Improved Differential Gray Wolf Algorithm Optimized Support Vector Regression Strip Thickness Prediction Method

Lijie Sun, Jing Li, Xuedong Xiao, Li Zhang, and Jianhua Li

***Abstract*—Strip thickness prediction has important contribution to solve the problems of accurate strip thickness control and saving raw materials. Support vector regression (SVR) is presented to apply for strip thickness prediction, and the key problem to be solved is determining parameters of SVR. A strip thickness prediction method is proposed based on SVR optimized by improved differential gray wolf algorithm (denoted as HGWO-SVR). Firstly, the feature of the strip data is extracted by mutual information calculation method. Next, the differential evolution algorithm is introduced to enrich the diversity of the gray wolf population in gray wolf optimizer (GWO) and avoid falling into the local optimum, and coefficient vector improved gray wolf optimizer (HGWO) is used to balance the ability of global search and local search. Then, HGWO is used to select optimal kernel coefficient σ and penalty factor C in SVR model. Finally, establish HGWO-SVR model and input the characteristics of strip data into the model to predict the strip thickness. The results state clearly that HGWO-SVR has better predictive performance than GWO-SVR and SVR.**

***Index Terms*—Thickness prediction, GWO, SVR, mutual information, differential evolution**

I. Introduction

WITH the products that the industry provides for people have improved in both performance and function, a wide range of industrial applications for steel products also put forward higher quality requirements. Strip thickness prediction has important contribution to solve the problems of accurate strip thickness control and saving raw materials. However, the rolling process has complex nonlinear characteristics, and the interaction between a series of process parameters makes it a challenging task to realize the accurate prediction of strip thickness. Therefore, it is very important to establish a good prediction model[1]-[3].

Many scholars at home and abroad combined traditional equipment with modern scientific and technological means to bid farewell to traditional production methods, and applied machine learning theory to the rolling industry. In the field of machine learning, for nonlinear regression problems, SVR provides a variety of kernel function options, which has strong generalization ability and is suitable for small sample prediction. Therefore, it is widely applied in many research fields, such as pattern recognition, fault diagnosis, regression prediction and text classification, and has achieved great results and good performance[4]-[7]. For example, in 2018, Iranian scholar Alireza Baghban et al used the least squares support vector regression (LS-SVR) to estimate solubility temperature, pressure and concentration range of ammonia in various ionic liquids. Compared with the traditional support vector regression (SVM) method, LS-SVM reduced the complexity for the optimization problem[8]; in the same year, Liu Hong, from South China University of Technology, combined SVM with clustering method and applied it to short-term load forecasting. In this method, the training set input to the SVM was clustered, and samples with high similarity were selected as the input value of the model, and then it was trained to construct a prediction model. The experimental results showed that the method significantly improved prediction accuracy and shortened prediction time[9]; in addition, in 2019, Malaysian scholar Ibrahim Olanrewaju Aladea et al used genetic algorithms to optimize the regularization parameters, loss function variables and kernel parameters of SVR, and proposed a GA-SVR model with smaller error and higher prediction accuracy[10].

In summary, SVR model may have good prediction performance, however it is difficult and critical to select proper kernel function and penalty factor in the SVR model, which directly affects the prediction performance. In 2014, Griffith University scholar Mirjalili et al proposed Gray Wolf Optimizer (GWO) by imitating the behavior of wolves in nature[11]. In terms of solving model parameter selection, compared with other similar algorithms, such as particle swarm optimization (PSO) [12], back propagation (BP) [13], genetic algorithm (GA) [14], GWO has faster convergence speed and higher optimization accuracy.

Considering the complex and non-linear nature of the strip thickness prediction and the existing research of domestic and foreign scholars, HGWO-SVR method is proposed in this paper, in which mutual information is adopt to extract original strip steel data, HGWO is used to determine the optimal kernel coefficient σ and penalty factor C in SVR model and HGWO-SVR model is built.

Manuscript received October 20, 2020; revised September 17, 2021. This work was supported by Research Project of Education Department of Zhejiang Province(Y201840245).

L. J. Sun is the corresponding author, a lecturer of Taizhou University, member of CCF, Taizhou, 317000, China (phone: 15757638496, e-mail: LIJIE198855@126.com)

J. Li is a master degree candidate from the School of Science, Liaoning University, Shenyang,110036, China (e-mail: 670110864@qq.com)

X. D. Xiao is a master degree candidate from the School of Science, Liaoning University, Shenyang, 110036, China (e-mail: 2685081678@qq.com)

L. Zhang is a professor and master tutor of Liaoning University, Shenyang, 110036, China (e-mail: zhang_li@lnu.edu.cn)

J. H. Li is a lecturer of Taizhou University, Taizhou, 317000, China (e-mail: 7422721@qq.com)

Experimental results clearly show that HGWO-SVR has better prediction performance than GWO-SVR and SVR.

II. Related Theoretical Basis

A. Support Vector Regression

SVR can solve the regression prediction problem by introducing an insensitive loss function, given a sample set $\Omega=\{(x_i,y_i)\,|\,i=1,2,\cdots,n\}$, Where x_i is the input column vector. The samples are divided into training data and test data, in which the training set data can be fitted to the linear model as much as possible and the test set is input into the model $f(x)$ to obtain the prediction result. Comparing the predicted result $f(x)$ with the actual result y , if they are the same, the loss is counted as zero, otherwise, the maximum acceptable error of the SVR is ε . When the absolute value of the difference between $f(x)$ and y is greater than ε , calculate the loss between $f(x)$ and y [15][16].

Fig.1 is a graphical illustration of the loss function, where each circle represents a piece of data, and the data loss inside the gray shade is zero, as shown in 1, 3, 5 and 7 in the figure; the loss of these data points is zero, and the loss of these data points 2, 4, and 6 is their length to gray zone.

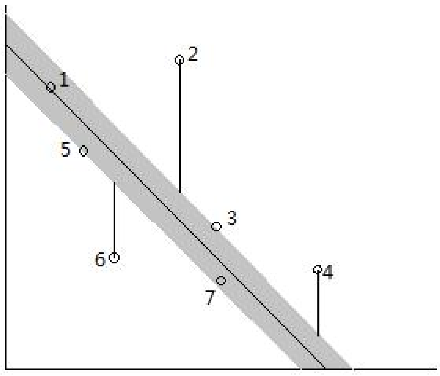

Fig. 1. Graphical illustration of loss function.

The linear regression function of SVR in the high-dimensional feature space is expressed as Eq. (1) :

$$f(x)=w^T x+b \tag{1}$$

Where w is the weight and b is the threshold.

By transforming and introducing relaxation variables, the problem of finding the hyperplane in the region is equivalent to finding the minimum value of Eq. (2):

$$\begin{cases} \min \dfrac{1}{2}\|w\|^2+C\sum_{i=1}^{n}(\xi_i+\xi_i^*) \\ s.t.\begin{cases} -\varepsilon-\xi_i \le y_i-wx_i-b \le \varepsilon+\xi_i^* \\ \xi_i \ge 0, \xi_i^* \ge 0 (i=1,2,...,n) \end{cases} \end{cases} \tag{2}$$

Where C is the penalty factor, ξ_i and ξ_i^* are relaxation variables, and ε is the loss function.

Due to the special structure of Eq. (2), through Lagrangian duality and introducing a kernel function, it is transformed into an optimization problem of dual variables and expressed as Eq. (3):

$$\begin{cases} \max\limits_{a,a^*}\left[-\dfrac{1}{2}\sum_{i,j=1}^{n}(a_i-a_i^*)(a_j-a_j^*)K(x_i,x_j)-\varepsilon\sum_{i=1}^{n}(a_i+a_i^*)+\sum_{i=1}^{n}y_i(a_i-a_i^*)\right] \\ s.t.\begin{cases} \sum_{i=1}^{n}(a_i-a_i^*)=0 \\ a_i,a_i^* \in [0,C] \end{cases} \end{cases} \tag{3}$$

Where a_i, a_i^*, a_j, a_j^* are Lagrange multipliers, $K(x_i,x_j)=\theta(x_i)\theta(x_j)$ is a kernel function, which is mapped to a high-dimensional feature space. Then solve the quadratic convex optimization problem to obtain SVR objective function, as shown in Eq. (4):

$$f(x)=\sum_{i=1}^{k}\left(a_i-a_i^*\right)K\left(x_i,x_j\right)+b \tag{4}$$

There are many kernel functions to choose from SVR model for $K(x_i,x_j)$. Gaussian kernel function is adopt as the kernel function in this paper. The formula is expressed as Eq. (5):

$$K\left(x_i,x_j\right)=\exp\left(\frac{-\left\|x_i-x_j\right\|^2}{2\sigma^2}\right) \tag{5}$$

Where σ is the kernel function parameter.

B. Gray Wolf Optimizer

The basic idea of GWO is to simulate the process of wolves capturing prey. In nature, they will not let all wolves go out to look for prey when wolves look for prey, but let a small part of the elite detective wolf searches for prey. The detective wolf judges the distance between itself and the prey according to the intensity of the smell left by the prey. If the prey is found, the detective wolf will summon the surrounding wolf to siege the prey, and the surrounding wolf receives the information from detective wolf, heading towards the detective wolf and rounding up the prey.

The GWO is divided into multiple steps of searching, tracking, surrounding, hunting, and attacking according to the behavior of the wolf pack hunting. The main steps are as follows:

Surrounding: The detective wolf determines the length of the distance between itself and the prey through the intensity of the smell left by the prey. During the search process.

The distance formula between the prey and the gray wolf is expressed as Eq. (6):

$$D=|C\cdot X_P(t)-X(t)| \tag{6}$$

Where t represents the current iteration number, $X_P(t)$ represents the location of the prey, $X(t)$ indicates the position of the gray wolf, C is the coefficient constant and the expression is shown in Eq. (7) :

$$C=2r_1 \tag{7}$$

Where r_1 is the random number between [0,1].

The gray wolf location is updated according to Eq.(8):

$$X(t+1)=X_p(t)-A\cdot D \tag{8}$$

Where A means the coefficient vector, and the expression of A is shown in Eq. (9):

$$A = 2ar_2 - a \tag{9}$$

$$a = 2 - 2\frac{t}{MaxIter} \tag{10}$$

Where r_2 is random number between [0,1], a is the convergence factor, and its expression is shown in Eq. (10), t is the current number of iterations, $MaxIter$ is the maximum number of iterations, and a continuously increases with linear decreasing of the iteration number t from 2 to 0 .

Hunting: First of all, the wolf pack is divided into four levels according to the ability of the self-adaptive, and the order is α wolf, β wolf, δ wolf, and ω wolf. Among them, the most self-adaptive is α wolf, next β wolf, then δ wolf, the last one ω wolf. Assuming that the α wolf, the β wolf and the δ wolf can find the potential position of the prey and guide ω wolf hunting by updating ω wolf's location with the position of three wolves.

$$D_\alpha = |C_1 \cdot X_\alpha(t) - X(t)| \tag{11}$$

$$D_\beta = |C_2 \cdot X_\beta(t) - X(t)| \tag{12}$$

$$D_\delta = |C_3 \cdot X_\delta(t) - X(t)| \tag{13}$$

$$X_1 = X_\alpha - A_1 \cdot D_\alpha \tag{14}$$

$$X_2 = X_\beta - A_2 \cdot D_\beta \tag{15}$$

$$X_3 = X_\delta - A_3 \cdot D_\delta \tag{16}$$

$$X_P(t+1) = \frac{X_1 + X_2 + X_3}{3} \tag{17}$$

In Eq. (11)~Eq. (16): $X_\alpha(t)$, $X_\beta(t)$, $X_\delta(t)$ represent the position of α wolf, β wolf and δ wolf when the number of iterations is t ; D_α , D_β and D_δ respectively represent the distance between α wolf, β wolf, δ wolf and gray wolf individual; A_1 , A_2 , A_3 and C_1 , C_2 , C_3 are all cooperative vectors. Eq.(17) is the position formula where the gray wolf individual moves next.

As the number of iterations increases, the location of the gray wolf is constantly updated, and the potential position of the prey is slowly predicted by α wolf, β wolf and δ wolf, so as to guide ω wolf to round up. When the maximum number of iterations is reached, the position X_α of the α wolf is the optimal solution.

C. Mutual Information Calculation

Mutual information refers to the correlation degree between two variables, which is the method to measure the internal dependence between the variable X and the variable Y.

Given two variables X and Y, the mutual information is calculated by Eq. (18):

$$I(X;Y) = \iint \mu_{XY}(x,y) \log \frac{\mu_{XY}(x,y)}{\mu_X(x)\mu_Y(y)} dxdy \tag{18}$$

Where $\mu_{XY}(x,y)$ means the joint probability density function of X and Y, $\mu_X(x)$ and $\mu_Y(y)$ mean the edge probability density function of X and Y respectively.

The size of mutual information value reflects the correlation strength of variables X and Y, and the two are positively correlated. When the mutual information value is zero, it means that X and Y are completely independent variables, and there is no intersection between them. Due to the characteristics of mutual information, it is widely applied in the field of machine learning and is the standard to judge the correlation of features[17][18].

D. Normalization Method

Because the unit dimension of each feature is different, it cannot be directly brought into the model for calculation, so it is necessary to standardize the data[19][20]. There are many common normalization methods, such as $\min-\max$ standardization, normalization methods, log function conversion and so on. The $\min-\max$ standardization method is adopted to process data in this paper, which is expressed as Eq.(19):

$$X^* = \frac{X - \min}{\max - \min} \tag{19}$$

Where X means the original value, X^* means the normalized value, $\min$ means the minimum value in the sample set X, and $\max$ means the maximum value in the sample set X.

III. Strip Thickness Prediction Model Based On HGWO-SVR

A. The Basic Idea of HGWO

Differential evolution (DE) is a heuristic global search algorithm simulating biological evolution[21]-[23]. Through continuous iteration, the individuals with strong adaptability in the population survive. In this paper, DE is used to retain the high-quality population to ensure the population diversity of GWO algorithm. DE mainly has three operations: mutation, crossover, and selection.

Mutation: The difference strategy is used to mutate the initial population. Two different individuals are randomly selected in the initial population, and the position vectors of the two individuals are made difference. The scaling factor is used to adjust the vector difference, and then the vectors are combined with the individuals to be mutated. The mutation process is expressed as Eq.(20):

$$D_i(T+1) = X_{r1}(T) + F \bullet (X_{r2}(T) - X_{r3}(T)) \tag{20}$$

Where F is the scaling factor between [0, 2], T is the current number of iterations, X_{ri} means the individual in the population, and $D_i(T+1)$ means the variant individual.

Crossover: The offspring population is obtained by cross operation between the mutant and the parent, and the formula of crossover operation on the $j-th$ dimension of the i-th individual is shown in Eq. (21):

$$U_{i,j}(T+1) = \begin{cases} D_{i,j}(T+1), r < CR \,||\, j = D_n \\ X_{i,j}(T) \end{cases} \tag{21}$$

Where CR is the crossover probability, which is a random value between [0,1], and D_n is the random dimension.

Selection: In the selection operation, DE uses greedy algorithm. Compare the adaptability of the offspring population and the parent population, and retain the highly adaptive population. The specific selection operation formula is shown in Eq.(22):

$$X_i(T+1)=\begin{cases}U_i(T+1), f(U_i(T+1))\le f(X_i(T))\\ X_i(T)\end{cases} \tag{22}$$

The global and local optimization ability of GWO is affected by the parameter C. The original expression of the parameter C is shown in Eq. (7), and it is random number between [0, 1]. When $C<1$, the difficulty of gray wolf approaching the prey is reduced, making it easier for gray wolf to find the prey; when $C>1$, the difficulty of gray wolf approaching the prey is increased, making it more difficult for gray wolf individuals to approach the prey. Through analyzing the optimization process of GWO, it is concluded that GWO has good population diversity in the early stage of iteration and should accelerate convergence, while in the late stage of iteration, its local accurate search ability should be enhanced. Therefore, the parameter C is improved and expressed as Eq. (23):

$$C=r_3+\frac{t}{MaxIter} \tag{23}$$

Where r_3 is a random number between [0,1], $MaxIter$ is the maximum number of iterations and t is the current number of iterations.

The value of the improved parameter C is still between [0, 1], but in the early stage of algorithm iteration, the probability of C value less than 1 is larger, which is conducive to the rapid convergence of GWO algorithm. As the number of iterations increases, the probability of C value greater than 1 is larger, and gray wolf has strong ability of accurate local search.

B. Basic Flow of HGWO

The basic process of HGWO is as follows:

Step 1. Set the parameter values, including the number of population, crossover probability, the maximum number of iterations, spatial dimension, scaling factor, the upper and lower limits of variables and other related parameters.

Step 2. Initialize the value of relevant parameters such as A, C and a, randomly generate the initial population.

Step 3. Calculate and sort the fitness values for all the gray wolves, then select the three wolves with the best fitness and mark them as α wolf, β wolf and δ wolf to record and save the positions of the three wolves;

Step 4. Update the positions of other gray wolf individuals except α wolf, β wolf and δ wolf according to Eq. (11) - Eq. (16).

Step 5. Respectively update the values of C, A and a according to Eq. (23), Eq. (9) and Eq. (10). Use Eq. (21) for cross processing to keep individuals with better fitness levels and carry out selection operations.

Step 6. Recalculate the fitness value for the population and update the fitness values and their corresponding positions of α wolf, β wolf, δ wolf as the next generation to optimized continuously. If reach the maximum number of iterations, HGWO stops and output the position X_α of the gray wolf with the best fitness value, otherwise return to step 3.

The operation flow of HGWO algorithm is shown in Fig. 2.

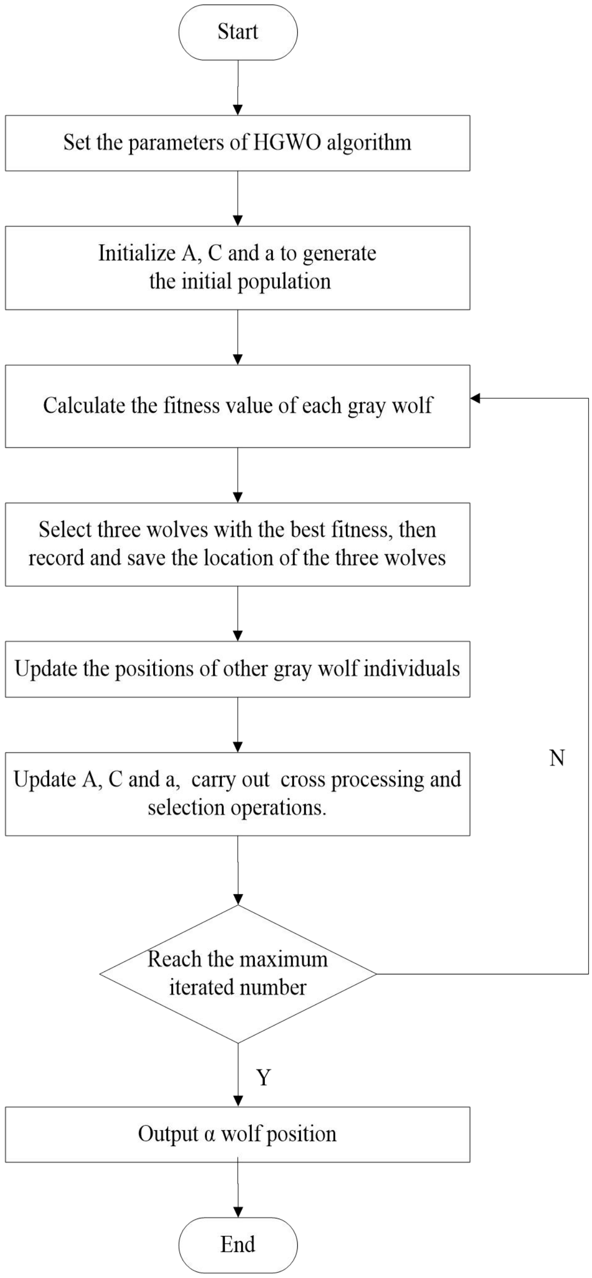

Fig. 2. Basic flowchart of HGWO.

C. Performance Analysis of HGWO

The benchmark function can check the algorithm performance. Four common benchmark functions including Quartic function, Sphere function, Ackley function and Rastrigin function are selected to test the performance of PSO, GWO and HGWO and their performance are compared in this paper.

The test dimension is 30, the population size is 30, the number of iterations is 1000, the crossover probability is 0.2, the lower bound is 0.2 and the upper bound is 0.8 for the scaling factor. The minimum values of these test functions are all 0, therefore, in the process of optimization, the smaller the optimization result, the better the convergence effect and the higher the accuracy of the algorithm. The algorithm fitness curves are as shown Fig.3 - Fig.6 under different test functions.

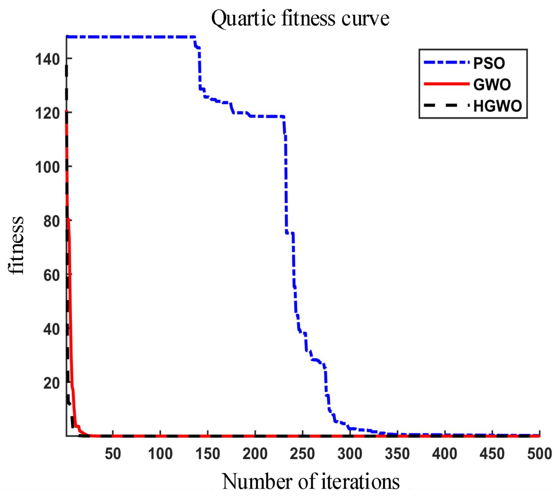

Fig. 3. Quartic function fitness change curve.

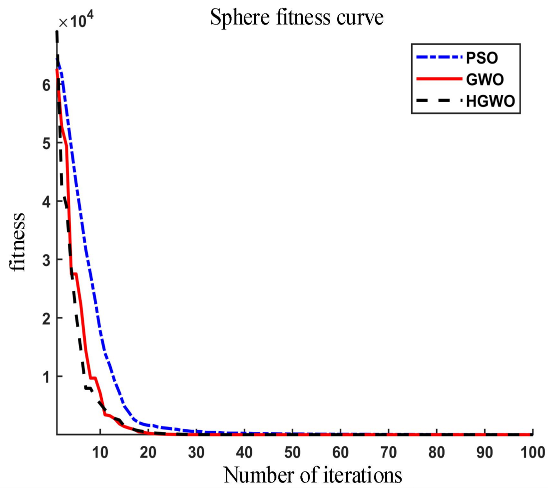

Fig. 4. Sphere function fitness change curve.

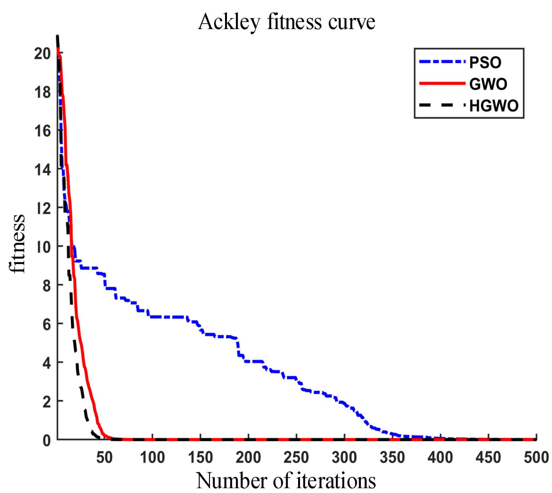

Fig. 5. Ackley function fitness change curve.

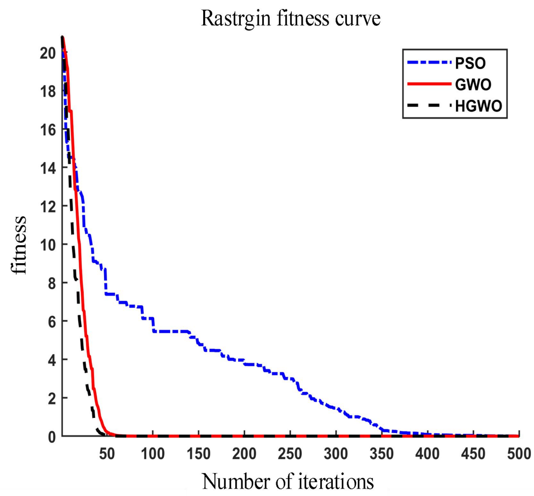

Fig. 6. Rastrigin function fitness change curve.

According to the fitness change graphs of four different test functions, under the simple unimodal function such as function, there is a small gap between PSO algorithm, GWO algorithm and HGWO algorithm in searching ability and convergence speed.Under several other test functions, it can be clearly seen that the convergence speed of HGWO is much better than that of the PSO, and has been improved compared with GWO. Therefore, it can be seen from the four test functions that HGWO performs best. The optimization results under each test function are as shown TABLE I.

TABLE Ⅰ
TEST FUNCTION OPTIMIZATION RESULT

Test function	algorithm	Best value	Worst value	average value	variance
	POS	0.000570625	0.148057443	0.082205427	0.034724416
Quartic	GWO	0.000302109	0.002116963	0.000947287	0.000533695
	HGWO	**0.000045426**	**0.000570625**	**0.000291798**	**0.000149382**
	POS	2.82726E-11	1.24847E-07	1.38688E-08	3.71472E-08
Sphere	GWO	6.47177E-11	9.811E-11	8.62407E-11	1.06333E-11
	HGWO	**6.71938E-73**	**2.54138E-68**	**3.69295E-69**	**7.36822E-69**
	POS	7.16435E-06	0.000156676	5.40079E-05	5.0759E-05
Ackley	GWO	8.08216E-11	9.97824E-11	9.20517E-11	5.94141E-12
	HGWO	**7.99361E-15**	**1.5099E-14**	**1.04805E-14**	**3.19744E-15**
	POS	24.8740971	70.64197896	43.99826272	14.46072777
Rastrigin	GWO	7.33849E-11	7.458364391	1.361755789	2.738961037
	HGWO	**4.2982E-14**	**7.99361E-14**	**5.87673E-14**	**1.16763E-14**

It can be seen that HGWO has superior performance under each test function in TABLE Ⅰ. The average value and variance of each algorithm under four different test functions show that HGWO is better than PSO and GWO in terms of optimization ability and stability.

D. HGWO-SVR Model Establishment

The performance of SVR is affected by the kernel function and penalty factor. If the values of the kernel function parameters and penalty factors are not appropriate, it will directly affect the generalization ability of the model, resulting in the failure to obtain the expected prediction results. The HGWO is applied to optimize the kernel function parameter and penalty factor of SVR. The kernel function and penalty factor are mapped to the optimal wolf position, and the optimal solution position is used as their values in SVR. The construction process of HGWO-SVR is as follows:

Step 1. Set the parameter values, including the number of population, crossover probability, the maximum number of iterations, spatial dimension, scaling factor, the upper and lower limits of variables and other related parameters.

Step 2. Initialize the related parameters such as A, C and a, randomly generate parent, mutation and offspring gray wolf population, and set the number of iterations to 1.

Step 3. Use the processed training samples to train SVR.

Step 4. Calculate and sort the objective function value of each gray wolf individual, select three wolves with the best fitness and set them as α wolf, β wolf and δ wolf, record and save their positions.

Step 5. Update the positions of other gray wolf individuals except α wolf, β wolf and δ wolf according to Eq. (11) ~ Eq. (16).

Step 6. Update the values of C, A and a respectively according to Eq. (23), Eq. (9) and Eq. (10). Use Eq. (20) to perform the mutation operation on the gray wolf individual in the DE to generate intermediate individuals. Carry out cross-processing according to Eq. (21) to retain the better gray wolf individuals in the population and perform selection operations according to Eq. (22).

Step 7. Recalculate the objective function value of each individual gray wolf and update the fitness value and corresponding position of α wolf, β wolf and δ wolf.

Step 8. Judge whether the maximum number of iterations is reached. If it is reached, HGWO stops and outputs the global optimal solution, namely the position X_α of α wolf, otherwise HGWO returns to step 3.

Step 9. Use the position X_α as the values for the penalty factor and kernel function in SVR, then the HGWO-SVR is constructed.

IV. Experimental Results And Analysis

A. Experimental Data Collection

This paper selects the actual data generated in the finishing rolling zone during the hot rolling process of a domestic steel factory, which consists of nine racks. For the strip thickness prediction, there are many influence factors, including rolling force, rolling speed , mill current, and rolling temperature and so on. The ibaAnalyzer is adopt to read and analyze data. The part of the rolling signals from the ninth rack are as shown Fig.7, in which the ordinate are respectively strip thickness (ABS_H/mm), rolling Force (F9_F/KN), rolling gap (F9_GAP/mm), rolling speed (F9_SPD_ACT/m·s^{-1}), Mill current (F9_CURRENT_ACT/A), rolling temperature (TEMP_OUT_FM/°C) , and the abscissa is time series.

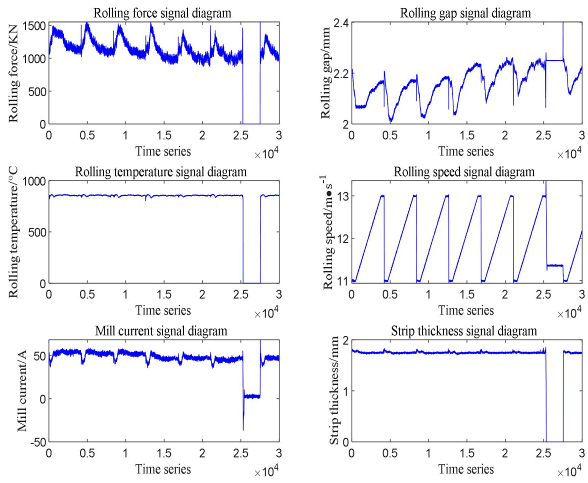

Fig. 7. The part of the rolling signals

B. Experimental Data Processing

There are many rolling parameters that affect the strip thickness, but it will reduce the generalization ability and increase the prediction time for the model that all parameters are used as inputs to the model. Therefore the mutual information method is adopt to select the feature set as the input of the model in this paper. The specific operations are as follows:

Step 1. According to Eq.(18), calculate the mutual information value between influence factors and the strip thickness. The results are shown as TABLE II.

TABLE II
MUTUAL INFORMATION VALUE OF CHARACTERISTIC PARAMETER AND THICKNESS

Parameters	Mill current	Rolling speed	Rolling gap	Rolling force	Rolling temperature	SONY value
I	0.7058	0.6131	0.5836	0.2524	0.1527	0.1029

Step 2. According to the principle of mutual information feature extraction, namely $N=6$, then $z=1/N=0.1667$.

Step 3. The parameters with larger mutual information value are selected, that is, mill current, rolling speed, rolling gap and rolling force are selected as a group of main factors to input into the model.

Due to not uniform of each feature unit dimension, it is necessary to normalize data. $\min-\max$ normalization processing method shown in Eq.(19) is used to scale the data set to the uniform range of [0,1]. TABLE III is the partial data before normalization, and TABLE IV is the partial data after normalization.

TABLE III
DATA BEFORE NORMALIZATION

Mill current	Rolling speed	Rolling gap	Rolling force	Strip thickness
47.83	10.98	2.09082	1368.82	1.77440
48.05	10.98	2.09007	1354.95	1.77348
47.58	10.99	2.09082	1346.41	1.77333
47.27	10.99	2.09149	1340.10	1.77125
46.67	10.99	2.09207	1345.62	1.77001
47.14	10.99	2.09382	1386.12	1.76689
47.52	10.98	2.09382	1359.04	1.76521
……	……	……	……	……

TABLE IV
NORMALIZED DATA

Mill current	Rolling speed	Rolling gap	Rolling force	Strip thickness
0.14625	0	0.88636	0.29080	0.60765
0.12536	0	0.90909	0.36131	0.61901
0.10161	0.01562	0.90909	0.24738	0.60580
0.07281	0.01562	0.92939	0.21530	0.58012
0.01519	10.99	2.09207	1345.62	1.77001
0.05982	10.99	2.09382	1386.12	1.76689
0.09591	10.98	2.09382	1359.04	1.76521
……	……	……	……	……

C. Analysis of Results

In this paper, HGWO-SVR is compared with GWO-SVR and SVR through the simulation experiments to verify the effectiveness of the proposed method. The experimental data include 1000 groups of training data and 50 groups of test data. The population size is 15, the maximum iteration number is 100, the dimension is 2, the crossover probability is 0.2, the lower limit of the scaling factor is 0.2, the boundary is 0.8. The root mean square error is used as a prediction performance evaluation index for HGWO-SVR, GWO-SVR and SVR. The prediction results are shown as TABLE V.

TABLE V
MODEL EXPERIMENT COMPARISON RESULTS

model	REMS
HGWO-SVR model	**0.0036685**
GWO-SVR model	0.0042379
SVR model	0.0078218

The results in TABLE V show that of HGWO-SVR has good prediction performance with smaller deviation and higher accuracy than GWO-SVR and SVR. For clearer observation of prediction effect under the three models, Fig.8-Fig.10 are HGWO-SVR prediction curve, GWO-SVR prediction curve, and SVR prediction curve respectively.

In GB/T 709-2019, the strip thickness is specified to be between 1.50mm and 2.00mm, and the strip thickness tolerance is ±0.19mm. According to the following figure, it can be seen intuitively that the deviations of HGWO-SVR is less than 0.19mm, so the prediction performance of HGWO-SVR model can meet the actual needs of users and reach the time standard of rolling strip thickness.

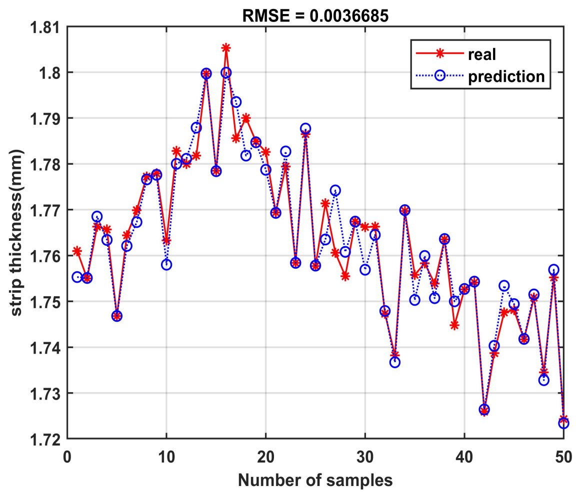

Fig. 8. HGWO-SVR prediction curve.

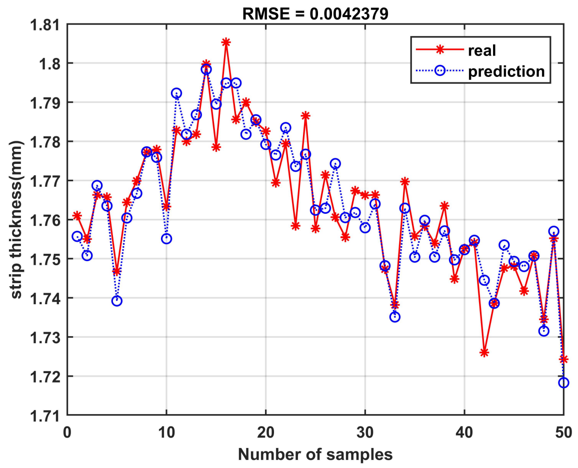

Fig. 9. GWO-SVR prediction curve.

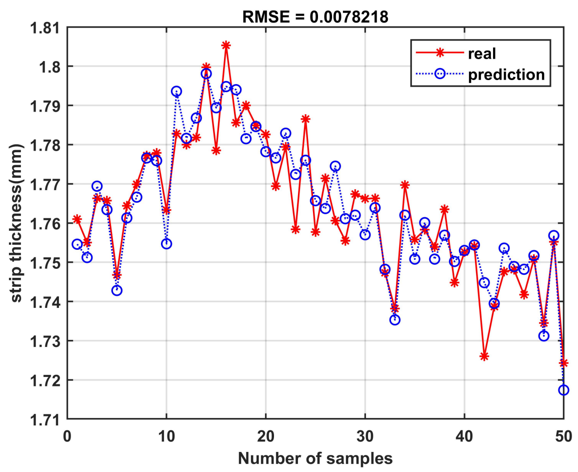

Fig. 10. SVR prediction curve.

V. CONCLUSION

The prediction performance of SVR is affected by the parameters of penalty factor and kernel function. If they are selected inappropriately, it can not achieve the expected prediction effect and there will be a big deviation in the prediction process of the data to be tested, which will cause serious economic losses to the steel industry in practical applications. In this paper, HGWO-SVR model can effectively improve the prediction accuracy of strip thickness by comparing with SVR model and GWO-SVR model, which can improve the strip quality and increase the benefits of the steel industry. However, further research is needed to apply HGWO-SVR model to strip thickness prediction in the steel industry.

REFERENCES

[1] Dema, R. R., et al. "Mathematical Model for Assessing the Management of Quality Parameters of Hot-Rolled Strips According to the Criterion of Local Thickness Variation." *Materials Today: Proceedings*, vol. 19, 2019, pp. 2417-2421.

[2] Rezaii, A., et al. "Experimental & Theoretical Investigation of Roll Bonding Process of Multilayer Strips by Finite Element Method." *Journal of Manufacturing Processes*, vol. 54, 2020, pp. 54-69.

[3] Hu, Y. J., et al. "Characteristic analysis and optimal control of the thickness and tension system on tandem cold rolling." *The International Journal of Advanced Manufacturing Technology*, vol. 101, 2019, pp.2297-2312.

[4] Wang, R., et al. "Automatic Blur Type Classification via Ensemble SVM." *Signal Processing-Image Communication*, vol. 71, 2019, pp. 24-35.

[5] Muthukumar, A., and A. Kavipriya. “A Biometric System Based on Gabor Feature Extraction with SVM Classifier for Finger-Knuckle-Print.” *Pattern Recognition Letters*, vol. 125, 2019, pp. 150-156.
[6] Shankar, K., et al. “Optimal Feature-Based Multi-Kernel SVM Approach for Thyroid Disease Classification.” *The Journal of Supercomputing*, vol. 76, no. 2, 2020, pp. 1128-1143.
[7] Yang, W. CH., and Si Y. W. “Research of Specific Field Ultra-Short Text Classification Based on Collaborative Filteringalgorithm.” *MATEC Web of Conferences*, vol. 189, 2018, pp. 3006.
[8] Baghban, A., et al. “Prediction of solubility of ammonia in liquid electrolytes using Least Square Support Vector Machines.”*Ain Shams Engineering Journal*, vol. 9,no. 4, 2018, pp. 1303-1312.
[9] Liu, H. “Research on iterative self-organizing data analysis algorithm in short-term power load forecast.” *South China University of Technology*, 2018.
[10] Alade, I. O., et al. “Modeling and Prediction of the Specific Heat Capacity of Al2 O3/Water Nanofluids Using Hybrid Genetic Algorithm/Support Vector Regression Model.” *Nano-Structures and Nano-Objects*, vol. 17, 2019, pp. 103-111.
[11] Liu, R. CH., et al. “A Dynamic Multiple Populations Particle Swarm Optimization Algorithm Based on Decomposition and Prediction.” *Applied Soft Computing*, vol. 73, 2018, pp. 434-459.
[12] Li SH. Y., et al. “An Optimal Resource Allocation Scheme for Elastic Applications in Multipath Networks via Particle Swarm Optimization.” *IAENG International Journal of Computer Science*, vol. 47, no. 2, 2020, pp. 278-283.
[13] Wu, L., et al. “Applying of GA-BP Neural Network in the Land Ecological Security Evaluation.” *IAENG International Journal of Computer Science*, vol. 47, no. 1, 2020, pp. 11-18.
[14] Islam T., et al. “A GA-Based Approach to Automatic Test Data Generation for ASP.NET Web Applications.” *IAENG International Journal of Computer Science*, vol. 47, no. 3, 2020, pp. 557-564.
[15] Lauren, J., et al. “Support vector regression.” *Machine Learning. Academic Press*, 2020, pp. 123-140.
[16] Sina, A., et al. “Short Term Load Forecasting Model Based on Kernel-Support Vector Regression with Social Spider Optimization Algorithm.” *Journal of Electrical Engineering & Technology*, vol. 15, no. 1, 2020, pp. 393-402.
[17] Bennasar, M., et al. “Feature selection using Joint Mutual Information Maximisation.” *Expert Systems with Applications*, vol. 42, no. 22, 2015, pp. 8520-8532.
[18] Wang, J. J., et al. “Multi-target feature selection method based on mutual information and particle swarm optimization.” *Computer Science and Exploration*, vol. 14, no. 1, 2020, pp. 83-95.
[19] Jain, S., et al. “Dynamic selection of normalization techniques using data complexity measures.” *Expert Systems with Application*, vol. 106, no. 1, 2018, pp. 252-262.
[20] Fritzler, M. J. “Autoantibody assays. performance, interpretation, and standardization.” *The Autoimmune Diseases. Academic Press*, 2020, pp. 1369-1389.
[21] Baioletti, M., et al. “Differential Evolution for Neural Networks Optimization.” *Mathematics*, vol. 8, no. 1, 2020, pp. 69.
[22] Choudhary, R., and Rahul G. “Recent Trends and Techniques in Image Enhancement Using Differential Evolution-A Survey.” *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 7, no. 4, 2017, pp. 106-112.
[23] Zhao, Y., and Wang G. SH. “A Dynamic Differential Evolution Algorithm for the Dynamic Single-Machine Scheduling Problem with Sequence-Dependent Setup Times.” *Journal of the Operational Research Society*, vol. 71, no. 2, 2020, pp. 225-236.

L. J. Sun was born in Liaoning, China. She received the B.Eng. degree from BOHAI University, China, in 2010, the M.Eng. degree from Liaoning University, China, in 2013, and the Ph.D. degree in control theory and control engineering from the Dalian University of Technology, China, in 2018. Since 2018, she has been with Taizhou University, where she is a lecturer. Her research interests include fault diagnosis and prediction control.

J. Li is a master degree candidate of computer application technology in Liaoning University, and her research interests include data analysis and data modeling.

X. D. Xiao is a master degree candidate of computer application technology in Liaoning University, and her research interests include data analysis and data modeling.

L. Zhang is a professor and master tutor of Liaoning University, and his Ph.D. is from Northeastern University, and his research interests include fault diagnosis and data modeling.

J. H. Li is a associate professor of Taizhou University, and his research interests is data mining.