2022 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2022, Virtual, December 12-15, 2022

Application of a multi-layer reservoir neural network to the prediction of spatiotemporal chaos

Shoki Yabu and Yo Horikawa

Faculty of Engineering, Kagawa University
2217 Hayashi-cho, Takamatsu 761-0396, Japan
Email: s21g482@kagawa-u.ac.jp, horikawa.yo@kagawa-u.ac.jp

Abstract –Reservoir Computing (RC) is one of recurrent neural networks that can process temporal information. In RC, the weights of input and reservoir layers are fixed and only the weights of an output layer are learned; RC has been used to predict various time series and spatiotemporal patterns. Recently, a parallel RC with multiple reservoirs assigned to different spatial domains has been proposed to predict spatiotemporal chaos, such as in the Kuramoto-Sivashinsky model [1, 2]. In this study, we apply a multi-layer reservoir neural network to predict spatiotemporally chaos. This model consists of the layers of reservoirs connected in parallel and a reservoir with spatial averages as input. It's performance is then compared to the parallel RC [1].

1 Introduction

Nonlinear dynamics appears in weather, economic, and audio data, and their prediction is a technique that contributes widely to our real lives. In recent years, there has been growing interest in reservoir computing as a machine learning model that can learn quickly. RC has self-feedback and has excellent properties for information processing that depends on past input. The greatest advantage of RC is that it requires less computation for training than other recurrent neural networks (RNNs). This allows for fast learning and can be handled well by standard computers.

Parallel reservoir computing systems have been applied as models for predicting large-scale spatiotemporal chaos with large degrees of freedom and spatially disordered structure in addition to temporal irregularities [1, 2]. In this study, we apply a multi-layer reservoir neural network to predict spatiotemporal chaos. We use it to predict the Lorenz-96 model [3]. Then, we evaluate its performance, and compare its performance with a parallel RC [1].

2 Model

2.1 Reservoir computing (RC)

RC is one of RNN and has one hidden layer called reservoir. Figure 1 shows the architecture of RC. The values of parameters for RC are listed in Table 1. Unless specified otherwise in this experiment, the parameter values in Table 1 are used.

Let $\boldsymbol{u}_t \in \mathbb{R}^N$ be the input vector at time t, $\boldsymbol{d}_t \in \mathbb{R}^N$ be the target vector, and D_r be the number of nodes in the reservoir. In the case of predicting a system consisting of N elements, in which $\boldsymbol{X}_t \in \mathbb{R}^N$ is the state of the system at time t, we let $\boldsymbol{u}_t = \boldsymbol{X}_t$ and $\boldsymbol{d}_t = \boldsymbol{X}_{t+1}$.

The state of the reservoir $\boldsymbol{r}_t$ is updated as

ORCID iDs Shoki Yabu: 0000-0001-8776-7252,
Yo Horikawa: 0000-0002-4298-2095

Eq. (1).

$$r_{t+1} = (1-\delta)r_t + \delta(\tanh(Ar_t + W_{in}u_t)) \quad (1)$$

Where W_{in} is the $D_r \times N$ input matrix that maps the input vector to the reservoir space and A is the $D_r \times D_r$ adjacency matrix that determines the reservoir dynamics.The components of W_{in} are sampled from a uniform distribution in $[-\alpha, \alpha]$. The matrix A is the non-zero component with random values sampled from the uniform distribution with proportion d, and is normalized to have the maximum eigenvalue ρ. The output vector $y_t \in \mathbb{R}^N$ can be obtained from the state of the reservoir as follows.

$$y_t = W_{out} r_t \quad (2)$$

where W_{out} is the $D_r \times N$ output matrix. The reservoir state series $\{r_0, r_1, \ldots, r_t\}$ is generated using Eq.(1) with the training data series $\{u_0, u_1, \ldots, u_t\}$. Using them, W_{out} is devived with ridge regression.

$$W_{out} = U_t R^T (RR^T + \beta I)^{-1} \quad (3)$$

where I is the unit matrix of $D_r \times D_r$ and R and U_t are matrices whose t-th columns are the states r_t of the reservoir and the target values d_t, respectively.

In forecasting, the output y_t is given as the next input u_{t+1}. Then, the forecast is made iteratively.

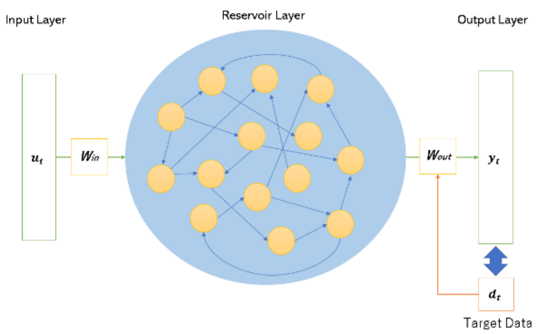

Figure1. Reservoir computing architecture

Table1. Parameter values of RC

PARAMETER	*DESCRIPTION*	*VALUE*
Dr	reservoir size	500
δ	leak rate	1.0
α	Input scale	1.0
d	adjacency matrix density	0.05
ρ	adjacency matrix spectral radius	0.95
β	ridge regression parameter	0.0001
g	Number of reservoir	8
l	Reservoir input overlaps	2

2.2 Parallel reservoir computing

A parallel RC connects g reservoirs in parallel as shown in Fig. 2, so that all of them contain $q = N/g$ number of elements. The following input is given to each reservoir.

$$h_t^i = \left(x_{t,(i-1)\times q+1-l}, x_{t,(i-1)\times q+1-l}, \ldots, x_{t,i\times q+l}\right)^T \quad (i = 1,2,\ldots,g), (4)$$

Then, the output weight $W_{out}^0, W_{out}^1, \ldots, W_{out}^i, \ldots$ for each reserver is learned. From the learned weight, q outputs can be obtained for each reserver.

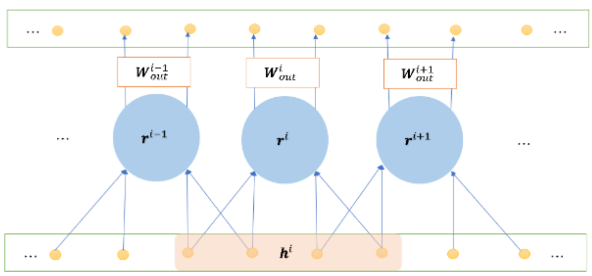

Figure2. Parallel reservoir computing

2.3 Multi-layer reservoir computing

In a multi-layer RC, a new reservoir $\boldsymbol{r}_A$ is connected to all the parallelized reservoirs as shown in Fig. 3. The reservoirs are given the following average $\boldsymbol{k}_t^0, \boldsymbol{k}_t^1, \ldots, \boldsymbol{k}_t^g$ for every q inputs.

$$\boldsymbol{k}_t^i = \frac{\left(x_{t,i\times q} + x_{t,i\times q+1} + \cdots + x_{t,i\times q+q-1}\right)}{q}$$

$$(i = 1, 2, \ldots, g) \quad (5)$$

This allows prediction to be made using global as well as local input.

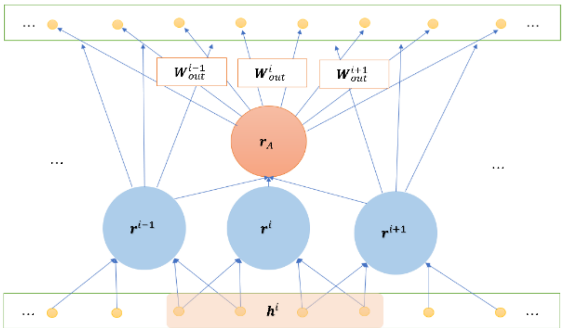

Figure3. Multi-layer reservoir computing

3 Experiment

The prediction of changes the state of the Lorenz-96 (L96) model [3] was carried out using the parallel RC and the multi-layer RC The L96 model is expressed by Eq. (6).

$$\frac{x_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F$$

$$(i = 1, \ldots, N) \quad (6)$$

Where $x_{-1} = x_{N-1}, x_0 = x_N$ and $x_{N+1} = x_1$. This model exhibits a chaotic behavior at $F = 8$.

The data were generated by calculating equation Eq. (6) with $F = 8$ and a time increment Δt=0.005. A total of 4000 steps of the data was used for training. The following RMSE was used to evaluate the prediction accuracy.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_t^i - y_t^i)^2} \quad (7)$$

The results of the prediction when $N = 64$ using the hierarchical reservoir are shown in Fig. 4.

Experiments were conducted to evaluate the performance of the multi-layer RC with $g = 8$ and varying N, and with $N = 96$ and varying g. Then, we compared the performance of the multi-layer RC with that of the parallel RC in each case. The RMSE per time step was employed for performance evaluation, and the number of time steps until the RMSE exceeds 1 was employed for performance comparison.

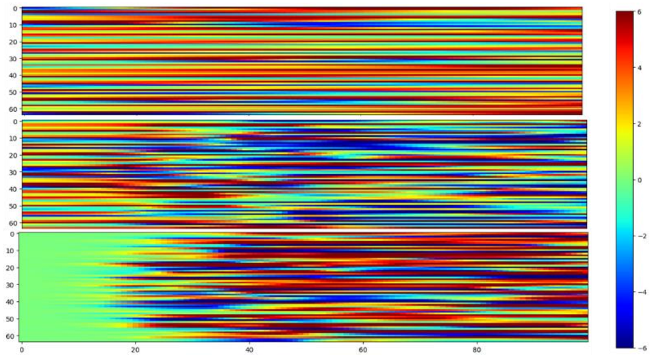

Figure4. (a)data (b)prediction by multi-layer reservoir (c) forecast errors

4 Result

The performance evaluation of the multi-layer RC and the performance comparison with the parallel RC when $N = 96$ is fixed and g is varied are shown in Figs. 5 and 6, respectively.

The Fig. 6 shows that the performance of multi-layer RC is better than the parallel RC as the number g of reservoirs increases.

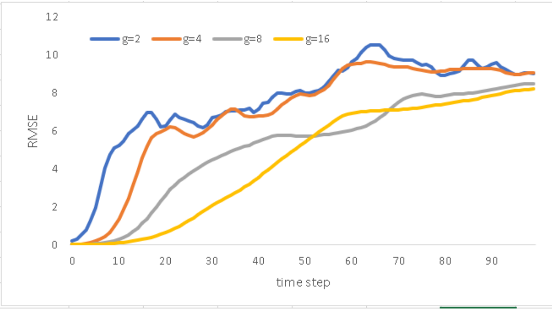

Figure5. Performance of the multi-layer RC ($N = 96$)

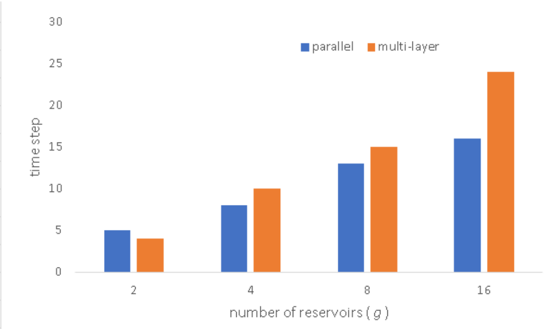

Figure6. Comparison between the multi-layer RC and the parallel RC ($N = 96$)

The performance evaluation of the multi-layer RC and the performance comparison with the parallel RC when $g = 8$ is fixed and N is varied are shown in Figs. 7 and 8, respectively. From Figs. 7 and 8, it can be seen that the performance of the multi-layer RC improves, especially when the input q for a single reservoir is small.

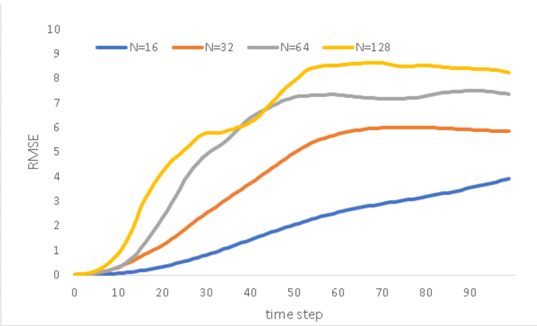

Figure7. Performance of the multi-layer RC ($g = 8$)

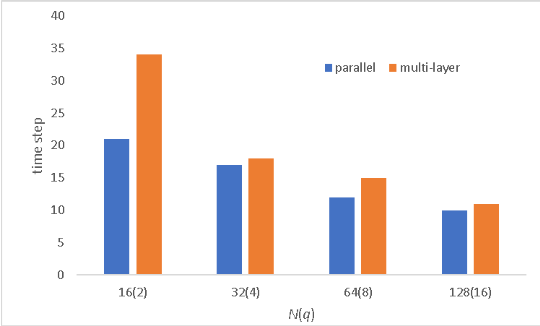

Figure8. Comparison between multi-layer RC and parallel RC ($g = 8$)

5 Conclusion

In this study, we designed the multi-layer RC and evaluated its prediction performance of spatiotemporal chaos in comparison with the parallel RC [1] using the L96 model. The results showed that the multi-layer RC has better prediction performance than the parallel RC. It was also found that the perfor-mance of the multi-layer RC improves more when the number of inputs of the reservoirs conn-ected in parallel is smaller.

References

[1] Pathak, J. et al. "Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing appro-ach." *Physical review letters* 120, 2 (2018) 024102.

[2] Tomizawa, F. and Sawada, Y. "Combin-ing ensemble Kalman filter and reservoir computing to predict spatiotemporal cha-otic systems from imperfect observations and models. " *Geoscientific Model Deve-lopment* 14, 9 (2021) 5623-5635.

[3] Lorenz, E. N. and Kerry A. E. "Optimal sites for supplementary weather observ-ations: Simulation with a small model. " *Journal of the Atmospheric Sciences* 55, 3 (1998) 399-414.