSSVR and PLSSVR-HGN. The second and third experiments are to predict the real wind-speed data-set at different times in the future to further verify the effectiveness of models LSSVR-HGN and PLSSVR-HGN. All experiments were carried out on a personal notebook with Inter Core i5-8700, 4GB memory, and windows 7 operation system in python 3.7 environment such that the same platform is provided for simulations.

In addition, parameter selection is one of the key issues affecting model evaluation, such as the regularization coefficient and the number of hidden layer nodes have a large impact on the generalization performance of the model. There have been many algorithms for selecting the optimal parameters, including Particle swarm optimization algorithm, grid search algorithm, etc. In this paper, the more popular and general grid search method is used to optimally select the parameters of the above models, which locate the optimal solution by traversing the specified parameters in the parameter space. In this section, the proposed model uses a polynomial kernel function, and the initial parameters of the proposed PSO method is $C \in [1, 201]$.

Meanwhile, to evaluate the performance of the aforementioned algorithms, the following five commonly used evaluation criterions are imported before presenting the experimental results [19], [21]. Namely, mean absolute error (MAE), root mean square error (RMSE), sum of error squares (SSE), total sum of squares (SST), and sum of squares of regression (SSR) to compare the learning performance of different models. Table 1 presents the evaluation indicator and its definitions of each metrics.

TABLE 1. Evaluation indicator and its definition.

	Calculation
MAE	$MAE=\frac{1}{M}\sum_{i=1}^{M}\left(\left\|y_i-y_i^*\right\|\right)$
$MAPE$	$MAPE=\frac{1}{M}\sum_{i=1}^{M}\frac{\left\|y_i-y_i^*\right\|}{y_i}$
$RMSE$	$RMSE=\sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(y_i-y_i^*\right)^2}$
SSE	$SSE=\sum_{i=1}^{M}\left(y_i-y_i^*\right)^2$
SST	$SST=\sum_{i=1}^{M}\left(y_i-\bar{y}\right)^2$
SSR	$SSR=\sum_{i=1}^{M}\left(y_i^*-\bar{y}\right)^2$
SSE/SST	$SSE/SST=\sum_{i=1}^{M}\left(y_i-y_i^*\right)^2/\sum_{i=1}^{M}\left(y_i-\bar{y}\right)^2$
SSR/SST	$SSR/SST=\sum_{i=1}^{M}\left(y_i^*-\bar{y}\right)^2/\sum_{i=1}^{M}\left(y_i-\bar{y}\right)^2$

Without loss of generality, assume that the mean value of the test sample is $\bar{y}=\frac{1}{M}\sum_{i=1}^{M}y_i$, while y_i represents the true value of the sample points, y_i^* indicates the prediction result, and M denotes the number of test samples. In addition, in general, a smaller MSE, MAE, SSE, SSE/SST indicates a better learning ability of the regression model. However, when the prediction sample contains noise, a smaller MSE may imply overfitting, and a smaller SSE/SST typically accompanies a larger SSR/SST. However, over-small is not necessarily the best, as it probably means mismatching.

A. UCI DATA-SET

In this section, in order to further verify the effectiveness of models, we apply the proposed model to UCI data-set, including Auto MPG(http://archive.ics.uci.edu/ml/index.php).

Select 200 training samples and 200 test samples from the Auto MPG data-set to verify the effectiveness of the model. The prediction results of the five regression models are shown in Figure 4, and the prediction error results are shown in Figure 5 and in Table 2. In the experimental part, we mainly use MAE, RMSE, SSE, SSE/SST and SSR/SST to evaluate the prediction results of five different models.

From Table 2, Figure 4 and Figure 5, among all the algorithms, the PLSSVR-HGN improves the learning effect and has the smallest evaluation standard, which is also the motivation to develop the algorithm in this paper. Specially,

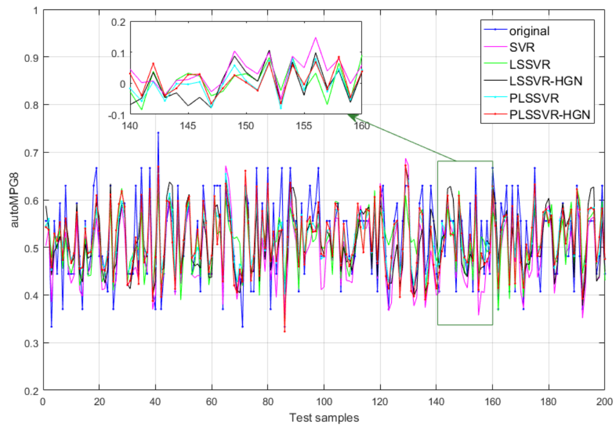


FIGURE 4. Prediction results of the five models on Auto MPG.

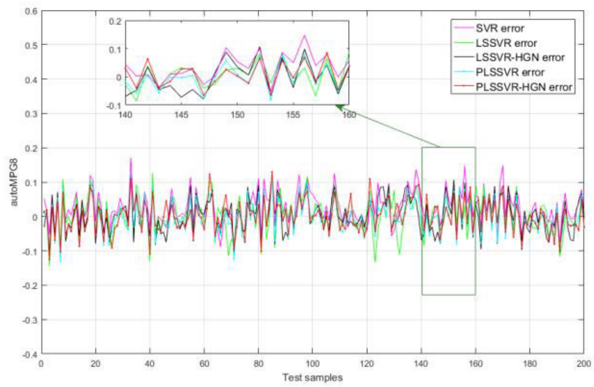


FIGURE 5. Error of the five models on Auto MPG.

TABLE 2. Prediction errors of five models on Auto MPG.

Model	MAE	RMSE	SSE	SSE/SST	SSR/SST
SVR	0.0439	0.0548	0.6004	0.3853	0.7633
LSSVR	0.0436	0.0531	0.5641	0.3620	0.7891
LSSVR-HGN	0.0391	0.0496	0.4929	0.3163	0.7912
PLSSVR	0.0388	0.0453	0.4844	0.3071	0.8218
PLSSVR-HGN	0.0379	0.0471	0.4454	0.2859	0.8787

the proposed model derives the smallest SSE and SSE/SST, and the largest SSR/SST among these five algorithms, which indicates the statistical information in the training datasets is well presented by the proposed model with fairly small regression errors. That is to say, the presented model not only obtain more accurate prediction but also owns good generalization performance.

B. SHORT-TERM WIND-SPEED FORECASTING

In the above subsection, PLSSVR-HGN has demonstrated its advantages on public data-set. To further proof the benefits of the model in practical applications, one-year wind-speed data of Heilongjiang province was gathered, which yields 62,466 samples with four attributes: mean, variance, minimum, and maximum. In this experiment, 432 training samples and 432 test samples are selected for analysis, respectively. The windspeed forecasting pattern is constructed as follows: the input vector $\overrightarrow{X_i} = (X_{i-11}, X_{i-10}, X_{i-9}, \cdots, X_{i-1}, X_i)$, $i = 1, 2, \cdots, 864$, the output value x_{i+step}, and x_i is the wind-speed value at a certain moment. The experimental setup $step = 3, 5$ in this section, where $step$ is the forecast scale. That is to say, the above models are used to forecast and analyze the wind-speed at 30 minutes and 50 minutes after a certain time in Heilongjiang Province in Summer. The forecasting results of five models on wind-speed after 30 minutes and 50 minutes are shown in Figure 6 and Figure 7. Table 3 and Table 4 show the result comparisons of five models on wind-speed after 30 minutes and 50 minutes, respectively.

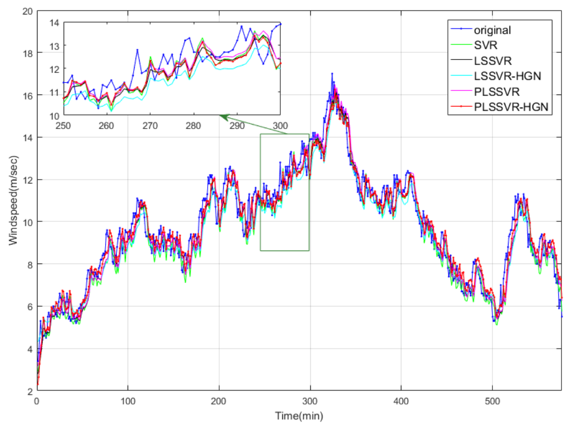


FIGURE 6. Forecasting results of five models on wind-speed after 30 min.

TABLE 3. Forecasting error of five models on wind-speed after 30 min.

Model	MAE	RMSE	SSE	SSE/SST	SSR/SST
SVR	0.5858	0.7371	312.9916	0.0831	0.9130
LSSVR	0.5747	0.7082	288.8970	0.0767	0.9170
LSSVR-HGN	0.5693	0.7058	286.9167	0.0761	0.9239
PLSSVR	0.5635	0.6971	279.9394	0.0743	0.9268
PLSSVR-HGN	0.5516	0.6786	265.2797	0.0704	0.9988

Table 3 and Table 4 demonstrate that the proposed models have distinct advantages over the other comparability models, particularly in wind-speed forecasting error statistics after 30 minutes, where model PLSSVR-HGN achieves smaller MAE, SSE, and SSE/SST, and the largest SSR/SST among these five algorithms. Furthermore, the prediction accuracy of the proposed model in terms of RMSE is always the strongest.

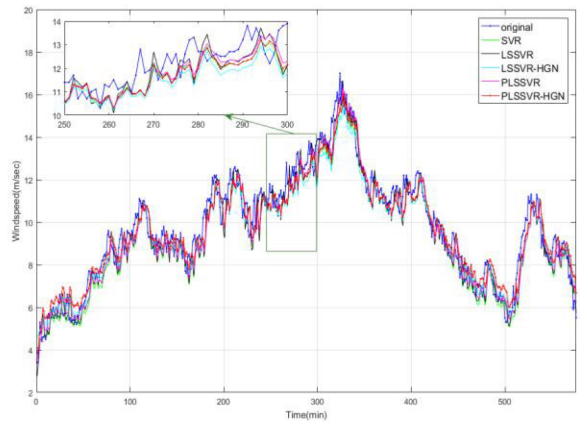


FIGURE 7. Forecasting results of five models on wind-speed after 50 min.

TABLE 4. Forecasting error of five models on wind-speed after 50 min.

Model	MAE	RMSE	SSE	SSE/SST	SSR/SST
SVR	0.7335	0.9114	478.4691	0.1286	0.8980
LSSVR	0.7017	0.8745	440.5329	0.1184	0.9008
LSSVR-HGN	0.6980	0.8781	444.1771	0.1194	0.9019
PLSSVR	0.6989	0.8676	433.6005	0.1166	0.9070
PLSSVR-HGN	0.6819	0.8419	412.6234	0.1109	0.9380

Also, as shown in Figure 6 and Figure 7, the regression curves obtained by these algorithms all deviate from the original equation to varying degrees, whereas the regression curve obtained by the proposed model is always the closest to the original system, indicating that the proposed model has a highest accuracy effect than several other models. As a result, the proposed models can be deemed an effective approach for predicting actual wind-speed.

VI. CONCLUSION

Proximal least squares support vector regression is a new regression machine designed by using regularization principle technology and least squares support vector regression. In the objective function, the bias term is introduced as the variable of the optimization problem, which makes the formula become a strongly convex objective function with equality constraints and improves the accuracy. From a variety of perspectives, it is not considered as the most adequate model and there is much room for improvements.

This section summarizes our main work: (1) Discover that the wind operation law obeys Gaussian distribution with zero mean Heteroscedasticity by investigating the properties of noise models in real wind-speed forecasting; Derive Heteroscedastic optimal empirical risk loss function by employing the Bayesian principle and maximizing posterior probability method; (2) Establish the regression models LSSVR-HGN and PLSSVR-HGN; (3) Using the Lagrange function, obtained the dual problems of LSSVR-HGN and

PLSSVR-HGN according to KKT conditions; (4) Solving models LSSVR-HGN and PLSSVR-HGN by ALM method, which guaranteed the effectiveness and stability of algorithms; (5) Experiments on Auto MPG, and wind-speed data-set demonstrate that the constructed models are more effective than other several three models recently released algorithms.

However, this work only solves the problem of Heteroscedastic Gaussian in least square regression models. In more practical situations, the true distribution of noise is complex and unknown. Considering the limited predictive ability of Heteroscedastic noise regression model to handle complex noise, authors will study using alternating mixed distributions to model noise distribution in practical problems. In addition, we can also develop issues similar to classification learning.

Author Contributions: Author Contributions: T. Zhou drafted the manuscript, conceived the algorithm and designed the experiments, G. Feng implemented the experiments; S.G. Zhang analyzed the results. All authors read and revised the manuscript.

Conflicts of Interest: The authors declare conflict of interest.

REFERENCES

[1] Z. D. Tian et al., "Short term wind-speed combination forecasting of wind farm based on wavelet transform," *J. Electrotechnics*, vol. 30, no. 9, pp. 112–120, 2015.

[2] *State Grid Corporation of China*, 2022. [Online]. Available: http://www.sgcc.com.cn/

[3] A. Gonen, F. Orabona, and S. Shalev-Shwartz, "Solving ridge regression using sketched preconditioned SVRG," in *Proc. 33rd Int. Conf. Mach. Learn.*, New York, NY, USA, 2016, p. 48.

[4] Z. Zhang, G. Dai, C. Xu, and M. I. Jordan, "Regularized discriminant analysis, ridge regression and beyond," *J. Mach. Learn. Res.*, vol. 11, pp. 2199–2228, Aug. 2010.

[5] H.-C. Zhang, Q. Wu, and F.-Y. Li, "Application of online multitask learning based on least squares support vector regression in the financial market," *Appl. Soft Comput.*, vol. 121, May 2022, Art. no. 108754.

[6] H.-C. Zhang, Q. Wu, F.-Y. Li, and H. Li, "Multitask learning based on least squares support vector regression for stock forecast," *Axioms*, vol. 11, no. 6, p. 292, 2022.

[7] C. Orsenigo and C. Vercellis, "Kernel ridge regression for out-of-sample mapping in supervised manifold learning," *Expert Syst. Appl.*, vol. 39, no. 9, pp. 7757–7762, Jul. 2012.

[8] X. Liu, X. Dong, L. Zhang, J. Chen, and C. Wang, "Least squares support vector regression for complex censored data," *Artif. Intell. Med.*, vol. 136, Feb. 2023, Art. no. 102497.

[9] H. Lei, Y. Guoxing, and H. Chao, "A sparse algorithm for adaptive pruning least square support vector regression machine based on global representative point ranking," *J. Syst. Eng. Electron.*, vol. 32, no. 1, pp. 151–162, Feb. 2021.

[10] Y. Yang, H. Zhou, Y. Gao, J. Wu, Y.-G. Wang, and L. Fu, "Robust penalized extreme learning machine regression with applications in wind speed forecasting," *Neural Comput. Appl.*, vol. 34, no. 1, pp. 391–407, Jan. 2022.

[11] S. Zhang, Q. Yuan, F. Yuan, and S. Liu, "Twin proximal least squares support vector regression machine based on heteroscedastic Gaussian noise," *J. Intell. Fuzzy Syst.*, vol. 44, no. 2, pp. 1727–1741, Jan. 2023.

[12] J. A. K. Suykens, L. Lukas, and J. Vandewalle, "Sparse approximation using least squares support vector machines," in *Proc. IEEE Int. Symp. Circuits Syst.*, Geneva, Switzerland, May 2000, pp. 757–760.

[13] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: Robustness and sparse approximation," *Neurocomputing*, vol. 48, pp. 85–105, Oct. 2002.

[14] G. Fung and O. L. Mangasarian, "Proximal support vector machine classifiers," in *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, Aug. 2001, pp. 77–86.

[15] K. Wang, H. Pei, X. Ding, and P. Zhong, "Robust proximal support vector regression based on maximum correntropy criterion," *Sci. Program.*, vol. 2019, pp. 1–11, Jan. 2019.

[16] D. Kumar, S. S. Meghwani, and M. Thakur, "Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets," *J. Comput. Sci.*, vol. 17, pp. 1–13, Nov. 2016.

[17] G. Li, L. Yang, Z. Wu, and C. Wu, "D.C. Programming for sparse proximal support vector machines," *Inf. Sci.*, vol. 547, pp. 187–201, Feb. 2021.

[18] Y. Xie, *Regression Analysis*. Beijing, China: Social Science Literature Press, 2010.

[19] Q. Hu, S. Zhang, M. Yu, and Z. Xie, "Short-term wind speed or power forecasting with heteroscedastic support vector regression," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 241–249, Jan. 2016.

[20] C. C. Onyekwena, Q. Xue, Q. Li, Y. Wan, S. Feng, H. I. Umeobi, H. Liu, and B. Chen, "Support vector machine regression to predict gas diffusion coefficient of biochar-amended soil," *Appl. Soft Comput.*, vol. 127, Sep. 2022, Art. no. 109345.

[21] Q. Hu, S. Zhang, Z. Xie, J. Mi, and J. Wan, "Noise model based ν-support vector regression with its application to short-term wind speed forecasting," *Neural Netw.*, vol. 57, pp. 1–11, Sep. 2014.

[22] W. Chu, S. S. Keerthi, and C. J. Ong, "Bayesian support vector regression using a unified loss function," *IEEE Trans. Neural Netw.*, vol. 15, no. 1, pp. 29–44, Jan. 2004.

[23] Y. Lean, "E-commerce credit risk warning based on least square approximation support vector regression model," *Theory Pract. Syst. Eng.*, vol. 32, no. 3, pp. 508–514, 2012.

[24] Q. Wu and R. Law, "The forecasting model based on modified SVRM and PSO penalizing Gaussian noise," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 1887–1894, Mar. 2011.

[25] X. X. Lu and H. D. Wang, "Research on the current situation and development strategies of genetic algorithm and particle swarm optimization," *Wireless Internet Technol.*, vol. 18, no. 21, pp. 108–109, 2021.

TING ZHOU received the B.S. degree from Shaanxi Normal University, in 1998, and the M.S. degree from Liaocheng University, in 2011. She is currently with the School of Information Engineering, Shandong Management University, China. She has authored more than ten peer-reviewed articles. Her research interests include machine learning and matrix analysis and application.

GE FENG received the B.Sc. degree in software engineering from the Pingdingshan College, in 2020. She is currently pursuing the master's degree with the College of Computer and Information Engineering, Henan Normal University. Her research interests include machine learning and knowledge discovery for regression.

SHIGUANG ZHANG received the B.Sc. degree in mathematics from Shaanxi Normal University, in 1998, the M.S. degree in mathematics from Guangxi University for Nationalities, in 2007, and the Ph.D. degree in applied mathematics from Hebei Normal University, in 2014. He is currently with the School of Information Engineering, Shandong Management University, China. He has authored more than 20 peer-reviewed articles and has served as a reviewer in several prestigious peer-reviewed international journals. His research interests include optimization algorithm, machine learning, and knowledge discovery for regression.