Improved echo state network with Graph Centrality Pruning Algorithm
Sudip Laudari¹, Dohee Jung² and Sang-Gu Lee²

Department of Mathematics, Sungkyunkwan University, Suwon 16419, Republic of Korea

Corresponding Author: sudip laudari, sudip.laudari2008@yahoo.com

Abstract
Reservoir Computing (RC) is an effective approach to design and train recurrent neural networks, which is successfully and widely applied in time series prediction tasks. Echo state network (ESN) is one of the most well-known types of reservoir computing because of its outstanding performance to a wide range of real-world applications. However, the random initialization of the reservoir structure may lead to poor generalization performance and high computational complexity. In this paper, we provide an effective algorithm for removing redundant connections inside the reservoir based on graph centrality. In order to improve generalization ability and reduce the computational complexity of the original ESN, we propose a Graph Centrality Pruning Algorithm (GCPA) to optimize the reservoir size and weights. Results obtained on several benchmarks and a real-world dataset of telephone call data records show the effectiveness of the proposed methods.

1. Introduction
Reservoir Computing (RC) is a well-established technique to design and train recurrent neural network (RNN). The reservoir is supposed to be sufficiently complex so as to capture a large number of features of the input stream that can be exploited by the reservoir-to-output readout mapping. Echo State Network (ESN) is one of the most popular reservoir computing approaches and is characterized by a fixed untrained reservoir and a simple linear readout. In this paper, we provide an effective algorithm GCPA for removing redundant connections inside the reservoir during training. The algorithm is based on the correlation of the states of the nodes, hence it depends only on the input signal, it is efficient to implement, and it is also local. By applying it, we can obtain an optimally sparse reservoir in a robust way. We present the performance of our algorithm on two synthetic datasets, which show its effectiveness in terms of better generalization and lower computational complexity of the resulting ESN. This behavior is also investigated for increasing levels of memory and non-linearity required by the task.

2. Echo state network
Echo state network was first introduced by Jaeger 2001 [28]. The main equation is presented below [26]:

\[ X(n+1) = f(W_x(n) + W^{in}u(n+1) + W^{back}y(n)) \]

where, \( f = (f_1, \ldots, f_N) \) is an activation function which is usually a sigmoid function. Finally, the outputs of the network, which can be determined by the following expression:
Figure 1: **Basic structure of ESN**

\[ y(n + 1) = f^\text{out}(W^i(outx(n + 1), x(n + 1), y(n))) \cdots (i) \]

where \( f^\text{out} = (f^\text{out}_1, \ldots, f^\text{out}_L) \) are output functions of the output units and \((u(n + 1), x(n + 1), y(n))\) is the concatenation of input, internal and previous output activation vectors. The spectrum of the \( W \) is less than 1 [28]. Similarly it is required that training input vectors \( u(n) \) belong to a compact interval \( U \) and training output vectors \( y(n) \) belong to a compact interval \( Y \) [26].

3. **Centrality in a weighted network**

The concept of node centrality is successfully applied in machine learning field. Some of the concepts of weighted centrality and our models which are applied in ESN are presented below.

1) \( \text{Closeness}(C_C(i)) = \left[ \sum_{j} d^w(i, j) \right]^{-1} \)

2) \( \text{Betweenness}(C_B(i)) = \frac{g^w_{jk}(i)}{g^w_{jk}} \)

3) \( C_1(i) = \frac{|I_i^+| - |I_i^-|}{|I_i^+| + |I_i^-|} \)

where \( |I_i^+| \) is the sum of positive weights toward \( i \) and \( |I_i^-| \) is the absolute value of the sum of negative weights toward \( i \). Based on centrality, other models will be presented in
full paper.

4. **Experiment design and network preparation**

4.1 **Mackey-Glass time series**

The Mackey-Glass time series is a standard benchmark for chaotic time series prediction model, on which ESNs have been successfully applied showing a good performance. The time series is defined by the following differential equation:

\[
\frac{\partial o(t)}{\partial t} = \frac{0.2 o(t - \alpha)}{1 + o(t - \alpha)^10} - 0.1 o(t) \ldots...
\]

Similarly, the normalized root mean square error at the 50th time step RMSE used here is

\[
RMSE = \sqrt{\frac{\sum_{j=1}^{N_r} (\hat{o}_i(30) - o_i(30))^2}{N_r}}
\]

where \(\hat{o}_i(30)\) and \(o_i(30)\) are the predicted output and the desire output at time step 30, and \(N_r\) is the number of independent simulations.

Figure 2: Mackey-Glass, n=200
Table 1: The reservoir size with testing error after pruning based on $C_{in}, C_{out}, C_B, C_C, C_1, C_2, C_3$ for Mackey-Glass Time Series prediction

<table>
<thead>
<tr>
<th>Methods</th>
<th>Initial N(RMSE)</th>
<th>Optimumal N(RMSE)</th>
<th>Reduced error(%)</th>
<th>Smallest N</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_B$</td>
<td>N=200(0.010477)</td>
<td>N=182(0.009978)</td>
<td>0.000499(4.8%)</td>
<td>N=165</td>
</tr>
<tr>
<td></td>
<td>N=300(0.010619)</td>
<td>N=293(0.010189)</td>
<td>0.00043(4.0%)</td>
<td>N=174</td>
</tr>
<tr>
<td></td>
<td>N=500(0.009047)</td>
<td>N=481(0.008799)</td>
<td>0.000248(2.7%)</td>
<td>N=414</td>
</tr>
<tr>
<td></td>
<td>N=700(0.009475)</td>
<td>N=619(0.009014)</td>
<td>0.000461(4.9%)</td>
<td>N=438</td>
</tr>
<tr>
<td>$C_C$</td>
<td>N=200(0.010477)</td>
<td>N=188(0.010193)</td>
<td>0.000284(2.7%)</td>
<td>N=153</td>
</tr>
<tr>
<td></td>
<td>N=300(0.010619)</td>
<td>N=284(0.009139)</td>
<td>0.00148(13.9%)</td>
<td>N=204</td>
</tr>
<tr>
<td></td>
<td>N=500(0.009047)</td>
<td>N=494(0.008798)</td>
<td>0.000249(2.8%)</td>
<td>N=393</td>
</tr>
<tr>
<td></td>
<td>N=700(0.009475)</td>
<td>N=671(0.008967)</td>
<td>0.000508(5.4%)</td>
<td>N=439</td>
</tr>
<tr>
<td>$C_1$</td>
<td>N=200(0.010477)</td>
<td>N=184(0.010032)</td>
<td>0.000445(4.2%)</td>
<td>N=142</td>
</tr>
<tr>
<td></td>
<td>N=300(0.010619)</td>
<td>N=279(0.009972)</td>
<td>0.000647(6.1%)</td>
<td>N=196</td>
</tr>
<tr>
<td></td>
<td>N=500(0.009047)</td>
<td>N=488(0.008417)</td>
<td>0.00063(7.0%)</td>
<td>N=364</td>
</tr>
<tr>
<td></td>
<td>N=700(0.009475)</td>
<td>N=666(0.009085)</td>
<td>0.00039(4.1%)</td>
<td>N=444</td>
</tr>
</tbody>
</table>

5. Conclusion

A GCPA method in which a large reservoir is assigned firstly and centrality method is applied to prune its nodes. Centrality models depended on the random weights between nodes of the reservoir. Mackey-Glass bench mark task was applied to illustrate the performance of GCPA. Among this, GCPA was applied to predict the electric load data and traffic load data. The pruning method was applied under the properties of ESN. pruning 0 nodes was the old ESN. All the above application showed that GCPA can significantly improve performance of old ESN. Every model shows significant effect on old ESN to generalized its performance. Besides this, it also shows that without managing dynamic reservoir can give worst performance. From our experiments and results, it shows that in the future we can use our models to improve the performance of other reservoir computing methods. Also we can apply more mathematical concept to improve those models.

References

[1] Michal Cernansky and Peter Tino, Predictive Modeling with Echo State Networks


[14] Lorenzo Livi, Member, IEEE, Filippo Maria Bianchi, and Cesare Alippi, Fellow, IEEE, Determination of the edge of criticality in echo state networks through Fisher information maximization, 2016


[16] Min Han, Senior Member, IEEE, and Meiling Xu, Laplacian Echo State Network for Multivariate Time Series Prediction, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, NO. 1, JANUARY 2018
[17] Xiao-chuan SUN1, Hong-yan CUI, Ren-ping LIU, Jian-ya CHEN, Yun-jie LIU, Modeling deterministic echo state network with loop reservoir, Sun et al. J Zhejiang Univ-ScC (Computand Electron) 2012 13(9):689-701


