

Research Article

Research on Sparsity of Output Synapses in Echo State Networks

Xiaohui Mu,¹ Lixiang Li ,² and Xiangyu He²

¹Information Security Center, State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

²National Engineering Laboratory for Disaster Backup and Recovery, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Lixiang Li; li.Lixiang2006@163.com

Received 15 September 2018; Accepted 27 November 2018; Published 18 December 2018

Academic Editor: Luiz H. A. Monteiro

Copyright © 2018 Xiaohui Mu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents an improved model of echo state networks (ESNs) and gives the definitions of energy consumption, energy efficiency, etc. We verify the existence of redundant output synaptic connections by numerical simulations. We investigate the relationships among energy consumption, prediction step, and the sparsity of ESN. At the same time, the energy efficiency and the prediction steps are found to present the same variation trend when silencing different synapses. Thus, we propose a computationally efficient method to locate redundant output synapses based on energy efficiency of ESN. We find that the neuron states of redundant synapses can be linearly represented by the states of other neurons. We investigate the contributions of redundant and core output synapses to the performance of network prediction. For the prediction task of chaotic time series, the predictive performance of ESN is improved about hundreds of steps by silencing redundant synapses.

1. Introduction

Artificial Intelligence is a branch of science and it studies and develops theories, methods, techniques, and applications for simulating and extending human intelligence [1]. Its applications cover a wide range including knowledge representation, search technology, and machine learning. Neural network is one of the hot issues in the field of artificial intelligence. Artificial neural network inspired by the human brain is a nonlinear adaptive dynamic system which consists of a large number of artificial neurons interconnected by synapses. Artificial neural network is a kind of artificial information processing system which aims to imitate the structure of human brain, the behavior, and the function of neural network to carry out distributed and parallel information processing. Synapses are connected between two neurons or between an neuron and an effector cell to deliver and convey information; i.e., synapses are the transmission channels of neuron signals and the basic constituent elements of human brain learning and memory [1, 2].

As a kind of artificial recurrent neural network, the echo state network was proposed by Jaeger et al. [2] in 2004.

ESN approach contains the learning and prediction processes. It is comprised by input neurons, sparse reservoir, and output neurons. It is characterized by feedback loops in their synaptic connection pathways. The reservoir can maintain an ongoing activation even in the absence of input and it can exhibit dynamic memory. Its large-scale RNN (called the reservoir) has been widely discussed [3–6]. The relationship between the connectivity structure of reservoir and its prediction performance was investigated [7–9]. Simplified models of the reservoir were studied [4, 10–12].

In the classical chaotic time series prediction, the precision of ESN improved several thousand times compared with traditional methods which solved the bottleneck problem in the previous research of neural network. ESN has rapidly become the research focus due to its excellent predictive performance in various areas [1]. The applications of ESN in image segmentation, image restoration, and wireless communication were studied [13, 14]. The dynamical behaviors of reservoir neuron were investigated [15]. Time series prediction of ESN was studied [16, 17]. The optimization performances of ESN from the aspects of learning mechanism

and the capacities of reservoir computing, prediction, and memory were studied [15].

Compared with the computer, human brain has higher energy efficiency with powerful capacity [18–21]. It has been discovered that the synaptic transmission dominated most of energy consumption of human brain [22]. The energy can be saved by removing redundant synapses and leaving only core synapses [23–27]. Any of the prior works [4, 15] have been proposed to reduce redundant synapses in the reservoir. However, few of the existing algorithms based on ESN studied the problems of energy consumption and energy efficiency and reduced redundant synapses in the output connections.

To find out and remove redundant synapses in the output connections, further to improve the predictive performance of ESN, this paper proposes an improved ESN model. The definitions of energy consumption, energy efficiency, etc. were given. We analyze the relationship between energy consumption and sparsity, the relationship between predicted steps and sparsity, and the relationship between energy efficiency and sparsity in ESN and then verify the existence of redundant synapses. We find that silent redundant synapses have little influence on the weights of other synapses. We also study the contributions of redundant and core output synapses to prediction performance in the improved ESN and discover that the contribution of redundant output synapses to predictive performance is close to zero. At the same time, the energy efficiency and the predictive steps are found to have the same variation trend when silencing different synapses. Thus, we propose a method to locate redundant output synapses based on energy efficiency of ESN system, which is computationally efficient. Numerical simulations of different chaotic systems are presented to demonstrate the feasibility and the effectiveness of the proposed approach in large-scale ESN. Compared to a fully connected network, the predictive performance is improved about hundreds of steps by silencing redundant synapses for the task of chaotic time series prediction.

2. Materials and Methods

2.1. The Traditional Model of ESN. An echo state network is an artificial recurrent neural network. It can maintain an ongoing activation even in the absence of input and thus it can exhibit dynamic memory. As shown in Figure 1(a), ESN consists of fore-end input layer, neuron reservoir, and output layer, and its corresponding input vector, state connection vector, and output vector can be expressed as

$$\begin{aligned} u(n) &= (u_1(n), u_2(n), \dots, u_k(n))^T, \\ x(n) &= (x_1(n), x_2(n), \dots, x_r(n))^T, \\ y(n) &= (y_1(n), y_2(n), \dots, y_l(n))^T, \end{aligned} \quad (1)$$

where $u(n)$ is k -dimensional input vector, $x(n)$ is r -dimensional state connection vector, and $y(n)$ is l -dimensional output vector.

At the sampling time n , the state update equation and the output equation of ESN are given by

$$\begin{aligned} x(n+1) &= \tanh(W^{in}u(n+1) + Wx(n) + W^{back}y(n) + v(n)), \end{aligned} \quad (2)$$

$$y(n+1) = W^{out}x(n+1), \quad (3)$$

where W is the $r \times r$ -dimensional matrix of internal connection weights, W^{in} is the $r \times k$ -dimensional matrix of input connection weights, W^{back} (optional) is the $r \times 1$ -dimensional weight matrix for feedback connections from the output neuron to the reservoir, $v(n)$ (optional) is the noise vector, and the hyperbolic tangent (\tanh) function is the activation function. Equation (3) is the output equation for a single-output network, and W^{out} is the $1 \times r$ -dimensional matrix of connection weights for the output neuron (bule dotted line in Figure 1(a)). W^{out} is adjusted by the linear regression $W^{out} = y_d X^{-1}$, and y_d is the teacher time series observed from the target system. Many applications of ESN aim at minimizing the training error $y_d(n) - y(n)$, where $y(n)$ is the network output and it is created from the internal neuron activation signals $x(n)$ through the trained connections W^{out} . They can maintain an ongoing activation even in the absence of input signal $u(n)$.

2.2. Our Improved Model of ESN. Synapses are the sites by which neurons can contact with each other and transmit information. The long-term transmission function of synapses can be depressed and potentiated. It has a particularly important influence on the advanced function of the brain while maintaining the computation, memory, and learning powers of the brain. It is a difficult problem to accurately find the redundant output synaptic connections and remove them in order to improve the network function in artificial neural networks.

To solve the above problem, the traditional model of ESN is improved. As shown in Figure 1(b), unlike the traditional model of ESN, we find out redundant output synapses and then we make the redundant output synapses silent in our improved model. The predicted steps and the energy efficiency are increased compared with the traditional ESN. As shown in Figure 1(b), our improved ESN with sparse output connections has the same reservoir structure as that in Figure 1(a).

We explain the improved ESN approach based on a prediction task of chaotic time series. The Mackey-Glass system (MGS) is a standard benchmark system for the study of time series prediction. It generates an irregular time series. The prediction task has two steps: (i) use an initial teacher sequence $y_d(n)$ generated by the original MGS to learn a black-box model M of the generating system, where $y_d(n)$ is fed into the reservoir through feedback connections W^{back} . Then the internal neurons are excited. After an initial transient period, they start to exhibit systematic individual variations of the teacher sequence. (ii) Use M to predict the values of the sequence some steps ahead.

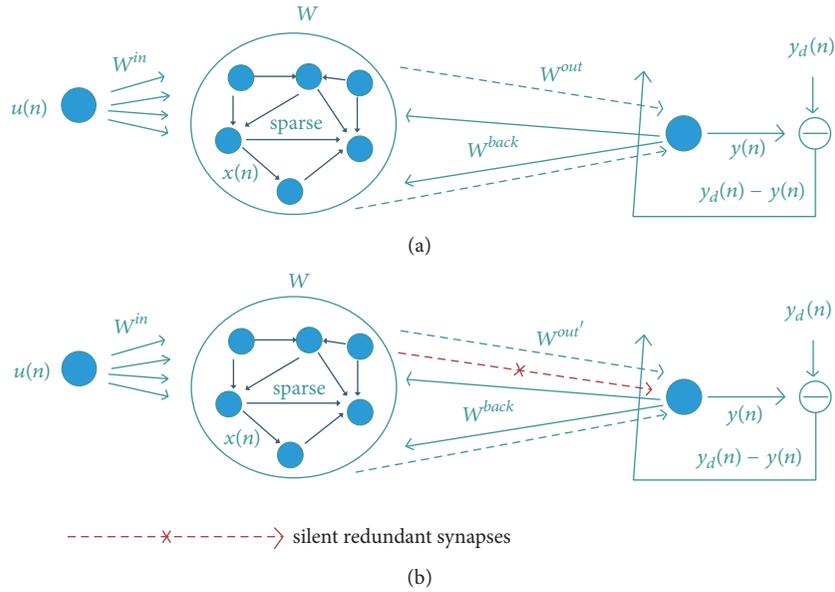


FIGURE 1: Schemas of the traditional ESN approach and our improved ESN approach. (a) The traditional ESN approach. Input signals $u(n)$, internal neuron activation signals $x(n)$, input weight matrix W^{in} , feedback matrix W^{back} , and internal connectivity matrix W . W^{in} , W , and W^{back} are equipped with fixed synaptic connections (solid line). The values of W^{in} , W , and W^{back} remain constants (sometimes W^{in} is 0). The output signal $y(n)$ is created from $x(n)$ through the trained connections W^{out} . W^{out} is adjusted by linear regression $W^{out} = y_d X^{-1}$ (blue dotted line). $y_d(n)$ is the teacher sequence observed from the target system. Many applications of ESN aim at minimizing the error $y_d(n) - y(n)$. (b) Our improved ESN approach. Fully connected output synapses are only part of the network that needed to be adjusted by the linear regression $W^{out'} = y_d X^{-1}$ (blue dotted line). ESN of high energy efficiency with sparse connections matrix $W^{out'}$ has the same reservoir structure as that in (a). Input neurons can be removed from the network. Redundant output connections (red dashed line with the mark "X") can be removed which will not affect the predictive performance, but will affect the energy consumption of the whole network.

At the sampling time n , the update equation and the output equation of our improved ESN model are given by

$$x(n+1) = \tanh(Wx(n) + W^{back}y_d(n) + v(n)), \quad (4)$$

$$y(n+1) = W^{out'}x(n+1), \quad (5)$$

where $W^{out'}$ is the output state matrix that makes the energy efficiency of ESN maximal after silencing an output synapse. Other variables have the same meanings as those in (2)-(3).

Since the reservoir can maintain an ongoing activation in the absence of input and it can exhibit dynamic memory. It is not necessary for the state update of the reservoir to have the input signal $u(n)$ in the improved ESN model.

2.3. Teacher Sequences Generated by Four Different Chaotic Systems. In the prediction of the classical chaotic sequence, the precision of ESN is several thousand times than those of traditional methods. So we select four typical chaotic systems determined by differential equations to generate the teacher sequence.

The Mackey-Glass system (MGS) is a standard benchmark system in the research on time series prediction. It generates a subtly irregular time sequence. Almost every available technique for nonlinear system modeling and prediction has been tested on the MGS system.

A teacher sequence $y_1(1), y_2(2), \dots, y_d(t)$ is generated from the following Mackey-Glass system

$$\dot{z} = \frac{0.2z(t-\tau)}{1+z(t-\tau)^{10}} - 0.1z(t) \quad (6)$$

where τ is the delay and z is state of MGS. 4000-step training data and 1000-step testing data are generated by solving the above MGS equation with the Runge-Kutta(4,5) method whose step size is 1.

The initialization process of Echo State Network by MGS is given as follows. We create a random partially connected ESN with 1000 neurons, and the sparsity of the reservoir is 1%. We generate a 1000×1000 -dimensional random weights matrix W whose random weights are drawn from a uniform distribution over the interval $[-1, 1]$, then they are rescaled to spectral radius 1.6. 1000×1 -dimensional output feedback connection weights matrix is randomly selected from a uniform distribution over the interval $[-1, 1]$. The strength of Gaussian white noise is 1×10^{-5} dBW.

The Lorenz system is governed by the following three-dimensional differential equation

$$\begin{aligned} \dot{z}_1 &= a(z_2 - z_1) \\ \dot{z}_2 &= -z_1z_3 + cz_1 - z_2 \\ \dot{z}_3 &= z_1z_2 - bz \end{aligned} \quad (7)$$

where a , b , and c are the parameters and z_1 , z_2 , and z_3 are the states of Lorenz system. We set the parameters of Lorenz system as $a = 10$, $b = 8/3$, $c = 28$, and now the Lorenz system is in a chaotic state. Training data (4000-step) and testing data (1000-step) are generated by solving the above Lorenz system equation with the Runge-Kutta(4,5) method whose step size is 0.01. Following most studies on the learning approaches of this system, we use only the trajectory of z_1 as the teacher sequence for the training and the testing of ESN.

The initialization process of Echo State Network by Lorenz system is given as follows. We create a random network with 1000 neurons. Its spectral radius is 1.6, and its sparsity is 0.02. Output feedback connection weights are sampled from a uniform distribution over $[-0.1, 0.1]$. The strength of Gaussian white noise is 1×10^{-8} dBW.

The three-dimensional differential equation of Rössler System is given as follows:

$$\begin{aligned}\dot{z}_1 &= -z_2 - z_3 \\ \dot{z}_2 &= z_1 + az_2 \\ \dot{z}_3 &= b - cz_3 + z_1z_3\end{aligned}\quad (8)$$

where a , b , and c are the parameters and z_1 , z_2 , and z_3 are the states of this system. When $a = 0.2$, $b = 0.2$, and $c = 5.7$, the Rössler system is in a chaotic state. We select the values of Rössler system parameters as $a = 0.2$, $b = 0.2$, and $c = 5.7$ to simulate. Similar with Lorenz system, 4000-step training data and 1000-step testing data are generated by solving the Rössler system equation with the Runge-Kutta(4,5) method whose step size is 0.1 (by employing Matlab ode45 solver in the Matlab software). We use only the trajectory of the first coordinate z_1 for training and testing in ESN.

The initialization process of Echo State Network by Rössler system is given as follows. This process is the same as the initialization setting of Lorenz system. We only should adjust the sparsity of ESN to be 0.01 to adapt to the Rössler chaotic system.

The equation of Chen chaotic system is presented as

$$\begin{aligned}\dot{z}_1 &= a(z_2 - z_1) \\ \dot{z}_2 &= (c - a)z_1 - z_1z_3 + cz_2 \\ \dot{z}_3 &= z_1z_2 - bz_3\end{aligned}\quad (9)$$

where a , b , and c are the parameters and z_1 , z_2 , and z_3 are the states of this system. When $a = 35$, $b = 3$, and $c = 28$, this system is in a chaotic state. We select the parameter values of Chen chaotic system as $a = 35$, $b = 3$, and $c = 28$ to simulate. By solving the Chen chaotic system equation, a 4000-step training sequence and a 1000-step testing sequence are generated. We use the trajectory of the first coordinate z_1 for training and testing of ESN.

The initialization process of Echo State Network by Chen system is given as follows. Similar with the network initialization process of Mackey-Glass system, we create a random network with 1000 neurons. The spectral radius of the network is 1.7 and its sparsity is 0.02. Output feedback

connection weights are sampled from the uniform distribution over $[-0.1, 0.1]$. The strength of Gaussian white noise is 1×10^{-5} dBW.

2.4. Definitions of Physical Quantities. In this paper, we present an improved model of ESNs and use many physical quantities. Here we give the definitions of these physical quantities.

Predicted step L symbolized *steps* is the minimal number of time n that meets $|(y_{predict}(n) - y_{original}(n))/y_{original}(n)| \leq \Delta$, where $\Delta = 1 \times 10^{-3}$, $y_{original}(n)$ is the value of the original signal sequence at time n , and $y_{original}(n)$ is the output of the corresponding predictive sequence at time n . It can be used as a reflection of the predictive performance in ESN.

Sparsity is the small connection probability between the neurons in the reservoir or between the neurons in the reservoir and the output neurons, and the sparsity is given by $\alpha = (N - S)/N$, where S is the number of silent synapses and N is the number of all adjustable synapses.

Energy consumption E is the total consumption energy of all activated output synapses which is given by $E = \sum_{i=1}^N (w_i^{out})^2$, where N is the number of neurons in the reservoir and w_i^{out} is weight of connections for the i th output neuron.

Energy efficiency C is the number of steps that can be predicted by a single neuron per unit of energy. That is, the contribution of each neuron to the total energy efficiency, unit energy efficiency, can be denoted by $C = I/nme$, where $I = L/E$ denotes the number of steps that can be predicted per unit of energy and nme is number of neurons that is not silenced (activated) when the reservoir is outputting, i.e., the number of nonzero elements in the output matrix W^{out} .

Contribution T is the contribution of energy consumption of i th neuron to the total energy consumption of network system at time $n + 1$ in the reservoir which is given by $T_i(n + 1) = |o_i(n + 1)| / \sum_{i=1}^N |o_i(n + 1)|$, where $o_i(n + 1) = w_i^{out} x_i(n + 1)$.

3. Results

3.1. There Are Redundant Output Synapses in ESN. Now we consider the experimental results. In Figure 2(a), the curves show the relationship between the sparsity of output connections and the predicted steps of ESN, and the corresponding teacher sequences are generated by three different chaotic systems (i.e., Mackey-Glass system, Lorenz system, and Rössler system) to train ESN. For the detailed explanations of these chaotic systems, please see Materials and Methods. The experimental results in Figure 2(a) were over 100 independent trials for every sparsity. The values of the parameters used in these experiments of Figure 2(a) have been given in Materials and Methods. In Figure 2(a), the abscissa is the sparsity of output connections, and the ordinate shows the predicted steps of ESN. The sparsity is related to the ratio between the silent output synapses and the whole output synapses. Different colors curves with different marks in Figure 2(a) represent the experimental results corresponding to three different chaotic systems, respectively. From Figure 2(a), we can see that the predicted

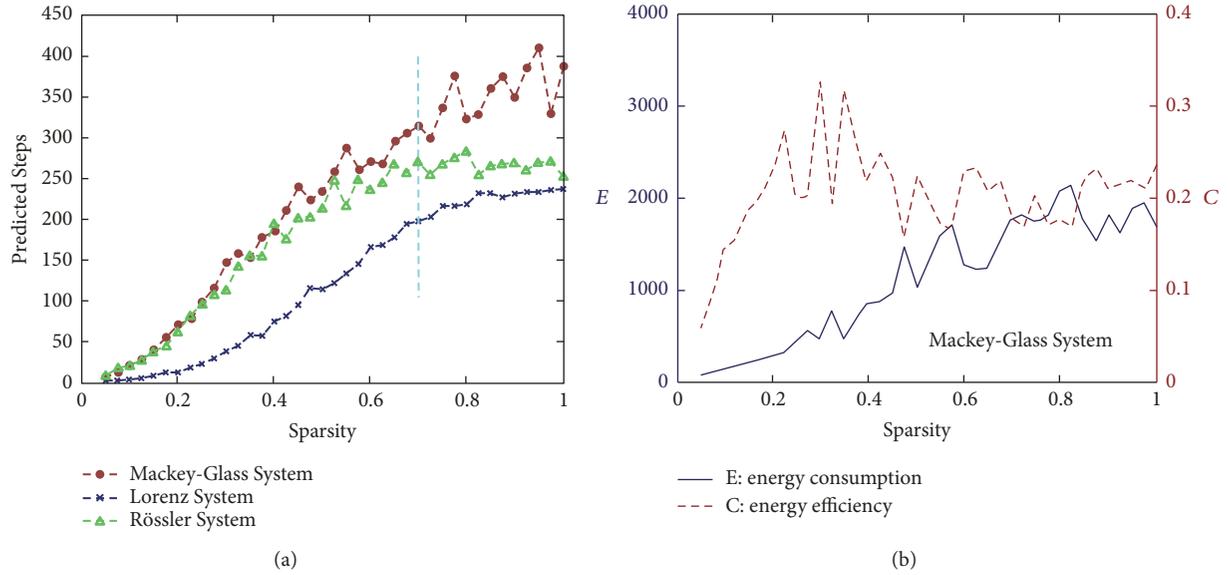


FIGURE 2: The experimental results verify that there are redundant synapses in the output connections of ESN. All the experimental results are averaged over 100 independent trials per sparsity. (a) Curves about the relationship between the predicted steps and the sparsity of ESN. The corresponding teacher sequences are generated by three different chaotic systems (Mackey-Glass system, Lorenz system, and Rössler system), respectively. (b) Curves about the effects of the sparsity on the energy consumption E (blue solid line) and the energy efficiency C (red dashed line), respectively. The left ordinate represents the energy consumption E of output synapses, and the right ordinate shows the energy efficiency C of output synapses.

steps L will change with the sparsity. For the definition of L , please see Materials and Methods. For the experimental results corresponding to three different chaotic systems, these curves are steadily rising, and their rising speeds are fast when $\alpha < 0.7$, but these curves become more gradual and present oscillations locally when $\alpha > 0.7$. From Figure 2(a), we also can see that when $\alpha > 0.7$, i.e., $S/N < 30\%$, the variation of the curve is more gentle. For the definitions of S and N , please see Materials and Methods. Figure 2(a) shows that with the increment of sparsity, the predicted steps do not have significant changes; i.e., from a certain point of view, the fact that the prediction performance has not been increased obviously means that there are some redundant synapses in ESN.

In order to further verify the existence of redundant output synapses in ESN, we analyze the influences of the sparsity of output synapses on energy consumption E and energy efficient C . For the detailed explanations of energy consumption E and energy efficient C , please see Materials and Methods. If there is not a redundant output synapse in ESN, then with the increment of sparsity the energy consumption will be higher and the energy efficiency will not decrease.

Figure 2(b) shows the effects of the sparsity of output synapses on energy consumption E and energy efficiency C of ESN, and the teacher sequence is generated by the Mackey-Glass system. In Figure 2(b), the abscissa represents the sparsity of output synapses, the left ordinate indicates the energy consumption E of output synapses, and the right ordinate is the energy efficiency C of output synapses. The parameters used in the experiment have been given in the Materials and Methods section, where the error precision

$\Delta = 1 \times 10^{-3}$. There are two curves in Figure 2(b). One curve (red dotted line) gives that the energy efficiency C of output synapses changes with the increment of the sparsity α of output synapses, and the other curve (blue solid line) shows that energy consumption E of output synapses varies with the increment of the sparsity α of output synapses. From Figure 2(b), we can see that with the increment of sparsity, when $0 < \alpha < 0.4$, the curve of energy efficiency C shows a significant rising trend, and C can reach its maximum. When $0.4 < \alpha < 1$, the curve of energy efficiency C shows a gentle trend and tends to oscillate locally. When $0 < \alpha < 0.8$, the curve of energy consumption E shows a significant rising trend. When $0.8 < \alpha < 1$, the curve of E shows a relatively gentle trend. The energy consumption E can reach its maximum when the value of α is around 0.8, while the energy efficiency C achieves its peak value when α is about 0.3. It can be seen from Figure 2(b) that with the increment of α , the curves of energy consumption and energy efficiency are not always rising. The reason is that there exist partial output synapses with low energy consumption and low energy efficiency in ESN. Thus, this phenomenon further demonstrates the existence of redundant output synapses in ESN.

Based on all the experimental results in Figures 2(a) and 2(b), we can see that in ESN that uses the data generated by Mackey-Glass system as teacher sequences for the training, when we pursue higher predicted steps (i.e., when $\alpha > 0.7$ in Figure 2(a)), the energy consumption E is in a state with high value, but the energy efficiency C is not in the state with high value (when $\alpha > 0.7$, please compare the blue solid line with the red dashed line in Figure 2(b)). When C reaches its highest value (i.e., when $\alpha = 0.3$ in Figure 2(b)), the prediction

step is smaller (when $\alpha = 0.3$, please see the red line marked with solid point in Figure 2(a)). In summary, the highest energy efficiency and more predicted steps are impossible to be obtained simultaneously. Obviously, no matter whether we pursue the highest energy efficiency or more predicted steps, the value of α corresponding to their peak values is less than 1. Therefore, there are redundant output synapses in ESN.

3.2. The Existence of Redundant Output Synapses in Small-Scale ESN. In the above subsection, we analyzed the existence of redundant output synapses in ESN. Now we verify the existence of redundant output synapses in small-scale ESN. To facilitate the observation about the variations of energy efficiency and the predicted steps, we randomly create an ESN with 10 neurons. Since the predictive performance of small-scale ESN is limited, we create the periodical sequence to adapt to the performance of ESN. A 500-step teacher sequence is generated from the equation $f(x) = \sin x + \cos x$ to predict about 1000 steps, where the error precision $\Delta = 1 \times 10^{-3}$. In this ESN, we silence a synapse each time and observe the variations of the energy efficiency and the predicted steps of ESN. The corresponding experimental results are shown in Figure 3(a), in which the abscissa representing the i th synapse is silent each time, the left ordinate is the energy efficiency C , and the right ordinate is the predicted steps. The blue solid line is the curve of energy efficiency C in ESN. The red dotted line is the curve of the predicted steps in ESN, and the horizontal red dotted line shows the predicted steps of ESN when all synapses are activated. We can see from Figure 3(a) that when we silence the 2nd or 4th synapse, the predicted step of ESN approaches that as all synapses are activated; i.e., the silence of the 2nd or 4th synapse has little effect on the predictive performance of ESN. It means that there are redundant synapses in the small-scale ESN. Obviously, the 2nd and 4th synapses are redundant ones. From Figure 3(a), we also can see that while the 1st, 5th, and 7th synapses are silenced, respectively, the energy efficiency and the predicted steps of ESN are close to 0; i.e., these synapses have great influence on the predictive performance of ESN. It means 1st, 5th, and 7th synapses are core synapses. At the same time, we discover from Figure 3(a) that the same variation trends are found between the energy efficiency (blue solid line) and predicted steps (red dotted line) when the corresponding synapses are silenced.

Figure 3(b) shows the contributions of 1st, 2nd, 4th, 5th, and 7th synapses to the predictive performance of ESN with the variation of time, where the time from 500 to 560 is randomly selected by us. In Figure 3(b), the abscissa represents the time, the ordinate represents the contribution of synapses to the predictive performance of ESN, and the curves with different colors and different marks show the energy contributions of 1st, 2nd, 4th, 5th, and 7th synapses to the predictive performance of ESN. From Figure 3(b), we can see that with the increment of the time, the energy contributions of 1st, 5th, and 7th synapses are varied, and the energy contributions of 2nd and 4th synapses are close to 0. It verifies that there are redundant synapses in the small-scale ESN, and the contribution of redundant

synapses to the predictive performance of ESN is close to 0.

Figure 3(c) shows the weight curves of all synapses when the 2nd or 4th synapse is silenced and all the synapses are activated (when the 2nd or 4th synapse is silenced, the weights of other synapses should be recalculated). In Figure 3(c), the abscissa denotes the i th synapse, and the ordinate is the weight of synapse. The curves with different colors and different marks represent the weight of all synapses when the 2nd or 4th synapse is silenced and all the synapses are activated. From Figure 3(c), we can see that silencing the 2nd or 4th synapse has little effect on the weights of other synapses. It means the 2nd and 4th synapses are redundant synapses, which further verifies there are redundant synapses in the small-scale ESN from the perspective of the synaptic weight.

Figure 3(d) shows the weight curves of all synapse when 1st, 5th, and 7th synapses are silenced, respectively, and all the synapses are activated. In Figure 3(d), the abscissa represents the i th synapse, and the ordinate gives the weight of the synapse. Four curves with different colors and different marks represent the weight curves of all synaptic weights when the 1st, 5th, and 7th synapses are silenced, respectively, and all the synapses are activated. We find that Figure 3(d) is different from Figure 3(c). In Figure 3(d), the differences of the curves about synaptic weight are very large; i.e., it has great influence on the weight of other synapses to silence the 1st, 5th, or 7th synapses. From Figure 3(a) we can also see that the predictive performance of ESN is decreasing. It means 1st, 5th, and 7th synapses are not redundant synapses and they cannot be removed. From the above experimental results, we can see the existence of redundant output synapses in small-scale ESN.

3.3. The Reason for the Existence of Redundant Output Synaptic Connections. From the perspective of mathematics, now we begin to analyze the reason for the existence of redundant output synaptic connections. Without loss of generality, we take the 4th synapse for example to illustrate this. x_i is the state of the i th neuron at all times. For example, x_4 is the state of the 4th neuron at all times. X_{4s} is the state matrix of all neurons when we silence the 4th neuron corresponding to the output synapse. $W^{out'}$ is the weight matrix of output synapses when we silence the 4th neuron corresponding to output synapse. $w_i^{out'}$ is the weight of the i th output synapse connected by the i th neuron when we silence the 4th neuron corresponding to output synapse. w_i^{out} is the weight of the i th output synapse connected by the i th neuron when all neurons are activated. Y_{4s} is the output matrix of ESN when we silence the 4th neuron corresponding to output synapse.

We can consider $X_{4s} = X$, and we know that

$$W^{out'} X_{4s} = Y_{4s} \quad (10)$$

$$W^{out} X = Y \quad (11)$$

Under the error Δ , in the interval S $Y_{4s} = Y$, (10)-(11) obtains

$$W^{out'} X_{4s} = W^{out} X \quad (12)$$

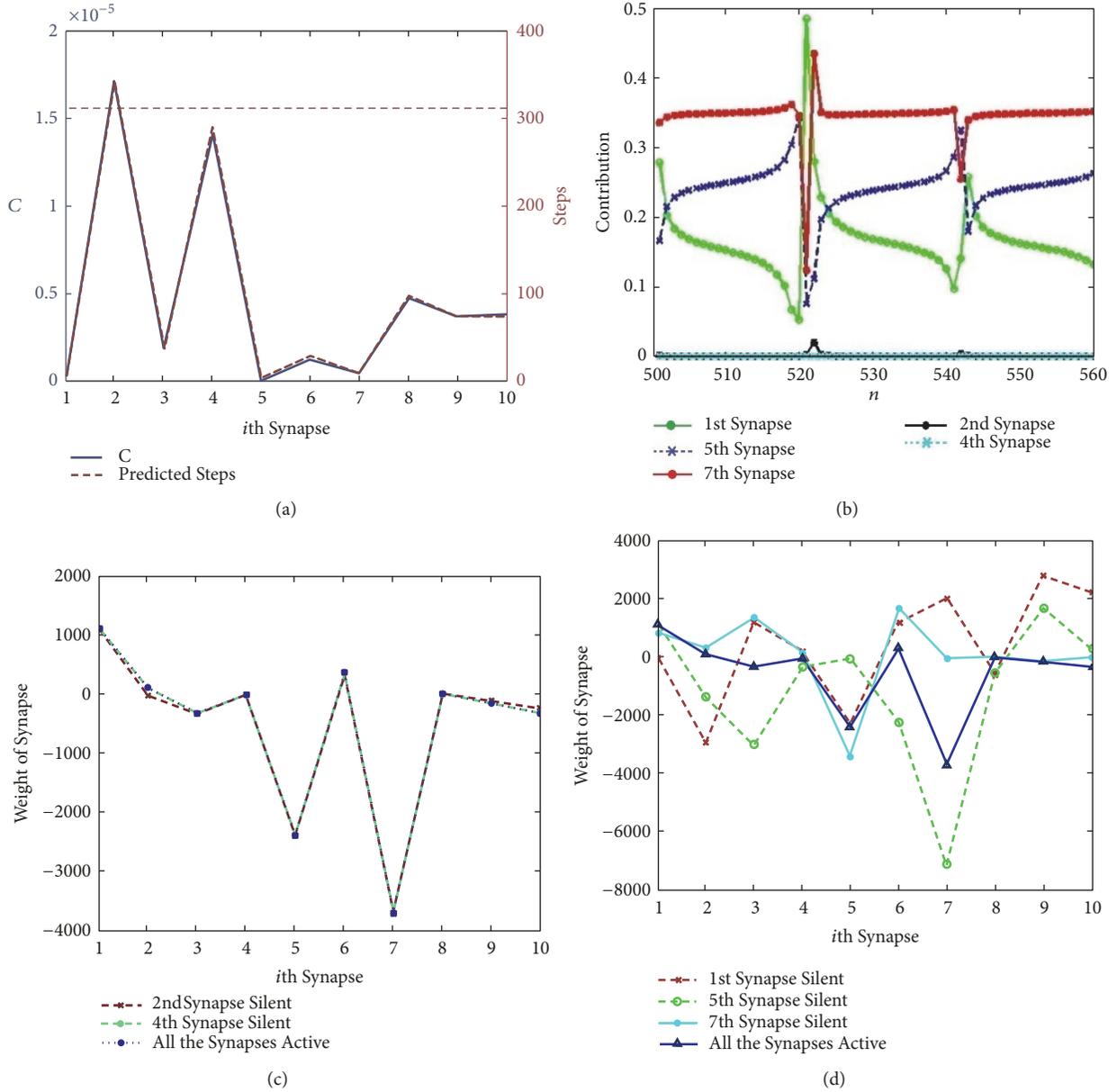


FIGURE 3: Experimental results to verify the existence of redundant output synapses in the small-scale ESN. (a) Curves of the energy efficiency (blue solid line) and the predicted steps (red dotted line) when we silence a synapse each time. The abscissa representing the i th synapse is silenced, the left ordinate is the energy efficiency C , the right ordinate is the predicted steps, and the horizontal dotted line represents the predicted steps of ESN when all synapses are activated. (b) Curves of energy contribution of 1st, 2nd, 4th, 5th, and 7th synapses to predictive performance of ESN as time increases. The abscissa is the time, the ordinate represents the energy contribution of synapses to the predictive performance of ESN, and the curves with different colors and different marks represent energy contributions of 1st, 2nd, 4th, 5th, and 7th synapses to predictive performances of ESN, respectively. (c) Curves of all synaptic weights when the 2nd and 4th synapses are silenced, respectively, and all the synapses are activated. The abscissa represents the i th synapse. The ordinate is the value of synaptic weight. Three curves with different colors and different marks represent the curves of all synaptic weights when the 2nd and 4th synapses are silenced, respectively, and all the synapses are activated. (d) Curves of all synaptic weights when the 1st, 5th, and 7th synapses are silenced, respectively, and all the synapses are activated.

In matrix $W^{out'}$, $W_4^{out'} = 0$ obtains

$$x_4 = \frac{1}{w_4^{out} \left((w_1^{out'} - w_1^{out}) x_1 + \dots + (w_{10}^{out'} - w_{10}^{out}) x_{10} \right)} \quad (13)$$

The form of (13) means that the states of the 4th neuron at all times can be linearly expressed by the states of other neurons. It indicates the output synapse connecting the 4th neuron can be silenced. Therefore, the output synapse connected to the 4th neuron is useless. It is a redundant synapse. Similarly, the states of the 2nd neuron at all moments

TABLE 1: Weights of output synapses when all output synapses are activated and the 2nd or 4th synapse is silenced, respectively. Here are the ratio between the weights of synapses when the 2nd or 4th synapse is silenced and the weights when all synapses are activated.

neurons	1	2	3	4	5	6	7	8	9	10
Weight when all are activated	1130.428	134.006	-308.004	-0.545	-2375.857	346.081	-3668.496	24.762	-135.418	-301.811
Weight when 2th synapse is silenced	1119.316	0	-313.068	6.100	-2349.637	343.812	-3680.391	20.728	-82.074	-231.207
$ \frac{W_{2th\text{silenced}}^{out}}{W_{allactive}^{out}} $	0.990	0	1.016	-11.183	0.988	0.993	1.003	0.837	0.606	0.766
Weight when 4th synapse is silenced	1130.079	131.702	-305.079	0	-2377.723	348.913	-3664.580	25.212	-135.811	-300.373
$ \frac{W_{4th\text{silenced}}^{out}}{W_{allactive}^{out}} $	0.999	0.982	0.990	0	1	1.008	0.998	1.018	1.002	0.995

can be linearly expressed by the states of other neurons. The output synapse connected to the 2nd neuron is a redundant synapse.

The states of the 2nd or 4th neuron at all moments can be linearly expressed by the states of other neurons in the state matrix. Other vector elements are independent and they cannot be linearly expressed by other appropriate elements. Similarly for the states of other redundant synapses, there exists the characteristic of linear representation in the long-term prediction.

Here we analyze the reason for the existence of redundant output synaptic connections by table data. We consider the circumstance that the 2nd or 4th synapse is silenced, respectively, and there are 10 neurons in reservoir. In Table 1, the first line is the sequence number of the neuron. The second line shows the weights of output synapses when all output synapses are activated. The third line gives the weights of all output synapses when the 2nd synapse is silenced. The fourth line presents the ratios between the weight of each synapse when the 2nd synapse is silenced and the weight of each synapse when all synapses are activated. The fifth line shows the weights of all output synapses when the 4th synapse is silenced. The sixth line gives the ratios of the weight of each synapse when 4th synapse is silenced and the weight of each synapse when all synapses are activated.

From Table 1, we can see that the ratio between the weights of synapses when the 2nd synapse is silenced and the weights when all synapses are activated is close to 1, which indicates the variation of other synaptic weights is very small when the 2nd synapse is silenced. It explains that the 2nd synapse has a little influence on other synapses. Similarly, we also can see that it has a little influence on other synapses to silence the 4th synapse. It also explains there are redundant output connections in ESN.

Based on the analytical results, we find that there are redundant output connections in ESN. Locating redundant synapses in ESN and then removing them are helpful to increase the predictive steps of ESN. Therefore, it is necessary to propose a method to find or locate redundant output synapses. According to the analytical results, the same variation trend is found between the energy efficiency and the predicted steps with the changes of sparsity.

Inspired by this rule, we propose a method to find redundant output synapses. The main mechanism of this method is to use the repeated iterations to search multiple redundant output synapses. If we silence these redundant output synapses, then ESN will have the highest energy efficiency. By iterations, potential redundant output synaptic connections are gradually reduced in ESN, thus the sparse output synaptic weight matrix $W^{out'}$ is obtained. We find that this method greatly improves the predictive performance of ESN compared with the output synaptic weight matrix W^{out} of the fully connected ESN. Inversely, if core output synaptic connections are gradually reduced to obtain sparse W^{out*} in ESN, then the energy efficiency of ESN will decrease; i.e., the predictive performance of ESN will be declining gradually.

The detailed steps to find out redundant output synapses are given as follows.

Step 1. Ergodic search: silence some output synapse, and record the energy efficiency of ESN after this synapse is silenced.

Step 2. Repeat Step 1, and record the value of energy efficiency of ESN.

Step 3. Locate redundant output synaptic connections. In the output result of Step 2, search for the output synapse that makes ESN has the highest energy efficiency after this synapse is silenced. Set this output synaptic connection as zero to silence it.

Step 4. Cycle repeatedly in order to find the optimum. Repeat Steps 1, 2, and 3, and record the variation of energy efficiency for ESN in order to find all the redundant synapses.

3.4. The Existence of Redundant Output Synapses in Large-Scale ESN. In the above section, we analyzed the reason for the existence of redundant output synapses in small-scale ESN. Now we verify the existence of redundant output synapses in large-scale ESN and further verify the effectiveness of the proposed method by analyzing redundant output synapses. We create an ESN with 1000 neurons with reservoir sparsity 0.9% to predict Mackey-Glass system, Lorenz system,

Rössler system, and Chen system, respectively. The parameter values of these four chaotic systems are given.

Figures 4(a)–4(d) show the variation of predicted steps of ESN with the increment of silent synapses. In Figures 4(a)–4(d), the abscissa is the number of silent synapses, and the ordinate represents the predicted steps of ESN. In each subgraph of Figure 4, there are five curves which represent different connection states of the output synapse, where the weight matrix $W^{out'}$ of output connections represents the condition that the redundant output synapses are located to make ESN obtain the maximal energy efficiency after they are silenced, W^{out} represents the weight matrix of output connections when output synapses are all activated, the weight matrix W^{out*} of output connections represents the condition that the redundant output synapses are located to make ESN obtain the minimal energy efficiency after they are silenced, the weight matrix W^{out**} of output connections represents the condition when we randomly select an output synapse and silence it, and the weight matrix $W^{out''}$ of output synaptic connections represents the condition that we silence redundant output synapses corresponding to the neurons with the minimal energy contribution. In Figures 4(a)–4(d), the horizontal line represents that the predicted steps are constant for the prediction of Mackey–Glass system, Lorenz system, Rössler system, and Chen system when the weight matrix W^{out} of output synaptic connections is in the fully connected state.

From Figures 4(a)–4(d), we can see the following facts. When $W^{out'}$ is the weight matrix of output connections (i.e., when ESN has the maximal energy efficiency after redundant output synapses are removed), with the increment of silent synapse, the curves of the predicted steps for the predictions of Mackey–Glass system, Lorenz system, Rössler system, and Chen system are over the horizon line that has the weight matrix W^{out} , and all these curves show an rising trend. It explains that the removal of redundant synapses can increase the predicted steps of ESN; i.e., it can increase the predictive performance of ESN. When W^{out*} is the weight matrix of output connections (i.e., ESN has the minimal energy efficiency after redundant output synapses are removed), with the increment of silent synapse, the curves of predicted steps for the predictions of Mackey–Glass system, Lorenz system, Rössler system, and Chen system are under the horizon line that has the weight matrix W^{out} , and these curves show a relatively fast decreasing trend. It explains that the removal of core synapses leads to the decrement of predicted steps for ESN; i.e., the predictive performance of ESN will be declined. While the ESN is in any state of the rest two weights, with the increment of silent synapse, the curves of predicted steps are around the horizon line. It explains the predicted steps of ESN cannot be improved when ESN is in any state of these two weights; i.e., the synapses corresponding to these two weights are not redundant synapses. We have verified the existence of redundant output synapses in large-scale ESN. From Figures 4(a)–4(d), we can also see that, compared to a fully connected ESN, the predicted steps are improved by at least a few hundred steps in the prediction of nonlinear sequence when we apply the proposed method to remove the

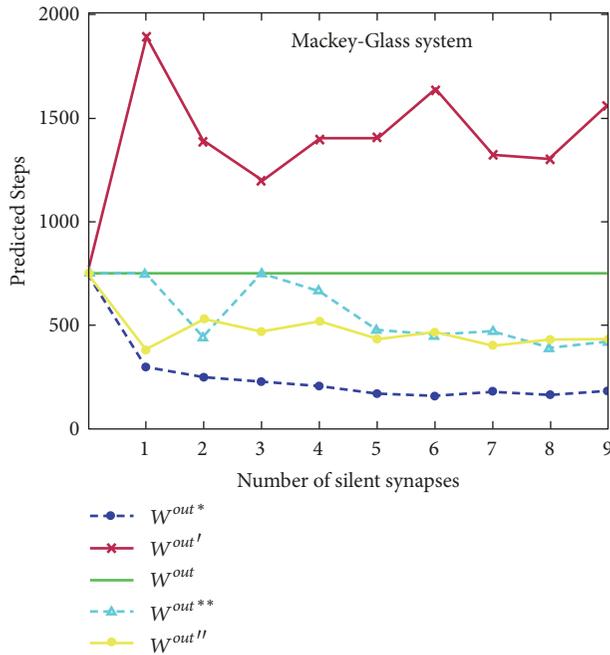
redundant synapses. It explains the predicted steps have been greatly increased when silencing redundant output synapses.

4. Discussion

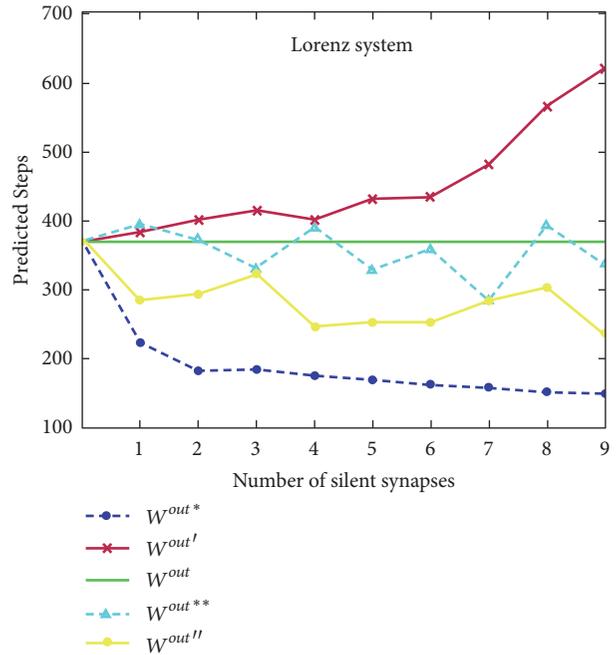
We emphasize the general character of the proposed method because this method can be used by other types of artificial neural networks to find out redundant synapses. Our approach is readily applicable to these situations, such as perception and BP network. It is worth considering whether other neural networks with synaptic connections also have such characteristic. Similarly, we can propose the corresponding method to improve the optimization of network structures and the predictive performance of neural networks based on the study of the energy efficiency of human brain. Furthermore, we can extend our discovery and the proposed method of ESN to the research on the networks of human brain. However, this important aspect has received little attention in the existing studies. The ideas of the proposed method inspired by the research on the energy of networks can be extended to the research processes of many human behaviors such as government service and cooperation layoff. For example, the government achieves interdepartmental data reduction and shares to improve the efficiency of government service by removing redundant data.

5. Conclusions

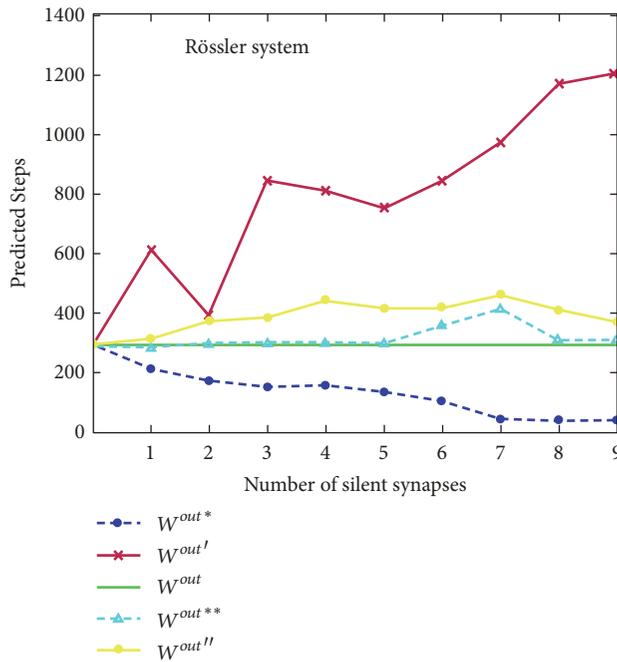
In summary, we have analyzed the reason for the existence of redundant output synaptic connections in ESN from the perspectives of mathematical analysis and numerical simulations. This paper presents an improved model of echo state network (ESN) and gives the definitions of energy consumption, energy efficiency, etc. In this paper, the data generated by four different chaotic systems (i.e., Mackey–Glass system, Lorenz system, Rössler system, and Chen system) are selected as teacher sequences to train ESN. We investigate the relationship between the sparsity of output connections and the predicted steps of ESN. We take the prediction of M–G system as example to investigate the relationship among energy consumption, energy efficiency, and the sparsity. And we discover that it is difficult to obtain high energy efficiency and more predicted steps simultaneously. For a small-scale ESN, we investigate the relationship among the predicted steps, the energy efficiency, and silent output synapses by numerical simulations, and we find that the energy efficiency of output synapse and the predicted steps of ESN have the same variation trend when the corresponding output synapses are silenced. We investigate the contributions of redundant and core output synapses to the predictive performance of ESN and find that the energy contribution of redundant synapses to predictive performance is close to 0 in ESN. We investigate the variations of all synaptic weights when redundant and core synapses are silenced and all synapses are activated, and we find that silencing of redundant synapses does not have much influence on the synaptic weights, while silencing of core synapses has a significant influence on other synaptic weights which can lead to lower predictive performance of ESN.



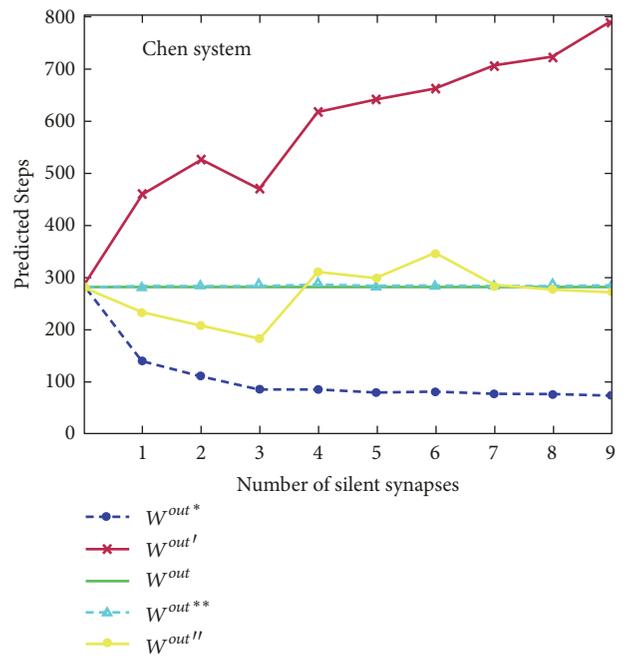
(a)



(b)



(c)



(d)

FIGURE 4: The variation of predicted steps for ESN with the increment of silent synapse. In each subgraph, there are five curves which represent different connection states of output synapses, where $W^{out'}$ is the weight matrix of output connections when ESN has the maximal energy efficiency after redundant output synapses are removed, W^{out} is the weight matrix of output connections when output synapses are all activated, W^{out*} is the weight matrix of output connections when ESN has the minimal energy efficiency after redundant output synapses are removed, W^{out**} is the weight matrix of output connections when we randomly select an output synapse and silence it, and $W^{out''}$ is the weight matrix of output connections when we silence redundant output synapses corresponding to the neurons with the minimal energy contribution. The horizontal line in each subgraph denotes that output synapses are all activated, and W^{out} is the weight matrix of output connections. (a) Prediction results of Mackey-Glass system. (b) Prediction results of Lorenz system. (c) Prediction results of Rössler system. (d) Prediction results of Chen system.

Based on the relationship between the energy efficiency and silent synapses, we have presented a general approach to find out the redundant synapses. The advantage of our approach is that redundant synapses can be accurately located in ESN. We have presented numerical results of the predictions for different chaotic systems to demonstrate the feasibility and the effectiveness of the proposed approach in large-scale ESN. With the increment of silent synapse, we give the variation curves of the predicted steps of ESN in five different connected states of output synapses. Compared to a fully connected network, the predicted steps are improved for the prediction of nonlinear time series when we apply the proposed method to remove the redundant synapses in ESN. Finding the redundant synapses in ESN to improve the predictive performance is one of the uttermost important problems in science and engineering and it is of extremely broad interest as well. We hope our work will stimulate further efforts in this challenging area.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work is supported by the National Key R&D Program of China (Grant Nos. 2016YFB0800604, 2016YFB0800602), the National Natural Science Foundation of China (Grant No. 61573067), and the “13th Five-Year” National Crypto Development Fund of China (Grant No. MMJJ20170122).

References

- [1] M. Hutson, “Artificial intelligence, in so many words,” *American Association for the Advancement of Science: Science*, vol. 357, no. 6346, 2017.
- [2] H. Jaeger and H. Haas, “Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication,” *Science*, vol. 304, no. 5667, pp. 78–80, 2004.
- [3] Y. An, Q. Song, and X. Zhao, “Short-term traffic flow forecasting via echo state neural networks,” *International Conference on Natural Computation*, pp. 844–847, 2011.
- [4] H. Cui, X. Liu, and L. Li, “The architecture of dynamic reservoir in the echo state network,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 3, 2012.
- [5] J. Qiao, F. Li, H. Han, and W. Li, “Growing echo-state network with multiple subreservoirs,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 2, pp. 391–404, 2017.
- [6] Q. Song and Z. Feng, “Effects of connectivity structure of complex echo state network on its prediction performance for nonlinear time series,” *Neurocomputing*, vol. 73, no. 10–12, pp. 2177–2185, 2010.
- [7] O. Fink, E. Zio, and U. Weidmann, “Fuzzy Classification With Restricted Boltzman Machines and Echo-State Networks for Predicting Potential Railway Door System Failures,” *IEEE Transactions on Reliability*, vol. 64, no. 3, pp. 861–868, 2015.
- [8] S. I. Han and J. M. Lee, “Output-tracking-error-constrained robust positioning control for a nonsmooth nonlinear dynamic system,” *IEEE Transactions on Industrial Electronics*, vol. 61, no. 12, pp. 6882–6891, 2014.
- [9] B. Zhang, D. J. Miller, and Y. Wang, “Nonlinear system modeling with random matrices: Echo state networks revisited,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 1, pp. 175–182, 2012.
- [10] E. A. Antonelo and B. Schrauwen, “On learning navigation behaviors for small mobile robots with reservoir computing architectures,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 4, pp. 763–780, 2015.
- [11] A. Rodan and P. Tiño, “Minimum complexity echo state network,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 22, no. 1, pp. 131–144, 2011.
- [12] T. D. Wang, X. Wu, and C. Fyfe, “Minimal echo state networks for visualisation,” *Discrete & Computational Geometry*, vol. 7, no. 3, pp. 219–226, 2011.
- [13] M. Chen, W. Saad, C. Yin, and M. Debbah, “Echo State Networks for Proactive Caching in Cloud-Based Radio Access Networks with Mobile Users,” *IEEE Transactions on Wireless Communications*, vol. 99, pp. 3520–3535, 2017.
- [14] H. Duan and X. Wang, “Echo state networks with orthogonal pigeon-inspired optimization for image restoration,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 11, pp. 2413–2425, 2016.
- [15] F. M. Bianchi, L. Livi, and C. Alippi, “Investigating echo-state networks dynamics by means of recurrence analysis,” *IEEE Transactions on Neural Networks Learning Systems*, vol. 99, pp. 1–13, 2016.
- [16] Z. Shi and M. Han, “Support vector echo-state machine for chaotic time-series prediction,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 18, no. 2, pp. 359–372, 2007.
- [17] M. Xu and M. Han, “Adaptive Elastic Echo State Network for Multivariate Time Series Prediction,” *IEEE Transactions on Cybernetics*, vol. 46, no. 10, pp. 2173–2183, 2016.
- [18] C. Chen, L. Li, H. Peng, and Y. Yang, “Finite time synchronization of memristor-based Cohen-Grossberg neural networks with mixed delays,” *PLoS ONE*, vol. 12, no. 9, pp. 1–19, 2017.
- [19] E. Roh, D. K. Song, and M.-S. Kim, “Emerging role of the brain in the homeostatic regulation of energy and glucose metabolism,” *Experimental & Molecular Medicine*, vol. 48, no. 3, article no. e216, 2016.
- [20] W. B. Levy and R. A. Baxter, “Energy Efficient Neural Codes,” *Neural Computation*, vol. 8, no. 3, pp. 531–543, 1996.
- [21] M. Zheng, L. Li, H. Peng et al., “Globally fixed-time synchronization of coupled neutral-type neural network with mixed time-varying delays,” *Plos One*, vol. 13, no. 1, 2018.
- [22] C. Howarth, P. Gleeson, and D. Attwell, “Updated energy budgets for neural computation in the neocortex and cerebellum,” *Journal of Cerebral Blood Flow & Metabolism*, vol. 32, no. 7, pp. 1222–1232, 2012.
- [23] S. Aliakbary, S. Motallebi, S. Rashidian, J. Habibi, and A. Movaghar, “Distance metric learning for complex networks: Towards size-independent comparison of network structures,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 2, 2015.
- [24] A. Knoblauch, G. Palm, and F. T. Sommer, “Memory capacities for synaptic and structural plasticity,” *Neural Computation*, vol. 22, no. 2, pp. 289–341, 2010.

- [25] A. Emin Orhan and W. J. Ma, “Efficient probabilistic inference in generic neural networks trained with non-probabilistic feedback,” *Nature Communications*, vol. 8, no. 1, p. 138, 2017.
- [26] S. Atasoy, I. Donnelly, and J. Pearson, “Human brain networks function in connectome-specific harmonic waves,” *Nature Communications*, vol. 7, 2016.
- [27] B. Sengupta, S. B. Laughlin, and J. E. Niven, “Balanced Excitatory and Inhibitory Synaptic Currents Promote Efficient Coding and Metabolic Efficiency,” *PLoS Computational Biology*, vol. 9, no. 10, pp. 200–203, 2013.



Hindawi

Submit your manuscripts at
www.hindawi.com

