Research on Extreme Learning Machine Algorithm and Its Application to El-Niño/La-Niña Southern Oscillation Model

De Xing, Weimin Zhang, Qunbo Huang, Bainian Liu
Academy of Ocean Science and Engineering
National University of Defense Technology
Changsha, China
xd_wony@icloud.com

Abstract—**Since it has the ability to give a faster result than traditional machine learning algorithms, Extreme Learning Machine (ELM) has become increasingly popular in various research fields recently. However, ELM has been worked on the research of computer science and other related areas except for the atmospheric field. This paper uses the ELM algorithm to simulate the forward integrating process of an ocean-atmosphere oscillator model -- El-Niño/La-Niña Southern Oscillation (ENSO). The results show that the ELM algorithm has a good accuracy and efficiency with a quick convergence speed and a strong resistance over observation noises.**

Keywords-Extreme Learning Machine; ocean–atmospheric oscillator; ENSO

I. INTRODUCTION

Since introduced the back-propagation (BP) algorithm, feedforward neural networks (FNN) have been well studied[1]. However, most traditional BP algorithms use first order gradient method to optimize parameters, which usually leads to a local optimal solution. Lots of efforts were made to improve the efficiency in training FNN, such as Hagan et al[2] and Li et al[3], but most methods still suffer the problems of time-consuming and a slow convergence speed. Therefore, Huang et al proposed a new kind of FNN called extreme learning machine (ELM)[4]. Compared to traditional FNN method, ELM has the ability to find a global optimum with less time, and thus drawn the attention of scientists in various fields.

Numerical Weather Prediction (NWP)solves atmospheric motion equations under the condition of given first guesses and boundary conditions, which are used for predicting future atmospheric states. How to solve these equations in a quick and accurate way has long been an important question for meteorologists. Since ELM has the advantages described above, this paper is aimed at introducing ELM to the simulation of ENSO and evaluate the accuracy and efficiency of the algorithm.

The structure of this paper is listed as follows. The following section introduces the structure and basic calculation principles of the ELM algorithm. Section III gives a brief introduction to the ENSO phenomenon. Section IV describes an ocean-atmosphere oscillator model of ENSO, which is used for our numerical experiments. Section V discusses and analyses the results of the numerical experiments. Finally, conclusions and plans for future work are given in section VI.

II. THE ELM ALGORITHM

The Extreme Learning Machine is a learning algorithm which randomly chooses hidden nodes and analytically determines the output weights of single-hidden layer feedforward neural networks (SLFN). Compared with traditional feedforward network learning algorithms (back-propagation algorithm, for instance), ELM has a much faster learning speed with a better performance. A typical ELM consists of three layers: the input layer, the hidden layer and the output layer. Fig. 1 demonstrates the structure of a typical SLFN.

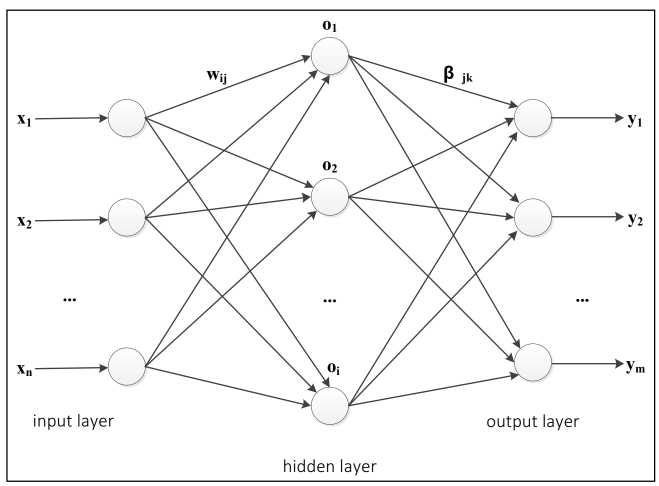

Figure 1. The structure of a typical SLFN

For N random distinct samples (x_i, t_i), where $= [x_{i1}, x_{i2}, \dots, x_{in}]^{\mathrm{T}} \in \mathcal{R}^n$ and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^{\mathrm{T}} \in \mathcal{R}^m$, a SLFN with $\widetilde{N}$ hidden nodes (where $\widetilde{N} \leq N$) and activation function $g(x)$ can be mathematically modeled as[5]:

$$\sum_{i=1}^{\widetilde{N}} \beta_i \, g\left(w_i \cdot x_j + b_i\right) = o_j, j = 1, \dots, N \qquad (1)$$

Where w_i and β_i are the weight vectors, b_i is the threshold of the i^{th} hidden node and o_j is the simulation result of the SLFN. If approximating that these N samples have zero error means, that is $\sum_{j=1}^{\tilde{N}}\|o_j - t_j\| = 0$, Eq. (1) can be rewritten as:

$$\sum_{i=1}^{\tilde{N}} \beta_i\, g(w_i \cdot x_j + b_i) = t_j, j = 1, \dots, N. \quad (2)$$

And combining these N equations we can get

$$\mathrm{H}\beta = \mathrm{T}. \quad (3)$$

Where H is called the hidden layer output matrix of the SLFN,

$$\mathrm{H} = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix} \quad (4)$$

and

$$\beta = \begin{bmatrix} \beta_1^{\mathrm{T}} \\ \vdots \\ \beta_{\tilde{N}}^{\mathrm{T}} \end{bmatrix}_{\tilde{N}\times m}, \mathrm{T} = \begin{bmatrix} t_1^{\mathrm{T}} \\ \vdots \\ t_N^{\mathrm{T}} \end{bmatrix}_{N\times m} \quad (5)$$

Unlike traditional first order gradient method that all the parameters need to be adjusted, the input weights w_i and hidden node threshold b_i can be given randomly and once they were set in the beginning of learning, the matrix H was determined and remained unchanged.

Since the minimum norm least-square solution of Eq. (3) is unique, it is apparent that

$$\hat{\beta} = \mathrm{H}^{\dagger}\mathrm{T}. \quad (6)$$

Where $\mathrm{H}^{\dagger} = (\mathrm{H}^{\mathrm{T}}\mathrm{H})^{-1}\mathrm{H}^{\mathrm{T}}$ is the Moore-Penrose generalized inverse matrix of H. Thus, the ELM algorithm can be summarized as follows:

Step 1: Generate the random (w_i, b_i) on each hidden node.

Step 2: Determine the hidden layer output matrix H using w_i and b_i.

Step 3: Calculate the output weight β using Eq. (6).

III. THE ENSO PHENOMENON

The El-Niño/La-Niña Southern Oscillation is a quasiperiodic climate pattern which happens in the ocean-atmosphere system in the region of tropical Pacific, resulting from the abnormal warming and cooling of the surface water of the central and eastern Pacific[6]. Southern Oscillation indicates the variations in equatorial eastern Pacific sea surface temperature (SST) and equatorial western Pacific sea surface pressure (SSP). These two events are interacted with each other: the warming of equatorial eastern Pacific SST, namely El-Niño, always comes with the increase of equatorial western Pacific SSP, and the cooling of equatorial eastern Pacific SST, namely La-Niña, always comes with the decrease of equatorial western Pacific SSP. China Meteorological Administration judges El-Niño and La-Niña events based on the SST anomalies in NINO1+2+3+4 zones. If the SST anomaly keeps≥0.5℃(≤-0.5℃) continuously for six months (can skip one month if its SST anomaly does not reach the level) or keeps ≥ 0.5 ℃ continuously for five months and the accumulated anomaly is ≥ 4 ℃ (≤ 4 ℃), the procedure is judged as an El-Niño(La-Niña) event.

The anomaly is the departure of an observation deviates from the mean, including positive and negative anomaly. In meteorology field, anomaly was mainly used to determine the data value retrieved among a certain time period is larger or smaller than the mean for a long time. The anomaly's accumulation among a certain period called the accumulated anomaly. Table. Ⅰ lists the five strength levels of ENSO:

TABLE I. THE STRENGTH LEVELS OF ENSO

ENSO level	Accumulated anomaly
extremely weak	≤3.8℃
weak	3.9-7.0℃
medium	7.1-13.2℃
strong	13.4-16.5℃
extremely strong	≥16.6℃

A large number of observations show that the energy of tropical Pacific ocean-atmosphere system releases intensely when ENSO happens, which can lead to a profound effect on global climate. ENSO usually causes storms and floods at coastal areas of the eastern Pacific as well as droughts at western Pacific and Indian Ocean rim countries and regions. So far, the 1997/98 Yangtze River flood triggered by El-Niño events is still fresh in Chinese people's memories.

However, to people's disappointment, ENSO is difficult to predict because of its complexity and uncertainty. During the first decade of the 21th century, the central location of ENSO has undergone a significant westward shift, causing the precipitation anomaly center to move westward to the western Pacific, which changed not only the affected areas but also the strength, frequency and transmission features of ENSO, making it more complex. This complexity directly leads to ENSO's uncertainty in its developing process and thus increases the difficulty in forecasting.

IV. AN OCEAN-ATMOSPHERIC OSCILLATOR MODEL OF ENSO

The ENSO phenomenon is very attractive among the international academic circles. Since the oscillatory nature of ENSO demands both positive and negative feedbacks[7] and a conceptual ENSO model should take both eastern and western Pacific abnormal patterns, an ocean-atmosphere oscillator model of ENSO can be denoted as:[8]

$$dT/dt = CT + Dh - \varepsilon T^3 \quad (7)$$

$$dh/dt = -ET - R_h h \quad (8)$$

$$T|_{t=0} = T_0,\ h|_{t=0} = h_0 \quad (9)$$

Where T is the equatorial eastern Pacific SST anomaly and h is the thermocline depth anomaly. C, D, E, R_h and ε are the positive model parameters, the explicit definitions of which are explained in [8].

To make it more convenient to verify the model accuracy, consider a special condition where $D = 0$, and then Eq. (7) can be rewritten as:

$$dT/dt = CT - \varepsilon T^3 \quad (10)$$

The nonlinear function above has the analytical solution of $T(t)$ at time t when $0 < \varepsilon \ll 1$[9-10]:

$$T(t) = [b + (1/{T_o}^2 - b)\exp(-2CT)]^{-1/2}, \quad (11)$$

where $b = \varepsilon/C$. Eq. (10) described the discharge and recharge of equatorial heat content. In the warm phase of ENSO, the equatorial eastern Pacific warm SST anomalies along with the divergence of Sverdrup transport associated with equatorial central Pacific westerly wind anomalies always lead to the discharge of equatorial heat content, which results in a transition phase where the entire equatorial Pacific thermocline depth becomes anomalously shallow. Based on this, the transition phase allows deep layer cold waters to be pumped into the sea surface and thus leads to the cold phase. Vice versa for the cold phase of ENSO. These recharge and discharge processes keep the ENSO system on internal time scale[11].

V. NUMERICAL EXPERIMENT

Two sets of experiments were taken, namely EXP.1 and EXP.2. The experiments try to find out how accurate the ELM method can predict the ocean-atmosphere oscillator model state and its efficiency.

EXP.1: First using a fourth-order Ronge-Kutta (RK4) method to integrate Eq. (10) forward 1500 time steps and get the output SST anomaly T at each time step. The model parameters are set to be: $T_0 = 0.5$, $C = 0.5$ and $\varepsilon = 0.01$. The sample quantity is 1000, and the time step length is 0.01. Each sample was assigned a Gaussian noise $\delta = N(0,1)$ to make it closer to real observation. Choose the output T at the 100th, 500th, 800th, 1000th, 1100th, 1200th, 1300th and 1500th time step as the train data set T_{train} and test data set T_{test} of the ELM (900 samples make up the train data set and the other make up the test data set). Use T_{train} to train the ELM and get the output T_{ELM}. Use Eq. (11) to get the analytical solution T_{ana}. Compare the ELM output T_{ELM} with T_{test} and T_{ana} and compare the time cost of ELM with that of RK4 to evaluate the accuracy and efficiency of ELM.

EXP.2: Amplify the Gaussian noise to 10* δ and repeat EXP.1 .

To evaluate the efficiency of the ELM method, the running times (in seconds) of the RK4 and ELM method after integrating 1500 time steps are shown in Table.Ⅱ:

TABLE II. THE RUNNING TIMES OF THE RK4 AND ELM METHOD

	RK4	ELM
EXP. 1	0.69432956	0.018574
EXP. 2	0.68562910	0.018363

It is clear that the ELM method has a much faster speed than the RK4 method.

The comparison of T_{test}, T_{ELM}, T_{ana} is shown in Fig. 2. It can be seen that the ELM method yields a rather accurate output in both experiments, but EXP.2 has a larger error (obviously at the 500th and 1000th time step).

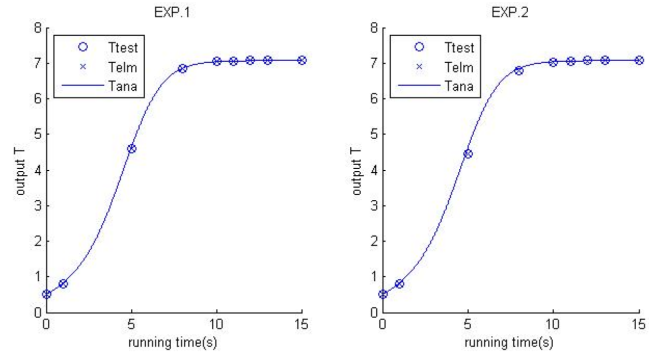

Figure 2. The comparison of T_{test}, T_{ELM}, T_{ana}

To make it more clear, we chose the absolute error sum $E_{test} = \sum_{n=1}^{100}|T_{ana(n)} - T_{test(n)}|$, $E_{ELM} = \sum_{n=1}^{100}|T_{ana(n)} - T_{ELM(n)}|$ and relative error of the ELM method $E_{rel} = (E_{ELM}/100)/T_{ana}$ of each experiment at time step 100, 500, 800, 1000, 1100, 1200, 1300, 1500. The results are shown in Table. III:

TABLE III. THE ABSOLUTE ERROR SUM AND RELATIVE ERROR OF EXP.1 AND EXP.2 AT CHOSEN TIME STEPS

Time step	EXP.1		
	E_{test}	E_{ELM}	$E_{rel}(\%)$
100	1.432240822400	1.432240742979	1.75
500	4.721987935999	4.721995191860	1.02
800	0.773136009999	0.773135664041	0.113
1000	0.113890560000	0.113890512076	0.0162
1100	0.042262100000	0.042262211781	0.00599
1200	0.015597289999	0.015597226880	0.00221
1300	0.005744590000	0.005744722832	0.000813
1500	0.000777939999	0.000777902502	0.000110
Time step	EXP.2		
	E_{test}	E_{ELM}	$E_{rel}(\%)$
100	14.640374036399	14.640375364431	17.9
500	50.673587896000	50.673609584708	10.1
800	10.114158048000	10.114087264106	1.48
1000	1.5503032479999	1.5501635436581	0.220
1100	0.5779303999999	0.5778635097216	0.0819
1200	0.2136572720000	0.2136315756423	0.0302
1300	0.0787431919999	0.0787332092790	0.0111
1500	0.0106665520000	0.0106652649766	0.00151

From Table. III we can come to two conclusions:

(1) ELM has a good accuracy, a fast convergence speed and can lead to a convergent solution. $\mathrm{E_{rel}}$ keeps decreasing in both experiments and from time step 800 in EXP.1 (1000 In EXP.2) it begins to be smaller than 1%, which indicates a good accuracy and a good convergence speed. Note that the sum of absolute error has a deceasing trend with an abnormal value at time step 500. This phenomenon can be explained by the fact that the output T has its maximum increasing speed at around time step 500 (see Fig.2), so a small disturbance can lead to a relatively bigger error. The

decreasing trend of $\mathrm{E_{ELM}}$ and $\mathrm{E_{rel}}$ shows ELM method can lead to a convergent solution.

(2) ELM has a strong resistance over observation noises. It is natural that the E_{ELM} and E_{rel} of EXP.2 are larger than that of EXP.1 for it has a larger Gaussian noise, but it is worthwhile to point out that the E_{ELM} and E_{rel} of EXP.2 decreases fast and keeps just a little larger than that of EXP.1 while the Gaussian noise of EXP.2 is 10 times of EXP.1's. This shows that ELM method has the ability to resist noises.

The errors between T_{ELM} and T_{ana} at time step 100, 500, 1000, 1500 of EXP.1 are given in figure 3, from which the accuracy of the ELM method is easily to be seen.

VI. CONCLUSIONS

Extreme learning machine has the ability to search for the global optimal solution in a quick way, and adapt it to the atmospheric and oceanic field has a promising future. This paper first introduces the ELM method and the ENSO phenomenon, and then adapts the ELM method to the ENSO model in the numerical experiments. The results present a faster calculation speed and a satisfying accuracy. Future work can be aimed at applying the ELM method to other more complex atmospheric and oceanic models.

ACKNOWLEDGMENT

Project supported by the Natural Science Foundation of China (Grant Nos. 41375113).

REFERENCES

[1] Huang, G., Huang G, B., Song, S., and You, K. "Trends in extreme leaning machines: a review." *Neural Networks,* 61(2015): 32-48.

[2] Hagan, M. T., and Menhaj, M. B., "Training feedforward networks with the Marquardt algorithm." *Neural Networks, IEEE Transactions on* 5.6 (1994):989-993.

[3] Li, K., Peng, J. X., and Irwin, G. W., "A fast nonlinear model identification method." *Automatic Control, IEEE Transactions on* 50.8 (2005): 1211-1216.

[4] Huang, G. B., Zhu, Q. Y., Siew, C. K., "Extreme learning machine: a new learning scheme of feedforward neural networks." *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*. Vol. 2. IEEE, 2004.

[5] Huang, G. B., Zhu, Q. Y., Siew, C. K., "Extreme learning machine: theory and applications." *Neurocomputing* 70.1 (2006): 489-501.

[6] Jia, M. O., "The variation iteration solving method for El-Niño/La-Niña southern oscillation model." *Advances in mathematics*. 35.2 (2006).

[7] Lin, W. T., Ji, Z. Z., Wang, B., Yang, X. Z., "A new method for judging the computational stability of the differences of linear and non-linear evolution qeuations.", *Chinese Science Bulletin*, 2000, 45(15): 1358-1361.

[8] Feng, G. L., Dong W. J., Jia X. J., "On the dynamics behavior and instability evolution of air-sea oscillator", *Chin. Phys. Soc*., 51 (2002): 1181-1185

[9] Mo, J. Q., Lin, W. T., Zhu, J., "The variational iteration solving method for El-Niño/La-Niña Southern Oscillation model." *Adv. Math.*, 35 (2006): 232

[10] He, J. H., Lee, E. W. M. "Variational principle for the differential-difference system arising in stratified hydrostatic flows." *Phys. Lett. A,* 373.18 (2009): 1644-1645

[11] Wang, C. Z., "On the ENSO mechanisms." *Advances in Atmospheric Sciences* 18.5 (2011): 674-691

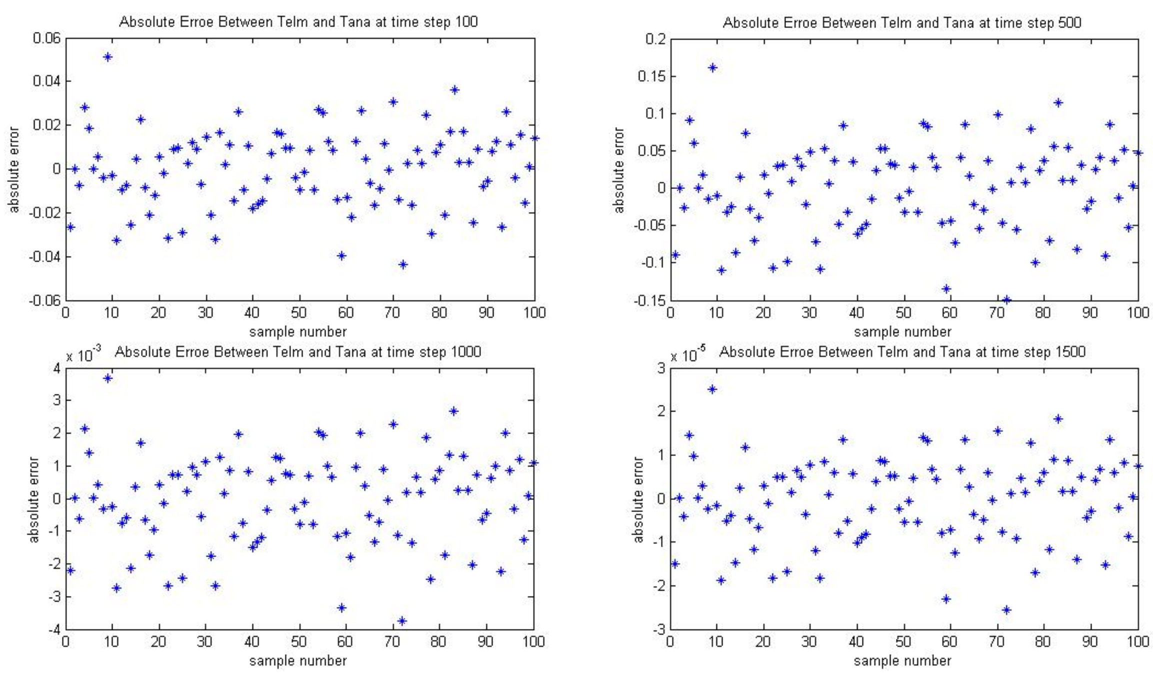

Figure 3. The errors between T_{ELM} and T_{ana} at time step 100, 500, 1000, 1500