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

Characteristics representation of reservoir set based on memory capacity and nonlinearity

Tomoya Kitamura†, Kazuyuki Yoshimura‡

†Graduate School of Sustainability Science, Tottori University
4-101 Koyama Minami, Tottori 680-8552, Japan
‡Faculty of Engineering, Tottori University
4-101 Koyama Minami, Tottori 680-8552, Japan
Email:m21j4012c@edu.tottori-u.ac.jp, kazuyuki@tottori-u.ac.jp

Abstract—Reservoir Computing has various degrees of freedom in terms of reservoir parameters, topology and activation functions; however, the design strategy of a good reservoir is still unclear. In this paper, we focus on memory capacity and nonlinearity, where the memory and nonlinearity indices represent the characteristics of the reservoir dynamics. And we show that the two indices correlate with the information processing performance of the reservoir for a particular benchmark task.

1. Introduction

Various dynamic systems such as electronic circuits, optical devices such as lasers, and fluid systems have the property of "obtaining the same output for the same input signal that is repeatedly input," and the information processing method that utilizes this property is reservoir computing (RC) [1]. Reservoir computing is mainly used for problems involving time series data, such as time series classification, time series generation, and time series prediction. Reservoir computing consists of three layers: an input layer, an intermediate layer (reservoir) using dynamical systems, and an output layer. The intermediate layer (reservoir) is fixed without learning, and only the coupling strength from the intermediate layer to the output layer is learned using the Least Squares Method (LSM). Since the reservoir does not need to be trained, various physical dynamical systems can be used as reservoirs, making it suitable for physical implementations. By making good use of the characteristics of physical phenomena in RC, an innovative "computer" that achieves high speed and energy saving is expected to be realized.

A RC that uses a recurrent neural network (RNN), which is a neural network with loops in the coupling between units, as a reservoir is called an Echo State Network (ESN) [1]. The numerical experiments in this paper deal with ESN.

When applying ESN to various tasks, such as time series forecasting, the optimal structure of RNN is considered to be task-specific: the structure of RNN includes various parameters, the functional form of the unit activation function, and the coupling topology among the units. When there are many elements to determine the reservoir structure, the reservoir structure to be searched becomes huge, and it becomes difficult to select the optimal structure [2]. From the viewpoint of clarifying guidelines for selecting an appropriate reservoir for each task from a huge set of reservoirs, a standard index that expresses reservoir characteristics is desired first.

In recent years, it has been revealed that there is a universal trade-off between memory capacity and nonlinear processing capacity in the computational performance of reservoirs, and it has been suggested that these are fundamental properties that characterize reservoirs [3, 4]. In this paper, we used memory capacity [5] and nonlinearity [3] as candidates for the aforementioned indices, and investigated whether these two indices can be used to express the characteristics of a reservoir for a mixed unit-type reservoir [4].

2. Reservoir model

2.1. Reservoir model

Although various structures can be considered for RNN to be used as reservoirs for ESN, in this paper, two types of RNN are used as reservoirs: RNN with random coupling and ring coupled RNN.

2.2. State variables and time evolution equations

The reservoir under consideration in this paper consists of one bias unit and N units, where the bias unit is numbered 0 and the N units are numbered $1, 2, 3, ..., N$. The i-th unit has a state variable x_i attached to it. The i-th unit has a state variable x_i, and the state variable x_i of each unit except the bias unit depends on the discrete time variable $t = 0, 1, 2 \ldots$. The time evolution of x_i is assumed to follow the following equation.

$$x_i(t) = \phi_i \left[g \left\{ \sum_{i=1}^{N} J_{ij} x_j(t-1) + \alpha \varepsilon_i S(t) \right\} + \beta J_{i0} x_0 \right] \quad (1)$$

ORCID iDs Tomoya Kitamura: 0000-0002-1708-1893, Kazuyuki Yoshimura: 0000-0002-4329-4999

This work is licensed under a Creative Commons Attribution NonCommercial, No Derivatives 4.0 License.

The state variable x_0 of the bias unit is always 1 without time evolution. The first combination of activation functions that we deal with here is $\phi_i(x) = \mathrm{sinc}(x), \mathrm{ReLU}^c_{-c}(x)$, and the second combination is $\phi_i[x] = \tanh(x), \mathrm{ReLU}^c_{-c}(x)$. where $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$. On the other hand, $\mathrm{ReLU}^c_{-c}(x)$ is a function that $x \to \min(\max(-c, x), c)$ for input x. $\mathrm{ReLU}^c_{-c}$is a linear function with a cutoff on the linear function to prevent divergence of the reservoir dynamics. The ratio of nonlinear function units to linear function units in the reservoir is determined by the mixing ratio p. That is, $p = A/N$, where A is the number of units that have nonlinear functions as their activating functions.

J_{ij} is the component of the coupling matrix J between units in the reservoir, and J_{ij} represents the coupling coefficient from the j-th unit to the i-th unit. The coupling matrix of a reservoir consisting of N units, not including the bias unit, is generally given by the following square matrix.

$$J = \begin{pmatrix} J_{11} & \cdots & J_{1i} & \cdots & J_{1N} \\ \vdots & \ddots & & & \vdots \\ J_{i1} & & J_{ii} & & J_{iN} \\ \vdots & & & \ddots & \vdots \\ J_{N1} & \cdots & J_{Ni} & \cdots & J_{NN} \end{pmatrix} \tag{2}$$

There are two types of reservoir topologies: random coupling and ring coupled. In the random coupled type reservoir, all the above matrix components J_{ij} follow a uniform distribution $U[-1, 1]$ with zero mean. In the ring coupled type reservoir, only the matrix component J_{ij} with adjacent unit number follows the uniform distribution $U[-1, 1]$ with zero mean, and the remaining matrix components J_{ij} are treated as zero, as in equation (3).

$$J = \begin{pmatrix} 0 & J_{12} & 0 & \cdots & 0 \\ 0 & 0 & J_{23} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & J_{N-1N} \\ J_{N1} & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{3}$$

The coupling coefficient J_{i0} of the bias units is randomly given by 1 or -1 for each i. $S(t) \in \mathbb{R}$ denotes the input from the input layer to each unit. g, $\alpha, \beta \in \mathbb{R}$ are parameters greater than or equal to zero.

3. Memory capacity and nonlinearity

Consider the case where a single task is defined for performance evaluation. That is, given an unknown functional relation $y(t)$ of the task, we assume the case where $C_T(X, y)$ is introduced according to the literature [3]. In this case, the capacity $C_T(X, y)$ is introduced according to the literature [3], where $C_T(X, y)$ is a measure of the accuracy with which the reservoir X estimates the function y, and is defined by the following equation (4).

$$C_T(X, y) = 1 - \frac{1}{\langle y^2 \rangle_T} \min_{\vec{w}} E(\vec{w}) \tag{4}$$

T is the number of data, $\hat{y}$ is the output of the reservoir, $E(\vec{w})$ is the mean squared error, and $\min E(\vec{w})$ is the weight $w_i, i = 0, 1, ..., N$ of the combination that gives the minimum value w_i,$i = 0, 1, ..., N$ such that the minimum value is given. $\langle y^2 \rangle_T$ denotes the mean of the square 1/T $\sum_{t=1}^{T}(y(t))^2$.

Storage capacity [5] and nonlinearity [3] are indices defined based on the ability of the reservoir to approximate a large number of functions. The function y used in this case is explained below. First, $\{d_i\}$ is an infinite sequence of numbers satisfying $d_i \in \mathbb{Z}, d_i \geq 0$. The function $y_{\{d_i\}}$ is defined as follows.

$$y_{\{d_i\}}(t) = \prod_{i=0}^{\infty} P_{d_i}(s(t-i)) \tag{5}$$

where $P_n(x)$ is an n-th order Legendre polynomial and is defined by the following equation.

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}(x^2 - 1)^n, \ n = 0, 1, 2, ... \tag{6}$$

The right-hand side of equation (5) is an infinite product of Legendre polynomial, and the input $s(t - i)$ to the reservoir at time $t - i$, i steps before time t, is used as the argument for each P_{d_i}. In the numerical experiments in the following section, $s(t)$ is assumed to be a random variable that is independent at each time and follows a uniform distribution $U[-1, 1]$. Next, we discuss the storage capacity L [5] and the nonlinear transformation capability NL [3]. Denote the reservoir under consideration by X. For this reservoir, L is defined by the following equation.

$$L[X] = \sum_{\{d_i\}, d_i \in \{0,1\}} \left(\sum_{i=0}^{\infty} d_i \right) C_T^a(X, y_{\{d_i\}}) \tag{7}$$

$C_T^a(X, y_{\{d_i\}})$, where $y_{\{d_i\}}$ in equation (5) is adopted as y in equation (4), and from this $C_T^a(X, y_{\{d_i\}})$ is defined as follows.

$$C_T^a(X, y_{\{d_i\}}) = \theta(C_T(X, y_{\{d_i\}}) - a) C_T(X, y_{\{d_i\}}) \tag{8}$$

where $\theta(x)$ is the Heaviside function defined by the following equation.

$$\theta(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{9}$$

On the other hand, for a reservoir X, NL is defined as follows.

$$NL[X] = \sum_{\{d_i\}} \left(\sum_{i=0}^{\infty} d_i \right) C_T^a(X, y_{\{d_i\}}) - L \tag{10}$$

Summing over all possible $\{d_i\}$ the summation in equation (10) includes sums over $\{d_i\}$ such that L is constructed. L is subtracted in equation (10) to eliminate the effect of these terms.

4. Benchmark Task

We investigate whether the (L, NL) pairs are useful as a measure to characterize the computational performance of a reservoir. For this purpose, we give a specific task and investigate the correlation between the value of (L, NL) and the reservoir's performance on that task. The following function approximation problem is used as the specific task. The function approximation problem is described below. The following function is used.

$$y(t) = \sin\left(\pi\nu \sum_{i=0}^{\tau} \frac{s(t-i)}{\sqrt{\tau+1}}\right) \tag{11}$$

where the input s is assumed to follow a uniform distribution $U[-1, 1]$. The parameter $\nu \in \mathbb{R}$ controls the degree of nonlinearity of the task; the larger the value of ν, the more nonlinear the task is. The parameter $\tau \in \mathbb{N}$ controls the time delay. By manipulating these parameters ν and τ, the nonlinearity and time delay of the task can be adjusted, and the performance of the reservoir can be studied in detail. The reservoir training for the above task is described below. The input signal to the reservoir at time t is $s(t)$, and the target value of learning is $y(t)$, obtained by equation (11). These many pairs $D = \{s(t), y(t)\}_{t=1}^{T}$ are the training data.

5. Numerical experiment results

In this section, we present the results of numerical experiments. For each reservoir X in the reservoir set, a value of (L, NL) can be computed. By plotting the points of this value on the (L, NL) plane, a scatter plot is obtained. Figures 1, 2, and 3 show the NMSE of a function approximation problem on a plane with L on the horizontal axis and NL on the vertical axis. Here, the number of units in the reservoir is set to $N = 100$. The number of training data $T = 4000$ for the task and function approximation problem used to find L and NL. The left and right sides of each figure correspond to the results for the randomly coupled and ring-coupled reservoirs, respectively.

In Figures 1, 2, and 3, the parameter τ, which determines the time delay of the function approximation problem, is reduced to 6, 3, and 1, respectively, while the parameter ν, which determines the degree of nonlinearity, is increased to $1/\sqrt{2}$, 1.0, and 2.0, respectively. Figure 1 shows the NMSE of the function approximation problem when the parameters $(\nu, \tau) = (1/\sqrt{2}, 6)$. Compared to Figures 2 and 3, this task has a larger time delay τ, which is a parameter of the function approximation problem, and a smaller nonlinearity ν. Under these parameter conditions, the NMSE tends to be low in the region where L is large. Next, Figure 2 shows the NMSE for the parameter $(\nu, \tau) = (1.0, 3)$ for the function approximation problem. Compared to Figures 1 and 3, both the time delay τ and the degree of nonlinearity ν are intermediate for this task. For this parameter condition, the NMSE tends to be low in regions where both L and NL are large. Finally, Figure 3 shows the NMSE for the parameter (ν, τ)=(2.0,1) for the function approximation problem. Compared to Figures 1 and 2, this task has a smaller time delay τ and a larger nonlinearity ν. Under these parameter conditions, the NMSE tends to be low in regions where L is small and NL is large. These results indicate that there is a certain correlation between the value of (L, NL) and the NMSE of function approximation problems. It can also be seen that the reservoirs corresponding to the boundaries of the colored regions on the (L, NL) plane tend to have a low NMSE for the function approximation problem.

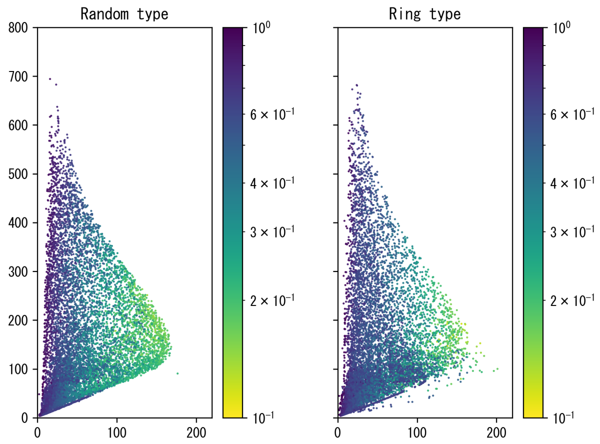

Figure 1: function approximation task. (ν, τ)=$(1/\sqrt{2}, 6)$

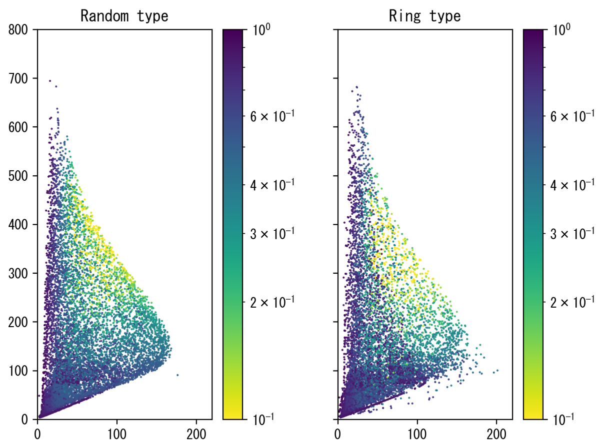

Figure 2: function approximation task. (ν, τ)=(1.0, 3)

6. Discussion

In this section, we discuss the possibility of utilizing the (L, NL)-valued plots of the reservoir set proposed in this paper.

The computational results in Section 5 show that there is a certain correlation between the indicator (L, NL) and

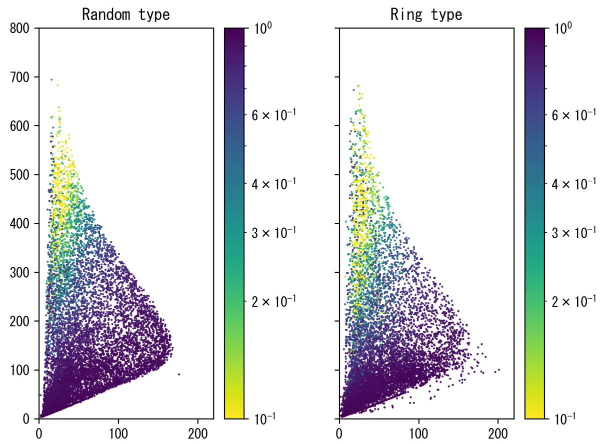

Figure 3: function approximation task. (ν, τ)=(2.0,1)

the reservoir computation performance in the function approximation task. It is expected that various tasks other than the function approximation task treated here also have a balance between memory capacity and nonlinear conversion capability suitable for their processing, i.e., the value of (L, NL).

Given an arbitrary task, we consider the case of constructing a new reservoir suitable for the task. In this case, the parameter values in the reservoir need to be appropriately selected. Once a plot of the (L, NL) values of the reservoir set corresponding to various changes in the parameter values is obtained, it is possible to determine the correspondence between changes in the parameter values and changes in the reservoir's basic properties: memory capacity and nonlinearity. In particular, we can find a method of changing parameters to change their balance while keeping the memory capacity and nonlinearity as large as possible, respectively. By adjusting parameters according to this method, it is possible to obtain parameter settings more efficient than brute-force search and more suitable for the task. This suggests that the method of understanding reservoir characteristics through parameters and indices L and NL is effective.

In addition, the (L, NL) value plots have the potential to be used to evaluate the characteristics of various tasks. There are various tasks to evaluate reservoir performance, but the required memory capacity or nonlinearity for each task is not known in advance. In such a case, one possible use is to analyze which reservoirs with low NMSE exist in which region on the index L,NL plane when one reservoir set is fixed, in order to clarify the degree of memory capacity and nonlinearity required by the task in question.

7. Conclusion

In this paper, we focused on two indices, memory capacity L and nonlinearity NL, in a RC with a mixed-unit type reservoir with random and ring coupling [4]. We then investigated whether these two indices can be used to characterize the reservoir performance based on the relationship between (L, NL) values and NMSE in function approximation problems. The results show that there is a certain correlation between the value of (L, NL) and the NMSE of function approximation problems. Therefore, the index L, NL can be used to characterize the performance of the reservoir set.

As future work, we would like to investigate whether the definition of the indices L and NL can be revised or new indices can be introduced to improve the generality of the indices.

References

[1] H. Jaeger, German National Research Center for Information Technology GMD Technical Report 148, (2001).

[2] Matthew Dale,Julian F. Miller,Susan Stepney,Martin A. Trefzer,Reservoir Computing,pp.141-166 (2021)

[3] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, Scientific Reports 2, 514 (2012).

[4] M.Inubush, K. Yoshimura,Scientific Reports vol.7 (2017) 10119.

[5] H. Jaeger, H. Haas, Science 304, 78-80 (2004).