Attention-Enhanced Reservoir Computing for Modeling Diverse Dynamical Systems

Felix Köster*, Kazutaka Kanno*, and Atsushi Uchida*
*Department of Information and Computer Sciences, Saitama University, Saitama City, Japan
Emails: felixk@mail.saitama-u.ac.jp, auchida@mail.saitama-u.ac.jp

***Abstract*—Reservoir computing (RC) has emerged as a powerful framework for time series prediction, offering versatile implementations. However, its application to modeling a broad spectrum of dynamical systems remains limited due to inflexibility of the training scheme. In this work, we introduce an attention-enhanced reservoir computing model that incorporates an attention mechanism at the output stage, enabling dynamic prioritization of important features. Experimental evaluations show that this approach achieves high prediction accuracy across multiple dynamical systems, underscoring its potential as a universal simulator for chaotic and complex systems.**

***Index Terms*—Reservoir computing, attention mechanism, dynamical systems, time-series prediction, machine learning.**

I. Introduction

Accurately predicting chaotic time series is a challenging problem due to their sensitivity to initial conditions and nonlinearity. Reservoir computing (RC) has shown promise in capturing complex temporal dependencies [1]–[3]. However, traditional RC frameworks are often limited by their inability to dynamically adjust to varying system behaviors.

Recent advancements in machine learning, such as attention mechanisms [4], offer the ability to focus selectively on relevant features within input sequences, significantly enhancing prediction accuracy [5]. Attention mechanisms have transformed sequence modeling in fields such as natural language processing and machine translation. Inspired by these developments, this study integrates an attention mechanism with the reservoir output layer to improve its predictive capabilities for a variety of chaotic systems [6].

We evaluated the proposed attention-enhanced reservoir computing (AERC) model on multiple well-known dynamical systems, demonstrating its ability to simulate diverse systems with high accuracy and stability. The results highlight the model's potential to serve as a universal simulator for complex time series data.

II. Methodology

A. Echo State Networks (ESNs)

Echo State Networks (ESNs) are a type of recurrent neural network (RNN) designed to efficiently handle sequential data using a fixed, randomly connected reservoir of neurons. This reservoir transforms input data into a higher-dimensional space, enabling the ESN to capture complex temporal dependencies without extensive training of recurrent connections.

The reservoir states, $\mathbf{x}_l$, are updated as:

$$\mathbf{x}_l = \tanh(\mathbf{W}_{res}\mathbf{x}_{l-1} + \mathbf{W}_{in}\mathbf{u}_l + \mathbf{b}), \tag{1}$$

where $\mathbf{W}_{res}$ is the reservoir weight matrix, $\mathbf{W}_{in}$ is the input weight matrix, and $\mathbf{b}$ is a bias vector. The output of the ESN, $\mathbf{y}_l$, is computed as:

$$\mathbf{y}_l = \mathbf{W}_{out}\mathbf{x}_l, \tag{2}$$

where $\mathbf{W}_{out}$ is trained using ridge regression, which ensures computational efficiency and robustness by regularizing the output weights.

B. Attention Mechanism

An attention mechanism is integrated at the output layer to dynamically assign importance to the reservoir states based on their relevance to the current input sequence. This enables the model to focus on the most critical features of the reservoir's representation, enhancing its adaptability and predictive accuracy for complex temporal patterns.

The attention weights, $\mathbf{w}_{att,l}$, are computed using a neural network:

$$\mathbf{w}_{att,l} = F(\mathbf{W}_{net}, \mathbf{r}_l), \tag{3}$$

where F is a neural network with parameters $\mathbf{W}_{net}$ and input $\mathbf{r}_l$ (the reservoir states at time l). These attention weights dynamically adjust to emphasize features that are most informative for the prediction task.

The final output is computed as a weighted sum of the reservoir states:

$$d_l = \mathbf{w}_{att,l}^\top \mathbf{r}_l. \tag{4}$$

This mechanism allows the model to change its focus dynamically for each time step, making it highly effective at handling systems with varying time-dependencies. By utilizing attention, the ESN can adaptively prioritize reservoir nodes, improving robustness and accuracy in modeling diverse dynamical systems. A flow chart diagram of the approach is shown in Fig. 1.

C. Dual-Objective Training of Attention-Enhanced Reservoir Computer

The AERC is trained with two simultaneous objectives: predicting the next time step of the dynamical system and classifying the attractor the system resides on. The input is the time series $\mathbf{x}_l$ which is processed by the AERC to produce

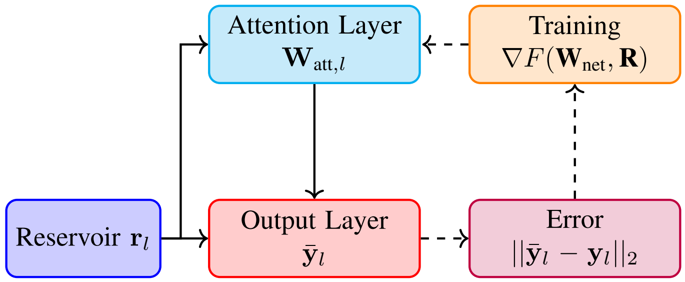

Fig. 1. Float chart diagram of the AERC layout. Reservoir states are collected. A gradient descent approach to optimize the attention layer is used. The dashed arrows show the training process, which are deactivated after training.

two output vectors, $\mathbf{y}_{l1}$ for prediction and $\mathbf{y}_{l2}$ for attractor classification.

The prediction vector $\mathbf{y}_{l1}$ is trained to minimize the Mean Squared Error (MSE) between the predicted and true next time step:

$$L_{\text{MSE}} = \frac{1}{L}\sum_{l=1}^{L}(\mathbf{y}_{l1} - \bar{\mathbf{y}}_{l1})^2,$$

where $\mathbf{y}_{l1}$ is the true next step, $\bar{\mathbf{y}}_{l1}$ is the predicted value, and L is the total number of data points.

Simultaneously, the classification vector $\mathbf{y}_{l2}$ is trained using cross-entropy loss:

$$L_{\text{CE}} = -\frac{1}{L}\sum_{l=1}^{L}\mathbf{y}_{l2}\log(\bar{\mathbf{y}}_{l2}),$$

where $\mathbf{y}_{l2}$ represents the true class probabilities, and $\bar{\mathbf{y}}_{l2}$ is the predicted class distribution.

The total loss combines these objectives:

$$L_{\text{total}} = L_{\text{MSE}} + L_{\text{CE}}.$$

The classification output is used only for performance evaluation and not revealed during training or prediction to ensure the model relies solely on the input dynamics. The network parameters $\mathbf{W}_{\text{net}}$ are updated via gradient descent to minimize L_{total}. After training, the AERC is evaluated in a closed-loop configuration, predicting system dynamics and identifying attractor classes autonomously.

III. Experimental Setup

A. Dynamical Systems Evaluated

The AERC model was tested on five well-known dynamical systems, each exhibiting unique characteristics and providing diverse challenges for time-series prediction:

1) **Lorenz System**:

$$\frac{dx}{dt} = \sigma(y - x), \tag{5}$$
$$\frac{dy}{dt} = x(\rho - z) - y, \tag{6}$$
$$\frac{dz}{dt} = xy - \beta z, \tag{7}$$

where $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$ [7].

2) **Rössler Attractor**:

$$\frac{dx}{dt} = -y - z, \tag{8}$$
$$\frac{dy}{dt} = x + ay, \tag{9}$$
$$\frac{dz}{dt} = b + z(x - c), \tag{10}$$

with parameters $a = 0.2$, $b = 0.2$, and $c = 5.7$ [8].

3) **Henon Map**:

$$x_{n+1} = 1 - ax_n^2 + y_n, \tag{11}$$
$$y_{n+1} = bx_n, \tag{12}$$

with $a = 1.4$ and $b = 0.3$ [9].

4) **Duffing Oscillator**:

$$\frac{dx}{dt} = v, \tag{13}$$
$$\frac{dv}{dt} = -\delta v - \alpha x - \beta x^3 + \gamma\cos(\omega t), \tag{14}$$

where $\delta = 0.5$, $\alpha = -1.0$, $\beta = 1.0$, $\gamma = 0.3$, and $\omega = 1.2$ [10].

5) **Mackey-Glass Equation**:

$$\frac{dx}{dt} = \beta\frac{x(t-\tau)}{1 + x(t-\tau)^n} - \gamma x(t), \tag{15}$$

with $\beta = 0.2$, $\gamma = 0.1$, $\tau = 17$, and $n = 10$ [11].

These systems were selected to cover a range of behaviors, from continuous-time chaotic attractors (Lorenz, Rössler) to discrete-time maps (Henon) and driven systems (Duffing), as well as delay differential equations (Mackey-Glass). This diversity enables a wide evaluation of the AERC model's ability to generalize across varying dynamical properties.

B. Performance Metrics

To evaluate the prediction quality of the AERC model, two key performance metrics were employed: the Valid Prediction Time (VPT) and the spectral and histogram similarity. These metrics provide a comprehensive assessment of the model's accuracy and stability in both open-loop and closed-loop configurations.

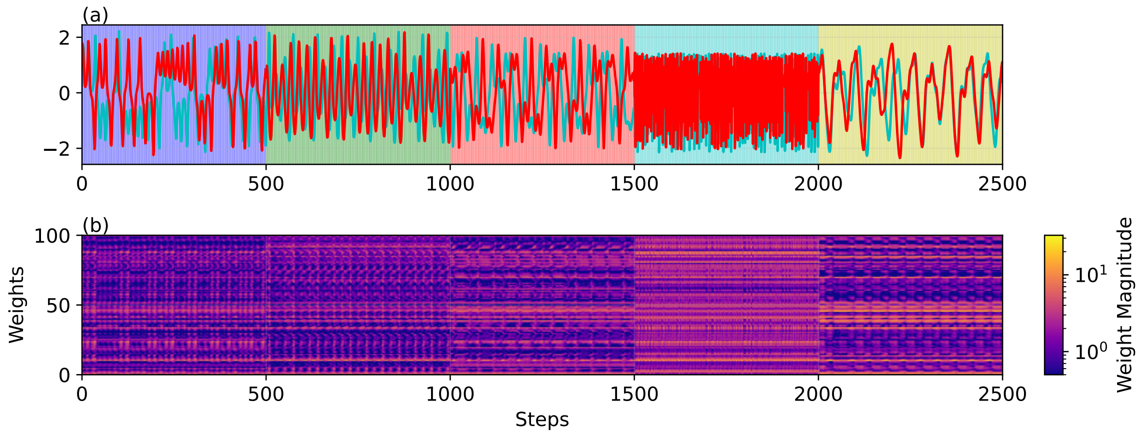

Fig. 2. (a) Time series of predictions for all five dynamical systems in a closed-loop configuration. The plots show the true and predicted values. (b) Output layer weight values for 100 example nodes over the same time period.

1) Valid Prediction Time (VPT): The VPT measures the duration over which the model's predictions remain within an acceptable error threshold in a closed-loop configuration. It is particularly relevant for assessing short-term stability and accuracy in chaotic systems. VPT is defined as the time T for which the normalized error δu_l satisfies:

$$\delta u_l = \frac{|y_l - d_l|^2}{\langle |y_l - \langle y_l \rangle|^2 \rangle} < \epsilon, \tag{16}$$

where, $\langle \cdot \rangle$ denotes the time average and ϵ is the error threshold, typically set to 0.4. The VPT metric is expressed in terms of the system's characteristic timescale, the Lyapunov time.

2) Power Spectra and Histogram Similarity: To compare two power spectra or histograms, we employ the Pearson correlation coefficient, which quantifies the linear relationship between them. The correlation coefficient is defined as:

$$\text{correlation} = \frac{\text{cov}(\mathbf{s}_1, \mathbf{s}_2)}{\sigma_{\mathbf{s}_1} \sigma_{\mathbf{s}_2}},$$

where $\text{cov}(\mathbf{s}_1, \mathbf{s}_2)$ represents the covariance, and $\sigma_{\mathbf{s}_1}$ and $\sigma_{\mathbf{s}_2}$ are the standard deviations of $\mathbf{s}_1$ and $\mathbf{s}_2$, respectively. A value near 1 indicates a strong linear correlation, suggesting a high degree of similarity between the spectra or histograms.

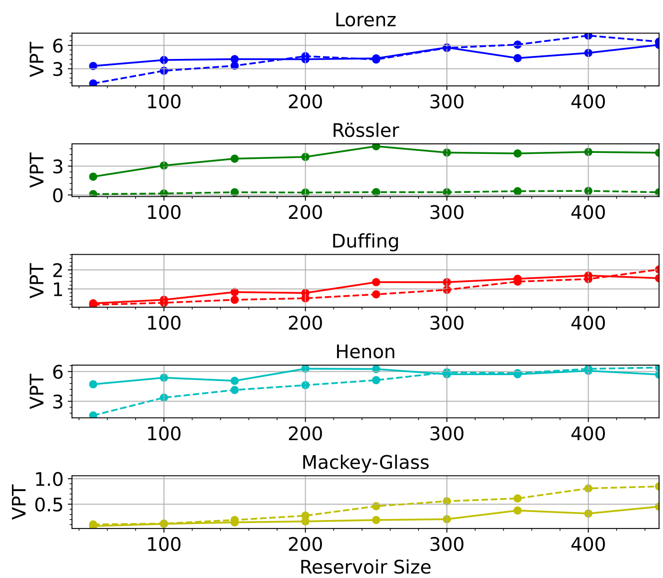

Fig. 3. Valid Prediction Time (VPT) measured in the characteristic timescale of the systems (Lyapunov time) across varying reservoir sizes for the five dynamical systems. The attention-enhanced ESN demonstrates on par VPT for all systems except the Mackey-Glass equation. Subfigures: (a) Lorenz System, (b) Rössler Attractor, (c) Duffing Oscillator, (d) Henon Map, and (e) Mackey-Glass Delay Differential Equation.

IV. Results

A. Prediction Accuracy

Figure 2 illustrates the AERC model's predictions for all five dynamical systems, demonstrating its ability to reconstruct complex trajectories with high accuracy all with one set of weights. The attention mechanism enables dynamic weighting of the reservoir states, allowing the model to focus on the most relevant features of the input data at each time step. This adaptability ensures robust performance even when simulating systems with significantly different dynamics.

For example, in the Lorenz system, the model captures the continuous chaotic patterns over extended time intervals, while in the Henon map, the discrete steps are reproduced. This is all done with one set of weights. These results highlight the model's versatility and its capacity to handle both continuous and discrete systems.

B. Valid Prediction Time

As shown in Figure 3, the AERC model is on par with traditional ESNs in Valid Prediction Time (VPT) for most dynamical systems, and even surpass it in the Rössler system. The attention mechanism's ability to dynamically adjust output weights is a key factor in extending VPT, as it allows the model to adapt to changing temporal dependencies. This is

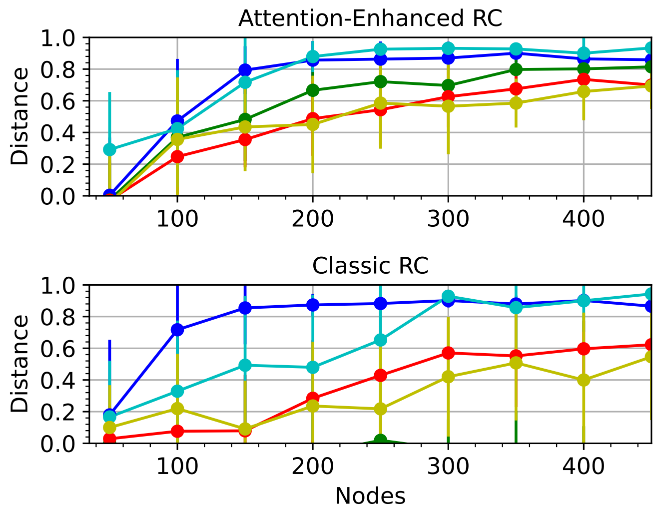

Fig. 4. Histogram similarity between the true and closed-loop time series measured by the correlation coefficient. (a) shows the AERC and (b) the classic RC.

especially interesting, as the classic RC is only trained on one of the tasks at once.

In the Lorenz system and Henon map, the AERC model demonstrates particularly long VPT, maintaining accurate predictions. For the Rössler system, which poses challenges to the classic reservoir approach, the AERC model achieves a VPT that is way higher.

The exception to this trend is the Mackey-Glass system, where the AERC model's VPT does not significantly surpass that of the traditional ESN and even falls behind a little bit. This could be attributed to the general bad performance of the reservoir for the Mackey-Glass system. It seems that the chosen base reservoir is not well suited for this particual prediction task.

C. Histogram and Power Spectra Similarity

Figure 4 presents a detailed comparison of the power spectra and histogram similarity between the true and predicted time series for all systems, using correlation coefficients. The AERC model achieves high correlation coefficients values across the board, indicating a strong alignment between the predicted and actual state distributions.

If the reservoir is big enough for most systems the correlation coefficient approaches 1.0, reflecting near-perfect reconstruction of the chaotic attractor's underlying structure. The classic RC approach on the other hand tends to fail to reconstruct specific attractors, especially the Roessler system.

V. Conclusion

This study demonstrates the significant advantages of incorporating an attention mechanism into the reservoir computing framework. The proposed AERC model improves in predicting chaotic systems, achieving improvements in both short-term accuracy and long-term stability. By dynamically prioritizing the most relevant reservoir states, the attention mechanism enhances the model's adaptability across a diverse range of dynamical systems.

The analysis of prediction accuracy, Valid Prediction Time, and histogram similarity underscores the AERC model's potential as a universal simulator for chaotic and complex systems. While the model performs exceptionally well for most systems, future work could explore further optimizations to reduce the number of weights and the chosen reservoir basis.

Acknowledgment

This work was supported by JSPS KAKENHI (Grant Nos. JP19H00868, JP20K15185, JP22H05195).

References

[1] H. Jaeger, "The echo state approach to analysing and training recurrent neural networks," GMD Tech. Rep. 148, 2001.
[2] M. Lukosevicius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," Comput. Sci. Rev., vol. 3, no. 3, pp. 127–149, 2009.
[3] G. Tanaka et al., "Recent advances in physical reservoir computing: A review," Neural Netw., vol. 115, pp. 100–123, 2019.
[4] A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), pp. 5998–6008, 2017.
[5] S. Siriwardhana, M. Epasinghe, S. Rajapakse, and J. Rajapakse, "Transformers in time series: A survey," arXiv:2111.12753, 2021.
[6] F. Köster, K. Kanno, and A. Uchida, "Attention-enhanced reservoir computing for chaotic systems," Phys. Rev. Appl., vol. 22, 014039, 2024.
[7] E. N. Lorenz, "Deterministic nonperiodic flow," J. Atmos. Sci., vol. 20, pp. 130–141, 1963.
[8] O. E. Rössler, "An equation for continuous chaos," Phys. Lett. A, vol. 57, no. 5, pp. 397–398, 1976.
[9] M. Hénon, "A two-dimensional mapping with a strange attractor," Commun. Math. Phys., vol. 50, no. 1, pp. 69–77, 1976.
[10] G. Duffing, "Erzwungene Schwingungen bei Veränderlicher Eigenfrequenz," Vieweg und Sohn, Braunschweig, 1918.
[11] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," Science, vol. 197, pp. 287–289, 1977.