Journal of Computers Vol. 32 No. 4, 2021, pp. 25-41
doi:10.53106/199115992021083204003

A Novel Data-Driven Rolling Force Predictive Model Based on PSO-BAS-HKSVR Approach

Ming-Hua Liu, Qiang Zhang, Ying-Hua Liu, Wen-Li Wang*

School of Metallurgical Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China, PRC
lmhxauat@163.com, {838619820, 1079055498}@qq.com, wangwl@nwpu.edu.cn

Received 20 November 2020; Revised 11 February 2021; Accepted 15 March 2021

Abstract. To remedy the defects of the single kernel function and PSO algorithm, a novel rolling force prediction model is proposed, combining particle swarm optimization (PSO) algorithm, beetle antennae search (BAS) algorithm and hybrid kernel function support vector regression (HKSVR), i.e., PSO-BAS-HKSVR model. Hybrid kernel function (HKF) is incorporated to reduce the defect of the single kernel function of support vector regression. In the meantime, PSO algorithm is improved and combined with BAS algorithm to optimize the HKSVR model parameters $(C, g, d, \varepsilon, m)$ Statistical indicators (R^2, RMSE, MAE and MAPE) are introduced to assess the comprehensive property of the model. The experimental data of the training and testing model originate from the actual production line of the steel plant. Rolling temperature, thickness reduction, initial strip thickness and width, front tension, back tension, roll diameter, and rolling speed are taken as the input variables. Under the identical experimental conditions, compared with the single SVR, PSO-SVR, PSO-HKSVR, BPNN, GRNN and RBF models, PSO-BAS-HKSVR model exhibits the highest prediction accuracy and the optimal generalization ability. As indicated from the results PSO-BAS-HKSVR method is suited for the rolling force prediction and the optimization of model parameters in the hot strip rolling process.

Keywords: beetle antennae search algorithm, hybrid kernel function, particle swarm optimization algorithm, rolling force, support vector regression

1 Introduction

The calculation of the rolling force parameter is critical to the production of the hot-rolled strip. The calculated precision of the rolling force immediately impacts the accuracy of strip thickness, strip shape quality, rolling stability, as well as mill set-up schedule [1-2]. The process of strip rolling is strongly coupled, nonlinear, multivariable and time-varying, and the working conditions are negative, thereby increasing the difficulty in building high-precision rolling force prediction models. The conventional rolling force prediction model is built based on the theoretical analysis [3], and the process of building the model should assume and simplify numerous practical factors in the rolling production site, resulting in large errors in the model, which often cannot satisfy the demands of modern high-precision rolling technologies. Accordingly, a novel method is urgently required to build a high-precision rolling force prediction model in the hot strip production process.

Unlike the theoretical analysis methods, the artificial intelligence methods [4] imitate the real process of human brain processing, capable of predicting the rolling force based on the field and experimental data, as well as avoiding the error attributed to the assumption deviating from reality and extremely rough simplification. When the rolling force prediction model is being built, common artificial intelligence methods mainly include artificial neural network (ANN) [5] and support vector machine (SVM) [6]. Chun et al. [7] adopted the back-propagation learning algorithm with artificial neural network (ANN), and then applied the gradient descent approach in the back-propagation learning arithmetic. The

* Corresponding Author

trained ANN model can efficiently and precisely forecast the flow stress, rolling force and rolling torque. Lee and Lee [8] introduced a novel long-term learning algorithm based on the ANN and combined the ANN method with conventional learning method in the pre-calculation phases to weaken the serious thickness deviation in the first-coil. Guo et al. [9] adopted a method to combine ANN method and FEM method to predict the rolling force. According to the deviations between the calculative values by FEM and the objective values, the reliability of the FEM model was verified. The FEM simulation results were adopted as the training and test samples of the ANN model, and the optimal network was generated by repeatedly training of the model. Moussaoui et al. [10] combined the ANN with the analysis model to facilitate the rolling force prediction on the hot-rolled finishing mill. Compared with the common experiential prediction model, the hybrid model exhibits a higher prediction accuracy. Furthermore, the ANN was employed in other aspects of the rolling field (e.g., mechanical properties prediction of microalloyed steel [11] and bending force prediction [12]). Though the ANN has been extensively used for rolling force prediction in hot rolled plates, it is also subject to many defects. In existing studies, back propagation neural network (BPNN) with gradient descent algorithm is adopted by most ANN models to predict the rolling force. All parameters of the ANN topology should be optimized iteratively, so it also tends to sink into local minima, with the slow convergence speed, and the consumption time. Moreover, the ANN model requires considerable training sample data, and the parameters of the hidden layer and others are difficult to confirm. Besides, the ANN employs an empirical risk minimization learning algorithm, limiting the generalization capacity of the ANN. The mentioned inherent defects have become the key factors restricting the development of the ANN.

SVM refers to a powerful machine learning approach by complying with the structural risk minimization principle [13-14]. Compared with ANN, the SVR is capable of easily addressing the problems of small sample size, nonlinearity and high dimension, as well as yielding the global optimal solution and avoiding the falling into the local minimum value point under the limited data samples [15]. Support vector regression (SVR) is the adhibition of SVM under the of regression analysis [16], which consists of rolling force prediction, equipment support capability prediction [17], corrosion rate prediction [18], etc. Specific to the rolling force prediction, Wei et al. [19] built the rolling force model by using SVR, optimized the parameters of SVR model by adopting genetic algorithm, and seven factors are selected as input variables of the model (e.g., entrance thickness, exit thickness and rolling speed). As indicated from the simulation results, SVR rolling force model outperforms the conventional rolling force model. Wu et al. [20] built a rolling force prediction model based on SVR by using particle swarm optimization (PSO) algorithm to optimize SVR model parameters. In addition, eight factors were selected as the input variables of the model (e.g., roll diameter, rolling temperature and rolling speed).As revealed from the simulation results, the SVR model exhibits a higher prediction accuracy than the BPNN model. Chen et al. [21] built a rolling force prediction model based on SVR by applying a chaotic optimization algorithm to optimize the parameters of the SVR model, and eleven factors were selected as input variables of the model (e.g., rolling speed, reduction ratio and rolling width). As suggested from the simulation results, compared with BPNN rolling force model, the rolling force model based on SVR exhibits the advantages of fast training convergence speed and strong generalization ability.

As suggested from the above research, compared with the rolling force prediction model based on ANN and the conventional rolling force prediction model, the rolling force prediction model based on SVR exhibits a faster training speed, a stronger generalization ability, as well as a higher prediction accuracy. However, the mentioned methods are subject to several limitations. (1) The SVR rolling force prediction model constructed with only a single kernel function may be difficult to solve the high-dimensional and non-linear strip rolling problems, since the selection of kernel functions significantly impacts the prediction accuracy of SVR model. Kernel functions can be distributed into two forms. One is local kernel function exhibiting the powerful learning capability and the feeble generalization ability, and the RBF is a typical local kernel function. The other is global kernel function with the powerful generalization ability and the feeble learning competence, and the Poly refers to a typical global kernel function. The SVR rolling force prediction model built by selecting any single kernel function cannot exhibit a powerful generalization ability and a learning ability simultaneously. Later, the Hybrid Kernel Function (HKF) theory was proposed by several scholars [22-23], whereas it is rarely applied to practical production problems. (2) The penalty factor (C), RBF parameter (g) and insensitive loss parameter (ε) of the SVR rolling force model are optimized with intelligent algorithms. To be specific, the PSO algorithm is the most extensively used, exhibiting several advantages (a strong generalization capability

and a simple structure). However, at the later stage of the search, the population diversity of the the PSO algorithm quickly disappears, and the algorithm is easy to fall into the local minimum, causing the optimization effect of the parameters to be unsatisfactory. Later, some scholars proposed a Beetle Antennae Search (BAS) algorithm with a high local search ability [24], which can jump out of the local minimum, whereas it has a weak global search ability and cannot obtain the global optimal solution.

To remedy the defects of the single kernel function and the PSO algorithm, a novel rolling force prediction approach (PSO-BAS-HKSVR) is proposed based on PSO algorithm, BAS algorithm and HKSVR. PSO algorithm and BAS algorithm are combined to optimize the HKSVR model parameters $(\mathrm{C}, \mathrm{g}, d, \varepsilon, m)$ and the HKF is composed of Poly and RBF kernel functions. The HKSVR model is built adopting data sets obtained from the actual production line of the steel plant. The accuracy of the PSO-BAS-HKSVR model is analyzed, which is further compared with the single SVR, PSO-SVR, PSO-HKSVR, BPNN, GRNN and RBF models. The research results demonstrate that the proposed model an effective method to predict the rolling force and optimize the model parameters.

The technical achievements of this work can be summarized as follows:

(1) The single kernel function in SVR rolling force prediction model is replaced by HKF, combining Poly kernel function with RBF kernel function. Compared with single kernel function, HKF can improve the defects of single kernel function.

(2) A hybrid algorithm is proposed, termed as PSO-BAS algorithm, combining PSO algorithm with BAS algorithm. Compared with PSO algorithm, the PSO-BAS algorithm can more effectively optimize the parameters of the model.

(3) A rolling force prediction model combining the PSO-BAS algorithm and the HKF is built, i.e., PSO-BAS-HKSVR model. Compared with the single SVR, PSO-SVR, PSO-HKSVR, BPNN, GRNN and RBF models, PSO-BAS-HKSVR model exhibits the highest prediction accuracy and the optimal generalization ability.

2 Support Vector Regression Theory

2.1 The Basic Principles of SVR

SVM is initially used to solve classification problems [25] and then adopted to address regression problems. The SVR refers to one of the forms of the SVM to solve regression problems. The experimental dataset is denoted by $\left\{(x_i, y_i)\middle|(i=1,2,...,l, x \in R^h)\right\}$, where x_i denotes the ith input in the hth dimension, and y_i is the corresponding actual output.

In terms of SVR applications under linear conditions, $y = f(x)$ can be defined as:

$$f(\boldsymbol{x}) = (\boldsymbol{w} \cdot \boldsymbol{x}) + b \ . \tag{1}$$

where $\boldsymbol{w}$ and b denote the weight vector and threshold value; and $\boldsymbol{w} \cdot \boldsymbol{x}$ represents the inner product of $\boldsymbol{w}$ and $\boldsymbol{x}$.

The optimization problem can be represented as [26]:

$$\begin{cases} \min\limits_{w,b,\ \xi_i,\xi_i^*} \dfrac{1}{2}\left\|\boldsymbol{w}^2\right\| + C\sum\limits_{i=1}^{l}(\xi_i + \xi_i^*) \\ s.t. \begin{cases} (\boldsymbol{w} \cdot \boldsymbol{x_i}) + b - y_i \le \varepsilon + \xi_i \\ y_i - (\boldsymbol{w} \cdot \boldsymbol{x_i}) - b \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, i = 1,2,3,...,l \end{cases} \end{cases} . \tag{2}$$

where $\boldsymbol{x_i}$ denotes the input vector; y_i is the matching objective value; ξ_i and ξ_i^* represent relaxation variables; C signifies the penalty parameter; ε represents the insensitive loss parameter; l expresses the number of sample data.

By importing the Lagrangian equation and based on the strong dual relationship, the optimization problem can be solved. Thus, Eq. (1) can be rewritten as:

$$f(\boldsymbol{x}) = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i)(\boldsymbol{x_i} \cdot \boldsymbol{x}) + b \,. \tag{3}$$

where α_i and α_i^* denote the Lagrange multiplier vectors.

2.2 Kernel Technique and HKF Method

In actual factories, experimental data samples are often considered to be nonlinear or linearly inseparable when being obtained. For this reason, this study adopts feature mapping functions, which are also termed as kernel functions. The fundamental of SVR is to nonlinearly map low-dimensional inseparable data into high-dimensional space through $\boldsymbol{\phi}(\boldsymbol{x})$, so it can achieve linear separable data in high-dimensional feature space. Accordingly, the SVR model function can be formulated as:

$$f(\boldsymbol{x}) = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i)(\boldsymbol{\phi}(\boldsymbol{x_i}) \cdot \boldsymbol{\phi}(\boldsymbol{x})) + b = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) K(\boldsymbol{x_i}, \boldsymbol{x}) + b \,. \tag{4}$$

where $K(\boldsymbol{x_i}, \boldsymbol{x})$ denotes a kernel function. Kernel functions denote classified according to typical, including Line, Poly, RBF, and Sigmoid, as listed in Table 1.

Table 1. Common kernel functions

Kernel function	Mathematical expression
Linear kernel	$K(x_i, x) = x_i \cdot x$
Poly kernel	$K(x_i, x) = ((x_i \cdot x) + 1)^d$
RBF kernel	$K(x_i, x) = \exp(-g * \|x_i - x\|^2)$
Sigmoid kernel	$K(x_i, x) = \tanh(\alpha x_i \cdot x + \upsilon)$

On the whole, the kernel function can display the distribution into two forms, i.e., local kernel and global kernel. For instance, the Poly is a typical global kernel function and the RBF refers to a typical local kernel function. However, neither the onefold global kernel function nor the local kernel function is capable of fully expressing the features of the experimental data distribution [27]. This result is explained as the local kernel function exhibits an excellent learning capability and a poor generalization competence, and the global kernel function achieves an excellent generalization ability and a poor learning capability. Hence, the mean of HKF is to merge two onefold kernel functions to cause the HKF more adaptive for practical project conditions. The integration of the two kernel functions is still the kernel function, since it satisfies the Mercer condition [28]. Thus, the hybrid kernel function is defined as:

$$\begin{cases} k_{Hybrid} = (1-m) \cdot k_{RBF} + m \cdot k_{Poly} \\ s.t. 0 \le m \le 1 \end{cases} . \tag{5}$$

where m signifies the control parameter. When m=0, the HKF executes the RBF kernel function. When m=1, the HKF performs the Poly kernel function. Thus, by regulating the control parameter m, the HKF is capable of adaptively expressing the characteristics of different training set distributions, so the optimal fitting results can be acquired in practical engineering problems.

3 Parameters of HKSVR Optimizatized by the PSO-BAS Algorithm

3.1 PSO Algorithm

PSO algorithm refers to a random swarm intelligence optimization algorithm proposed by Kennedy and Eberhart in 1995 [29]. The algorithm derives from the predatory behavior of birds, and each particle signifies a latent solution. If the population scale is $n(i = 1, 2, 3, ..., n)$, searching in h-dimensional space, the position of the ith particle can be signified as $x_i = (x_{i1}, x_{i2}, ..., x_{ih})$, and the flight speed of the ith particle can be described as $v_i = (v_{i1}, v_{i2}, ..., v_{ih})$. The optimal position of particle search is the optimal solution, and individual extreme value is expressed as $p_i = (p_{i1}, p_{i2}, ..., p_{ih})$. Accordingly, optimal position of the population is expressed as $p_g = (p_{g1}, p_{g2}, ..., p_{gh})$.

The standard 2011 PSO algorithm has been employed, and the basic parameters setting of PSO [2] are listed in Table 2, in this study. The flowchart of the PSO algorithm is plotted in Fig. 1. The specific flow of the algorithm is signified as shown below:

Table 2. Basic parameters of PSO algorithm

Parameters	Values
Population scale	100
Inertia weight	$\frac{1}{2\ln 2}$
Acceleration factors	$c_1 = c_2 = \ln 2 + 0.5$
Number of iterations	50

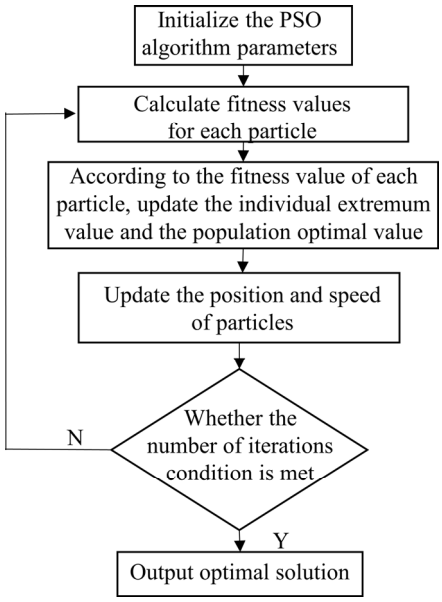


Fig. 1. Flowchart of PSO algorithm

Step 1: Initialize the parameters of the PSO algorithm. n is the population scale of the PSO, x_{ih} and v_{ih} denote the position and speed of the particle, c_1 and c_2 are termed as the acceleration factors, β signifies the inertia weight, t signifies the number of iterations,.

Step 2: Calculate fitness values for each particle. With MSE as the fitness function, the MSE formula is expressed as:

$$F_{fitness} = \frac{1}{l}\sum_{i=1}^{l}(y_i' - y_i)^2 \ . \tag{6}$$

where $y_i^{'}$ and y_i denote the prediction and target value; l represents the number of samples.

Step 3: By comparing the fitness function values of the respective particle, the individual extremum p_i and population optimal value p_g are obtained.

Step 4: Updating the position and speed of particle, the equations are defined as:

$$v_{ih}^{(t+1)} = \beta \times v_{ih}^{t} + c_1 \times rand() \times (p_{ih}^{t} - x_{ih}^{t}) + c_2 \times rand() \times (p_{gh}^{t} - x_{ih}^{t}) \ . \tag{7}$$

$$x_{ih}^{(t+1)} = x_{ih}^{t} + v_{ih}^{(t+1)} \ . \tag{8}$$

where β denotes the inertia weight; c_1 and c_2 are termed as the acceleration factors; $rand()$ represents uniform distribution random numbers belong to 0-1; t signifies the number of iterations.

Step 5: Whether the iteration number condition is met and the exit if it is contented, otherwise, return to step 2.

Step 6: Export the optimal solution of population, that is to get the optimal parameters.

3.2 PSO Algorithm Combined with BAS Algorithm (PSO-BAS)

In 2017, Jiang and Li [24] proposed a novel bionic optimization algorithm-Beetle Antennae Search Algorithm (BAS). The bionic nature of the algorithm is that beetles exploit their two antennae to sense the position with higher odor concentration of food, and gradually approach the food. In the BAS algorithm, the individual is determined by the individual's judgment of environment space in each iteration and does not consider the relationship between groups. BAS algorithm exhibits a strong learning capacity, but a weak generalization ability. By combining PSO algorithm with BAS algorithm, the PSO-BAS algorithm is capable of retaining the strong generalization ability of PSO algorithm, while exhibiting a strong local learning capacity of BAS algorithm, so the optimization ability of the algorithm is noticeably optimized.

In this study, a novel algorithm is proposed, i.e., an incorporation of the PSO and BAS algorithm. In the iterative process of the algorithm, the PSO algorithm is used to quickly find the appropriate solution in the searching space, and then the solution acts as the initial parameter of the BAS algorithm. BAS algorithm is used to continue searching in the space till the optimal solution is obtained. After PSO-BAS algorithm is adopted, the particles can jump out of the local minima, and then they can detect the global optimal solution. The detailed processes of PSO-BAS algorithm are expressed as shown below:

Step 1: Initialize parameters of PSO-BAS algorithm. Specific to the particle swarm, n denotes the population scale of the PSO, x_{ih} and v_{ih} indicate the position and speed of the particle, respectively, c_1 and c_2 express two the acceleration factors, β signifies the inertia weight, and t indicates the number of iterations. For the beetle group, n' represents the scale of the BAS, $x^{t'}$ is the coordinate of the centroid of beetle, t' denotes the number of iterations; d_0 signifies the antennae length, and $step$ expresses the step length.

Step 2: Perform steps 2-6 of the PSO algorithm to obtain an optimized solution for the PSO algorithm search.

Step 3: The optimal solution searched by the PSO algorithm is regarded as the initial position of the beetle group.

Step 4: Calculate the search behavior of the left and right sides of the beetle.

The beetle head searches randomly in any direction, so the orientation of the vector from the left antennae to the right antennae is also arbitrary. A random vector can be yielded, as written by $dir_0 = rands(h,1)$. Normalizing the random vector can get $dir = dir_0 / \|dir_0\|$, so the relationship between the two antennae can be represented as $xl - xr = d_0 * dir$. Obviously, the dimensional coordinates

of the two antennae are written as:

$$xl = x^{t'} + d_0^{t'} * dir / 2 . \tag{9}$$

$$xr = x^{t'} - d_0^{t'} * dir / 2 . \tag{10}$$

where xl denotes the left antennae of coordinate; xr is the right antennae of coordinate; $x^{t'}$ represents the centroid coordinate of beetle under the number of iterations of t'; $d_0^{t'}$ expresses the antennae length of the beetle when the number of iterations is t'; dir denotes the random unit vector; h represents dimensions of position.

Step 5: Update the position of the beetle.

Referring to the Eq. (6), the odor concentration of the left antennae is compared with the odor concentration of the right antennae to find the minimum value $f(x)$. If $f(xl) < f(xr)$, the beetle travels a distance $step$ towards the direction of the left antennae; if $f(xl) > f(xr)$, the beetle travels a distance $step$ towards the direction of the right antennae. Using a sign function, the mentioned two cases can be written as an equation:

$$x^{t'} = x^{t'-1} - step^{t'} * dir * sign(f(xl) - f(xr)). \tag{11}$$

where *step* represents the step size of each iteration; *sign* represents a sign function.

Step 6: Compare $f(x^{t'})$ with f_{best}.

If $f(x^{t'}) < f_{best}$, $x_{best} = x^t$; otherwise, the former optimal position will remain unchanged. x_{best} is the present optimal position, and f_{best} is the fitness function value of the present best position.

Step 7: Updating the antennae length and step size of the beetle as follows:

$$d_0^{t'} = 0.95 d_0^{t'-1}. \tag{12}$$

$$step^{t'} = 0.95 * step^{t'-1}. \tag{13}$$

Step 8: Whether the iteration number condition is met and the exit if it is contented; otherwise, go back to step 4.

Step 9: Export the best solution with PSO-BAS algorithm, i.e., to get the best parameters.

3.3 Parameters of HKSVR Optimizatized by the PSO-BAS Algorithm

Before adopting the HKSVR model to predict rolling force, the parameters (C ,g , d , ε , m) should be optimized. In this study, when the parameters are optimized by the PSO-BAS algorithm, its will be limited to a certain range, C = [0.01, 100], g = [0, 1], d = [1,3], ε = [0.01, 1], and m = [0, 1]. The steps of the parameters are optimized adopting the PSO-BAS algorithm, as presented below:

Step 1: Perform steps 1-9 of the PSO-BAS algorithm.

Step 2: The best parameters are assigned to the HKSVR model, and the HKSVR prediction model is built. The workchart of the parameters of HKSVR optimized by the PSO-BAS algorithm is presented in Fig. 2.

Through the mentioned analysis, the parameters to be optimized are different for the single SVR model, PSO-SVR model, PSO-HKSVR model and PSO-BAS-HKSVR model, as listed in Table 3. The single SVR model does not set optimization initial parameters. PSO-SVR sets optimization parameters (C, g), and PSO-HKSVR and PSO-BAS-HKSVR models set optimization parameters (C , g , d , ε , m) .

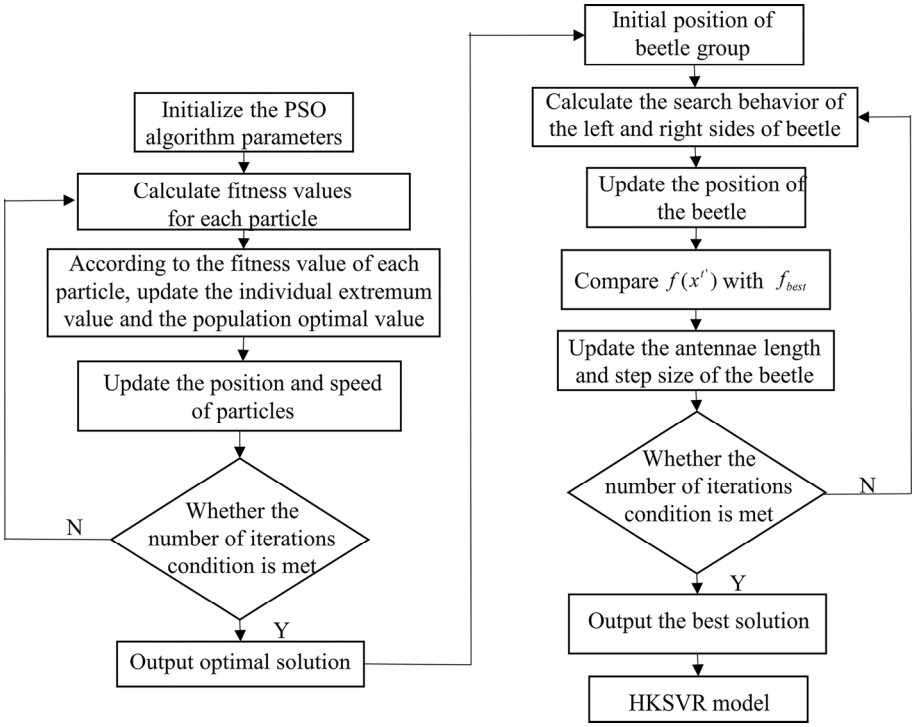


Fig. 2. The workchart of the parameters of the HKSVR optimizatized by the PSO-BAS algorithm

Table 3. Parameters optimized by different models

Prediction model	C	g	ε	d	m
Single SVR	⇓	⇓	⇓	⇓	⇓
PSO-SVR	√	√	⇓	⇓	⇓
PSO-HKSVR	√	√	√	√	√
PSO-BAS-HKSVR	√	√	√	√	√

4 Analysis and Discussion

To verify the advantage of the proposed PSO-BAS-HKSVR model for handling problem on rolling force prediction, predictive accuracy of the proposed PSO-BAS-HKSVR model is compared with that of the single SVR, PSO-SVR, PSO-HKSVR, BPNN, GRNN and RBF models.

4.1 Criteria for Evaluating Performance

To comprehensively verify whether the predictive results of the proposed PSO-BAS-HKSVR model have enhanced as compared with the other models, coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) are employed to assess the prediction performance of various models (Table 4). The value of R^2 is closer to 1, and the values of RMSE, MAE and MAPE are relatively smaller, demonstrating that this model exhibits a better prediction accuracy.

4.2 Experiment Data Preprocessing

Rolling force data of 228 pairs originate from the hot rolling production line of a steel plant as experimental data. The experiment data are listed in Table 5. The experiment data acquired from the steel plant inevitably contains abnormal and noisy data, probably affecting the performance of the model. Hence, data processing should be performed to achieve reliable analysis results, instead of being used

directly. The present study uses T test criterion to complete bad data cleaning. The T test criterions are written as:

Table 4. Using different statistical indicators to evaluate model performance

Statistical indicators	Mathematical expression
R^2	$R^2 = \frac{(l\sum_{i=1}^{l} y_i y_i' - \sum_{i=1}^{l} y_i \sum_{i=1}^{l} y_i')^2}{\left[l\sum_{i=1}^{l} y_i^2 - (\sum_{i=1}^{l} y_i)^2\right]\left[l\sum_{i=1}^{l} y_i'^2 - (\sum_{i=1}^{l} y_i')^2\right]}$
RMSE	$RMSE = \sqrt{\frac{1}{l}\sum_{i=1}^{l}(y_i - y_i')^2}$
MAE	$MAE = \frac{1}{l}\sum_{i=1}^{l}\left\|y_i - y_i'\right\|$
MAPE	$MAPE = \frac{1}{l}\sum_{i=1}^{l}\left\|\frac{y_i - y_i'}{y_i}\right\| \times 100\%$

Table 5. Experiment data of hot rolling

Sequence	H_0 [1] (mm)	ε [2] (%)	T_0 [3] (℃)	v [4] (m/s)	D [5] (mm)	B [6] (mm)	T_b [7] (KN)	T_f [8] (KN)	F [9] (KN)
1	1.8444	12.15	850.6	11.0125	676.800	1249	26.50965	15.90579	8736.312
2	2.341	11.28	871.5	11.0382	647.220	1319	22.89735	13.73841	6221.147
3	2.669	12.29	872.5	10.8127	652.056	1500	26.71631	16.02979	8725.214
…	…	…	…	…	…	…	…	…	…
227	27.086	34.04	975.9	1.8200	792.217	1250	4.97665	9.00272	19345.82
228	27.750	32.96	997.1	1.9000	738.700	1250	5.09481	8.24848	17793.96

Note. [1] Initial strip thickness, [2] Thickness reduction, [3] Rolling temperature, [4] Rolling speed, [5] Roll diameter, [6] Initial strip width, [7] Front tension, [8] Back tension, [9] Rolling force.

$$\bar{x} = \frac{1}{l}\sum_{i=1}^{l} x_i \ . \tag{14}$$

$$\sigma = \sqrt{\frac{1}{l}\sum_{i=1}^{l}(x_i - \bar{x})^2} \ . \tag{15}$$

$$\left|x_i - \bar{x}\right| > k(\alpha, n)\sigma \ . \tag{16}$$

$$k(\alpha, n) = t_\alpha(n-1)\sqrt{1 + \frac{1}{n-1}} \ . \tag{17}$$

where x_i denotes the raw data; $\bar{x}$ represents the mean value of x_i; σ expresses the standard error of the original data. When the error of data satisfies Eq. (16), the data will be eliminated. Then, the remaining data are recalculated until all data do not satisfy Eq. (16). The data of output variable are processed in this way.

The experimental results eliminated data points of 10 pairs and select data of 218 pairs as experimental data. 80% (175) of the experimental data are adopted as the training set for the model training, and the rest of experimental data are adopted for the model reliability test. To be specific, rolling temperature, thickness reduction, initial strip thickness and width, front tension, back tension, roll diameter, and rolling speed act as the input variables, and rolling force is selected as output variable. However, different variables of input often have the order of magnitude differences, thereby affecting the prediction accuracy and the training speed of the model. To remove the effect of the order of magnitude difference,

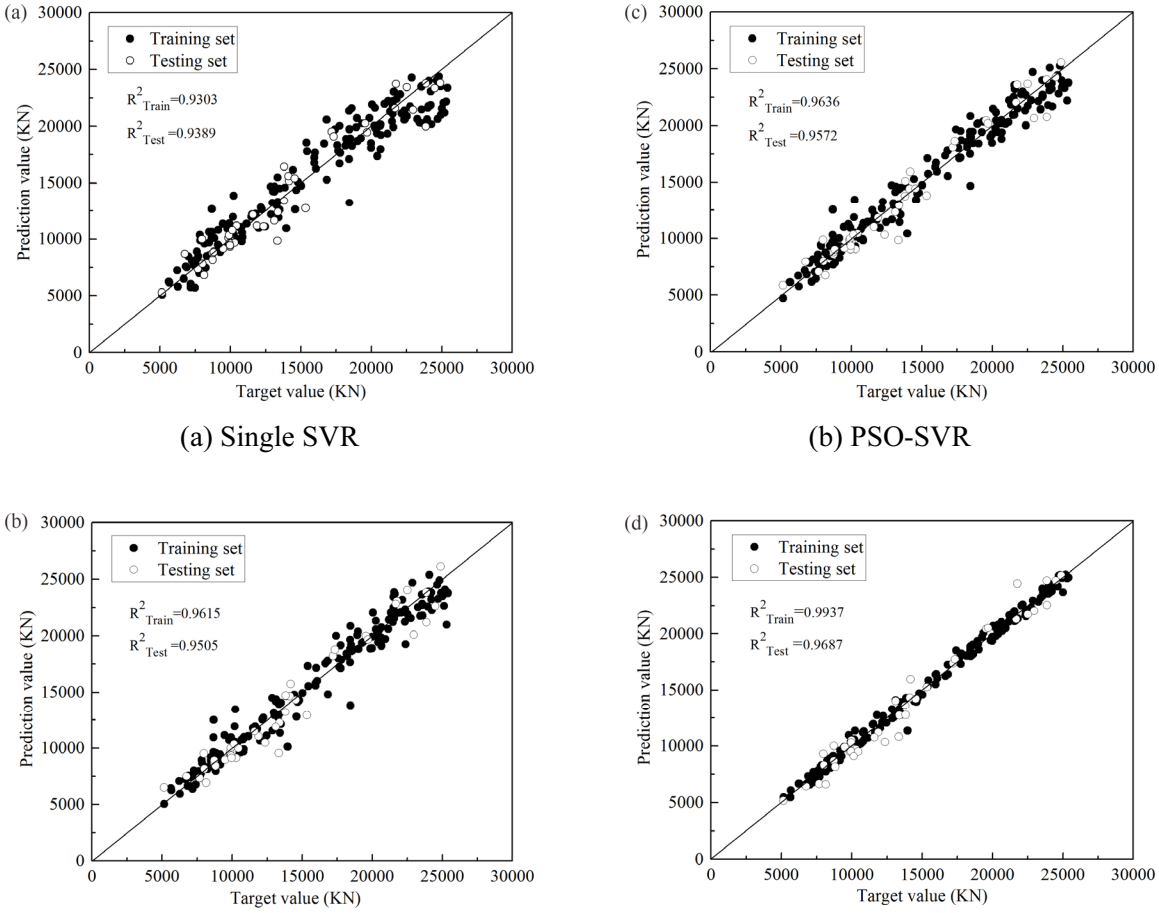
the sample data should be normalized. Before establishing the model, experimental data are normalized to [-1, 1], so the difference of the order of magnitude can be eliminated, and the prediction precision and the training velocity of the model can be improved. The normalized formula is employed as:

$$x_i^{'} = 2 * \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} - 1, i = 1, 2, 3, ..., l. \quad \textbf{(18)}$$

where $\max(x_i)$ and $\min(x_i)$ express the maximal and minimal values of the data sequences, respectively.

4.3 Results and Analysis

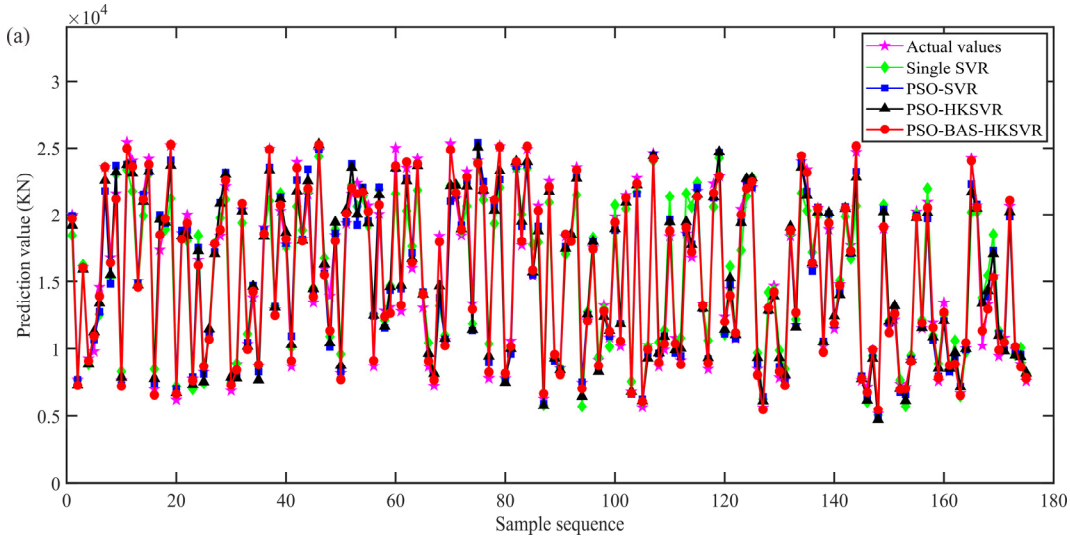
Comparison between PSO-BAS-HKSVR Model and PSO-HKSVR, PSO-SVR and Single SVR Models. To order to comprehensively compare the prediction accuracy and generalization ability of the proposed PSO-BAS-HKSVR model with PSO-HKSVR, PSO-SVR and single SVR models, the evaluation indicators in section 4.1 are employed to verify the prediction results of each model. The regression results of the proposed PSO-BAS-HKSVR model with the PSO-HKSVR, PSO-SVR and single SVR models are plotted in scatter diagram as Fig. 3, and Fig. 4 demonstrates the comparison of prediction value and target value of the four models for rolling force, on training and testing set. According to Fig. 3 and Fig. 4, the R^2 of the four models are all above 0.93 in the training and the testing sets, and the predicted value of the four models highly fits with the target value, demonstrating that the four models have a higher prediction accuracy. Hence, the rolling force prediction model based on SVR achieves a higher generalization ability.



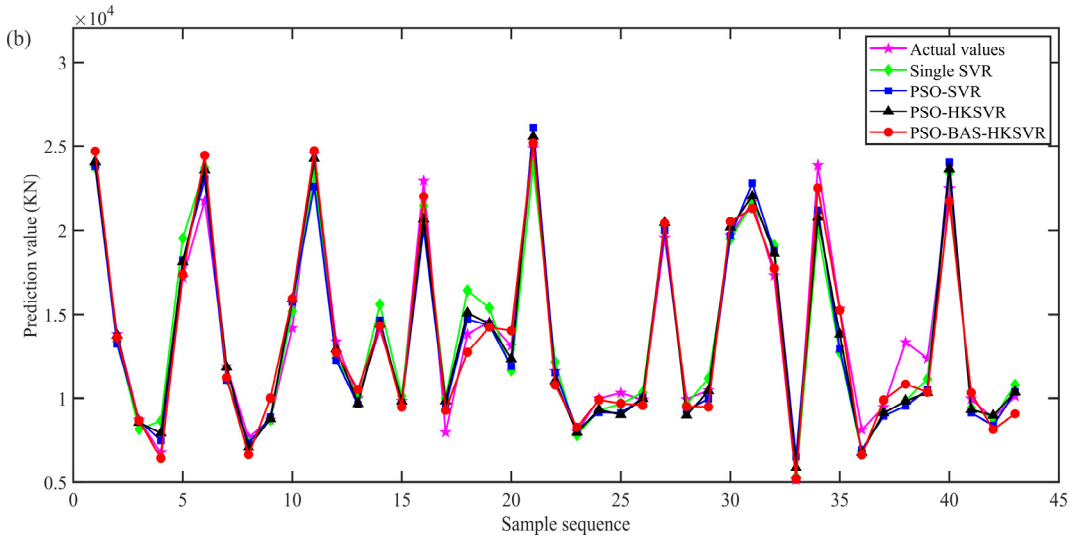
(a) Single SVR (b) PSO-SVR

(c) PSO-HKSVR (d) PSO-BAS-HKSVR

Fig. 3. The regression results of models on training and testing set



(a) training set



(b) testing set

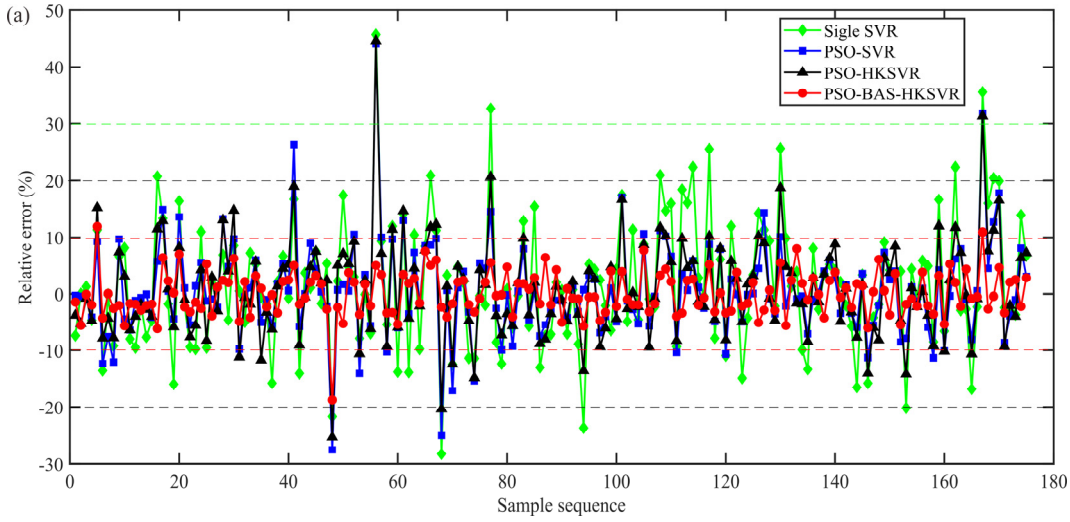
Fig. 4. Comparison of predicted values and target values for different models

Fig. 3 indicates that the black solid circle and the white hollow circle represent the prediction results of the model in the training set and the testing set respectively. If more data points in the model are clustered near the y = x line, the prediction results are closer to the target data. The following conclusions can be drawn from Fig. 3: (1) The prediction performance of PSO-SVR model is compared with that of single SVR model, and that of the former is significantly better than that of the latter. Whether in the training set or the testing set, the R^2 of the single SVR model is not more than 0.9400, while the R^2 of the PSO-SVR model reaches over 0.9500, fully indicating that the prediction accuracy and generalization ability of single SVR model are improved by PSO algorithm compared with that without the use of the optimization algorithm. The main reasons for the mentioned phenomenon are presented below. When the single SVR model is built, the initial parameters (C, g) of the single SVR model are artificially set, so a large prediction error of the single SVR model is generated. When the PSO-SVR model is built, the initial parameters (C, g) of the SVR model are optimized by PSO algorithm, and the optimized parameters (C, g) of PSO algorithm are taken as the initial parameters of the SVR model. (2) In the training set or testing set, the R^2 of PSO-HKSVR model is slightly higher than that of PSO-SVR model, demonstrating that the HKF can improve the prediction accuracy and generalization ability of the model

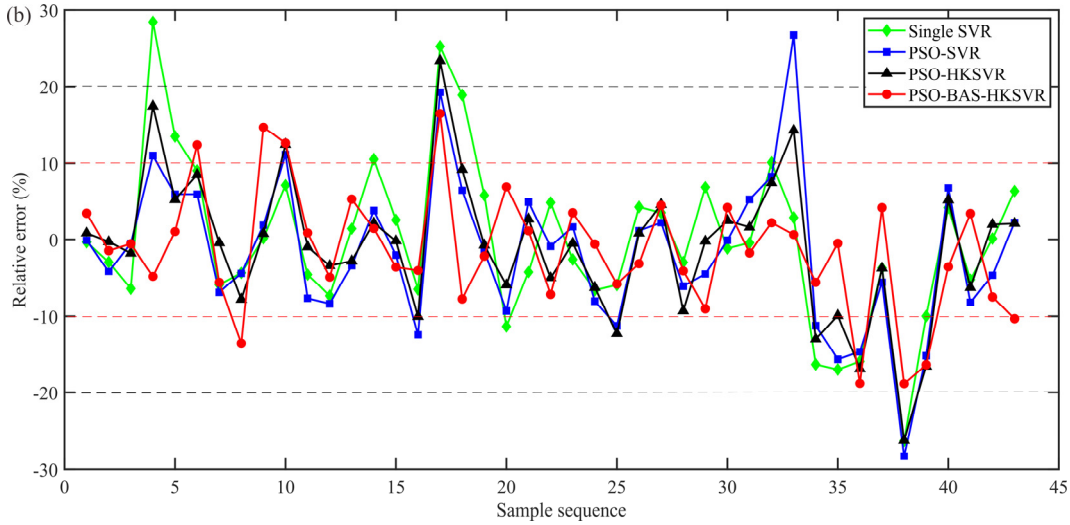
compared with the single kernel function. The main reasons for the mentioned phenomenon are presented below. Local kernel function has a strong learning ability but a weak generalization ability, whereas global kernel function has strong generalization ability and weak learning ability. Neither single global kernel function nor local kernel function is capable of fully describing the characteristics of experimental data distribution. However, the HKF combines the advantages of local kernel function and global kernel function, so HKF has a strong learning ability and generalization ability, and improves the defects of the single kernel function. Thus, the HKF can fully describe the characteristics of experimental data distribution, and is suitable for solving the problem of rolling force prediction in plate rolling process. (3) Compared with the PSO-HKSVR model, the R^2 of PSO-BAS-HKSVR model is 0.9937, which is much larger than that of PSO-HKSVR model (R^2 is 0.9636), in the training set. Similarly, in the testing set, the R^2 of PSO-BAS-HKSVR model is 0.9687, which is much larger than that of PSO-HKSVR model (R^2 is 0.9572). Consequently, the prediction precision and generalization ability of the PSO-BAS-HKSVR model far exceed those of the PSO-HKSVR model. The main reason for this phenomenon: the PSO algorithm has weak local search ability, and it is easy to fall into the local minimum. Once it falls into the local minimum, it is difficult to jump out, so the optimal parameters of the model cannot be obtained, reducing the prediction accuracy and generalization ability of the model. However, the PSO-BAS algorithm retains the powerful global search ability of PSO algorithm, while exhibiting the powerful local search capability of BAS algorithm. Accordingly, the PSO-BAS algorithm noticeably improves the local search capability of PSO algorithm and reduces the possibility of PSO algorithm falling into local minimum. (4) Under the mentioned research background, compared with PSO-HKSVR, PSO-SVR and single SVR models, PSO-BAS-HKSVR model exhibits the maximum prediction accuracy and the optimal generalization ability.

The prediction accurate differences of the four models are unlikely to be directly deduced from Fig. 4. Consequently, the relative errors in the prediction model are compared. Fig. 5 demonstrates the relative error deviations of the predictive and objective values of the four models. The differences in the relative error indexes of the four models are obviously presented in Fig. 5. In the training set, the relative error of all data except for the two data of PSO-BAS-HKSVR model is less than 10%, that of most data of the PSO-HKSVR model and PSO-SVR model is less than 20%, that of most data of the PSO-HKSVR model is slightly lower than that of the PSO-SVR model, and that of most data of the single SVR model is less than 30%. In the testing set, the relative error of all data of PSO-BAS-HKSVR model is within 20%, that of most data is within 10%; the relative error of all data of the PSO-HKSVR, PSO-SVR and single SVR models is within 30%, and that of most data is within 20%. The relative error of most data of the PSO-HKSVR model is slightly lower than that of PSO-SVR model, and that of most data of PSO-SVR model is slightly lower than that of single SVR model. In the training and testing set, the relative error of PSO-BAS-HKSVR model is distinctly less than that of the other three models in the rolling force prediction. As a result, under the mentioned research background, the rolling force prediction ability of PSO-BAS-HKSVR model is higher than the other three models.

Besides, to comprehensively evaluate the precision of the rolling force prediction of the four models, RMSE, MAE and MAPE are also act as the error criterions to assess the performance of models. In addition, RMSE is termed as standard deviation. It exhibits high sensitivity to the large or small error in a set of measurement results, and it is capable of effectively reflecting the measurement precision. MAE can accurately reflect the actual prediction error, and MAPE has the capability of measuring the accuracy of prediction. Table 6 lists the computational results of the three error criterions of the four models. Fig. 6 presents a column diagram of the RMSE and MAE error distribution plotted from the consequences of the computation. Moreover, Fig. 7 gives a histogram of the MAPE error distribution plotted from the consequences of the computation. According to Fig. 6, Fig. 7 and Table 6, as opposed to PSO-SVR model, PSO-HKSVR model is not conspicuously superior. In the testing set, the three error criterions of PSO-HKSVR model are noticeably lower than those of PSO-SVR model. In the training set, though the RMSE error criterion of the PSO-HKSVR model remains better than that of the PSO-SVR model, the PSO-HKSVR model has opposite MAE and MAPE error criterions. Accordingly, compared with single kernel function, HKF is capable of improving the generalization ability of single kernel function without losing the learning ability of single kernel function. The prediction precision of PSO-BAS-HKSVR model far exceeds that of PSO-HKSVR model. PSO-BAS-HKSVR model has lower RMSE and MAE error criterions than PSO-HKSVR model on the training and testing set, the MAPE error criterion of PSO-BAS-HKSVR model also displays the identical pattern. For this reason, PSO-BAS algorithm



(a) training set



(b) testing set

Fig. 5. The relative error values for different models

Table 6. Statistical analysis of error criterion for different models

Prediction model	Training set			Testing set		
	RMSE	MAE	MAPE(%)	RMSE	MAE	MAPE(%)
Single SVR	1624.9765	1241.4747	8.6716	1375.9640	1028.7568	7.7727
PSO-SVR	1183.6423	833.2698	5.8595	1277.0700	973.7705	7.4431
PSO-HKSVR	1154.8952	919.6922	6.7141	1179.3299	848.9232	6.6027
PSO-BAS-HKSVR	476.2226	396.3731	3.0533	978.6133	756.2589	6.0644

outperforms PSO algorithm in optimizing the parameters of the model. Whether in the training or the testing set, the three error indexes of PSO-BAS-HKSVR model are at the lowest level compared with those of PSO-HKSVR, PSO-SVR and single SVR models, which fully verifies that PSO-BAS-HKSVR model exhibits the highest prediction accuracy and the optimal generalization ability.

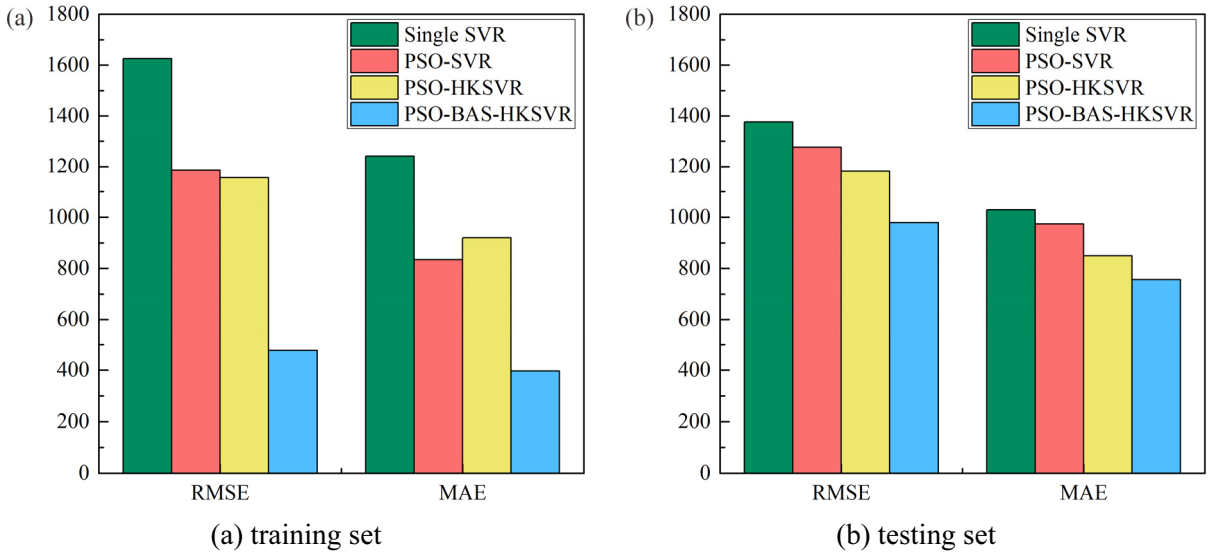


Fig. 6. The RMSE and MAE error column diagram for different models

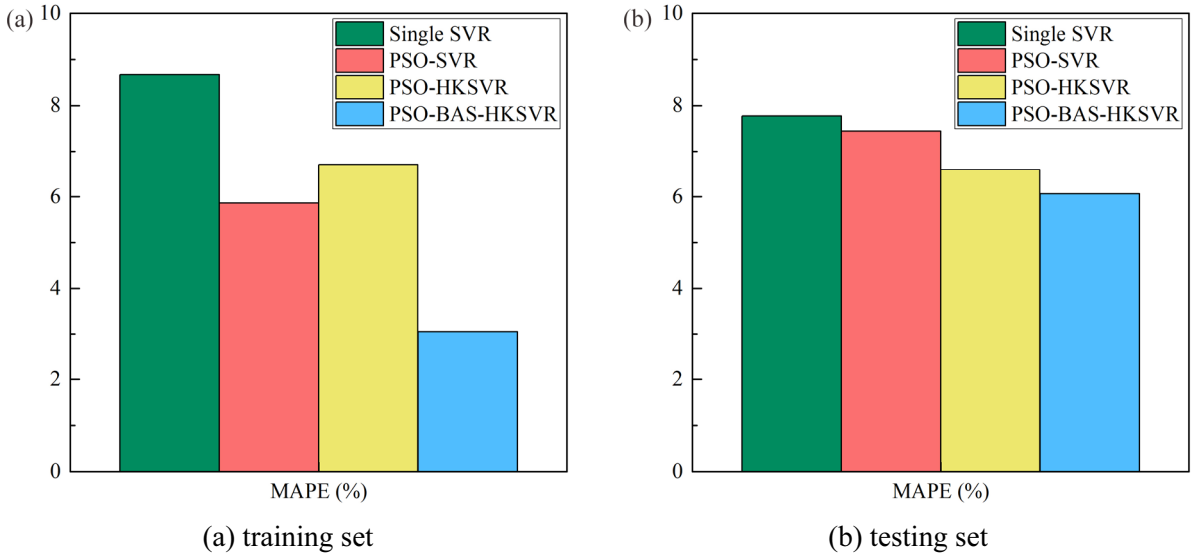


Fig. 7. The MAPE error histogram for different models

Comparison between PSO-BAS-HKSVR Model and BPNN, GRNN and RBF Models. Artificial neural networks have been commonly adopted for building the regression forecasting model. Three types of the ANN are adopted to verify the effectiveness of the PSO-BAS-HKSVR model, i.e., back propagation neural network (BPNN), general regression neural network (GRNN), and radial basis function (RBF). All parameters of the mentioned three models are set as default, except that BPNN complies with a single hidden layer structure with ten neural. The regression performance of PSO-BAS-HKSVR, BPNN, GRNN and RBF models on the test is drawn in the scatter diagram in Fig. 8. As obviously suggested from the figure, most of the prediction results of PSO-BAS-HKSVR model cluster near the y=x line, demonstrating that the generalization ability of the PSO-BAS-HKSVR model outperforms that of the other three ANN models. Calculation results of model error-indexes with PSO-BAS-HKSVR, BPNN, GRNN and RBF models are listed in Table 7, and their column errors are plotted in Fig. 9. Table 7 and Fig. 9 clearly present that the PSO-BAS-HKSVR model achieves the minimum three error indexes compared with the other three ANN models. Based on the mentioned experimental results, the PSO-BAS-HKSVR model is sufficiently demonstrated to be more suitable for rolling prediction than the BPNN, GRNN and RBF models.

Table 7. Statistical analysis of the performance for different models

	PSO-BAS-HKSVR	BPNN	GRNN	RBF
RMSE	978.6133	1600.4806	2524.2044	3809.3486
MAE	756.2589	1184.9589	2123.5831	2655.4228
MAPE(%)	6.0644	9.8055	17.6758	25.5150

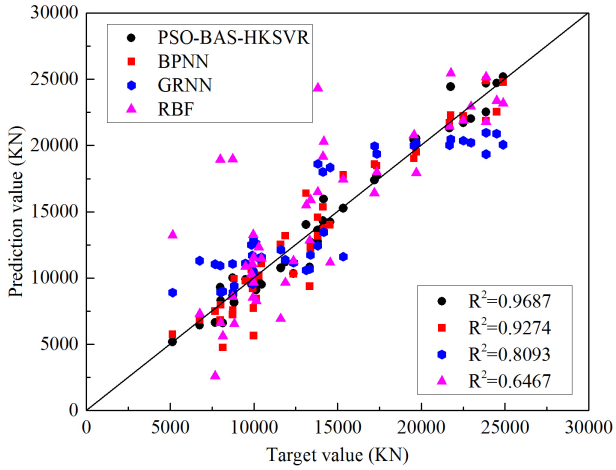


Fig. 8. Scatter diagram of PSO-BAS-HKSVR, BPNN, GRNN and RBF models in the prediction of test set

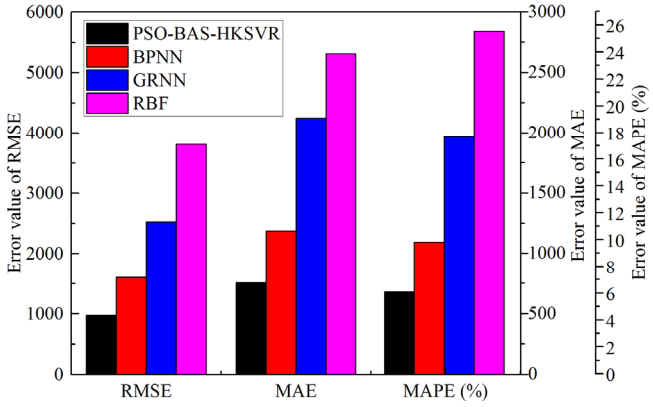


Fig. 9. Error histogram of different models

5 Conclusions

To remedy the defects of the single kernel function and the PSO algorithm, an original rolling force prediction model is built in the present study with the PSO-BAS-HKSVR method in the hot rolling strip process. First, the single kernel function in SVR rolling force prediction model is substituted with HKF that combines Poly kernel function with RBF kernel function. Second, a hybrid algorithm combining PSO algorithm and BAS algorithm is presented, i.e., PSO-BAS algorithm, which is employed to optimize the parameters $(C, g, d, \varepsilon, m)$ of HKSVR model. Third, the PSO-BAS-HKSVR calculation consequences are compared with those of the SVR, PSO-SVR and PSO-HKSVR models. As indicated from the research results, the HKF is capable of remedying the defects of single kernel function, the PSO-BAS algorithm can enhance the ability of the PSO algorithm to optimize the parameters of the model, and the PSO-BAS-HKSVR model exhibits the highest prediction precision and the optimal generalization capacity. Lastly, under the identical conditions, the calculation results of PSO-BAS-HKSVR are compared with those of BPNN, GRNN and RBF models, and the superiority of PSO-BAS-HKSVR model is verified again. In brief, PSO-BAS-HKSVR method can be successfully employed to predict the rolling force and optimize the parameters of the model in the hot strip rolling process. Rare

theoretical basis has been laid for the rolling force prediction model based on the machine learning methods, so the future direction is to combine machine learning methods with theoretical analysis methods, which may be more suitable for rolling force prediction.

References

[1] M. Bagheripoor, H. Bisadi, Application of artificial neural networks for the prediction of roll force and roll torque in hot strip rolling process, Applied Mathematical Modelling 37(7)(2013) 4593-4607.

[2] Z.-H. Wang, D.-H. Zhang, D.-Y. Gong, W. Peng, A new data-driven roll force and roll torque model based on FEM and hybrid PSO-ELM for hot strip rolling, ISIJ International 59(9)(2019) 1604-1613.

[3] S. Li, Z.-G. Wang, C.-M. Liu, J.-H. Ruan, Z.-B. Xu, A simplified method to calculate the rolling force in hot rolling, The International Journal of Advanced Manufacturing Technology 88(5)(2016) 2053-2059.

[4] X.-H. Liu, Q.-L. Zhao, Z.-Y. Huang, Prospect of artificial intelligent application in rolling, Steel Rolling 34(4)(2017) 1-5.

[5] A. Parvizi, H.R.R. Raftar, Application of artificial neural network and genetic algorithm to predict and optimize load and torque in T-section profile ring rolling, Journal of Mechanical Engineering Science 233(17)(2019) 5966-5976.

[6] Z.-K. Zhang, F. Luan, D.-Y. Li, J.-T. Xu, H.-Z. Wang, J.-M. Geng, Prediction of rolling force in the hot strip rolling using support vector regression with principal components analysis, in: Proc. 2019 IEEE 2nd International Conference on Information Systems and Computer Aided Education, 2019.

[7] M.-S. Chun, J. Biglou, J.-G. Lenard, J.-G. Kim, Using neural networks to predict parameters in the hot working of aluminum alloys, Journal of Materials Processing Technology 86(1-3)(1999) 245-251.

[8] D. Lee, Y. Lee, Application of neural-network for improving accuracy of roll-force model in hot-rolling mill, Control Engineering Practice 10(4)(2002) 473-478.

[9] Z.-Y. Guo, J.-N. Sun, F.-S. Du, Application of finite element method and artificial neural networks to predict the rolling force in hot rolling of Mg alloy plates, Journal of the Southern African Institute of Mining and Metallurgy 116(1)(2016) 43-48.

[10] Y.S.A. Moussaoui, H.A. Abbassi, Hybrid hot strip rolling force prediction using a bayesian trained artificial neural network and analytical models, American Journal of Applied Sciences 3(6)(2006) 1885-1889.

[11] S.-W. Wu, X.-G. Zhou, G.-M. Cao, N.-A. Shi, Z.-Y. Liu, High-dimensional data-driven optimal design for hot strip rolling of microalloyed steel, Steel Research International 89(7)(2018) 1-11.

[12] Z.-H. Wang, D.-Y. Gong, X. Li, G.-T. Li, D.-H. Zhang, Prediction of bending force in the hot strip rolling process using artificial neural network and genetic algorithm (ANN-GA), The International Journal of Advanced Manufacturing Technology 93(9)(2017) 3325-3338.

[13] V.N. Vapnik, The nature of statistical learning theory, IEEE Transactions on Neural Networks 8(6)(1997) 1564-1564.

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

[15] L.-Y. Wang, L. Li, Z.-H. Zhang, Accurate descriptions of hot flow behaviors across β transus of Ti-6Al-4V alloy by intelligence algorithm GA-SVR, Journal of Materials Engineering and Performance 25(9)(2016) 3912-3923.

[16] H. Wang, E. Li, G.-Y. Li, The least square support vector regression coupled with parallel sampling scheme metamodeling technique and application in sheet forming optimization, Materials and Design 30(5)(2009) 1468-1479.

[17] Y. Shuai, T.-L. Song, J.-P. Wang, H. Shen, W.-B. Zhan, A integrated IFCM-MPSO-SVM model for forecasting equipment support capability, Journal of Computers 28(1)(2017) 233-245.

[18] C. Xuan, F. Dan, L. Jing, Computational intelligence for corrosion rate prediction of refinery cooling water plant, Journal of Computers 29(3)(2018) 1-11.

[19] L.-X. Wei, X.-Y. Wei, H. Sun, R. Fan, B. Feng, Rolling force prediction of SVM based on improved genetic algorithm, in: Proc. 2018 37th Chinese Control Conference, 2018.

[20] D.-S. Wu, Q. Yang, D.-Z. Wang, Rolling force prediction based on PSO optimized support vector regression, in: Proc. 2011 7th International Conference on Natural Computation, 2011.

[21] Z.-M. Chen, F. Luo, X.-H. Huang, Y.-G. Xu, Rolling force prediction based on chaotic optimized support vector machine, Control and Decision, 24(6)(2009) 808-812.

[22] X. Wu, W.-M. Tang, X. Wu, Support vector machine based on hybrid kernel function, Lecture Notes in Electrical Engineering 154(2012) 127-133.

[23] S.-F. Ding, Y.-N. Zhang, X.-Z. Xu, L.-N. Bao, A novel extreme learning machine based on hybrid kernel function, Journal of Computers 8(8)(2013) 2110-2117.

[24] X.-Y. Jiang, S. Li, Bas: beetle antennae search algorithm for optimization problems, International Journal of Robotics and Control 1(1)(2018) 1-5.

[25] Y. Liu, Y.-K. Shi, M.-W. Xu, A convenient classification system for face orientations recognition based on support vector machine, Journal of Computers 30(5)(2019) 88-97.

[26] V.N. Vapnik, Statistical learning theory, John Wiley & Son, New York, 1998.

[27] X.-G. Gu, W. Wang, L. Xia, P. Jiang, A system optimisation design approach to vehicle structure under frontal impact based on SVR of optimised hybrid kernel function, International Journal of Crashworthiness (2019) 1-15.

[28] Y. Tan, J. Wang, A support vector machine with a hybrid kernel and minimal vapnik-chervonenkis dimension, IEEE Transactions on Knowledge and Data Engineering 16(4)(2004) 385-395.

[29] R. Eberhart, J. Kennedy, A new optimizer using particle swarm theory, in: Proc. 1995 6th International Symposium on Micro Machine and Human Science, 1995.