Research on the Application of Echo State Network in Data Prediction

Cao Peng

Bank of Tibet Limited by Share Ltd, Lasa, Xizang, 850000, China
E-mail: yscls2013@163.com

Abstract

Echo state network is often used in the analysis of industrial data. The main drawback of this approach is that the prediction results are often locally optimal. In order to further improve the prediction accuracy of the echo state network, this paper designs a global optimization strategy based on the three level network model. In the optimization process, the core parameters and optimization methods of the echo state network are improved to achieve the goal of the global optimum. The improvement measures proposed in this paper can reduce the prediction error of echo state network in the practical application of data prediction.

Keywords: Data processing, core parameters, global optimization, local optimization

1. Introduction

In the field of industry and economy, it is difficult to find out whether there is correlation and causality between different problems directly from the surface. By means of the token variable corresponding to different problems to analyze the correlation and causality in the time series data, which is always a strong evidence for the final conclusion. At present, the analysis and prediction of time series data has become an important method to solve the problems of industrial technology and economic management[1-2].

Analysis and prediction of time series data , using statistical methods to abstract a set of data on the time dimension of the problem to be studied. and then build a applicable to the arrangement of the time series analysis model. This model can be used for the analysis of historical data, and can also be used to predict the future data[3]. In practical application, the analysis and prediction of time series data can be realized by curve fitting method. According to the historical data of the research object fitting a curve over time which can analyze the changes more intuitive in the rules, and then extend the curve in the future, which formed a prediction of the future of the data[4]. In recent years, various kinds of time series data prediction method is extracted, according to the different programming models can be divided into: prediction method based on support vector machine, prediction method based on self organizing feature map, prediction method based on extended Calman filter, prediction method based on artificial neural network[5-8].

Regardless of industrial technical problems or economic management problems, there is a certain degree of nonlinearity. The excellent performance of artificial neural network is that it can find a suitable neural network to approximate the system even if it is a very serious nonlinear system. Echo State Network (referred to as ESN), is a new type of neural network structure. Compared with classical neural networks, such as BP, the iteration process of ESN has better stability, and the approximation effect is more close to the optimal[9]. In particular, ESN does not fall into the local minimum as the classical

neural network method[10]. In addition, the ESN training process is simple and quick, and it is easier to use in the solution of complex relation.

ESN method has been applied in many fields since its appearance, but it also has some problems. The performance of ESN is mainly depended on the configuration of key parameters in practical application. From the existing situation, the parameter configuration of ESN mainly depends on the experience method and the test method, which not only has the problem of efficiency but also is difficult to reach the best. To this end, this paper proposes an improved method to enhance the application of ESN in time series data prediction.

## 2. Improved ESN Prediction Method

2.1. ESN Prediction Model Structure

From a structural perspective, ESN is the same as the classical neural network, including the input layer, the output layer and the hidden layer. Three levels of neuronal vector expression as shown in the formula (1).

$$I(k) = (i_1(k), i_2(k), \cdots, i_l(k))^T$$
$$C(k) = (c_1(k), c_2(k), \cdots, c_m(k))^T \qquad (1)$$
$$O(k) = (o_1(k), o_2(k), \cdots, o_n(k))^T$$

Here, $I(k)$ represents the input layer, contains a total of l neurons; $O(k)$ represents the output layer, a total of m neurons; $C(k)$ represents the hidden layer, a total of n neurons; k represents the iterative step of the network training process.

A typical ESN network structure is shown in Figure 1.
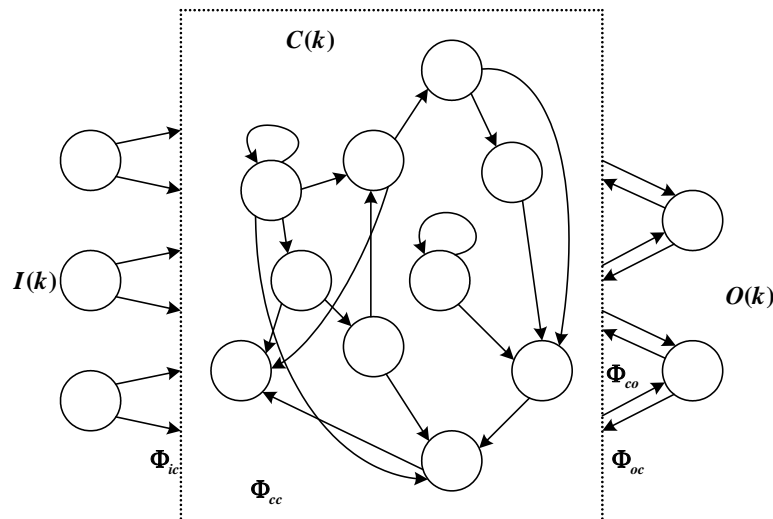
Figure 1. Echo State Network

In ESN, The relationship between input layer and the hidden layer is represented by the weight matrix Φ_{ic}; The relationship between the implicit and the output layer is represented by the weight matrix Φ_{co}; The feedback relationship between the output layer and the hidden layer is represented by the weight matrix Φ_{oc}; The relationship between each neuron in the hidden layer is represented by the weight matrix Φ_{cc}. The biggest difference between ESN and classical neural network lies in the configuration of

the hidden layer. The hidden layer of ESN contains a large size of neurons, and these neurons are connected in a sparse way. This makes ESN has a stronger approximation ability to a variety of complex relationships between input and output, and has great flexibility and short-term memory function.

2.2. ESN Forecast Process

Setting the process of training ESN can use M samples, including M input data samples $I(k)$, M output data samples $O(k)$, then the process of training ESN is as follows:

The first step is to initialize the entire ESN network. This initialization work includes initialization of the hidden layer neuron vector, the formation of $C(0)$, but also includes the connection matrix initialization, the formation of $\Phi_{ic}(0)$, $\Phi_{oc}(0)$, $\Phi_{cc}(0)$. It is necessary to point out that, $\Phi_{co}(0)$ finally calculated through the calculation, do not have to initialize.

The second step, update iteration operation. By means of the input data sample $I(k)$ and the output data sample $O(k)$, the state of the hidden layer $C(k)$ is continuously updated iteratively, and the iterative operation is updated as shown in the formula (2):

$$C(k+1) = f(\Phi_{ic}I(k+1) + \Phi_{cc}C(k) + \Phi_{oc}O(K)) \tag{2}$$

Here, $f(\cdot)$ represents a nonlinear function, which makes the ESN have the ability of strong approximation to nonlinear. The third step, vector data is processed. After waiting for ESN to enter the stationary state, the hidden layer neuron vector and the input layer neuron vector are sorted, and the state matrix is formed as follows:

$$S = \begin{bmatrix} C_{new}(M_0) \\ C_{new}(M_0+1) \\ \dots \\ C_{new}(M) \end{bmatrix} \tag{3}$$

Here, $C_{new}(M) = [C(M); I(M)]^T$ is expressed by the vector of the input layer neuron and the vector of the hidden layer neurons, and M_0 represents a certain iteration after the ESN enters the stationary state.

In the same way, the output matrix is sorted as follows:

$$D = \begin{bmatrix} O(M_0)^T \\ O(M_0+1)^T \\ \dots \\ O(M)^T \end{bmatrix} \tag{4}$$

Here, $T = M - M_0 + 1$.

Fourth step, calculate the connection matrix Φ_{co}. Here using the matrix of the pseudo inverse method, the formula is as follows:

$$(\Phi_{co})^T = (S^T S)^{-1} S^T D \tag{5}$$

The fifth step, use ESN to complete the forecast. After training to determine the various parameters of the ESN, connected to the weight matrix, you can predict the output according to the input data, the prediction formula is as follows:

$$C(k) = f(\Phi_{ic}I(1) + \Phi_{cc}C(k-1) + \Phi_{oc}O(K-1)) \tag{6}$$

$$O'(k) = \Phi_{co}[C(k); I(k)] \tag{7}$$

Here, $O'(k)$ is the predicted results, can be calculated by comparing the actual results of $O(k)$ to predict the error, and then through the error analysis to determine the merits of the ESN prediction effect.

2.3 Proposed Improvement Measures

According to the above process, it can be known that the prediction results of ESN are closely related to the four matrixes ESN, Φ_{ic}, Φ_{oc}, and Φ_{cc} are obtained by first initializing and then executing the ESN training , and Φ_{co} is calculated after the determination of Φ_{ic}, Φ_{oc}, and Φ_{cc}. Among them, the initialization of Φ_{ic}, Φ_{oc}, and Φ_{cc}, often depend on the method of experience and test. If the value of these initialization choices is not good, it will lead to the implementation of the training efficiency is reduced, and the results are difficult to achieve the best training. In view of the situation, this paper puts forward two improvement measures.

2.3.1 Optimization of Φ_{ic}, Φ_{cc}, and Φ_{oc}

First of all, make Φ_{ic}, Φ_{cc}, and Φ_{oc} further expressed as $\Phi_{ic} = s_{ic}\Phi'_{ic}$, $\Phi_{cc} = \rho\Phi'_{cc}$, $\Phi_{oc} = s_{oc}\Phi'_{oc}$, then the hidden layer of ESN will be updated as:

$$C(k+1) = f(s_{ic}\Phi'_{ic} I(k+1) + \rho\Phi'_{cc} C(k) + s_{oc}\Phi'_{oc} O(k)) \tag{8}$$

In the formula, The spectral radius of Φ'_{ic}, Φ'_{cc}, and Φ'_{oc} are 1, through the regulation of s_{ic}, ρ, and s_{oc}, it can be formed to meet the requirements of the spectral radius of Φ_{ic}, Φ_{cc}, and Φ_{oc}. In this way, the optimization problem of Φ_{ic}, Φ_{cc}, and Φ_{oc} is transformed into the optimization of s_{ic}, ρ, and s_{oc}.

The optimal allocation effect of the three matrices is to make the error of the prediction result and the actual result as small as possible. To this end, set the two optimization objectives are as follows:

$$\varepsilon(k) = O(k) - O'(k) \tag{9}$$

$$E(k) = \frac{1}{2}\|\varepsilon(k)\|^2 \tag{10}$$

Next, the random gradient descent method is used to optimize the three parameters of s_{ic}, ρ, and s_{oc}, which can be used in a unified formula:

$$p(k+1) = p(k) + \eta\frac{\partial E(k)}{\partial p} \tag{11}$$

p can take any one of these three parameters in s_{ic}, ρ, and s_{oc}, η represents the update speed.

In this way, the final determination of the Φ_{ic}, Φ_{cc}, and Φ_{oc} is the best results in line with the expected requirements through the sample training.

2.3.2 Calculation of Φ_{co}

The final determination of Φ_{co} is still use stochastic gradient descent method to calculate, and its iterative updating strategy is shown in the formula (12):

$$\Phi_{co}(k+1) = \Phi_{co}(k) + \lambda\frac{\partial E(k)}{\partial\Phi_{co}} \tag{12}$$

Here, λ also indicates the update speed. Because of $\dfrac{\partial E(k)}{\partial \Phi_{co}} = -\varepsilon(k)[C(k); I(k)]^T$, the above formula can be rewritten as follows:

$$\Phi_{co}(k+1) = \Phi_{co}(k) - \lambda\varepsilon(k)[C(k); I(k)]^T \tag{13}$$

So far, Φ_{ic}, Φ_{cc}, Φ_{oc} and Φ_{co} are all determined. Compared with the traditional experience method or test method, the four matrix parameters which are determined by this kind of improvement measures have the best characteristic, and have strong adaptability to different sample data.

3. Improved ESN Prediction Results and Error Analysis

3.1 Experimental Data

In order to verify the effectiveness of the improved ESN prediction method proposed in this paper, the relationship between the two sets of time series variable data is selected as the research object. One group is the inflation rate CPI monthly data of 2005-2014 as the input of the improved ESN method; the other group is the 2005-2014 M2 monthly data as the output of the improved ESN method. There is a large difference in the absolute value of the two groups data, so the log expression of the two groups of data is used as the experimental data, as shown in table 1.

Table 1. Original Analysis Data

| time	I(k)	O(k)	time	I(k)	O(k)
	L(CPI)	L(M2)		L(CPI)	L(M2)
2005.1	4.633077	12.48593	2010.1	4.626932	13.38465
2005.2	4.622322	12.52738	2010.2	4.633758	13.42087
2005.3	4.618383	12.56876	2010.3	4.639572	13.45378
2005.4	4.618777	12.60738	2010.4	4.651099	13.49501
2006.1	4.617099	12.64591	2011.1	4.657763	13.53861
2006.2	4.618777	12.68465	2011.2	4.667206	13.56810
2006.3	4.617790	12.71248	2011.3	4.662495	13.57650
2006.4	4.625267	12.75297	2011.4	4.645352	13.65486
2007.1	4.632104	12.80520	2012.1	4.640537	13.70525
2007.2	4.640537	12.84220	2012.2	4.626932	13.73755
2007.3	4.664382	12.88182	2012.3	4.624973	13.75756
2007.4	4.669365	12.90769	2012.4	4.626932	13.78937
2008.1	4.682409	12.95526	2013.1	4.625953	13.85097
2008.2	4.679999	13.00164	2013.2	4.628887	13.86858
2008.3	4.656528	13.02342	2013.3	4.629863	13.89004
2008.4	4.630253	13.07142	2013.4	4.630838	13.91672
2009.1	4.599253	13.18181	2014.1	4.630838	13.96453
2009.2	4.599957	13.25149	2014.2	4.633758	14.00535
2009.3	4.592389	13.28006	2014.3	4.625953	14.01201
2009.4	4.612146	13.32158	2014.4	4.635699	14.02122

3.2 Predict Results

From 40 groups of data in Table 1, select 20 sets of data to train the ESN, according to the improved ESN method to get the ESN of the key matrix parameters, so as to

determine the best ESN structure of the two groups experimental data. Then, the 40 L (CPI) data as input into ESN, and can obtain 40 L(M2)'s predicted values according to ESN. Draw a curve from the predicted value and the true value of the L (M2), the formation of the comparison results shown in figure 2.

From the results shown in Figure 2, the predicted value and true value of L(M2) has a lot of deviation at the beginning , but it adjusted to the more consistent state of the true value of L(M2) quickly in ESN approximation capability, and has maintained a very good approximation effect.

After 40 data, ESN further predict the next 20 data in the future time range, this extension of the prediction results can provide a statistical criterion for the future relationship of input and output.

The result in Figure 2 shows that the improved ESN method proposed in this paper is effective for the prediction of time series data.
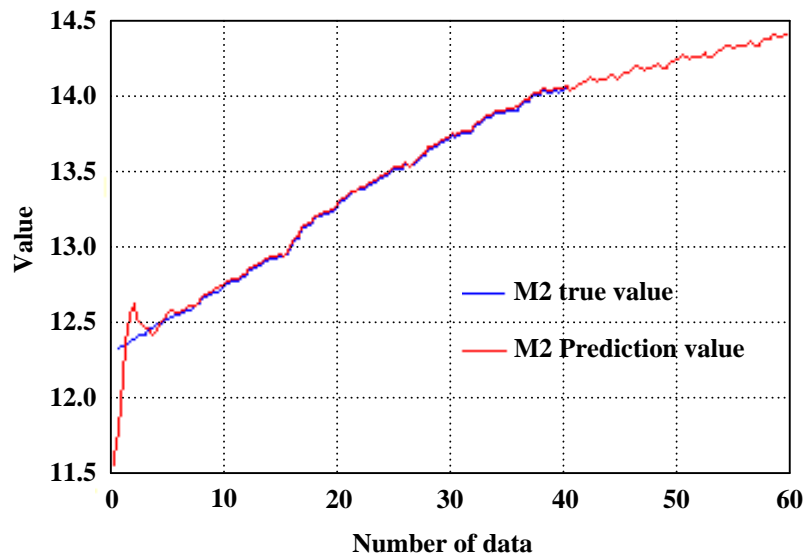
Figure 2. Experimental Results

3.3 Error Analysis

In order to form the horizontal contrast between the proposed method and the traditional ESN method, the prediction error analysis is performed. The prediction of the field of time series data , generally through the 4 error indexes to evaluate the forecasting effect, that are root mean square error, mean absolute error, mean absolute percentage error, standard root mean square error.

The root mean square error can be represented by:

$$\delta_{rmse} = \sqrt{\frac{1}{b-a}\sum_{n=a}^{b}(\hat{o}(n) - o_p(n))^2} \tag{14}$$

Here, $\hat{o}(n)$ represents the true value of the data , $o_p(n)$ represents the predicted value of the data , a represents the starting point for the execution of the predicted data sequence and b represents the end point for the execution of the predicted data sequence.

The mean absolute error can be represented by :

$$\delta_{mad} = \frac{1}{b-a}\sum_{n=a}^{b}\left|\hat{o}(n) - o_p(n)\right| \tag{15}$$

The mean absolute percentage error can be represented by:

$$\delta_{mape} = \frac{1}{b-a} \sum_{n=a}^{b} \frac{\left|\hat{o}(n) - o_p(n)\right|}{\left|o_p(n)\right|} \tag{16}$$

The standard root mean square error can be represented by:

$$\delta_{nrmse} = \sqrt{\frac{1}{(b-a)*\sigma^2} \sum_{n=a}^{b} (\hat{o}(n) - o_p(n))^2} \tag{17}$$

Here, σ^2 represents time series variance.

The experimental objects in the 3.1 section are performed by the traditional ESN prediction and improved ESN prediction respectively, the prediction error of the two methods are compared with the results shown in table 2.

Table 2. Predictive Performance Analysis Results

Error evaluation index	Traditional ESN method	Improved ESN method
δ_{rmse}	0.803	0.451
δ_{mad}	0.811	0.477
δ_{mape}	10.29%	5.01%
δ_{nrmse}	0.089	0.045

From the results of Table 2, we can see that the prediction effect of the improved ESN method is much better than the traditional ESN method, and the evaluation indexes of the 4 prediction errors are far lower than the traditional ESN method.

4. Conclusion

According to the prediction of time series data, an improved ESN prediction method is proposed. Based on the traditional ESN prediction, the key parameters of ESN are optimized. The optimization process maps the spectral radius of the key parameters again, and the optimization calculation is done with the method of stochastic gradient descent. The improved ESN method avoids the experience of key parameters and sets of tests, which makes the ESN have the best approximation properties. Experimental results on time series data show that the improved ESN method can obtain high accuracy of prediction, and the degree of agreement between the predicted value and the true value curve is very high.

References

[1] L. E. Vargas, R. D. Torres and T. C. Homero, "A study of wavelet analysis and data extraction from second-order self-similar time series", Mathematical Problems in Engineering, vol. 22, no. 7, (2013), pp. 111-120.

[2] K. B. Olafsdottir, M. Schulz and M. Mudelsee, "REDFIT-X: Cross-spectral analysis of unevenly spaced paleoclimate time series", Computers and Geosciences, vol. 91, (2016), pp. 11-18.

[3] Z. Q. and J. Lu, "The elements of statistical learning: data mining inference, and prediction", Journal of the Royal Statistical Society, vol. 173, no. 3, (2010), pp. 693-694.

[4] T. Hastie, R. Tibshirani, J. H. Friedman and J. Franklin, "The elements of statistical learning, second edition: data mining, inference, and prediction", Mathematical Intelligencer, vol. 27, no. 2, (2004), pp. 83-85.

[5] S. Zhu, J.-Z. Zhou, L. Ye and C.-Q. Meng, "Streamflow estimation by support vector machine coupled with different methods of time series decomposition in the upper reaches of Yangtze River", Environmental Earth Sciences, vol. 75, no. 6, (2016), pp. 209-215.

[6] M. Czyzewska, J. Szkola and K. Pancerz, "Self-organizing feature maps in correlating groups of time series: experiments with indicators describing entrepreneurship", Proceedings of the 21th International Workshop on Concurrency", Specification and Programming, vol. 928, (2012), pp. 73-78.

[7]   J.-C. Shen, C.-H. Chang, S.-J. Wu, C.-T. Hsu and H.-C. Lien, "Real-time correction of water stage forecast using combination of forecasted errors by time series models and Kalman filter method", Stochastic Environmental Research and Risk Assessment, vol. 29, no. 7, **(2015)**, pp. 1903-1920.

[8]   G.-L. Hou, T.-Y. Sun and F.-R. Huang, "Short-term wind speed prediction based on BP neural network with wavelet and time-series", ICIC Express Letters, vol. 8, no. 3, **(2014)**, pp. 867-874.

[9]   M. L. Alomar, V. Canals and N. P. Mora, "FPGA-based stochastic echo state networs for time-series forecasting", Computational Intelligence and Neuroscience, vol. 36, no. 3, **(2016)**, pp. 78-85.

[10] B. Meftah, O. Lezoray and A. Benyettou, "Novel approach using echo state networks for microscopic cellular image segmentation", Cognitive Computation, vol. 8, no. 2, **(2016)**, pp. 237-245.