

# A combined prediction model based on double echo state network with improved immune genetic algorithm(IIGA)

Yuan Xu, Jianlong Yin, Yanlin He\*, Qunxiong Zhu\*

*College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China;*

*Engineering Research Center of Intelligent PSE, Ministry of Education of China, Beijing 100029, China;*

*\*Corresponding author: Tel.: +86-10-64413467; Fax: +86-10-64437805. Email: [hey1@mail.buct.edu.cn](mailto:hey1@mail.buct.edu.cn), [zhuqx@mail.buct.edu.cn](mailto:zhuqx@mail.buct.edu.cn)*

**Abstract:** To improve the prediction accuracy of key variables in the industrial process, a combined prediction model based on a double-reservoir echo state network with improved immune genetic algorithm(IIGA) is proposed. Firstly, a double-reservoir ESN(DES) model is established, then partial least squares (PLS) is introduced to calculate the output weight of DES model to enhance the ability to deal with multiple correlation issues; Secondly, in order to reduce the parameters randomness in double-reservoir and prevent the local optimization. IIGA is proposed to optimize the reservoir parameters, which introduces the opposite operator to expand the search range. Based on the above, a combined prediction model IIGA-DES-PLS is proposed; Third, the proposed IIGA is verified by the Griewank function. And the proposed combined prediction model is applied to the purified terephthalic acid (PTA) solvent system. The comparison results show that the combined prediction model has higher prediction accuracy and provides certain guidance for predicting key variables of industrial process.

**Keywords:** Combined Prediction Model, Double-reservoir echo state network, Partial Least Squares, Immune genetic optimization algorithm, Purified Terephthalic Acid

## 1. INTRODUCTION

With the increasing complexity of industrial process, more demand has put forward to a higher-precision prediction model to monitor and predict the key process variables. In recent years, with the wide application of Echo State Networks (ESN) [1-3], more achievements have been made to predict complex time series. ESN uses the reservoir to replace the hidden layer in the network structure [4], contributing to a simpler model and less training calculation time [5]. While, as the reservoir is similar to a 'black box', the final result is not stable [6]. Zhang et al. proposed a high-precision self-organizing deep belief module ESN (SDBMESN) [7] with good

prediction performance and robustness. However, the efficiency is significantly reduced because the deep network needs a lot of data and long training time. Hua et al. transformed ESN into multi-input multi-output ESN (MIMO-ESN) [8], which significantly improved the prediction accuracy of ESN, but this method did not fundamentally solve unstable factors of ESN reservoir. Li et al. [9] proposed that multiple ESN reservoirs were connected in series to improve the prediction accuracy, but it did not ameliorate the training process failed to optimize the results. Zhong et al. proposed a genetic algorithm(GA)-based DESN with higher accuracy. However, GA algorithms often trap into local extrema, the optimization efficiency is always not satisfactory [12]. Wang et al. [13] proposed that the ESN-PLS model trained the output weights by partial least squares(PLS), obtaining a better output weight matrix. However, due to the constant parameters, the prediction accuracy would be limited.

Therefore, this paper proposes a combined prediction model based on a double-reservoir echo state network with improved immune genetic algorithm(IIGA), namely IIGA-DES-PLS, in which partial least squares (PLS) is introduced to calculate output weights of DESN and improved immune genetic algorithm (IIGA) is proposed to optimize the reservoir parameters. First, the DESN model is established. Second, PLS is introduced to calculate the output weights of DESN, so as to reduce the coupling effect in reservoir. Third, in order to prevent the problem of trapping into local extreme values, IIGA is proposed to generate the opposite operators to expand the search range of parameters. Finally, simulation experiment is made by Griewank function and Purified Terephthalic Acid (PTA) solvent system [14], the comparison results show that the combined prediction model has higher accuracy and provides an effective way to monitor and predict the key variables in industrial process.

## 2. PRELIMINARIES

### 2.1. DESN

DESN is a kind of recurrent neural network, which has

two input layers, double reservoirs and one output layer. This structure can enrich the network's diversity, and the double-reservoir can reduce the influence of random generation of reservoir weight matrix. The structure chart of DESN is shown in Fig.1.

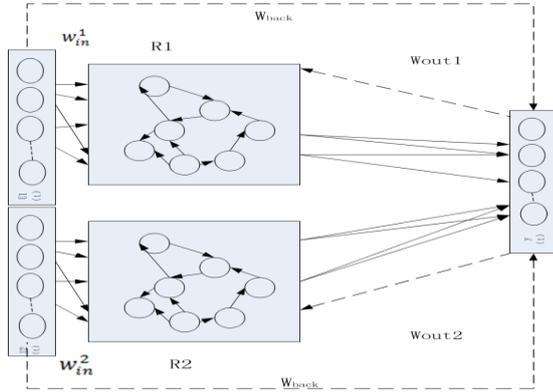


Fig.1 Structure chart of DESN

In Fig.1,  $u(t)$  is the input value, R1 and R2 are the two reservoirs.  $W_{out}$  is the output weight matrix,  $W_{out1}, W_{out2} \in R^{L(N_1 * K)}$ .  $w_{in}^1$ , and  $w_{in}^2$  are the input weight matrix,  $w_{in}^1, w_{in}^2 \in R^{N_1 * K}$ .  $W_{back}$  is the output feedback weight matrix.  $K$  is the input dimension.  $L$  is the number of output layers.  $N_1$  and  $N_2$  are the number of neurons in the reservoir R1 and R2. Assuming that system input is  $U$  and the system output is  $Y$ .  $Y = X * W_{out}$ ,  $X$  is the state collection matrix,  $W_{out}$  is the output weight matrix. At  $t$  time, the input is expressed by formula (1), the state matrix is expressed by formula (2), the output is expressed by formula (3).

$$\begin{cases} u1(t) = [u1_1(t), u1_2(t), \dots, u1_q, \dots, u1_K(t)]^T \\ u2(t) = [u2_1(t), u2_2(t), \dots, u2_q, \dots, u2_K(t)]^T \end{cases} \quad (1)$$

In the formula,  $u(t)$  is the input value. Set time  $q$  as the forgetting point, and set time  $(q + 1)$  as the starting point of the system.

$$\begin{cases} x1(q+1) = f_1(IS \cdot w_{in}^1 \cdot u1(q+1) + w_1 \cdot x1(q) + w_{back}^1 \cdot y(q)) \\ x2(q+1) = f_2(IS \cdot w_{in}^2 \cdot u2(q+1) + w_2 \cdot x1(q) + w_{back}^2 \cdot y(q)) \end{cases} \quad (2)$$

where  $x1$  and  $x2$  are the state matrix,  $f$  is the excitation function,  $IS$  is the unit scale,  $W_1, W_2 \in R^{N_1 * N_1}$  are the weight matrices of the reservoirs.

$$y(t) = f_o(W_{out1}(u1(t), x1(t)) + W_{out2}(u2(t), x2(t))) \quad (3)$$

where  $y(t)$  is the output value,  $f_o$  is the output excitation function. DESN only needs to train the output weights.

## 2.2. Immune genetic algorithm(IGA)

Immune genetic algorithm(IGA) [15] is an improved genetic algorithm based on the biological immune mechanism. Its antigen is equivalent to the objective function, and the antibody is equivalent to the problem to

be solved. The IGA is implemented through operators which includes affinity evaluation operator, antibody concentration evaluation operator, incentive calculation operator, immune selection operator, mutation operator, clone suppression operator and population update operator, etc.

IGA retains the characteristics of genetic algorithm's random global parallel search and avoids immature convergence to a considerable extent. Assuming that the immune system is composed of  $N$  antibodies, the length of each antibody gene is  $M$ , and the symbol set size is  $S$ . The two basic concepts of IGA are as below:

### (1) Diversity

The difference between individuals is expressed by Shannon's average information entropy  $H(N)$ ,

$$H(N) = \frac{1}{M} \sum_{j=1}^M H_j(N) \quad (4)$$

where  $H_j(N)$  is the  $j$ th gene's information entropy.

### (2) Similarity

Similarity  $A_{ij}$  is the similarity degree between the  $i^{th}$  and  $j^{th}$  antibodies.

$$A_{ij} = \frac{1}{1+H(2)} \quad (5)$$

where  $H(2)$  is the average information entropy of the  $i$ th and  $j$ th antibodies. Extend the concept of similarity between two antibodies to the entire group, and take its group similarity as  $A(N)$  and its definition as follows:

$$A(N) = \frac{1}{1+H(N)} \quad (6)$$

$A(N)$  represents the overall similarity of the entire group. The larger the  $A$  is, the lower the diversity of the group is. The smaller the  $A(N)$  is, the higher the diversity of the group is.

## 3. A COMBINED PREDICTION MODEL BASED ON DESN-PLS WITH IIGA

### 3.1. Prediction model of DESN-PLS

The traditional DESN trains the output weight by using the generalized inverse method. It is noted that the output weights obtained often make the final result lower in accuracy. In this paper, PLS is introduced to DESN which can effectively solve the multiple correlation problem well during the calculation of output weight matrix.

The model uses PLS to find the connection matrix ( $W_{out}$ ) between the reservoir and the output matrix. In a general regression model, if there is a set of dependent variables  $Y = (y_1, \dots, y_q)$  and a set of independent variables  $X = (x_1, \dots, x_q)$ , PLS is described by formula (7).

$$\tilde{Y} = X(X^T X)^{-1} X^T Y \quad (7)$$

where  $\tilde{Y}$  is the estimation of  $Y$ .

The specific steps of PLS to get the output weight matrix are as follows:

- i. The state matrix  $X$  and the output matrix  $Y$  are used to construct the matrix  $pz$ , and the standardized

matrix  $pz$  extracts matrices of the input  $e0$  and the output  $f0$ .

- ii. Construct the least square matrix  $V$ , let the number of matrix columns  $n$  is the number of iterations, and obtain the transformation matrix  $W$ .
- iii. By obtaining and sorting the eigenvalues of the matrix  $pz$ , the eigenvectors and coefficients corresponding to the eigenvalues are assigned to the transformation matrix  $W$ . The transformation matrix  $W^*$  is obtained after  $n$  iterations.
- iv. Denormalize the obtained transformation matrix to obtain  $W_{out}$ .

### 3.2. IIGA for parameters optimization of DESN-PLS

To some extent, the reservoir parameters in DESN-PLS will determine the final performance of the network. The number of neurons would affect the fitting situation in the reserve pool. If the sparseness SD is too large, the internal disorder of the reservoir will affect the prediction effect. Therefore, network parameters need to be optimized. Although the traditional IGA is very mature, it will still easily trap into the local extreme value. Therefore, this paper introduces the idea of cooperative search with multiple operators. Four parameters (Scale factor, Sparsity, Unit size, Number of reservoir's neurons) of the reservoir are selected as the optimization target. The root mean square error (RMSE) obtained from training is used as the affinity evaluation operator. Moreover, inspiring by the mutation operator and adding opposite operator, which make the system generate a new population to prevent the occurrence of local optimal. Totally, there are clone operator, mutation operator, opposite operator and population update operator searching for the optimal values of the four parameters. In this way, it obviously expands the search range of the population and prevents the algorithm from trapping into local extremum. The flowchart of IIGA-DESN-PLS model is exhibited in **Fig.3**. The algorithm steps are as follows:

- i. Initialize the population, calculate the affinities of  $N$  individuals (root mean square error), then calculate the antibody concentration and incentive degree by formula (8) and formula (9) respectively.

$$ben(ab_i) = \frac{1}{N} \sum_{j=1}^N H(ab_i, ab_j) \quad (8)$$

$$sim(ab_i) = \alpha aff(ab_i) - bben(ab_i) \quad (9)$$

where  $N$  is the population size,  $sim(ab_i)$  is the excitation degree of the antibody  $ab_i$ ,  $\alpha$ ,  $b$  is the calculation parameter,  $H(ab_i, ab_j)$  indicates the similarity among antibodies, which can be expressed by formula (10):

$$H(ab_i, ab_j) = \begin{cases} 1, & aff(ab_i, ab_j) < \delta_s \\ 0, & aff(ab_i, ab_j) \geq \delta_s \end{cases} \quad (10)$$

where  $ab_i$  is the  $i$ -th antibody in the population,

$aff(ab_i, ab_j)$  is the affinity between the  $i$ th and  $j$ th antibody.  $\delta_s$  is the similarity threshold.

- ii. Sort the incentives of the population in ascending order. The conventional IGA is to perform mutation operations on the first half of the population. The IIGA introduces the opposite operator. The most significant difference between the opposite operator and the mutation operator is that the opposite operator generates new individuals through mutation, which will perform randomness on each component of the individual. It can be found that the opposite operator expands the search scope of the parameters, and avoids the population falling into the local optimal value. The formula for using mutation operator to generate new parameters is shown in formula (11), and the formula for using the opposite operator to generate new parameters is shown in the following formula (12):

$$T_c(ab_i) = \begin{cases} ab_{i,j} + (rand - 0.5)\delta, & rand < P_m \\ ab_{i,j}, & else \end{cases} \quad (11)$$

where  $ab_{i,j}$  is the  $j^{th}$  individual of the antibody  $ab_i$ ,  $\delta$  is the defined neighborhood range which can be set randomly,  $rand$  is the random number function generated in the range of  $(0,1)$ , and  $P_m$  is the mutation probability.

$$T_c(ab_i) = ab_{i,j,m} + (rand - 0.5)\delta, rand < 1$$

Boundary setting

$$\text{if } T_c(ab_i) > ab_{\max} \quad (12)$$

$$T_c(ab_i) = ab_{\max}$$

$$\text{if } T_c(ab_i) < ab_{\min}$$

$$T_c(ab_i) = ab_{\min}$$

where  $ab_{i,j,m}$  is the  $j$ th parameter of the  $m$ th individual of the antibody  $ab_i$ . By setting the dimension, each component of the individual can be changed randomly. In addition, the mutation probability is set to 1, so that the individual mutation will occur. In order to ensure the effectiveness of each obtained operator, set the boundary. If it exceeds the maximum, the new parameter generated is equal to the maximum of the parameter, and if it is lower than the minimum, the new parameter generated is equal to the minimum of the parameter.

- iii.  $N$  new populations are obtained through mutation operator and opposite operator in the proposed IIGA algorithm, and the population incentive degree is calculated.
- iv. Refresh the population  $N$ , and the corresponding update formula for refreshing is formula (13). The incentive degree of  $N$  individuals is obtained by formula (9), the population retains the first  $N/2$  individuals with high incentives, refreshes the  $N/2$  individuals with low incentives to expand the new feasible solution field.

$$bf(i) = \text{rand} \left( 1, \frac{N}{2} \right) * Xs(i) \quad (13)$$

where  $bf(i)$  is the  $i^{th}$  component of the population, and  $Xs(i)$  is the upper limit of the  $i^{th}$  component.

- v. Calculate the incentive degree of the refreshed population, sort the incentive degree with the previously obtained population, and keep the individual with the highest incentive degree recorded as  $bfbest(i)$ .
- vi. Repeat steps 2 to 5 for an iterative loop and keep the most motivated individual recorded as  $bfbest$ .
- vii. Bring the final optimized parameter ( $bfbest$ ) into the DESN-PLS network model for regression generalization prediction.

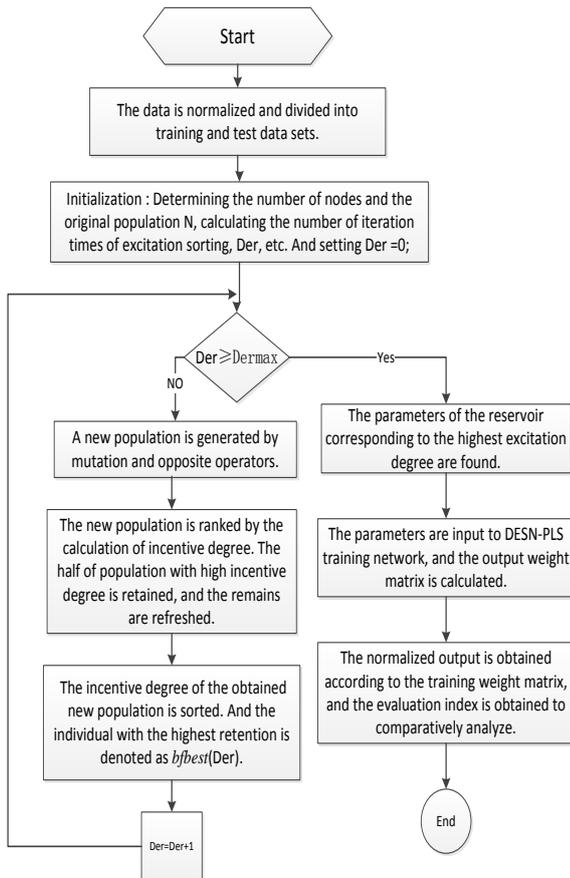


Fig.3 Flow chart of IIGA-DESN-PLS prediction model

#### 4. CASE STUDY

In order to verify the effectiveness of the improved immune genetic algorithm (IIGA), this study uses the Griewank function for verification. The Griewank function is a commonly used function to verify the efficiency of the optimized function. Since this function has multiple local minima, it can be used to test whether the optimized function will trap into local extremum or not. The minimum of the function is 0. The formula of Griewank function is shown in formula (14):

$$\min f(x_i) = \sum_{i=1}^N \frac{x_i^2}{4000} - \prod_{i=1}^N \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1 \quad (14)$$

$$x_i \in [-600, 600]$$

First, standard Immune Genetic Algorithm (IGA), Standard Particle Swarm Algorithm (PSO) and IIGA are used to optimize the Griewank function iteratively. The number of immune individuals and the population size of the particle swarm both set as 50. The maximum immune algebra and particle swarm evolution number are set to 30 times, and the fitness function is set to the Griewank function. The e results of experiment are shown in Fig.4.

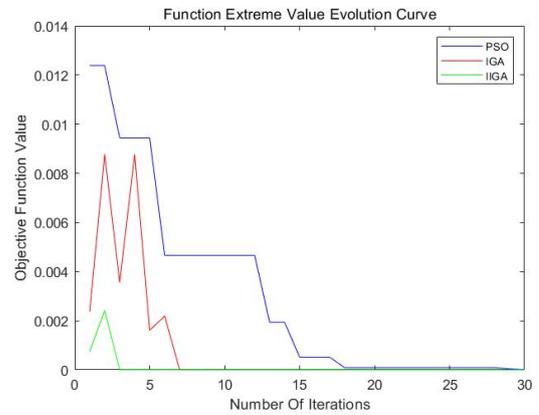


Fig.4 Comparison chart of optimization results

As shown in Fig.4, PSO has a slower convergence speed due to its small search range. Comparing with PSO, IGA converges faster, but it is still slower than IIGA. Obviously, the proposed IIGA can converge faster to obtain the optimal value and effectively prevent the algorithm from trapping into local extrema.

#### 5. EXAMPLE APPLICATION

To further verify the validity of the IIGA-DESN-PLS model, PTA solvent system is used to make the simulation experiment. The flow chart of PTA solvent system is shown in Fig.5. PTA system is mainly composed of a raw material dehydration device (solvent dehydration tower), a product n-butyl acetate recovery device and a recycling device. In the PTA system, the primary substance is acetic acid. Due to its volatility, we can not measure the change of acetic acid content directly. in general, the conductivity of top tower is measured to represent the acetic acid content.

In this experiment, a total of 260×18 sets of PTA solvent system data are collected. 174 sets of data are the model training data, and 86 sets are the model test data. Based on the IIGA-DESN-PLS, IGA-DESN-PLS, DESN-PLS, ESN methods, simulation experiments are carried out, and the results are compared and analyzed. For the double-reservoir, the 17 sets of inputs of PTA solvent system data are divided into flow parameters (the first 9 columns) and Temperature parameters (the last 8 columns) respectively, the mutual influence of the two types of data is weakened. The IIGA population size is 50. The experimental results are shown in Fig.6. RMSE compares the accuracy of these prediction models and the comparison results of model indicators are shown in Table 1.

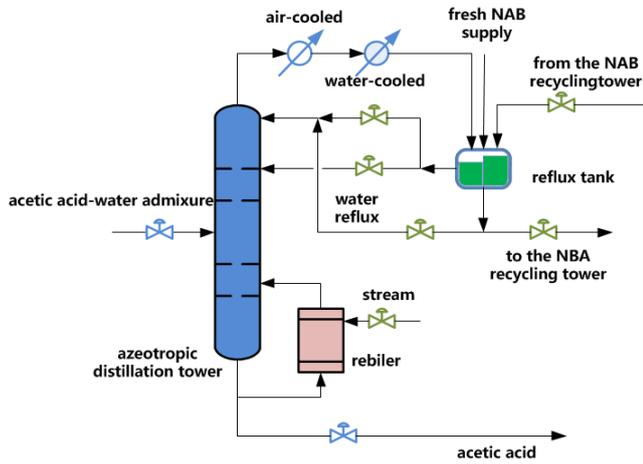


Fig.5 Flow chart of PTA solvent system

Table 1. Comparison result of prediction accuracy

Model \ Indicator	IIGA-DESN-PLS	IGA-DESN-PLS	DESN-PLS	ESN
RMSE	0.277	0.306	0.348	0.535

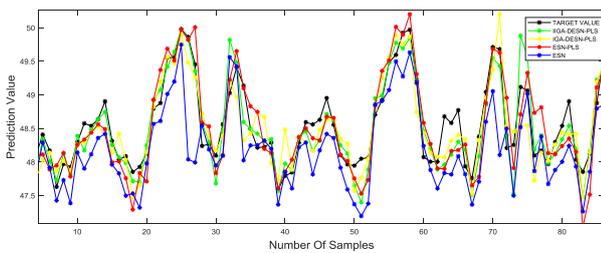


Fig.6 Comparison result of different prediction models

From **Table 1**, the proposed IIGA-DESN-PLS model has higher prediction accuracy. Form **Fig.6**, the model output by the proposed IIGA-DESN-PLS model is closer to real data.

### 6.CONCLUSION

This paper proposes a combined prediction model based on a double-reservoir echo state network(DESN) with improved immune genetic algorithm(IIGA) for industrial processes. Firstly, a DESN model is established. And PLS is introduced to calculate the output weights of DESN to resolve the multiple correlations of variables. Secondly, in order to optimize the parameters of reservoir, IIGA is proposed, which introduces the opposite operator to expand the parameter search range. Then, a combined IIGA-DESN-PLS prediction model is formed. Finally, the performance of IIGA is evaluated by Griewank function. And a simulation experiment on PTA solvent system is made to illustrate the effectiveness of the proposed combined prediction model. The comparison results show the IIGA-DESN-PLS combined prediction model has higher accuracy and can provide the guidance to the prediction of key variable for complex industrial process.

### Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 61973022, 62073022) and the Fundamental Research Funds for the Central Universities (buctrc202123).

### References

- [1] Huang Z , Yang C , Chen X , et al. Functional deep echo state network improved by a bi-level optimization approach for multivariate time series classification[J]. Applied Soft Computing, 2021(1):107314.
- [2] Rodan A , Tino P . Minimum Complexity Echo State Network[J]. IEEE Transactions on Neural Networks, 2011, 22(1):131-44.
- [3] Li D , Min H , Wang J . Chaotic Time Series Prediction Based on a Novel Robust Echo State Network[J]. IEEE Transactions on Neural Networks & Learning Systems, 2012, 23(5):787-799.
- [4] Jaeger H . Echo state network[J]. Scholarpedia, 2007, 2(9):1479 - 1482.
- [5] Wang Y , Cai T . Wireless spectral prediction by the modified echo state network based on leaky integrate and fire neurons[J]. Metallurgical and Mining Industry, 2015, 7(12):260-265.
- [6] Peng L , Lv S X , Wang L , et al. Effective electricity load forecasting using enhanced double-reservoir echo state network[J]. Engineering Applications of Artificial Intelligence, 2021, 99(4):104132.
- [7] Zhang H , Hu B , Wang X , et al. Self-organizing deep belief modular echo state network for time series prediction[J]. Knowledge-Based Systems, 2021, 222(17):107007.
- [8] Hua Z , Zheng Z , Gao F , et al. Remaining useful life prediction of PEMFC systems based on the multi-input echo state network[J]. Applied Energy, 2020, 265.
- [9] Li Q , Wu Z , Ling R , et al. Multi-reservoir echo state computing for solar irradiance prediction: A fast yet efficient deep learning approach[J]. Applied Soft Computing, 2020, 95:106481.
- [10] Zhong S , Xie X , Lin L , et al. Genetic algorithm optimized double-reservoir echo state network for multi-regime time series prediction[J]. Neurocomputing, 2017, 238:191-204.
- [11] A U M , B S B . Genetic algorithm-based clustering technique[J]. Pattern Recognition, 2000, 33( 9):1455-1465.
- [12] Zhu Y , Wang D , Li P , et al. Optimization of exergy efficiency of a cascaded packed bed containing variable diameter particles[J]. Applied Thermal Engineering, 2021, 188(6):116680.
- [13] Y Wang , Li H , Yang B . Modeling of furnace operation with a new adaptive data echo state network method integrating block recursive partial least squares[J]. Applied Thermal Engineering, 2020, 171:115088.
- [14] Karayannidis G P , Chatziavougstis A P , Achilias D S . Poly(ethylene terephthalate) recycling and recovery of pure terephthalic acid by alkaline hydrolysis[J]. Advances in Polymer Technology, 2002, 21(4):250-259.
- [15] Zhou Z , Luo D , Shao J , et al. Immune genetic algorithm based multi-UAV cooperative target search with event-triggered mechanism[J]. Physical Communication, 2020, 41:101103.