

RESEARCH PAPER

Time Series Prediction Model of Grey Wolf Optimized Echo State Network

Huiqing Wang¹, Yingying Bai¹, Chun Li¹, Zhirong Guo¹ and Jianhui Zhang²¹ College of Information and Computer, Taiyuan, University of Technology, Taiyuan, Shanxi, CN² Postal Savings Bank of China, CN

Corresponding author: Chun Li (429092823@qq.com)

As a novel recursion neural network, Echo State Networks (ESN) are characterized by strong nonlinear prediction capability and effective and straightforward training algorithms. However, conventional ESN predictions require a large volume of training samples. Meanwhile, the time sequence data are complicated and unstable, resulting in insufficient learning of this network and difficult training. As a result, the accuracies of conventional ESN predictions are limited. Aimed at this issue, a time series prediction model of Grey Wolf optimized ESN has been proposed. W^{out} of ESN was optimized using the Grey Wolf algorithm and predictions of time series data were achieved using simplified training. The results indicated that the optimized time series prediction method exhibits superior prediction accuracy at a small sample size, compared with conventional prediction methods.

Keywords: Echo State Network; Grey Wolf optimization; time series prediction

1. Introduction

The time series are a group of random variables arranged in time order and has been widely applied in our daily life and industry, including commerce, meteorology, finance, and agriculture. To fully understand universal laws and provide references to optimized decision-making, great attentions have been invested in time series predictions (Rojas and Pomares 2016) (Jiang et al. 2017).

Owing to effects by various factors, time series is usually characterized by significant randomness and nonlinearity (Chen et al. 2017). For accurate predictions of time sequence data, various time series prediction models have been proposed. For instance, ARIMA is a conventional linear time series prediction model with considerable prediction capability (Liu et al. 2016). However, continuously increasing data size leads to increasing size of nonlinear time series data and applications of linear prediction models have been limited. In virtue of rapid developments in machine learning and artificial intelligence, novel time series prediction models such as neural networks (Chandra 2015) (Wen et al. 2012) and support vector machine (SVM) (Yaseen et al. 2016) (Misaghi and Sheijani 2017) (Nieto et al. 2017) (Jaramillo et al. 2017) have been proposed and widely applied in nonlinear time series predictions. Ren et al. achieved rapid collections of receptor temperature using the reverse propagation neural network and rapid prediction of collector tube temperature based on the data collected (Ren et al. 2016). However, traditional neural networks are prone to fall into local optimum and dimensionality disasters due to the uncertainty of their structure. Huang et al. reported effective predictions of mammary cancer using the SVM algorithm (Huang et al. 2017). Although the algorithm overcomes the problems of traditional neural network prediction, its sequence suitable for processing is limited. Yao et al. proposed a RNN-based double layer mechanism model (denoted as DA-RNN) (Qin et al. 2017). This model can obtain the long-term time sequence dependency relationship and predict by selecting relevant time sequence driving series. This model can obtain the long-term time sequence dependency relationship and predict by selecting relevant time sequence driving series. The network needs to calculate the error gradient in the training process. Because the error ladder method is difficult to train the network, the inherent disadvantage of this difficult training limits the wide application of the recurrent neural network in practice (Egrioglu et al. 2015). In addition, since the weighting requirements of the training algorithm are

continuously updated, and the update process is computationally intensive, the training time of the RNN network is increased.

The Echo State Network (ESN) (Sacchi et al. 2007) is a novel recursion network proposed by Jaeger in 2001 and has been applied in time series predictions. ESN is improved on the basis of the traditional recurrent neural network. The network structure is unique. The concept of “reservoir pool” is introduced, which can better adapt to the application of nonlinear systems. The network training process uses linear regression and has short-term memory function and the network model is simple and fast, has high prediction performance, overcomes the problems of large computational complexity, low training efficiency and local optimization in traditional recurrent neural networks, and can be adapted to the processing of time series data in practical problems. However, ESN requires a large volume of training samples owing to its unique structure and training is difficult in ESN. Meanwhile, reservoir applicability and prediction accuracy in complicated situations needs to be improved. Afterwards, various optimized ESN models have been proposed. Qin et al. proposed a novel E-KFM model by combining the KFM algorithm and ESN and applied it for multi-step prediction of time sequence data (Xiao et al. 2017). The E-KFM model exhibits excellent effectiveness and robustness, but did not achieve good optimization results. Qiao et al. reported accurate predictions by ESN via optimization of W^{out} of ESN using particle swarm algorithm (Qiao et al. 2016), and proposed the PSO_ESN prediction model. The model improves the prediction performance to a certain extent, but the model training time is longer due to the evolution of particles and the number of iterations. Zhong et al. reported optimization of double layer ESN by genetic algorithms and applied the optimized model in multi-area time series prediction (Zhong et al. 2017). The prediction model optimizes the echo state network, improves the accuracy of time series prediction, and shortens the prediction time to some extent. However, the genetic algorithm has complex coding, many parameters and choices rely on experience, which cannot solve the problem of large-scale calculation.

Aimed at this issue, a time series prediction model of Grey Wolf optimized ESN is proposed by introducing the Grey Wolf algorithm, a swarm intelligence optimization algorithm. First, significance of time series predictions and the state-of-the-other studies in this field were introduced. Then, the GWO time series prediction method for ESN was proposed and described in details. Finally, the proposed model is verified based on different data sets.

2. Time series prediction method of Grey Wolf optimized ESN

In order to solve issues (e.g., difficult training) in ESN predictions, time series prediction model of Grey Wolf optimized ESN is proposed. This method eliminates the issue of difficult training by optimizing W^{out} using the Grey Wolf algorithm and improves the accuracy of ESN prediction. Additionally, experiments demonstrated significant enhancements of prediction accuracy of the proposed prediction method over different time series data sets.

2.1. Echo State Network

ESN is a novel recursion neural network consisting of input layer, hidden layer, and output layer (Lun et al. 2015) (Han and Mu 2011). As shown in **Figure 1**, layers are connected to each other via different weight matrices. Herein, W^{in} refers to the input weight connection matrix, which is the connection between input

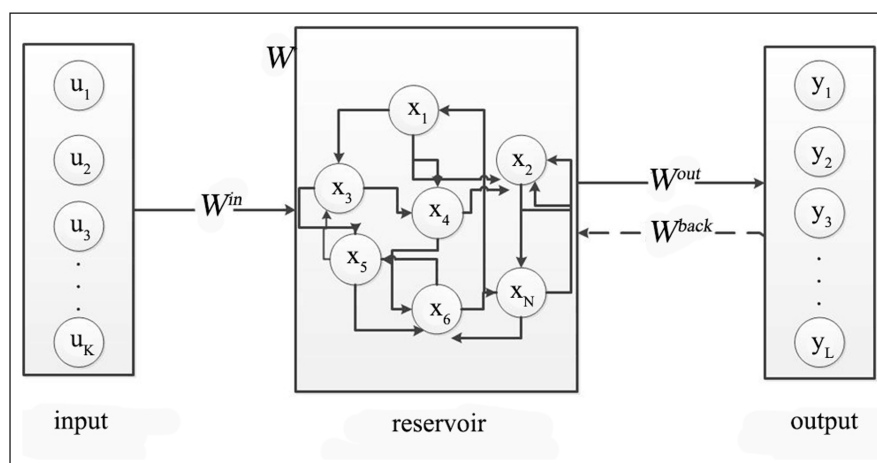


Figure 1: Echo State Network diagram.

layer and reservoir; W refers to the internal weight connection matrix, which is the connection between reservoir and internal neuron; W^{back} refers to the feedback weight connection matrix, which is the connection between output layer and the next output layer; W^{out} refers to the output weight connection matrix, which is the connection between reservoir and output layer. W^{out} is the only key parameter that requires training.

Unlike other neural networks, the hidden layer in this network is replaced by reservoir. Herein, the reservoir consisting of various sparse neurons dynamically connected to each other and it exhibits memory capability via performance of weight storage system between neurons. The reserve pool is the core part of the ESN network, and its parameters are of great significance to the network, including the size of the reserve pool N , the internal connection weight spectrum radius SR of the reserve pool, the input unit scale IS and the sparsity degree SD . Among them, the size of the reserve pool N is reflected by the number of neurons. The size of the scale N affects the predictive power of the ESN network. In general, the size of N is adjusted by the number of data sets. The internal connection weight spectrum radius of the reserve pool is a key parameter of the reserve pool, affecting its memory capacity. In general, an ESN network can have a stable echo state attribute when $0 < SR < 1$. Due to the different types of neurons in the reserve pool and the different characteristics of the data, the input signal needs to be scaled by the reserve cell input unit size IS to be transported from the input layer to the reserve pool. The size of the input unit scale is related to the nonlinear data to be processed. The stronger the nonlinearity, the larger the input unit scale. The sparsity of the reserve pool SD specifically refers to the proportion of neurons connected in the reserve pool to the total number of neurons. In general, when the SD is 10%, the reserve pool can maintain certain dynamic characteristics.

The basic equations of ESN are:

$$x(n+1) = f(W^{in}u(n+1) + W_x(n)W^{back}y(n)) \quad (1)$$

$$y(n+1) = f^{out}(W^{out}u(n+1), x(n), y(n)) \quad (2)$$

where $u(n) = u_1(n), u_2(n), \dots, u_k(n)$, $x(n) = x_1(n), x_2(n), \dots, x_N(n)$, $y(n) = y_1(n), y_2(n), \dots, y_L(n)$ are input vector, state vector, and output vector of ESN, respectively; f and f^{out} are activation functions for internal neurons of processing unit and output unit of the reservoir, respectively, and they are generally tanh functions.

2.2. Grey Wolf Optimizer

The Grey Wolf Optimizer (GWO) is a novel swarm intelligence algorithm proposed by Mirjalili in 2014 (Saremi et al. 2015) (Rezaei et al. 2018). This algorithm is based on mimicking of social hierarchy and hunting activities of grey wolf herd. Grey wolves are social animals with strict hierarchy, including α , β , δ , and ω . Herein, α is the leader who distribute different tasks (surrounding, hounding, attacking) to individuals of different levels to achieve global optimization. In virtue of its simple structure, negligible parameter adjustment required, and high effectiveness, the GWO algorithm has been widely applied for function optimizations.

For a population consisting of N grey wolves ($X = X_1, X_2, \dots, X_N$), the location of the i th wolf is defined as $X_i = X_i^1, X_i^2, \dots, X_i^d$ and X_i^d refers to the location of the i th wolf is d -dimensional space. The specific hunting activity is defined as follows:

$$D = |C * X_j(t) - X(t)| \quad (4)$$

$$X_i(t+1) = X(t) - A * D \quad (5)$$

where A and C are coefficient vectors, t is the iteration number, $X(t)$ is the location vector of a grey wolf, $X_j(t)$ is target location vector of the grey wolf, D is the distance between the grey wolf and the prey.

The coefficient vector is defined as follows:

$$A = 2 * a * r_1 - a \quad (6)$$

$$C = 2 * r_2 \quad (7)$$

$$a = 2 - i * \left(\frac{2}{\text{Max_iteration}} \right) \quad (8)$$

where r_1 and r_2 are random vectors with values in $[0, 1]$ and a is the iteration factor.

Grey wolves have a strong prey search capability. α is the leader who command all activities and β and δ may participate occasionally. In the GWO algorithm, α is defined as the optimal solution, while β and δ can also provide effective target information to α . Therefore, α , β , and δ are the three optimal solutions currently and their updated locations are as follows:

$$D_\alpha = |C_1 * X_\alpha(t) - X(t)| \tag{9}$$

$$D_\beta = |C_2 * X_\beta(t) - X(t)| \tag{10}$$

$$D_\delta = |C_3 * X_\delta(t) - X(t)| \tag{11}$$

$$X_1 = X_\alpha - A_1 * D_\alpha \tag{12}$$

$$X_2 = X_\beta - A_2 * D_\beta \tag{13}$$

$$X_3 = X_\delta - A_3 * D_\delta \tag{14}$$

$$X(t+1) = \frac{(X_1 + X_2 + X_3)}{3} \tag{15}$$

where X_α , X_β , and X_δ are current locations of α , β , and δ , respectively; $X(t)$ is the target location of grey wolf; D_α , D_β , and D_δ are distances from the prey to α , β , and δ , respectively; $X(t + 1)$ refers to the location vector with updated searching factor; C and A are random vectors.

2.3. The optimized Echo State Network algorithm

As a key parameter in ESN, W^{out} is selected by a series of linear regressions of data in the training set. Owing to its unique structure, ESN requires a large volume of training samples, making its training highly challenging. Therefore, W^{out} was optimized using the Grey Wolf algorithm and a Grey Wolf optimized echo state network algorithm (denoted as the GWO_ESN algorithm) is proposed.

Procedures of the GWO_ESN algorithm are as follows:

- a) Establish ESN as shown in **Figure 1** and initialize the parameters of this network.
- b) Initialize parameters and location functions and target location functions of α , β , δ , as shown in **Figure 2**. Herein, the initial locations of α , β , δ are fixed and the corresponding parameters are C and a , respectively. ω is the wolf at the bottom hierarchy and prey is located in the middle part.
- c) Calculate the value of fitness function using Eq (16) and compare it with the value of target function in Step b. Herein, y_i is predicted value based on W^{out} and Eq (2) and y is the practical value.

$$fitness = \frac{1}{n} \sum_{i=1}^n (y_i - y)^2 \tag{16}$$

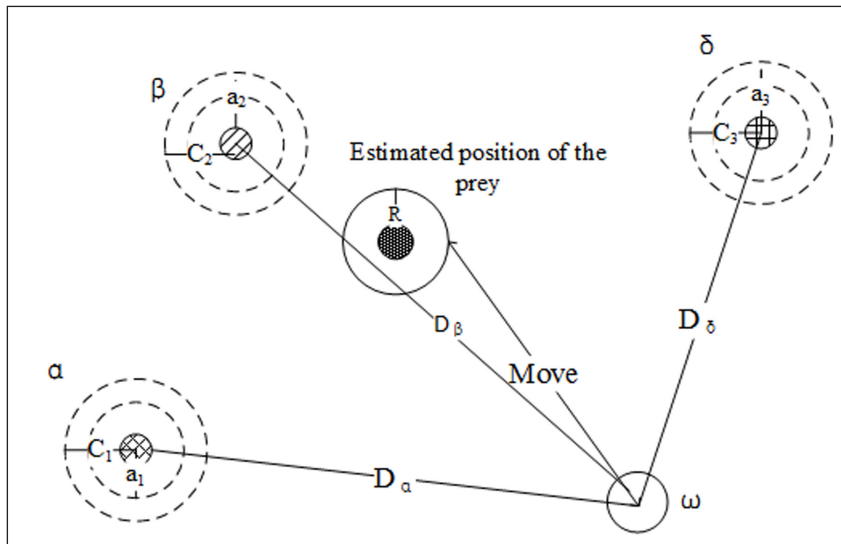


Figure 2: Grey wolf's initial position.

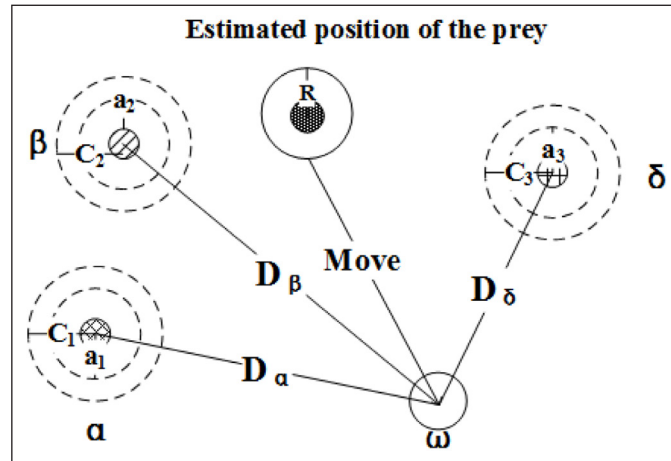


Figure 3: Grey wolf location update.

- d) If the values of fitness function obtained in Step c are lower than target function values of α , β , δ , target function values of α , β , δ are updated to fitness function values.
- e) Calculate Parameter a in each iteration using Eq (8) and coefficient coefficients (A and C) corresponding to α , β , δ using Eq (6) and (7).
- f) Execute time series transversal and update locations of α , β , δ using Eq (4) and (5). The specific updating equations are Eq (9), (10), (11), (12), (13), and (14). **Figure 3** shows updated locations of grey wolves.
- g) If the maximized iteration number is not achieved, go back to Step b and repeat the process; if the maximized iteration number is achieved, obtain updated locations of α , β , δ and calculate the ultimate optimization result (W^{out}) using Eq (15).

The pseudo code of GWO_ESN:

Algorithm GWO_ESN

Optimize W^{out}

function GWO_ESN ($X_i, a, A, C, N, W^{in}, W, W^{back}$)

position = initialization (m, dim);

W^{out} = position

do

fitness = ESN (U, Y, M, W^{out});

If fitness < X_α

fitness = X_α

X_1 = position

end

if fitness > X_α && fitness < X_β

fitness = X_β

X_2 = position

end

if fitness > X_α && fitness > X_β && fitness > X_δ

fitness = X_δ

X_3 = position

end

$t=0$

for X_1, X_2, X_3

update by Equation (12 13 14)

end for

update a, A , and C

update $X_\alpha, X_\beta, X_\delta$

$t=t+1$

until ($t > \text{Max_iteration}$)

$W^{out} = 3 / \text{sum}(X_1 + X_2 + X_3)$

return W^{out}

end function

2.4. Time series prediction model of GWO_ESN

In this article, a time series prediction model of Grey Wolf optimized ESN (denoted as the GWO_ESN model) combining ESN and the Grey Wolf algorithm is proposed. Herein, W^{out} of ESN is optimized using the Grey Wolf algorithm and the proposed GWO_ESN algorithm is applied in time series predictions. This model eliminates the issue of over-large volume of training samples in ESN and improves the prediction accuracy.

- Step1: Pre-process the original sequence and obtain de-noising and dimensionality reduced normalized data.
 Step2: Initialize parameters in ESN and Grey Wolf algorithm.
 Step3: Optimize W^{out} of ESN using the GWO algorithm.
 Step4: Predict using ESN based on W^{out} .

3. Results and discussion

3.1. Background and data

The experiment environment includes Matlab R2014b, Windows 7 Basic, 8G memory, Intel(R) Core(TM) i7-4790 CPU @ 3.60 GHz.

In order to better verify the performance of the time series prediction model, this experiment selected seven sets of data, of which the first five groups are nonlinear data., including the EEG public EEG data EEG, China Statistical Yearbook official website 1999–2008 different influencing factors The Shanghai Railway Index in the historical stock index data of the railway passenger traffic volume, China's 1985–2011 grain production data 1, 2 and Netease Financial Network 1990/12/20–1991/1/24. The latter two groups are chaotic time series data, mainly Lorenz chaotic sequence and Mackey-Glass chaotic sequence. The specific nonlinear data set information is shown in **Table 1**. The chaotic time series is defined as follows:

- (1) The Mackey-Glass chaotic time series is defined by the following time delay differential equation:

$$\frac{dx}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \quad (17)$$

Where $x(0) = 1.2$, $\tau = 17$, iteratively generates chaotic time series using the fourth-order Runge-Kutta method.

- (2) Lorenz chaotic time series

The Lorenz chaotic time series is described by the following three-dimensional ordinary differential equations:

$$\begin{cases} dx/dt = a(y-x) \\ dy/dt = (c-z)x - y \\ dz/dt = xy - bz \end{cases} \quad (18)$$

When the parameters $a = 10$, $b = 8/3$, $c = 28$, the initial value $x(0) = y(0) = z(0) = 1$, the system generates chaos, which is iteratively generated by the fourth-order Runge-Kutta method. Chaotic time series. The delay time and the embedding dimension of the sequence are set as: $\tau_1 = 19$, $\tau_2 = 13$, $\tau_3 = 12$, $m_1 = 3$, $m_2 = 5$, $m_3 = 7$.

Table 1: Data set information.

No.	Datasets	Data Length	Training set	Testing set
1	Separation of EEG data	5001*1	2000	500
2	Railway passenger traffic	34*8	16	16
3	Food production 1	27*8	13	13
4	Food production 1	10*10	5	5
5	The Shanghai Composite Index	400*7	200	199
6	Mackey-Glass	400*1	200	199
7	Lorenz	600*1	300	299

3.2. Evaluation standards

To compare accuracies of different prediction models and evaluate performance of the proposed GWO time series prediction method for ESN, two evaluation parameters are involved: comparison of fitting of predicted sequence and actual sequence and mean square error (MSE) of predicted values and actual values. The MSE as an evaluation parameter in this study is defined as:

$$\text{MSE} = \frac{\sum_{i=1}^n (\hat{y} - y)^2}{n} \quad (19)$$

where \hat{y} refers to the prediction value, y refers to the measured value, n refers to the data length.

3.3. Results and analysis

3.3.1. Experiment 1: Fitting of prediction curves by different time series prediction models and practical curves

The BP neural network model (Zhai and Cao 2016), the Elman neural network model (Liang et al. 2017), and the ESN model (Li et al. 2012), ESN prediction model based on recursive least squares (denoted as RLS_ESN) (Chouikhi et al. 2017), PSO optimization based ESN model (denoted as PSO_ESN) (Zhang et al. 2015), and the proposed GWO_ESN model were involved in this prediction experiment. The prediction results by these five models over the five time series data sets were compared with practical results and to each other in the way of fitting graphs.

Figure 4 shows fitting of practical results and prediction results by the five prediction models over five data sets. As observed, fitting of practical results and prediction results by the five prediction models

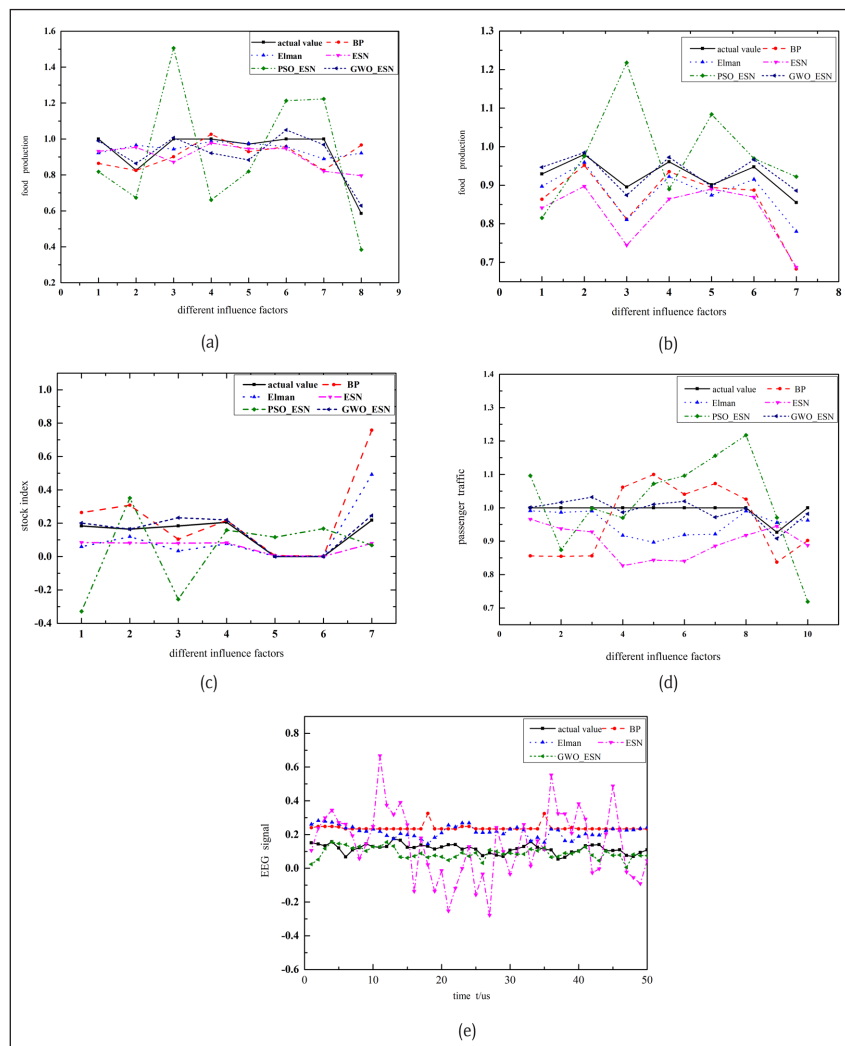


Figure 4: Comparison of different model predictions.

follows: GWO_ESN model > PSO_ESN model > RSN model > Elman model > BP model. Herein, prediction results by the proposed GWO_ESN model were highly fitted with practical results and the amplitudes were relatively small, indicating high prediction accuracy of the proposed model. In terms of different data sets, performances of the BP network and the Elman network varied significantly with the data set due to their poor structural stabilities. Meanwhile, the Elman network shows advantages in applicability to time series cases over the BP network, thus presenting good prediction performance.

Compared with the Elman model and the BP model, the ESN model exhibits excellent prediction accuracy in stock data set and EEG data set. As shown in **Figure 4** (d) and (e), stock data set and EEG data set are characterized by large volume of training samples and predictions by ESN are based on sufficient training in these cases. For the other three data sets with relatively small training sample sizes, sufficient training cannot be achieved and prediction accuracy of ESN was limited in these cases. On the other hand, the GWO_ESN model exhibited good prediction performances in all five cases, indicating strong generalization capability of this model. In other words, the GWO_ESN model is applicable for predictions of various time series data. Therefore, two conclusions can be drawn. First, fitting effectiveness of the ESN model varies significantly with the data set. Second, the GWO-ESN model shows excellent fitting effectiveness for all data sets, while the PSO-ESN model is inferior for history data sets but its effectiveness still satisfies the requirement. Additionally, predictions by the PSO-ESN model are significantly deviated from practical results for certain data sets. This can be attributed to data characteristics and parameter setting.

Experiment 1 demonstrated that fitting efficiency of the proposed GWO_ESN time series prediction model is significantly improved compared with the ESN model and the prediction results by the proposed GWO_ESN time series prediction model are perfectly aligned with practical results. Therefore, the prediction performance of the proposed GWO_ESN time series prediction model is considered to be optimized. Meanwhile, the proposed GWO_ESN time series prediction model is characterized by low time complexity, less parameters required, and highly effective algorithm compared with other models.

3.3.2. Experiment 2: MSEs of different prediction models for different data sets

Table 2 summarizes MSE of different data sets by the five prediction models. A low MSE indicates good model performance. As observed, MSE of data by the GWO-ESN model is significantly lower than that by other models, indicating excellent prediction performance of the GWO-ESN prediction model. Additionally, prediction accuracies of the PSO-ESN model and the ESN model are significantly higher than those of the BP neural network and the Elman network. Moreover, performances of the BP neural network model and the Elman network model are unstable and their prediction performances may surpass the ESN model and the PSO-ESN model in certain cases, but never the GWO-ESN prediction model.

In summary, the proposed GWO_ESN model exhibited excellent prediction performance even at small training sample size and it is superior to other models in terms of prediction accuracy. Meanwhile, due to its superior structural stability, the ESN network structure shows advantages in prediction based on nonlinear data over the BP neural network model and the Elman network model. Additionally, involvement of the GWO algorithm makes the proposed model leads to enhanced overall performance in all cases compared to the BP neural network model and the Elman network model. Sufficient learning of fluctuating data avoids performance degradation induced by any individual parameter.

Table 2: Mean square error comparison.

Number	BP	Elman	ESN	RLS_ESN	PSO_ESN	GWO_ESN
1	0.0357	0.0164	0.0250	0.0303	0.0217	0.0019
2	0.0413	0.0058	0.0306	0.0224	0.0272	6.2226e-5
3	0.0240	0.0253	0.0221	0.0189	0.0189	0.0013
4	0.0464	0.1284	0.0207	0.1023	0.0266	3.84e-6
5	0.2834	0.0887	0.0086	0.0005	0.1241	1.6817e-6
6	0.0362	0.0214	0.0122	0.0056	0.0435	0.0011
7	0.0413	0.0326	0.0237	0.0147	0.1267	2.65e-4

Table 3: Running time comparison(s).

Number	BP	Elman	ESN	RLS_ESN	PSO_ESN	GWO_ESN
1	5.0357	8.9908	3.1273	6.0547	240.4544	30.3024
2	4.1332	4.8048	2.0346	4.3509	80.0272	20.3445
3	3.0233	3.4559	0.2234	2.0465	100.3323	14.3445
4	2.1347	5.5456	0.4563	2.1342	90.2314	10.8436
5	3.4536	6.1877	1.1386	3.2432	130.3213	34.2564
6	4.5434	7.2331	2.0454	4.0989	205.4512	38.0921
7	4.6564	6.7789	1.8732	3.9807	180.3455	45.6733

3.3.3. Experiment 3: Run time of each prediction model under different data sets

Table 3 shows the comparison of the running time of the six predictive models on different datasets. It can be seen from the table that the GWO_ESN predictive model has relatively few running times under seven different datasets, although in some datasets the model running time is not dominant compared to the BP, Elman, and ESN prediction models, but it can be seen from **Table 2** that, in the case of ensuring higher prediction accuracy, the model has a relatively small running time compared to other optimization models.

4. Conclusions

In this paper, we proposed a GWO_ESN time series prediction model in which W^{out} of ESN is optimized using the Grey Wolf algorithm to solve difficult training issues in ESN induced by. Meanwhile, this model allows sufficient learning of fluctuating and nonlinear time series data. Compared with the PSO_ESN model, the RLS_ESN model, the ESN model, the BP neural network model, and the Elman network model, the proposed model exhibits advantage in prediction accuracy and reliability. In addition, parameters of the reserve pool in the ESN network in this experiment are mainly selected through empirical summary and multiple experimental results, and these parameters have certain influence on the experimental results, so find more suitable parameters to achieve better. The experimental effect is worthy of further study and discussion. Besides, performances of the proposed model for prediction of data distributions in other cases need to be verified.

Funding Information

This work was financially supported by the National Youth Science Foundation of China (No.61503272), the Scientific and technological project of Shanxi (No.201603D22103-2).

Competing Interests

The authors have no competing interests to declare.

References


- Chandra, R.** 2015. Competition and Collaboration in Cooperative Coevolution of Elman Recurrent Neural Networks for Time-Series Prediction. *IEEE Transactions on Neural Networks & Learning Systems*, 26(12): 3123. DOI: <https://doi.org/10.1109/TNNLS.2015.2404823>
- Chen, C, Twycross, J and Garibaldi, J.** 2017. A new accuracy measure based on bounded relative error for time series forecasting. *Plos One*, 12(3): 1–23. DOI: <https://doi.org/10.1371/journal.pone.0174202>
- Chouikhi, N, Ammar, B, Rokbani, N and Alimi, AM.** 2017. PSO-based analysis of Echo State Network parameters for time series forecasting. *Applied Soft Computing*, 55: 211–225. DOI: <https://doi.org/10.1016/j.asoc.2017.01.049>
- Egrioglu, E, Yolcu, U, Aladag, CH and Bas, E.** 2015. Recurrent multiplicative neuron model artificial neural network for non-linear time series forecasting. *Neural Processing Letters*, 41(2): 249–258. DOI: <https://doi.org/10.1007/s11063-014-9342-0>
- Han, M and Mu, DY.** 2011. LM algorithm in echo state network for chaotic time series prediction. *Control & Decision*, 26(10): 1469–1472.
- Huang, MW, Chen, CW, Lin, WC, Ke, SW and Tsai, CF.** 2017. SVM and SVM ensembles in breast cancer prediction. *Plos One*, 12(1): e0161501. DOI: <https://doi.org/10.1371/journal.pone.0161501>

- Jaramillo, J, Velasquez, JD and Franco, CJ.** 2017. Research in financial time series forecasting with SVM: Contributions from literature. *IEEE Latin America Transactions*, 15(1): 145–153. DOI: <https://doi.org/10.1109/TLA.2017.7827918>
- Jiang, P, Dong, Q, Li, P and Lian, L.** 2017. A novel high-order weighted fuzzy time series model and its application in nonlinear time series prediction. *Applied Soft Computing*, 55: 44–62. DOI: <https://doi.org/10.1016/j.asoc.2017.01.043>
- Li, D, Han, M, Wang, J.** 2012. Chaotic time series prediction based on a novel robust echo state network. *IEEE Trans Neural Netw Learn Syst*, 23(5): 787–799. DOI: <https://doi.org/10.1109/TNNLS.2012.2188414>
- Liang, Y, Qiu, L, Zhu, J and Pan, J.** 2017. A Digester Temperature Prediction Model Based on the Elman Neural Network. *Applied Engineering in Agriculture*, 33(2): 142–148. DOI: <https://doi.org/10.13031/aea.11157>
- Liu, C, Hoi, SCH, Zhao, P and Sun, J.** 2016. Online arima algorithms for time series prediction. In: *Thirtieth AAAI Conference on Artificial Intelligence*. AAAI Press, 1867–1873.
- Lun, SX, Yao, XS, Qi, HY and Hu, HF.** 2015. A novel model of leaky integrator echo state network for time-series prediction. *Neurocomputing*, 159(1): 58–66. DOI: <https://doi.org/10.1016/j.neucom.2015.02.029>
- Misaghi, S and Sheijani, OS.** 2017. A hybrid model based on support vector regression and modified harmony search algorithm in time series prediction. In: *2017 5th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*. IEEE, 54–60. DOI: <https://doi.org/10.1109/CFIS.2017.8003657>
- Nieto, PJG, García-Gonzalo, E, Fernández, JRA and Muñoz, CD.** 2017. A hybrid wavelet kernel SVM-based method using artificial bee colony algorithm for predicting the cyanotoxin content from experimental cyanobacteria concentrations in the Trasona reservoir (Northern Spain). *Journal of Computational & Applied Mathematics*, 309(1): 587–602. DOI: <https://doi.org/10.1016/j.cam.2016.01.045>
- Qiao, J, Li, R, Chai, W and Han, HJ.** 2016. Prediction of BOD based on PSO-ESN neural network. *Control Engineering*, 23(4): 463–467. DOI: <https://doi.org/10.15407/fm23.03.463>
- Qin, Y, Song, D, Chen, H, Cheng, W, Jiang, G and Cottrell, GJ.** 2017. A dual-stage attention-based recurrent neural network for time series prediction. *International Joint Conferences on Artificial Intelligence Organization*, 2627–2633. DOI: <https://doi.org/10.24963/ijcai.2017/366>
- Ren, T, Liu, S, Yan, G and Mu, HJ.** 2016. Temperature prediction of the molten salt collector tube using BP neural network. *IET Renewable Power Generation*, 10(2): 212–220. DOI: <https://doi.org/10.1049/iet-rpg.2015.0065>
- Rezaei, H, Bozorg-Haddad, O and Chu, X.** 2018. Grey Wolf Optimization (GWO) Algorithm. In *Advanced Optimization by Nature-Inspired Algorithms*. Springer, 81–91. DOI: https://doi.org/10.1007/978-981-10-5221-7_9
- Rojas, I and Pomares, H.** 2016. Time Series Analysis and Forecasting. *Contributions to Statistics*, 43(5): 175–197. DOI: <https://doi.org/10.1007/978-3-319-28725-6>
- Sacchi, R, Ozturk, MC, Principe, JC and Carneiro, AAFM.** 2007. Water Inflow Forecasting using the Echo State Network: a Brazilian Case Study. In: *International Joint Conference on Neural Networks*. DOI: <https://doi.org/10.1109/IJCNN.2007.4371334>
- Saremi, S, Mirjalili, SZ, Mirjalili, SM.** 2015. Evolutionary population dynamics and grey wolf optimizer. *Neural Computing and Applications*, 26(5): 1257–1263. DOI: <https://doi.org/10.1007/s00521-014-1806-7>
- Wen, L, Liang, XM, Long, ZQ, Qin, HY.** 2012. RBF neural network time series forecasting based on hybrid evolutionary algorithm. *Control & Decision*, 27(8): 1265–1268+1272.
- Xiao, Q, Chu, C, Zhao, L.** 2017. Time series prediction using dynamic Bayesian network. *Optik International Journal for Light and Electron Optics*, 135: 98–103. DOI: <https://doi.org/10.1016/j.ijleo.2017.01.073>
- Yaseen, ZM, Allawi, MF, Yousif, AA, Jaafar, O, Hamzah, FM, El-Shafie, A.** 2016. Non-tuned machine learning approach for hydrological time series forecasting. *Neural Computing & Applications*, 1–13.
- Zhai, J and Cao, J.** 2016. The combined prediction model based on time series ARIMA and BP neural network. *Statistics and Decision*, 3(4): 29–32.
- Zhang, Y, Yu, D, Seltzer, ML and Droppo, J.** 2015. Speech recognition with prediction-adaptation-correction recurrent neural networks. In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5004–5008. DOI: <https://doi.org/10.1109/ICASSP.2015.7178923>
- Zhong, S, Xie, X, Lin, L and Wang, F.** 2017. Genetic algorithm optimized double-reservoir echo state network for multi-regime time series prediction. *Neurocomputing*, 238: 191–204. DOI: <https://doi.org/10.1016/j.neucom.2017.01.053>

How to cite this article: Wang, H, Bai, Y, Li, C, Guo, Z and Zhang, J. 2019. Time Series Prediction Model of Grey Wolf Optimized Echo State Network. *Data Science Journal*, 18: 16, pp.1–11. DOI: <https://doi.org/10.5334/dsj-2019-016>

Submitted: 12 May 2018 **Accepted:** 03 April 2019 **Published:** 07 May 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 