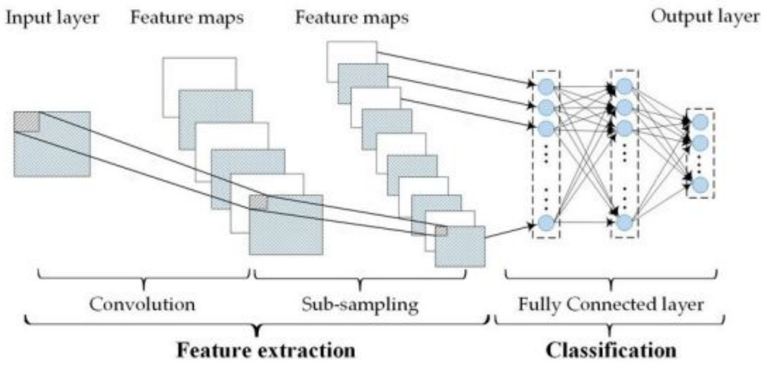
Genetic Algorithm In Echo State Network

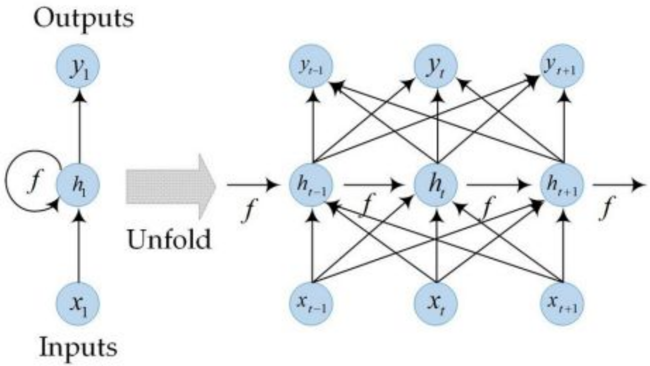
Soumyadeep Jana
Roll - 21D070075
Guide - Prof. Udayan Ganguly
Mentor - Anmol Biswas

Convolutional vs Recurrent Neural Network



Convolutional Neural Net (CNN)

Spatial Structure matters

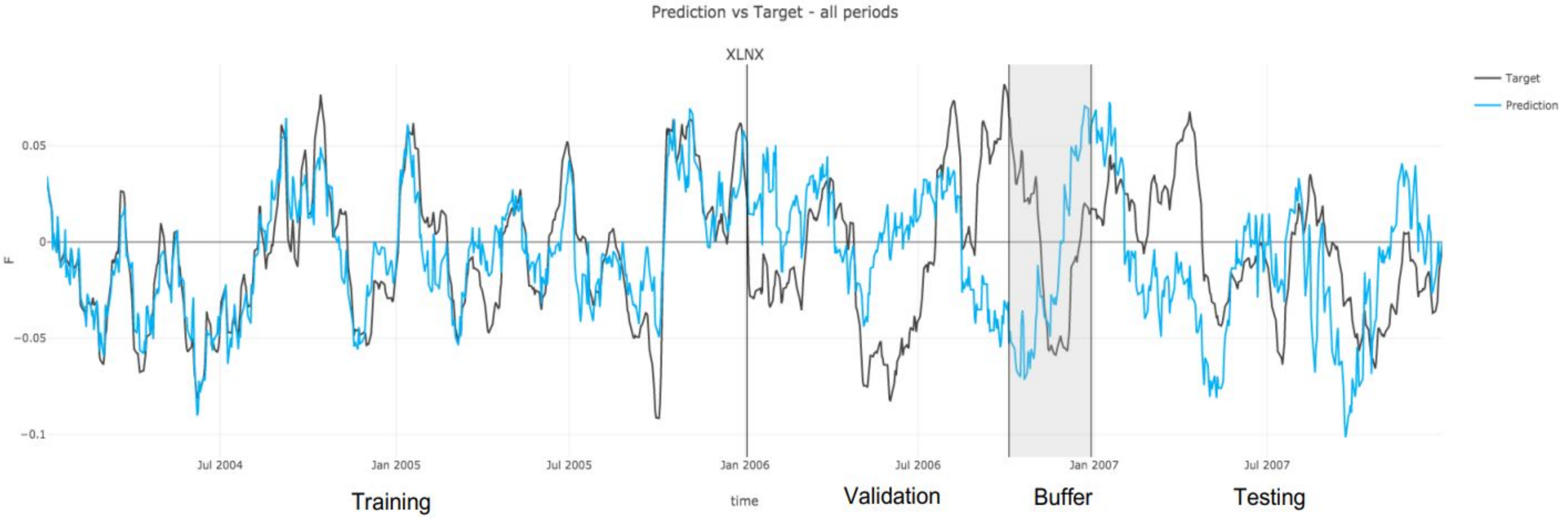


Recurrent Neural Net (RNN)

Sequential Context matters

Source:- https://github.com/tguidici/ODSC_2020_ESN/blob/master/teal_ODSC_2021_EAST_ESN.pdf

Future Trend...



Time Series Prediction

Source:- https://github.com/tguidici/ODSC_2020_ESN/blob/master/teal_ODSC_2021_EAST_ESN.pdf

Limitations of RNN

- Vanishing Gradient = Gradient becomes too small → Hard to learn long-term patterns.
- Exploding Gradient = Gradient becomes too big → Training becomes chaotic. These are like trying to balance between not losing your message and not making it overwhelming!
- Computationally intensive and hard to tune

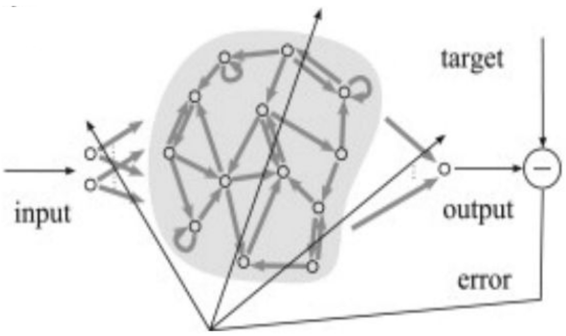


Fig - Traditional RNN. Bold arrows indicate trained weights

Source:- https://github.com/tguidici/ODSC_2020_ESN/blob/master/teal_ODSC_2021_EAST_ESN.pdf

Reservoir Computing

- Fast! Accurate! Still challenging to tune
- Only output weights are adapted, no need of long back propagation

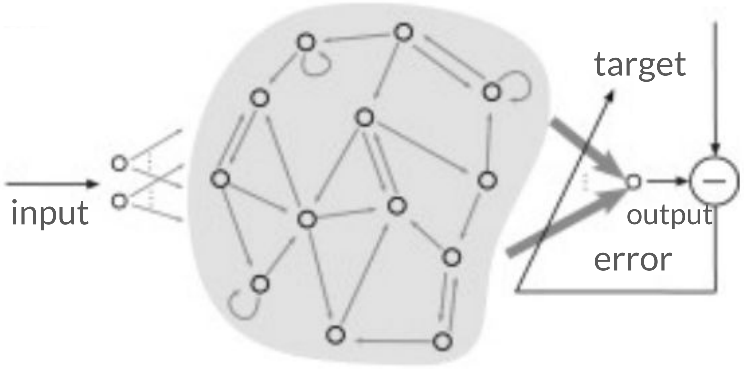


Fig - Traditional Reservoir computing .Bold arrows indicate trained weights

Source:- https://github.com/tguidici/ODSC_2020_ESN/blob/master/teal_ODSC_2021_EAST_ESN.pdf

Types of Reservoir Computing

- Liquid State Machine
- **Echo State Network**
- Context Reverberation
- Nonlinear Transient Computation
- Deep Reservoir Computing

Source:- https://youtu.be/uF4i9_7IQlI?si=IGY6-e5zHBVT1Ycn

Echo State Network

ESNs are a recurrent neural network

- Captures dynamic behavior of a time series or sequential data.
- Internal state (memory) to process sequences of inputs and remember context

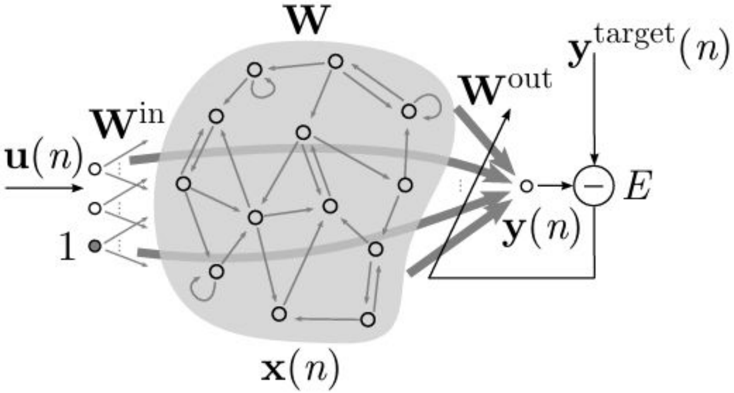
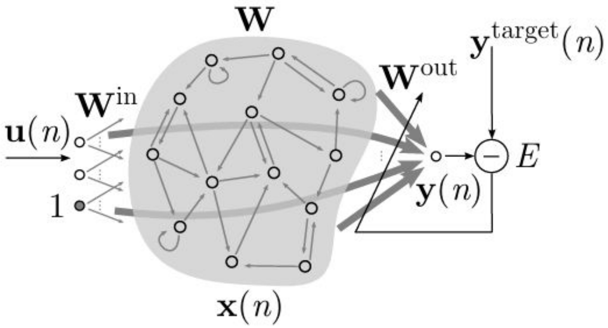


Fig :- An Echo State Network

Source:- Lukoševičius, M. (2012). A Practical Guide to Applying Echo State Networks. In: Montavon, G., Orr, G.B., Müller, KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol 7700. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35289-8_36

Overview



For each time step n
u(n): input
x(n): reservoir activations
y(n): network output
ytarget(n): desired output

Randomly Initialized & fixed
Win: input weights . Linearly map input into reservoir
W: recurrent weights. How are the reservoir neurons connected to each other?

Learned parameters
Wout: output weights

- Linearly map (u(n), x(n)) to y(n)
- Minimize Error(ytarget(n), y(n))

Source:- Lukoševičius, M. (2012). A Practical Guide to Applying Echo State Networks. In: Montavon, G., Orr, G.B., Müller, KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol 7700. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35289-8_36

Dimensions and Update Equation

$$W^{in} \in \mathbb{R}^{N_x \times (1+N_u)}\text{: input weights}$$
$$W \in \mathbb{R}^{N_x \times N_x}\text{ : recurrent weights}$$
$$\alpha \in (0,1]\text{: leak rate}$$

$$x(n) \in \mathbb{R}^{N_x}\text{ : neuron activations}$$
$$\tilde{x}(n) \in \mathbb{R}^{N_x}\text{ : neuron updates}$$

$$x(n) = (1-\alpha)x(n-1) + \alpha\tilde{x}(n)$$
$$\tilde{x}(n) = tanh(W^{in}[1; u(n)] + Wx(n-1))$$

Source:- Lukoševičius, M. (2012). A Practical Guide to Applying Echo State Networks. In: Montavon, G., Orr, G.B., Müller, KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol 7700. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35289-8_36

Reservoir Parameter:- Input Weights

$$W^{in} \in \mathbb{R}^{N_x \times (1+N_u)}$$
$$W^{in}_{ij} \sim Unif(-a, a)$$
Tune/scale input weights by varying a

- Input scaling determines how nonlinear reservoir responses are
  - $a < 1$ (generally)
- Scaling input weights helps remove the influence of outliers
- Taken $a = 0.5$

Source:- Lukoševičius, M. (2012). A Practical Guide to Applying Echo State Networks. In: Montavon, G., Orr, G.B., Müller, KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol 7700. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35289-8_36

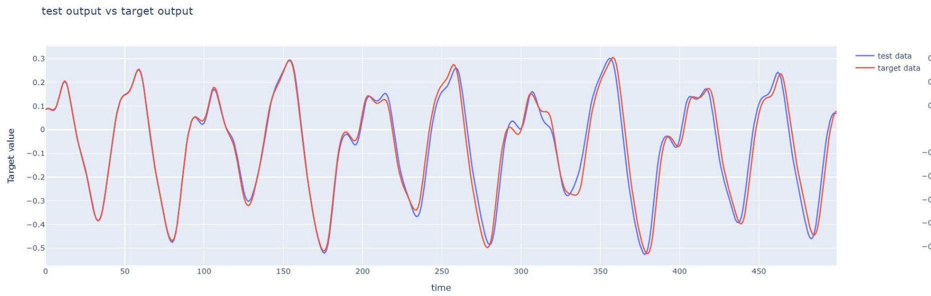
Reservoir Parameter:- Spectral Radius

$$\boldsymbol{W} \in \mathbb{R}^{N_x \times N_x}$$
$$\rho(\boldsymbol{W}) = \text{ spectral radius of W}$$

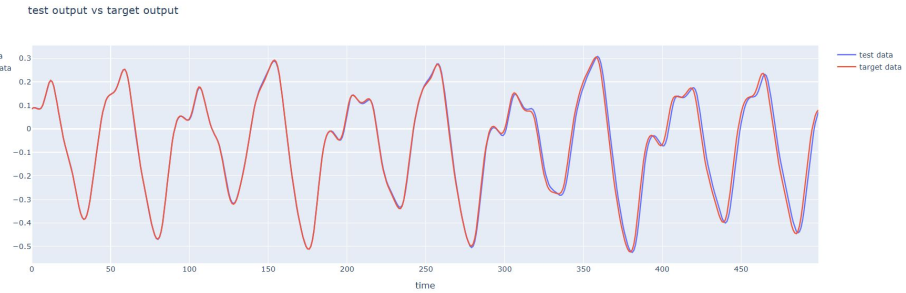
- Spectral radius = eigenvalue of W with greatest magnitude
- The value of ρ is related to the variable memory length and the degree of contractivity of reservoir dynamics, with larger values of $\rho < 1$ resulting in longer memory length
- Taken $\rho = 0.95$

Source:- Lukoševičius, M. (2012). A Practical Guide to Applying Echo State Networks. In: Montavon, G., Orr, G.B., Müller, KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol 7700. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35289-8_36

Effect of Spectral Radius



Spectral radius = 0.75
MSE = 0.001528



Spectral radius = 0.95
MSE = 0.000137

Reservoir Parameter:- Leaky Rate

$$\alpha \in (0, 1]\text{: leak rate}$$
$$x(n) = (1 - \alpha)x(n-1) + \alpha\tilde{x}(n)\text{: neuron activations}$$
$$\tilde{x}(n) = tanh(W^{in}[1; u(n)] + Wx(n-1))\text{: neuron updates}$$

- Captures temporal dynamics of system
- ”speed of reservoir update dynamics discretized in time”
- Taken as 0.5

Source:- Lukoševičius, M. (2012). A Practical Guide to Applying Echo State Networks. In: Montavon, G., Orr, G.B., Müller, KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol 7700. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35289-8_36

Output Weight:-

$$\mathbf{y}(t) = \mathbf{W}_{out}\mathbf{x}(t)$$

$y(t) \in R^{Ny}$ $W_{out} \in R^{Ny \times Nx}$

$$W_{out} = Y^T X^T (XX^T + \lambda I)^{-1}$$

where $y = [y(1), y(2), ..., y(N_y)]^T \in R^{Ny}$ is the target vector
$X = [x(1), x(2), \ldots, x(N_x)] \in R^{Ny \times Nx}$

Source:- Lukoševičius, M. (2012). A Practical Guide to Applying Echo State Networks. In: Montavon, G., Orr, G.B., Müller, KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol 7700. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35289-8_36

Ensemble ESN

- Combines outputs of multiple ESNs for more accurate predictions
- Less sensitive to noisy or corrupted data compared to a single ESN
- Each ESN in the ensemble can have different reservoir parameters or connectivity, capturing a wider range of dynamics

But in what pattern should reservoirs be connected ?

Let's Take some examples...

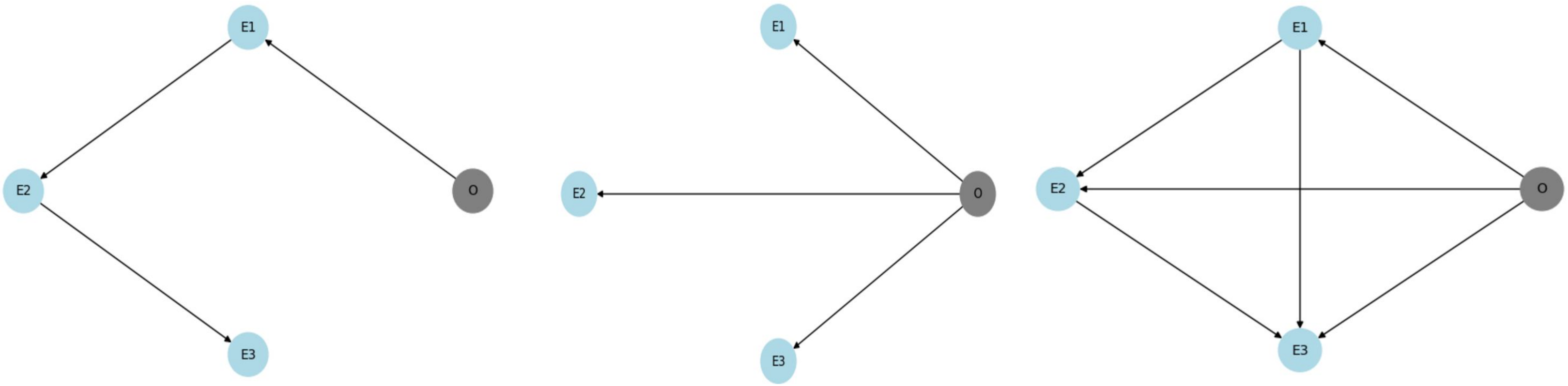


Fig - Some possible ensemble topology. Here 0(input) and Ei(Reservoir)

Ensemble Case

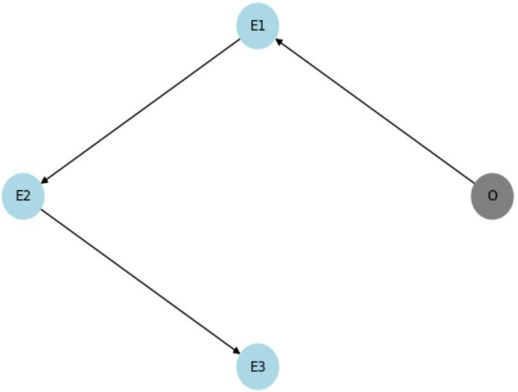
$$\mathbf{x}^{(l)}(t) = (1 - a^{(l)})\mathbf{x}^{(l)}(t-1) + a^{(l)}\tanh(\mathbf{W}_{in}^{(l)}\mathbf{i}^{(l)}(t) + \boldsymbol{\theta}^{(l)} + \hat{\mathbf{W}}^{(l)}\mathbf{x}^{(l)}(t-1)),$$

$$\mathbf{y}(t) = \mathbf{W}_{out}\left[\mathbf{x}^{(1)}(t)\ \mathbf{x}^{(2)}(t)\ \dots\ \mathbf{x}^{(N_L)}(t)\right]^T$$

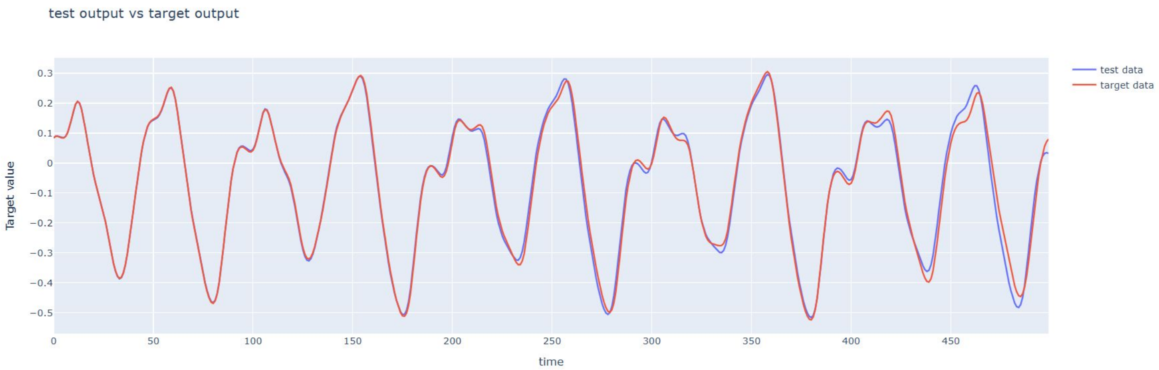
$$\mathbf{i}^{(l)}(t) = \begin{cases} \mathbf{u}(t) & if\ \ l = 1 \\ \left[\mathbf{u}(t)\ \mathbf{x}^{(l-1)}(t)\right]^T & if\ \ l > 1 \end{cases}$$

Source:- Claudio Gallicchio, Alessio Micheli, Luca Pedrelli, Deep reservoir computing: A critical experimental analysis, Neurocomputing, Volume 268, 2017, Pages 87-99, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2016.12.089

Series Connection

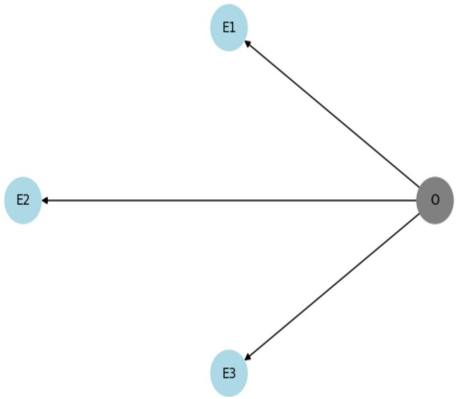


Series Connection topology: Deep ESN

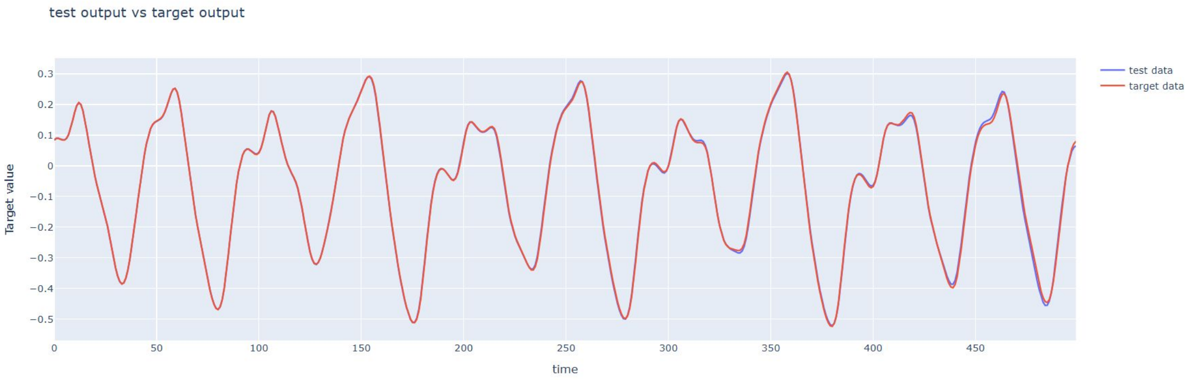


MSE: 0.000462

Parallel Connection

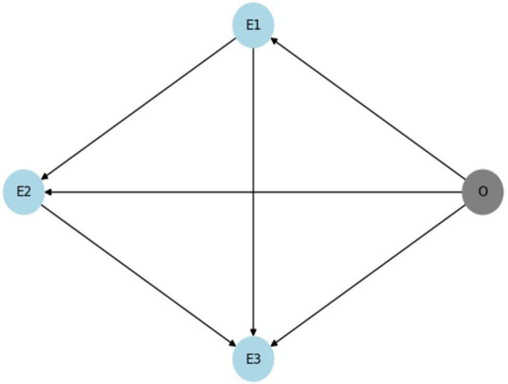


Parallel Connection Topology: GroupedESN

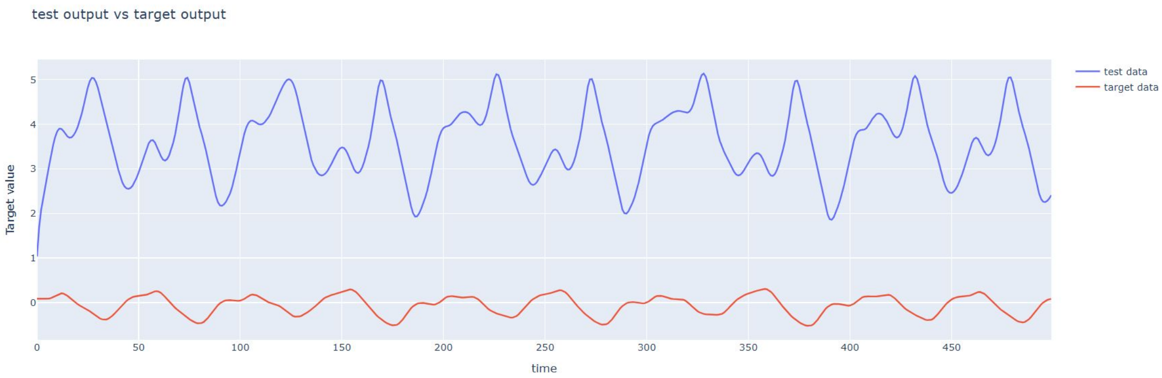


MSE: 3.8e-05

Cross Connection



Cross Connection Topology



MSE: 14.205438

Genetic Algorithm

- Represent reservoir combination as binary sequence

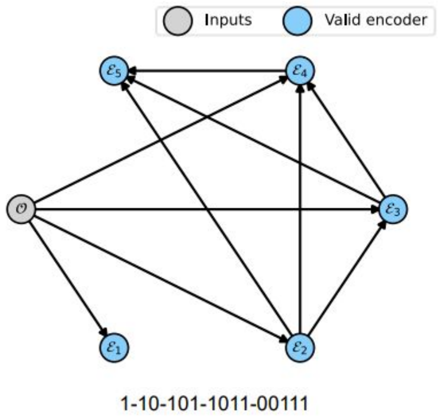


Fig: An example network topology of an MRESN (upper) and the corresponding binary sequence (bottom)

Z. Li, K. Fujiwara and G. Tanaka, "Designing Network Topologies of Multiple Reservoir Echo State Networks: A Genetic Algorithm Based Approach," *2024 International Joint Conference on Neural Networks (IJCNN)*, Yokohama, Japan, 2024, pp. 1-9, doi: 10.1109/IJCNN60899.2024.10650945

Valid Encoder

Not all possible combination will be valid.
The encoder node Ei is invalid

- its in-degree is zero or
- all ancestor nodes of Ei are invalid encoder nodes

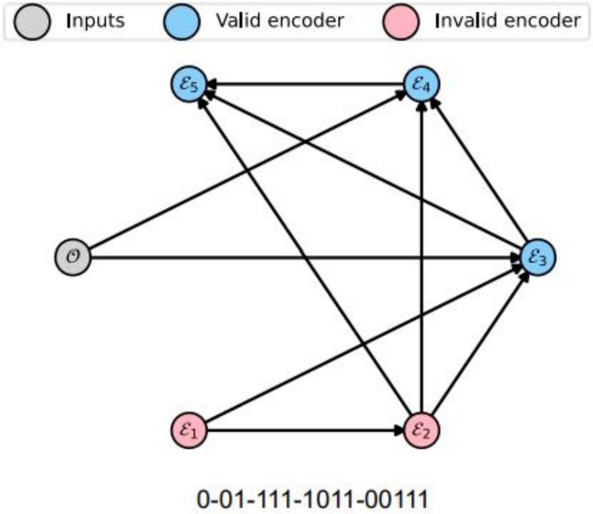
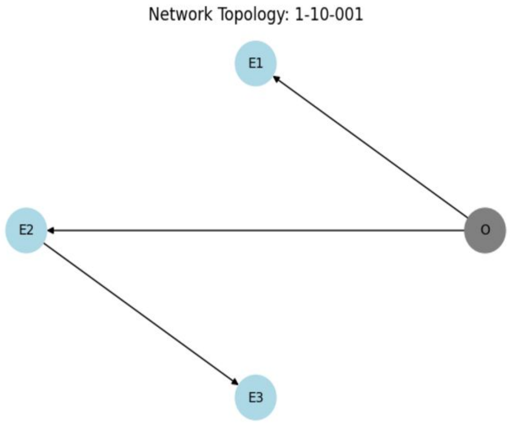


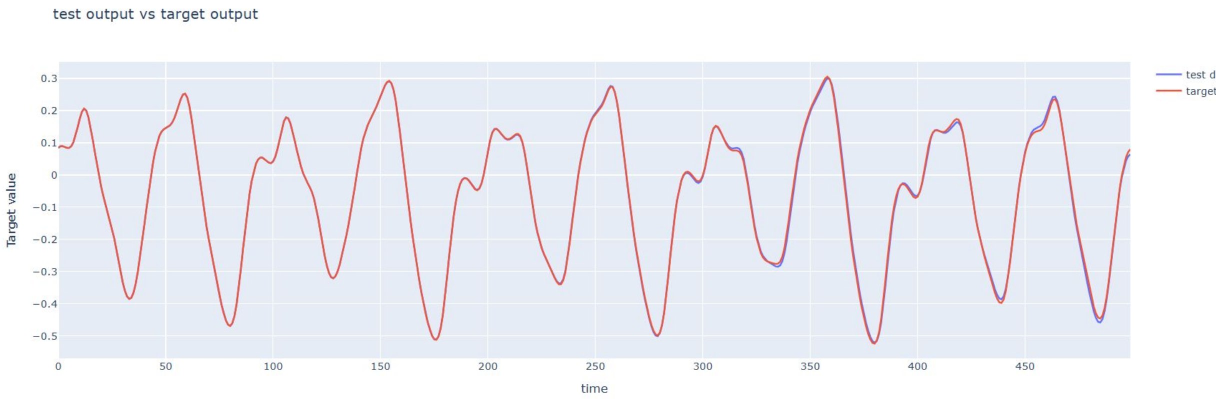
Fig - An example network topology of an MRESN with three valid encoders and two invalid encoders

Z. Li, K. Fujiwara and G. Tanaka, "Designing Network Topologies of Multiple Reservoir Echo State Networks: A Genetic Algorithm Based Approach," *2024 International Joint Conference on Neural Networks (IJCNN)*, Yokohama, Japan, 2024, pp. 1-9, doi: 10.1109/IJCNN60899.2024.10650945

Results:



Most optimal connection from Genetic Algorithm



MSE: 9.8e-06

My Work

https://github.com/Soumyadeepj/Genetic_ESN.git

Future Work:

- Effect of increment of the number of neurons and reservoir
- Hyperparameter tuning for better MSE
- Application of Genetic Algorithm in Liquid State Machine

References

1. https://github.com/tguidici/ODSC_2020_ESN/blob/master/teal_ODSC_2021_EAST_ESN.pdf

2. Lukoševičius, M. (2012). A Practical Guide to Applying Echo State Networks. In: Montavon, G., Orr, G.B., Müller, KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol 7700. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35289-8_36

3. Claudio Gallicchio, Alessio Micheli, Luca Pedrelli, Deep reservoir computing: A critical experimental analysis, Neurocomputing Volume 268, 2017, Pages 87-99, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2016.12.089

4. Z. Li, K. Fujiwara and G. Tanaka, "Designing Network Topologies of Multiple Reservoir Echo State Networks: A Genetic Algorithm Based Approach," *2024 International Joint Conference on Neural Networks (IJCNN)*, Yokohama, Japan, 2024, pp. 1-9, doi: 10.1109/IJCNN60899.2024.10650945

Thank You