ISSN 0543-5846
METABK 61(2) 325-328 (2022)
UDC / UDK 669.162.2:536.45/669.01:669.782:001.8:168.2=111

# PREDICTION OF SILICON CONTENT IN HOT METAL BASED ON GOLDEN SINE PARTICLE SWARM OPTIMIZATION AND RANDOM FOREST

Received – Primljeno: 2021-07-07
Accepted – Prihvaćeno: 2021-10-15
Original Scientific Paper – Izvorni znanstveni rad

Particle Swarm Optimization (PSO) algorithm quickly falls into local optimum, low precision. In this paper, add the golden sine operation to the particle position update. The results show that the improved PSO algorithm has better optimization ability. The main parameters affecting the silicon content in hot metal are selected. Then, calculate the correlation coefficient and significance level between parameters and silicon content in hot metal. Finally, the prediction model of silicon content in hot metal is established based on the Random Forest (RF) optimized by improved PSO. The results show that the hit rate is 87,17 %.

*Keywords:* blast furnace, hot metal, silicon, particle swarm optimization, golden sine algorithm, random forest

## INTRODUCTION

Furnace temperature significantly influences blast furnaces, and it isn't easy to measure by instrument. Silicon content in the hot metal can indirectly reflect furnace temperature. The prediction of silicon content can provide an essential reference for blast furnace operators. Researchers have established various prediction models, including neural network [1], support vector machine (SVM) [2] and RF [3]. The recurrent neural network with a long short-term memory structure predicts silicon content in [1]. The SVM optimized by an improved particle swarm optimizer predicts silicon content in [2]. The prediction model is RF optimized by the grid search in [3]. However, the hit rate is not high enough. The main reasons are parameter optimization and data set quality. Eberhart and Kennedy propose PSO [4] in 1995. Tanyildizi and Demir propose the Golden Sine Algorithm (Gold-SA) [5] in 2017. This paper proposes a Golden Sine Particle Swarm Optimization (GSPSO). Add the golden sine operation to the local part of PSO and improve the inertia weight to exponential decreasing. Experiments show that the search capability of GSPSO is enhanced. Firstly, found out the main parameters which affect the silicon content fluctuation. Then, calculate the correlation coefficient and significance level between parameters and silicon content in hot metal. In this way, form a high-quality data set according to the correlation coefficient. Finally, the RF optimized by GSPSO is used to predict the silicon content. The results show that the GSPSO-RF proposed in this paper has better performance than SVM and GSPSO-SVM.

Ch. Hu, K. Yang, School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China. Corresponding author: K. Yang (asyangkai@126.com)

## PARTICLE SWARM OPTIMIZATION

In PSO, $x_i = \left(x_i^1, x_i^2, \cdots, x_i^N\right)$ and $v_i = \left(v_i^1, v_i^2, \cdots, v_i^N\right)$ are the position and velocity of the particle. $p_i = \left(p_i^1, p_i^2, \cdots, p_i^N\right)$ and g = ($g^1$,$g^2$,...,$g^N$) are the optimal individual position and the optimal global position, respectively. Show the velocity and position update formulas with inertia weight in (1) and (2) [4].

$$v_i^d(t+1) = \omega v_i^d(t) + c_1 r_1\left(p_i^d(t) - x_i^d(t)\right) + c_2 r_2\left(g^d(t) - x_i^d(t)\right) \quad (1)$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \quad (2)$$

In the formula, i is the i-th particle; d is the d-th dimension; $c_1$ and $c_2$ are local and global learning factors; t is the iteration; $r_1$ and $r_2$ are random numbers in the range [0, 1]; ω is the inertia weight. Decreasing inertia weight is the basic strategy.

## GOLDEN SINE ALGORITHM

Gold-SA introduces golden section coefficients $x_1$ and $x_2$, which can narrow the search space. Show the formula of $x_1$ and $x_2$ in (3) and (4) [5].

$$x_1 = a * (1-t) + b * t \quad (3)$$

$$x_2 = a * t + b * (1-t) \quad (4)$$

$t = \left(\sqrt{5}-1\right)/2$ is the golden section ratio; The initial values of a and b are $-\pi$ and $\pi$; Update $x_1$ and $x_2$ with a and b. The update rules are as follows [5].

*if current_fitnnes < best_fitnnes*

$$b = x_2$$
$$x_2 = x_1$$
$$x_1 = a * t + b * (1-t)$$

*else*

$$a = x_1$$
$$x_1 = x_2$$
$$x_2 = a * (1-t) + b * t$$

*if* $x_1 == x_2$

$$a = random_1$$
$$b = random_2$$
$$x_1 = a * t + b * (1 - t)$$
$$x_2 = a * (1 - t) + b * t$$

$V_i(t) = (V_i^1, V_i^2, \cdots, V_i^N)$ is the position of the particle. $D(t) = (D^1, D^2, ..., D^N)$ is the optimal global position. Show the position update formula of Gold-SA in (5) [5].

$$V_i^d(t+1) = V_i^d(t) * |\sin(r_1)| - r_2 * \sin(r_1) * |x_1 * D^d(t) - x_2 * V_i^d(t)| \quad (5)$$

In the formula, $r_1$ and $r_2$ are random numbers, $r_1 \in [0, 2\pi]$, $r_2 \in [0, \pi]$.

## Golden sine particle swarm optimization

This paper changes the linear decreasing to exponential decreasing. Shown the decreasing formula in (6).

$$\omega = \exp(-t/T_{max}) \quad (6)$$

t is the current number of iteration. $T_{max}$ is the maximum number of iteration.

Calculate the new position $x(t+1)$ according to formulas (1) and (2), and calculate the fitness of the $x(t+1)$. If the fitness is better than the current optimal fitness, the optimal fitness is updated, go to the next iteration; Otherwise, calculate the new position $V(t+1)$ according to formula (5), and calculate the fitness of the $V(t+1)$. The optimal fitness is updated according to the new fitness quality and go to the next iteration. The detailed steps are as follows.

**Step 1:** Initialize the population. Calculate the fitness, and record the optimal fitness;

**Step 2:** If $t < t_{max}$, go to step 3; Otherwise, go to step 8;

**Step 3:** Update the velocity and position according to (1) and (2).

**Step 4:** Calculate the fitness of the new position;

**Step 5:** Check whether the fitness of the new position is better than the optimal fitness. If it is better, update the optimal fitness, and go to step 2; Otherwise, go to step 6.

**Step 6:** Update the position of particles according to (5);

Table 1 **Benchmark functions**

| Functions | Range | $f_{min}$ |
|---|---|---|
| $f_1(x) = \sum_{i=1}^{n} x_i^2$ | [-100, 100] | 0 |
| $f_2(x) = \sum_{i=1}^{n} \lvert x_i \rvert + \prod_{i=1}^{n} \lvert x_i \rvert$ | [-10,10] | 0 |
| $f_3(x) = \max_i \{\lvert x_i \rvert, 1 \le i \le n\}$ | [-100, 100] | 0 |
| $f_4(x) = \sum_{i=1}^{n} (x_i^2 - 10\cos(2x_i\pi) + 10)$ | [-5,12, 5,12] | 0 |
| $f_5(x) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos(\frac{x_i}{\sqrt{i}}) + 1$ | [-600, 600] | 0 |
| $f_6(x) = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + e$ | [-32, 32] | 0 |

Table 2 **Experimental results**

| | Algorithm | Best | Mean | Worst |
|---|---|---|---|---|
| $f_1$ | GWO | 5,25E-178 | 1,2296E-168 | 2,4436E-167 |
| | PSO | 1,27046E-21 | 3,21169E-15 | 5,99251E-14 |
| | GSPSO | 0 | 0 | 0 |
| $f_2$ | GWO | 6,53945E-99 | 1,94349E-94 | 2,23189E-93 |
| | PSO | 4,85973E-08 | 0,024391209 | 0,317523335 |
| | GSPSO | 2,6302E-191 | 1,8254E-184 | 2,0277E-183 |
| $f_3$ | GWO | 2,89404E-51 | 1,42659E-46 | 1,45609E-45 |
| | PSO | 0,000122538 | 0,070499468 | 0,498645697 |
| | GSPSO | 4,0674E-192 | 8,7421E-181 | 1,7453E-179 |
| $f_4$ | GWO | 0 | 0,418757575 | 5,346141264 |
| | PSO | 2,002504621 | 6,047133723 | 11,93950029 |
| | GSPSO | 0 | 0 | 0 |
| $f_5$ | GWO | 0 | 0,024481762 | 0,104378212 |
| | PSO | 0,039375088 | 0,168153111 | 0,539093075 |
| | GSPSO | 0 | 0 | 0 |
| $f_6$ | GWO | 3,9968E-15 | 4,17444E-15 | 7,54952E-15 |
| | PSO | 1,28342E-13 | 0,414134233 | 2,013315236 |
| | GSPSO | 4,44089E-16 | 4,44089E-16 | 4,44089E-16 |

**Step 7:** Calculate the fitness of the new position. The optimal fitness is updated according to the new fitness quality, and go to step 2.

**Step 8:** Output optimal position and fitness.

## BENCHMARK FUNCTIONS VALIDATION

This paper selects PSO, Grey Wolf Optimizer (GWO) [6], GSPSO for comparative experiments. The experiment uses Python language. The iterations are 1 000, the population size is 10, the dimension is 10, and it runs 20 times. This paper sets $c_1 = 0,2$, $c_2 = 0,3$, and the inertia weight ω=1. Show the specific information of benchmark functions in Table 1. Show the experimental results in Table 2. Show the convergence graphs in Figures 1 - 6.

It can be concluded from the results that GSPSO has higher search accuracy than PSO and GWO. The optimal value found is closer to the actual global optimal value. Only on $f_4$ and $f_5$, the best values found by GSPSO are the same as those found by GWO, but both are the actual optimal values. The worst values found by GSPSO are also better than those found by PSO and GWO. The results demonstrate the feasibility of the improved idea and the apparent advantages of the GSPSO.

## PARAMETER SELECTION

Direct reduction of iron is the most critical factor affecting silicon content in hot metal. The reduction of iron oxide with *CO* is a mainly exothermic reaction, which is called indirect reduction. The reduction of iron oxide with *C* is an endothermic reaction, which is called direct reduction. $H_2$ partly replaces *CO* in reduction and primarily replaces *C* [7]. Calculate the indirect reduction degree of *CO* by the content of $CO_2$. The content of $CO_2$ indicates the utilization rate of coal gas and the reduction of iron ore [8]. The gas flow condition has a powerful influence on the reduction reaction.

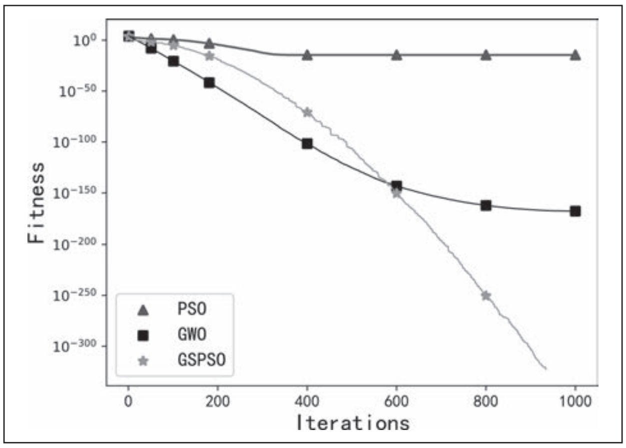


**Figure 1** Convergence graph of $f_1$

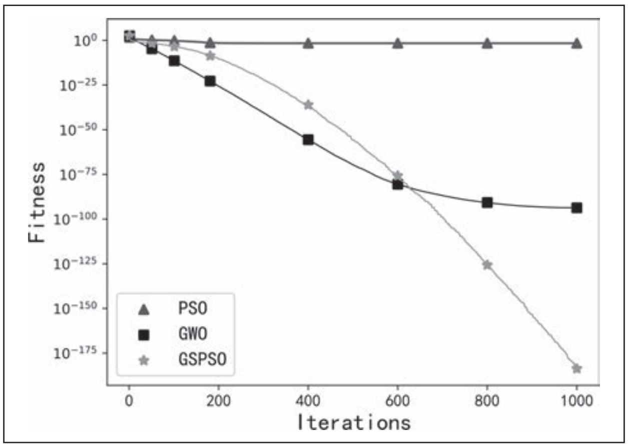


**Figure 2** Convergence graph of $f_2$

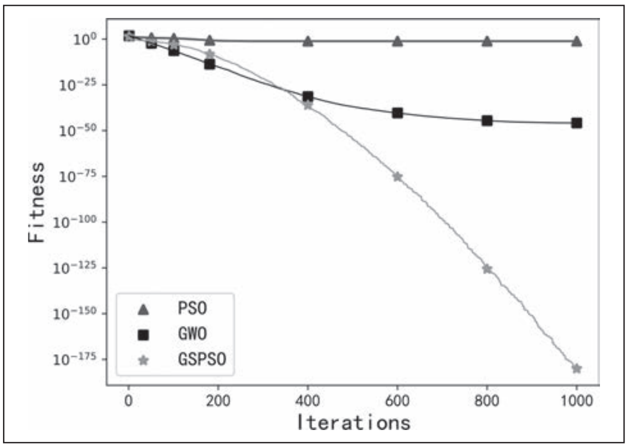


**Figure 3** Convergence graph of $f_3$

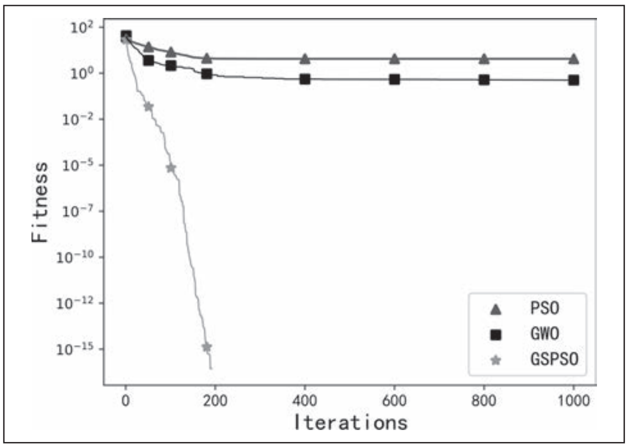


**Figure 4** Convergence graph of $f_4$

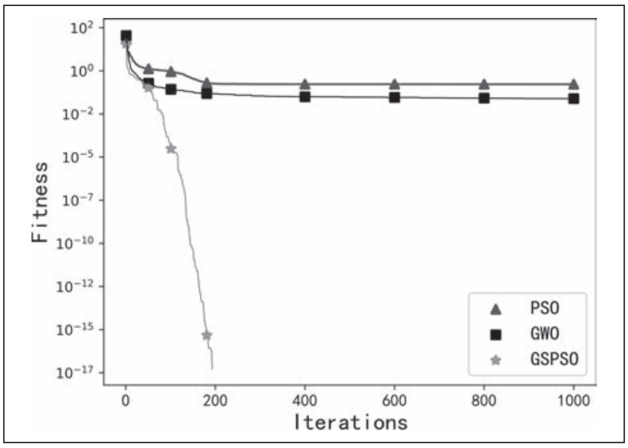


**Figure 5** Convergence graph of $f_5$

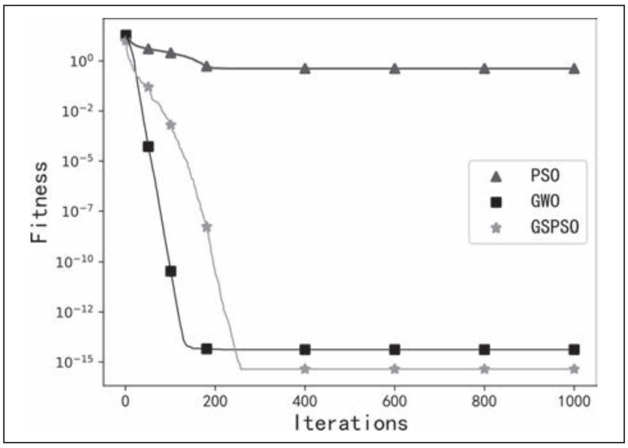


**Figure 6** Convergence graph of $f_6$

Based on the above analysis, the selected input parameters are fuel ratio, gas temperature, gas flow rate, gas pressure, gas utilization rate and gas composition. Show Pearson correlation coefficient and significance level between parameters and silicon content in Table 3.

## GSPSO-RF

The training subset is generated by random sampling with put back. Suppose X and Y are input and output variables. Select the j-th input variable $x^j$ and its value s as the partition. Define two sets, $R_1(j,s)=\{x|x^j \le s\}$ and $R_2(j,s)=\{x|x^j > s\}$. Calculate

$$\min_{j,s}\left[\min_{c_1}\sum_{x_i\in R_1(j,s)}(y_i-c_1)^2+\min_{c_2}\sum_{x_i\in R_2(j,s)}(y_i-c_2)^2\right]$$

Find out the best partition s. $c_1$ and $c_2$ are the averages of y in $R_1$ and $R_2$.

$$c_1 = average(y_i|x_i \in R_1(j,s))$$
$$c_2 = \mathrm{average}(y_i|x_i \in R_2(j,s))$$

Traverse all input variables to find the optimal j and s, and divide the training set into two subsets. The above process is carried out recursively for the two subsets until the stop condition is satisfied [9]. The optimization results show that the maximum depth of the decision

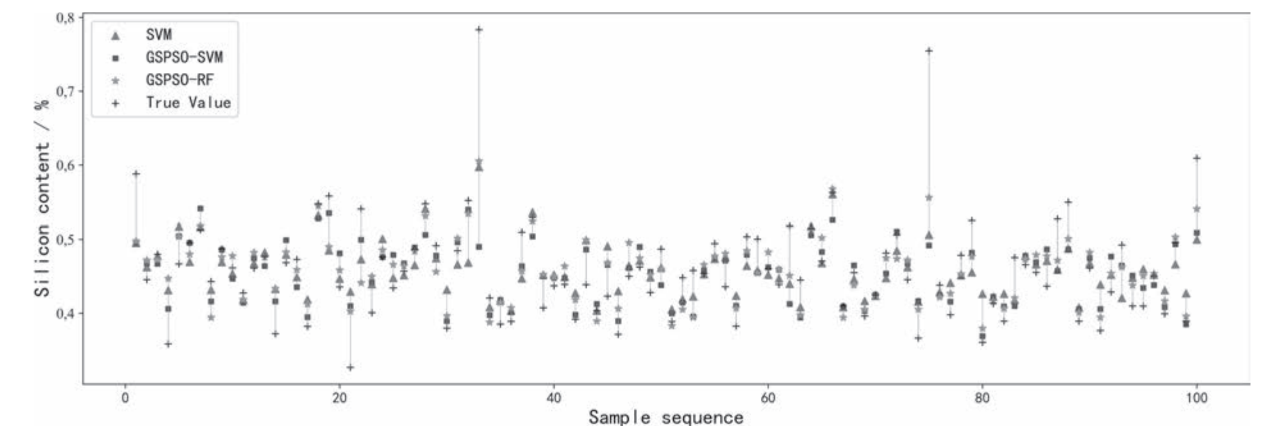


**Figure 7** Predictive values of test samples 1-100

Table 3 **Correlation coefficient r and significance level P**

| Parameters | r | P |
|---|---|---|
| fuel ratio | 0,5262 | 4,514587653834121e-75 |
| gas pressure | -0,4511 | 3,0654299753089757e-53 |
| gas utilization rate | -0,5454 | 1,330712895334096e-81 |
| gas flow rate | -0,2791 | 4,725816809725324e-20 |
| gas temperature | -0,5208 | 2,6274060485132683e-73 |
| CO | -0,1724 | 2,2199494013726574e-08 |
| $CO_2$ | -0,6082 | 3,89883665925288e-106 |
| $H_2$ | -0,3681 | 1,080525070256233e-34 |
| $N_2$ | 0,5535 | 1,7573579035605091e-84 |

Table 4 **Prediction results**

| Model | $H_{rate}$ | MAE |
|---|---|---|
| SVM | 80,76 % | 0,03418691684509517 |
| GSPSO-SVM | 85,57 % | 0,03271539068568513 |
| GSPSO-RF | 87,17 % | 0,029821164874625872 |

tree is 17, the maximum number of features is 5, and the number of decision trees is 133.

## RESULTS AND DISCUSSION

This paper evaluates the performance of the model from two aspects: hit rate and mean absolute error.

(1) Hit rate

$$\mathrm{H}_{\mathrm{rate}} = \frac{1}{312}\sum_{i=1}^{312} I\left(|y_i - \hat{y}_i| \le 0{,}05\right) \times 100\% \qquad (7)$$

$y_i$ is the true value. $\hat{y}_i$ is the predictive value. I is an indicator function. If the input is true, output 1; otherwise, output 0.

(2) Mean Absolute Error

$$\mathrm{MAE} = \frac{1}{312}\sum_{i=1}^{312} |y_i - \hat{y}_i| \qquad (8)$$

Show the prediction results of each model in Table 4.

See from Table 4 that the GSPSO-SVM are better than SVM optimized by grid search. It shows that GSPSO has obvious advantages over grid search. The GSPSO-RF has the best prediction performance. It shows that the feasibility and correctness of the research method. Show the predictive values of test samples 1 - 100 in Figure 7.

## CONCLUSION

In this paper, the GSPSO-RF is proposed. By analyzing the reason for silicon content change, form a high-quality data set. The data set is an important reason for the better results. Finally, the production data of a steel plant verify the GSPSO-RF. The results show that the prediction method proposed in this paper has apparent advantages.

## Acknowledgements

The work was supported by the Basic Scientific Research Project of Colleges and Universities of Liaoning Province Education Department.

## REFERENCES

[1] Z. L. Li, C. J. Yang, W. H. Liu, et al, Research on hot metal Si-content prediction based on LSTM-RNN, Journal of Chemical Industry and Engineering, 69 (2018) 3, 992-997

[2] X. Xu, C. C. Hua, Y. G. Tang, et al, Modeling of the hot metal silicon content in blast furnace using support vector machine optimized by an improved particle swarm optimizer, Neural Computing and Applications, 27 (2016) 6, 1451-1461

[3] W. H. Wang, X. G. Liu, X. Y. Liu, Prediction model of silicon content in blast furnace hot metal based on random forests algorithm, Metallurgical Industry Automation, 38 (2014) 5, 33-38

[4] Y. Liu, L. Ma, H. Z. Zhang, et al, Intelligent optimization algorithm, Shanghai People's Publishing House, Shanghai, 2019, pp. 51, 53 and 57

[5] E. Tanyildizi, G. Demir, Golden Sine Algorithm: A Novel Math-Inspired Algorithm, Advances in Electrical and Computer Engineering, 17 (2017) 2, 71-78

[6] S. Mirjalili, S. M. Mirjalili, A. Lewis, Grey Wolf Optimizer, Advances in Engineering Software, 69 (2014), 46-61

[7] Z. Y. Liang, Ironmaking, Metallurgical Industry Press, Beijing, 2009, pp. 134

[8] S. R. Na, Analysis of ironmaking calculation, Metallurgical Industry Press, Beijing, 2010, pp. 377

[9] H. Li, Statistical learning methods 2nd Edition, Tsinghua University Press, Beijing, 2019, pp. 81-82

**Note:** The responsible translators for English language is Qinghai Pang – University of Science and Technology Liaoning, China