

A Prediction Method Based on Improved Echo State Network for COVID-19 Nonlinear Time Series

Banteng Liu¹, Wei Chen², Yourong Chen¹, Ping Sun¹, Heli Jin², Hao Chen²

¹College of Information Science and Technology, Zhejiang Shuren University, Hangzhou, China

²School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China

Email: 513016159@qq.com

How to cite this paper: Liu, B.T., Chen, W., Chen, Y.R., Sun, P., Jin, H.L. and Chen, H. (2020) A Prediction Method Based on Improved Echo State Network for COVID-19 Nonlinear Time Series. *Journal of Computer and Communications*, 8, 113-122. <https://doi.org/10.4236/jcc.2020.812011>

Received: November 10, 2020

Accepted: December 17, 2020

Published: December 24, 2020

Abstract

This paper proposes a prediction method based on improved Echo State Network for COVID-19 nonlinear time series, which improves the Echo State Network from the reservoir topology and the output weight matrix, and adopt the ABC (Artificial Bee Colony) algorithm based on crossover and crowding strategy to optimize the parameters. Finally, the proposed method is simulated and the results show that it has stronger prediction ability for COVID-19 nonlinear time series.

Keywords

COVID-19, Nonlinear Time Series, Prediction, Echo State Network

1. Introduction

The pneumonia caused by a novel coronavirus infection that was erupted in Wuhan, Hubei Province, China in December 2019, which is known as COVID-19. There are a large number of nonlinear time series which hide the development trend of the epidemic, including newly confirmed cases, newly suspected cases, cumulative confirmed cases, existing suspects cases, cured rate and mortality rate, etc. The development trend of COVID-19 can be learned in time through the prediction of the above-mentioned nonlinear time series, and then relevant personnel can take corresponding measures and strengthen the prevention and control for the upcoming trend of the epidemic before it becomes severe, which is of vital importance to guard the lives of the people.

The prediction method based on the Echo State Network (ESN) [1] is currently one of the main methods to predict nonlinear time series. The ESN is

characterized by using a large-scale random sparsely connected network called “reservoir” as the hidden layer to deal with nonlinear and unstable time series. The training process only needs to train the output weight from the reservoir to the output layer, which simplifies the training process of the network, and avoids the problems of being prone to falling into local optimality and complex training algorithms of traditional neural networks [2]. At present, some scholars have begun to research the improved methods of ESN, mainly focusing on reservoir topology optimization and output weight optimization [3] [4].

In terms of the reservoir topology, the reservoir of traditional ESN is a random network, which leads to model training purposeless. In order to solve this problem, Li *et al.* [5] proposed to adopt a NW small world Echo State Network with both randomness and regularity as the reservoir of ESN to predict nonlinear time series, it improves the adaptability and prediction accuracy of the prediction model, but the node connections of the NW small world Echo State Network are deterministic connections, and the prediction accuracy of time series with time-varying and ambiguity is limited.

In terms of the output weight, the output weight of traditional ESN calculation adopts pseudo inverse method, but it is prone to appear the question of multicollinearity when solving high-dimensional linear regression [6]. In order to solve this problem, Wang *et al.* [7] [8] proposed to use Ridge regression, Lasso regression and other linear regression methods to calculate the output weight by adding L_2 norm and L_1 norm, but Ridge regression and Lasso regression belonging to biased estimation to impose greater punishment to larger output weight, and over fitting problem is prone to exist in model prediction [9]. The asymptotically unbiased regularization method is needed to improve the prediction accuracy and generalization performance of the prediction model to solve the above problems. The common asymptotically unbiased regularization methods include SCAD (Smooth Clipped Absolute Deviation) regularization method [10] and MCP (Minimax Concave Penalty) regularization method [11]. At present, some scholars have successfully applied SCAD regularization method to the output weight optimization of small world Echo State Network [12], which improves the prediction accuracy of small world Echo State Network for nonlinear time series. However, the MCP regularization method has not been proposed to optimize the output weight of small world Echo State Network, and the penalty function of MCP regularization method has the minimum maximum convexity, which can make more appropriate punishment for the large or small output weight, and it is more suitable for processing multi-dimensional nonlinear data [13].

Therefore, this paper proposes a prediction method based on improved Echo State Network for COVID-19 nonlinear time series, which improves the ESN from the optimization of the reservoir topology and the optimization of the output weight to improve the prediction accuracy for the COVID-19 nonlinear time series.

2. A Prediction Method Based on Improved Echo State Network for COVID-19 Nonlinear Time Series

The prediction method based on improved Echo State Network for COVID-19 nonlinear time series includes three parts: optimization of the reservoir topology, optimization of output weight and optimization of parameters.

2.1. Optimization of the Reservoir Topology

The improved small world network is used as the reservoir of ESN to obtain the Small World Echo State Network (SWESN). Its topological structure is shown as in **Figure 1**.

It can be seen from **Figure 1** that the SWESN has three layers, including input layer, hidden layer (reservoir) and output layer. The internal weight matrix W_x of the reservoir in the improved small world network is obtained by establishing the function relationship between the edge probability and the distance between nodes, and it will not change after determination. The edge probability value p decreases exponentially as the distance between nodes increases, that is:

$$p = \alpha \times e^{(-\beta \times d)} \quad (1)$$

where p denotes the connection weight between nodes and its value range is $[0, 1]$, d denotes the Euclidean distance between nodes, α is used to adjust the distance sensitivity, and β is used to adjust the overall density of the network. The internal weight matrix W_x is obtained by solving p value. Then, the state equation and output equation of the SWESN are as follows:

$$x(t) = f(W_{in}u(t) + W_x x(t-1)) \quad (2)$$

$$y(t) = x^T(t)W_{out} \quad (3)$$

where $u(t) \in R^L$, $x(t) \in R^M$ and $y(t) \in R$ respectively denote the input variable, state variable and output variable at the time t ; the activation function f usually takes the hyperbolic tangent function \tanh ; $W_{in} \in R^{M \times L}$, $W_x \in R^{M \times M}$ and $W_{out} \in R^M$ respectively denote the input weight matrix, the internal weight matrix of the reservoir and the output weight matrix. The input weight matrix is randomly generated and will not change after determination.

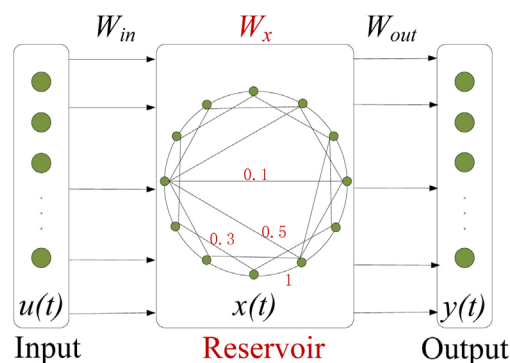


Figure 1. Small world echo state network topology.

2.2. Optimization of Output Weight

The output weight matrix W_{out} is obtained during training, that is, W_{out} is the matrix corresponding to the minimized objective function, as shown in Equation (4), and it is obtained by the least square method, as shown in Equation (5):

$$W_{out} = \arg \min \|Y - XW_{out}\|^2 \tag{4}$$

$$W_{out} = X^\dagger Y = (X^T X)^{-1} X^T Y \tag{5}$$

where (X, Y) denotes the training sample and X^\dagger denotes the pseudo-inverse of X . The process of optimizing the output weight is shown as in **Figure 2**.

It can be seen from **Figure 2** that the output weight matrix optimization process has three steps. Firstly, the objective function of output weight with MCP penalty term is obtained by MCP (Minimax Concave Penalty), and then the objective function of derivable origin is obtained by LQA (Local Quadratic Approximation). Finally, the optimized output weight is obtained by Ridge regression.

MCP generates singular values at the origin and can produce sparse solutions. And, when $|\theta| > \gamma\lambda$, the variable is directly set to 0, which satisfies the approximate unbiased estimation of the variable θ , and the MCP function is shown in Equation (6):

$$\rho_{\lambda,\gamma}(|\theta|) = \begin{cases} \lambda|\theta| - \frac{|\theta|^2}{2\gamma} & |\theta| \leq \gamma\lambda \\ \frac{1}{2}\gamma\lambda^2 & |\theta| > \gamma\lambda \end{cases} \tag{6}$$

where γ, λ are adjustable hyperparameters ($\gamma > 2, \lambda > 0$). θ is a parameter vector, which denotes the output weight W_{out} in this paper. Then the objective function of output weight with MCP penalty term is obtained, as shown in Equation (7), and the estimated matrix \widehat{W}_{out} is the matrix corresponding to the minimized objective function with MCP penalty term, as shown in Equation (8):

$$L = \|Y - XW_{out}\|^2 + \sum_{j=1}^J \rho_{\lambda,\gamma}(|W_{out_j}|) \tag{7}$$

$$\widehat{W}_{out} = \arg \min(L) \tag{8}$$

where J denotes the number of variables, and $\rho_{\lambda,\gamma}$ denotes the penalty function.

LQA is chosen to approximately decompose the MCP penalty function to obtain an approximate solution of the model since the MCP penalty function is not directable at the origin. Assuming that $W_{out}^{(0)}$ is known, the approximate decomposition of the MCP is shown in Equation (9):

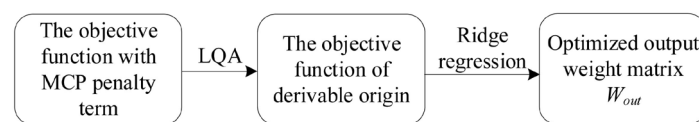


Figure 2. Output weight matrix optimization process block diagram.

$$\rho_{\lambda,\gamma}(|W_{out}|) \approx \rho_{\lambda,\gamma}(|W_{out}^{(0)}|) - \frac{1}{2} \frac{\rho'_{\lambda,\gamma}(|W_{out}^{(0)}|)}{W_{out}^{(0)}} (W_{out}^{(0)})^2 + \frac{1}{2} \frac{\rho'_{\lambda,\gamma}(|W_{out}^{(0)}|)}{W_{out}^{(0)}} (W_{out})^2 \quad (9)$$

where $W_{out}^{(0)}$ denotes the adjacent point of W_{out} , which is obtained by Equation (5) in this paper, $\rho'_{\lambda,\gamma}(|W_{out}^{(0)}|)$ denotes the first derivative of the penalty function. The first two items in the equation are not associated with W_{out} and can be regarded as a constant C . Then the estimated output weight matrix W_{out} is expressed as:

$$\widehat{W}_{out} = \arg \min \left(\|Y - XW_{out}\|^2 + \sum_{d=1}^D \frac{1}{2} \frac{\rho'_{\lambda,\gamma}(|W_{out_d}^{(0)}|)}{W_{out_d}^{(0)}} (W_{out_d})^2 + C \right) \quad (10)$$

where D denotes the number of non-zero elements in W_{out} , and the estimated output weight matrix W_{out} can be obtained by repeatedly executing the Ridge regression solution to Equation (10), as shown in Equation (11):

$$\widehat{W}_{out} = \left[X^T X + \sum_{d=1}^D \frac{\rho'_{\lambda,\gamma}(|W_{out_d}^{(0)}|)}{W_{out_d}^{(0)}} \right]^{-1} X^T Y \quad (11)$$

Finally, the optimized output weight matrix W_{out} is obtained through iterative Equation (11).

2.3. Optimization of Parameters

In this paper, an ABC (Artificial Bee Colony) algorithm based on crossover and crowding strategy is adopted to optimize the γ and λ of MCP. ABC algorithm is a kind of global optimization algorithm based on swarm intelligence with fast convergence speed. Its intuitive background comes from bee colony's honey gathering behavior. Bees with different divisions of labor find the best solution to the problem by sharing and exchanging information. The crossover strategy is integrated to expand the search range of the whole parameter solution and the crowding strategy is integrated to eliminate the similar solutions within the population. The training process of the ABC algorithm based on crossover and crowding strategy are shown in **Figure 3** and the steps are as follows:

Step 1: Initialize. Set the number of food source population as Z , the maximum number of iterations of the population as Q , the maximum evolution threshold as H , the crossover probability as C , the crowding factor as P , the crowding number as P_a , the current iteration times $q = 0$, the current evolution threshold $h = 0$, and initialize γ and λ as the initial food source (Randomly generate a group of uniform distribution and adjustable hyperparameter combination);

Step 2: Calculate the fitness value F according to the training error of the sample;

$$F = 1 / f_{\lambda,\gamma}(\widehat{W}_{out}) \quad (12)$$

Step 3: Generate a new food source by searching the unknown solution space near each food source until the number of food source population reaches $2 * Z$;

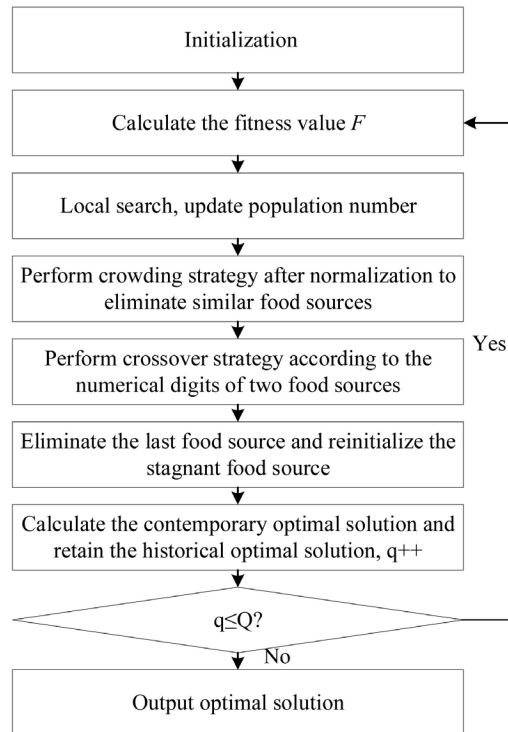


Figure 3. Parameters optimization flow chart.

Step 4: Perform crowding strategy. First, randomly select P food sources as crowding factors after normalizing γ and λ , and then calculate and sort the difference between other food sources and crowding factors in ascending order. Finally, eliminate the former P_a food sources and randomly generate P_a food sources to ensure that the population size does not change;

Step 5: Perform crossover strategy. First, randomly select two food sources from the population to calculate the numerical digits, and then circularly perform the following crossover operations until each bit of the food source is crossed to obtain two new food sources: select the i -th bit of the food source and randomly generate $j \in [0,1]$. If $j > C$, it will not be changed, otherwise perform the crossover operation (Exchange the i -th value of two food sources);

Step 6: Eliminate the last food source and reinitialize the stagnant food source. First, sort the food sources in the current population in descending order according to the fitness value, and then select a certain proportion of the food sources with the fitness value in the last in the current population to initialize. Finally, record and store the contemporary food sources in the set. If the food source already exists, set $h = h + 1$. When $h = H$, set $h = 0$ and retain the contemporary optimal solution. If the contemporary optimal solution is better than the historical optimal solution, it will be replaced. Otherwise, the historical optimal solution will be retained and entered into the next generation population;

Step 7: Let $q = q + 1$, if $q \leq Q$, jump to step 2, otherwise, jump to step 8;

Step 8: Output the historical optimal solution and end the training.

3. Simulation Experiment

3.1. Test Data

The proposed prediction method is applied to the COVID-19 nonlinear time series for one-step prediction. The COVID-19 nonlinear time series is derived from the National Epidemiological Map Platform, including newly confirmed cases, newly suspected cases, cure rate and mortality rate in China.

3.2. Simulation Analysis

Simulation test and analysis were carried out after normalization of COVID-19 nonlinear time series and reconstruction of phase space. This paper randomly selects 50 days for training and 30 days for testing. The test results are inverse normalized. The fitting results of predicting the newly confirmed cases, newly suspected cases, cure rate and mortality rate are shown in **Figures 4-7**.

It can be seen from **Figures 4-7** that the prediction method proposed in this paper can better predict the curve trend of the four kinds of COVID-19 nonlinear

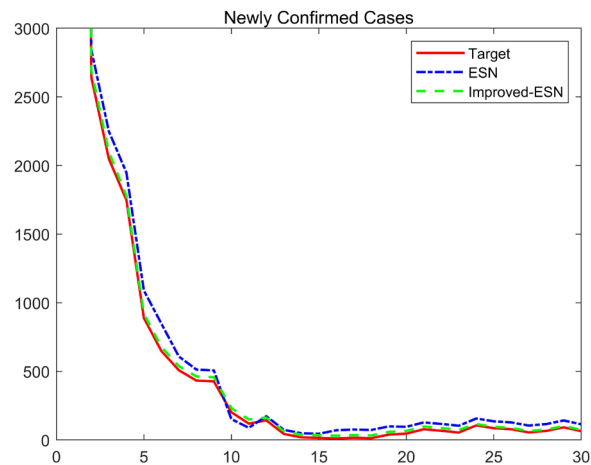


Figure 4. Newly confirmed cases ($\gamma = 3.06; \lambda = 0.08$).

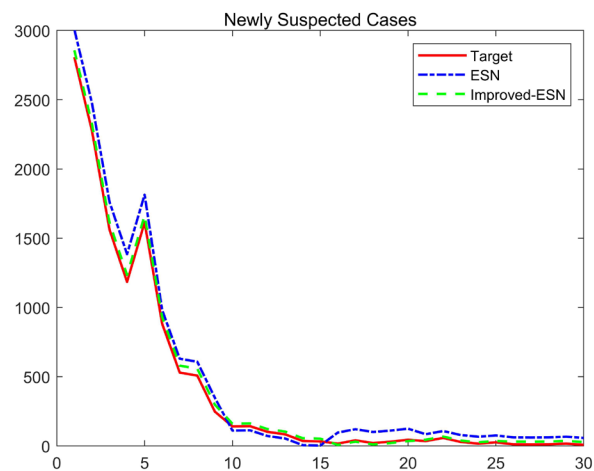


Figure 5. Newly suspected cases ($\gamma = 3.11; \lambda = 0.12$).

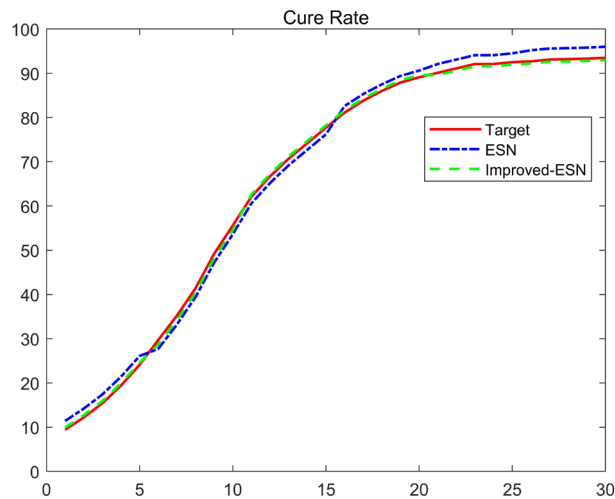


Figure 6. Cure rate ($\gamma = 2.86; \lambda = 0.05$).

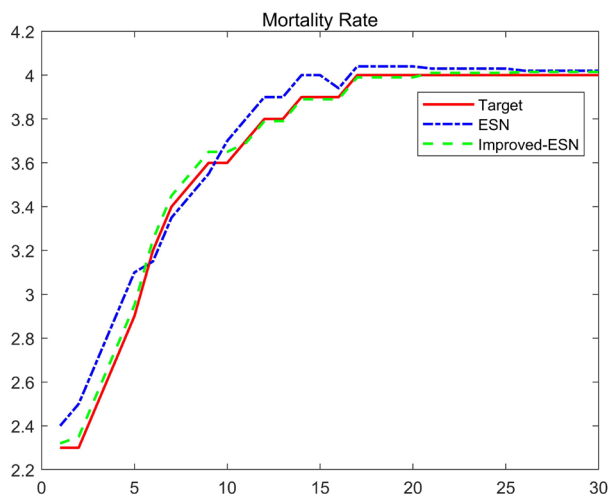


Figure 7. Mortality rate ($\gamma = 3.29; \lambda = 0.16$).

time series, and the prediction accuracy has been improved compared with the traditional ESN.

In order to further analyze the prediction error, compare the prediction results of ESN, SWESN, Ridge-SWESN, Lasso-SWESN, SCAD-SWESN, MCP-SWESN and the proposed Improved-ESN method in this paper for the COVID-19 non-linear time series, the Standard Root Mean Square Error (NRMSE) is used as the performance indicator for all simulation predictions:

$$NRMSE = \sqrt{\frac{\sum_{n=1}^N (Y_n - \hat{Y}_n)^2}{\sum_{n=1}^N (Y_n - \bar{Y}_n)^2}} \tag{13}$$

where N denotes the length of the predicted time series, Y_n denotes the target data, \bar{Y}_n denotes the average of target data and \hat{Y}_n denotes the predicted data. The above-mentioned ESN and its improved seven methods were used for 30 simulation experiments, and the error results were averaged. The predicted NRMSE results are shown in **Tables 1-4**.

Table 1. Error results of predicting newly confirmed cases.

Newly Confirmed Cases	Methods						
	ESN	SWESN	Ridge-SWESN	Lasso-SWESN	SCAD-SWESN	MCP-SWESN	Improved-ESN
NRMSE	0.0410	0.0215	0.0129	0.0148	0.0113	0.0099	0.0087

Table 2. Error results of predicting newly suspected cases.

Newly Suspected Cases	Methods						
	ESN	SWESN	Ridge-SWESN	Lasso-SWESN	SCAD-SWESN	MCP-SWESN	Improved-ESN
NRMSE	0.1506	0.1022	0.0832	0.0686	0.0492	0.0421	0.0410

Table 3. Error results of predicting cure rate.

Cure Rate	Methods						
	ESN	SWESN	Ridge-SWESN	Lasso-SWESN	SCAD-SWESN	MCP-SWESN	Improved-ESN
NRMSE	0.0674	0.0469	0.0322	0.0316	0.0288	0.0212	0.0205

Table 4. Error results of predicting mortality rate.

Mortality Rate	Methods						
	ESN	SWESN	Ridge-SWESN	Lasso-SWESN	SCAD-SWESN	MCP-SWESN	Improved-ESN
NRMSE	0.1812	0.1021	0.0815	0.0782	0.0635	0.0565	0.0542

It can be seen from **Tables 1-4** that the prediction method proposed in this paper has lower NRMSE than the other six prediction methods for newly confirmed cases, newly suspected cases, cure rate and mortality rate, which shows that the proposed method has stronger prediction ability for COVID-19 nonlinear time series.

4. Conclusion

In this paper, a prediction method based on improved Echo State Network for COVID-19 nonlinear time series is proposed. The improved small world is used as the reservoir of Echo State Network to shorten the training time of the model and improve the prediction accuracy for COVID-19 nonlinear time series, the MCP regularization method is used to optimize the output weight of Echo State Network to solve the over fitting problem and improve the nonlinearity of the prediction model fitting ability for COVID-19 nonlinear time series; the ABC algorithm based on crossover and crowding strategy is adopted to optimize the parameters of MCP to improve MCP performance. Finally, seven prediction methods of ESN, SWESN, Ridge-SWESN, Lasso-SWESN, SCAD-SWESN, MCP-SWESN and the proposed Improved-ESN method in this paper were simulated to predict

COVID-19 nonlinear time series. The results show that the prediction method based on improved Echo State Network for COVID-19 nonlinear time series has stronger prediction ability.

Acknowledgements

The Public Welfare Technology Research Projects of Zhejiang Province of China under Grant No. LGG20F010009, the Zhejiang Shuren University Basic Scientific Research Special Funds (2020XZ009).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Jaeger, H. (2007) Echo State Network. *Scholarpedia*, **2**, 1479-1482. <https://doi.org/10.4249/scholarpedia.2330>
- [2] Lukoševičius, M. and Jaeger, H. (2009) Reservoir Computing Approaches to Recurrent Neural Network Training. *Computer Science Review*, **3**, 127-149. <https://doi.org/10.1016/j.cosrev.2009.03.005>
- [3] Kawai, Y., Park, J. and Asada, M. (2019) A Small-World Topology Enhances the Echo State Property and Signal Propagation in Reservoir Computing. *Neural Networks. The Official Journal of the International Neural Network Society*, **112**, 15-23. <https://doi.org/10.1016/j.neunet.2019.01.002>
- [4] Chouikhi, N., Ammar, B., Rokbani, N. and Alimi, A. (2017) PSO-Based Analysis of Echo State Network Parameters for Time Series Forecasting. *Applied Soft Computing*, **55**, 211-225. <https://doi.org/10.1016/j.asoc.2017.01.049>
- [5] Li, H. (2013) Prediction of Nonlinear Time Series Based on Echo State Network. Dalian University of Technology, Dalian.
- [6] Qiao, Y.Q., Xiao, J.H., Huang, Y.H. and Yin, K.Y. (2015) Randomized Hough Transform Straight Line Detection Based on Least Square Correction. *Computer Applications*, **35**, 3312-3315.
- [7] Wang, G.T., Li, P. and Su, C.L. (2011) ELM Ridge Regression Learning Algorithm of Ridge Parameter Optimization. *Information and Control*, **40**, 497-506.
- [8] Lu, Z.X., Huang, J.S., Tu, L.Y., Xu, X.J. and Zhang, D.Q. (2018) Tensor-Based Regularized Multilinear Regression and Its Application. *Journal of Frontiers of Computer Science and Technology*, **118**, 121-129.
- [9] Han, M. and Xu, M. (2015) Predicting Multivariate Time Series Using Subspace Echo State Network. *Neural Processing Letters*, **41**, 201-209. <https://doi.org/10.1007/s11063-013-9324-7>
- [10] Fan, J.Q. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. <https://doi.org/10.1198/016214501753382273>
- [11] Zhang, C.H. (2010) Nearly Unbiased Variable Selection under Minimax Concave Penalty. *The Annals of Statistics*, **38**, 894-942. <https://doi.org/10.1214/09-AOS729>
- [12] Zhang, G.G., Xu, Z., Zeng, B. and Chen, X.T. (2017) Time Series Prediction Model Based on SCAD-ESN. *Engineering Science and Technology*, **49**, 129-134.
- [13] Yao, L. and Ma, X.J. (2016) Variable Selection in Variable Coefficient Model. *Statistics and Decision Making*, No. 12, 10-12.