

# 延迟回声状态神经网络用于 复杂系统分析和应用

徐一宸<sup>1</sup>, Eric Li<sup>2,3</sup>

(1. 中国人民大学 信息学院, 北京 100872;

2. Teesside 大学 计算、机械和数字技术学院, 英国北约克郡 米德尔斯堡 TS1 4;

3. 吉林大学 数学学院, 长春 130012)

**摘要:** 提出一种改进的回声状态神经网络模型, 用于复杂系统的长期行为分析和预测。模型通过引入隐层状态的延迟反馈体现系统过去时刻的信息对当前状态的影响, 避免了传统回声状态网络方法记忆能力弱的缺点以及获得最优参数的困难。

**关键词:** 回声状态网络; 混沌时间序列; 储备池计算; 稳定性; 长期预测

**中图分类号:** O193 **文献标志码:** A **文章编号:** 1671-5489(2024)05-1017-05

## Delayed Echo State Neural Network for Analysis and Application of Complex Systems

XU Yichen<sup>1</sup>, Eric Li<sup>2,3</sup>

(1. School of Information, Renmin University of China, Beijing 100872, China;

2. School of Computing, Engineering & Digital Technologies, Teesside University,  
Middlesbrough TS1 4, North Yorkshire, United Kingdom;

3. College of Mathematics, Jilin University, Changchun 130012, China)

**Abstract:** We proposed an improved echo state neural network model for the analysis and prediction of long-term behavior of complex systems. The model introduced the delayed feedback of hidden layer state to reflect the influence of the past time information on the current state of the system, avoiding the shortcomings of weak memory ability and difficulty of obtaining optimal parameters in traditional echo state network methods.

**Keywords:** echo state network; chaotic time series; reservoir computing; stability; long-term prediction

回声状态网络(ESN)是一种新型的回归人工神经网络(RNN)模型, 由于其在复杂动态系统预测问题上性能优异, 因此在科学和工程领域已引起广泛关注<sup>[1-4]</sup>。ESN 的核心思想是先通过使用储备池(reservoir)将训练数据投影到高维特征空间中, 然后通过读出权重将高维状态映射到输出数据。读出权重可通过线性岭回归算法计算, 克服了使用梯度下降算法带来的梯度爆炸和梯度消失的困难<sup>[5-12]</sup>。但现有的 ESN 模型所体现的系统特征随着时间的推移逐渐消失, 在当前时间附近的输入特征显著, 而对于远离当前时间的输入特征太弱而无法回忆。ESN 模型这种重视短期行为导致记忆能力不足的特

收稿日期: 2023-07-12.

第一作者简介: 徐一宸(2002—), 男, 汉族, 硕士研究生, 从事机器学习和数据挖掘的研究, E-mail: yichenruc@126.com. 通信

作者简介: Eric Li(1980—), 男, 英国, 博士, 教授, 从事动力系统的研究, E-mail: lieric2023@126.com.

基金项目: 国家自然科学基金(批准号: 12072128).

点限制了其解决复杂系统长期预测问题的能力. 基于此, 本文提出一种具有延迟反馈的回声状态网络模型(delay-ESN), 通过考虑储备层中神经元的延迟效应, 引入隐层状态的延迟反馈体现系统过去时刻信息对当前状态的影响, 从而提高系统的记忆性能.

### 1 回声状态网络

回声状态网络是一种特殊的回归神经网络模型, 由输入层、输出层和储备池层组成. 储备池层由大量随机相连的神经元构成, 将输入信号映射到更高维度的状态空间. 假设一个 ESN 的输入层包含  $D$  个单元, 储备池层有  $N_{res}$  个单元, 输出层有  $d$  个单元. 若  $u(t) \in \mathbb{R}^D$ ,  $x(t) \in \mathbb{R}^{N_{res}}$ ,  $y(t) \in \mathbb{R}^d$  分别表示外界的输入、储备池的输出及整个网络的输出, 则储备池层节点的动力学可以表示<sup>[1]</sup>为

$$x(t+1) = \tanh(W_{in}u(t) + W_{res}x(t)), \tag{1}$$

其中  $W_{in} \in \mathbb{R}^{N_{res} \times D}$ ,  $W_{res} \in \mathbb{R}^{N_{res} \times N_{res}}$  分别是输入-储备池层的权值和储备池层权值, 这些权值随机产生并且在运行过程中不需要更新. 整个 ESN 网络的输出为

$$y(t) = f_{out}(W_{out}[x(t); u(t)]),$$

这里  $[\cdot; \cdot]$  表示纵向串联. 在回归任务中,  $f_{out}(\cdot)$  通常为单位函数,  $W_{out} \in \mathbb{R}^{d \times (N_{res} + D)}$  可以通过岭回归算法计算<sup>[2]</sup>. 矩阵  $W_{res}$  的设计是 ESN 是否成功的关键, 其主要受稀疏度  $\eta$  和谱半径  $r_{max}$  等参数的影响, 其中稀疏度通常为 10%~20%, 谱半径通常被归一化为  $r_{max} < 1$  以获得回声状态特性<sup>[2]</sup>.

Leaky-ESN(L-ESN)是一种改进的 ESN 模型<sup>[5]</sup>, 储备池层动力学描述为

$$x(t+1) = \alpha x(t) + (1 - \alpha) \tanh(W_{in}u_{t+1} + W_{res}x(t)), \tag{2}$$

L-ESN 通过引入参数  $\alpha \in [0, 1]$  建立系统当前与其上一步状态之间的关系.

### 2 延迟反馈回声状态网络

本文提出一种具有延迟自反馈效应的 ESN 模型(delay-ESN), 若引入隐层状态的延迟反馈  $\gamma x_{t-m}$ , 则其隐层输出可表示为

$$x_{t+1} = f(W_{in}u_{t+1} + W_{res}x_t + \gamma x_{t-m}), \tag{3}$$

这里  $f(\cdot)$  表示隐层节点的激活函数, 满足

$$0 \leq |f(\cdot)| \leq 1,$$

且

$$f(\hat{x}) = 0 \Leftrightarrow \hat{x} = 0,$$

$\gamma x_{t-m}$  表示延迟自反馈,  $\gamma$  为反馈强度,  $m \geq 0$  为反馈延迟. 延迟自反馈建立了储备层状态的当前和先前  $m$  时间步之间的关系, 提供了反映系统动态特性的有效途径, 其模型结构如图 1 所示, 其中虚线箭头表示延迟反馈,  $m$  表示  $m$  步延迟,  $\gamma$  为反馈增益.

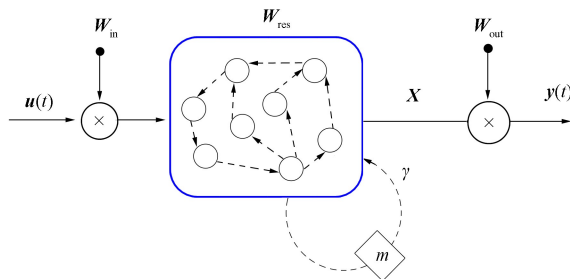


图 1 延迟反馈 ESN 模型结构

Fig. 1 Structure of delayed feedback ESN model

在训练阶段, 回声状态  $x(t)$  按列堆叠到状态  $X$  中, 对应的教师输出值  $y_s(t)$  按列堆叠成向量  $Y$ . 当输出单元的激活函数是单位函数时, 输出层变成线性组合器. 因此, 网络训练的目的在于通过求解如下优化问题:

$$\min_{\mathbf{W}_{out}} (\| \mathbf{W}_{out} \mathbf{X} - \mathbf{Y} \|^2 + \xi \| \mathbf{W}_{out} \|^2), \tag{4}$$

找到最优的输出权值:

$$\mathbf{W}_{out} = \mathbf{YX}^T (\mathbf{XX}^T + \xi \mathbf{I})^{-1}, \tag{5}$$

这里  $\| \cdot \|$  表示 Euclid 范数,  $\mathbf{I}$  表示单位矩阵,  $\xi$  是正则化参数.

当获得了最优的权值  $\mathbf{W}_{out}$  后, 将网络的输出反馈到输入层作为输入信号, 即令  $\mathbf{u}(t) = \mathbf{y}(t)$ , 则可以对系统的动态行为进行预测, 表达式为

$$\begin{cases} \mathbf{x}(t+1) = \tanh(\mathbf{W}_{in} \mathbf{y}(t) + \mathbf{W}_{res} \mathbf{x}(t) + \gamma \mathbf{x}_{t-m}), \\ \mathbf{y}_{pre}(t) = \mathbf{f}_{out}(\mathbf{W}_{out} [\mathbf{x}(t); \mathbf{y}(t)]). \end{cases} \tag{6}$$

delay-ESN 具有以下特点:

- 1) 引入延迟反馈可以更好地模拟神经元之间信号的有限传输速度和延迟交互作用;
- 2) 通过引入延迟状态反馈, 考虑了系统过去状态对系统的影响, 提高了记忆能力;
- 3) 如果没有自反馈, 则 delay-ESN 退化为标准 ESN;
- 4) L-ESN 只考虑了隐含层与其上一个时刻的状态的线性组合, 而 delay-ESN 考虑了隐含层与其之前  $m$  个时间步的状态之间的非线性关系, 具有更好的动态性能.

### 3 回声状态特性分析

ESN 模型中最重要的特征是回声状态特性(ESP), 即驱动的储备池需要对小扰动具有鲁棒性<sup>[1,3]</sup>. 当不满足该条件时, 两个非常接近的输入信号可能导致两个非常不同的储层表示, 进而导致 ESN 对监督学习任务毫无用处. 事实上, 在扰动或噪声存在的情况下, 没有 ESP 的网络准确性会很差: 一个小扰动即可能使网络达到前所未有的状态, 破坏网络的预测能力.

**假设 1** 非线性激活函数  $f(u, x, z)$  在集合  $D \in \mathbb{R}^N$  上满足局部 Lipchitz 条件:

$$\| f(u, x, z) - f(u, x', z') \| \leq \alpha \| x - x' \| + \beta \| z - z' \|,$$

这里  $\alpha$  和  $\beta$  是已知常数.

**定理 1** 如果 delay-ESN 模型满足下列条件:

- 1)  $f(\cdot)$  满足假设 1;
- 2)  $\alpha \rho_{max} + \beta |\gamma| < 1$ , 这里  $\rho_{max}$  是矩阵  $\mathbf{W}_{res}$  的范数.

则 delay-ESN 模型具有回声状态特性.

证明: 设  $\mathbf{x}_t$  和  $\mathbf{z}_t$  分别是不同初始条件下储备池的输出状态, 这里  $\mathbf{x}_t(0) \neq \mathbf{z}_t(0)$ , 则有

$$\begin{aligned} \| \mathbf{x}_{t+1} - \mathbf{z}_{t+1} \| &= \| f(\mathbf{W}_{in} \mathbf{u}_t + \mathbf{W}_{res} \mathbf{x}_t + \gamma \mathbf{x}_{t-m}) - f(\mathbf{W}_{in} \mathbf{u}_t + \mathbf{W}_{res} \mathbf{z}_t + \gamma \mathbf{z}_{t-m}) \| \leq \\ &\alpha \| \mathbf{W}_{res} (\mathbf{x}_t - \mathbf{z}_t) \| + \beta \| \gamma (\mathbf{x}_{t-m} - \mathbf{z}_{t-m}) \| \leq \\ &\alpha \| \mathbf{W}_{res} \| \cdot \| \mathbf{x}_t - \mathbf{z}_t \| + \beta |\gamma| \cdot \| (\mathbf{x}_{t-m} - \mathbf{z}_{t-m}) \| \leq \\ &\alpha \rho_{max} (\mathbf{W}_{res}) \cdot \| \mathbf{x}_t - \mathbf{z}_t \| + \beta |\gamma| \cdot \| (\mathbf{x}_{t-m} - \mathbf{z}_{t-m}) \| \leq \\ &(\alpha \rho_{max} (\mathbf{W}_{res}) + \beta |\gamma|) \max(\| \mathbf{x}_t - \mathbf{z}_t \|, \| (\mathbf{x}_{t-m} - \mathbf{z}_{t-m}) \|), \end{aligned}$$

因此根据文献[1,3]的结论, 如果满足条件 1) 和 2), 则 delay-ESN 模型有回声状态特性.

### 4 数值实验

下面通过混沌时间序列预测实验验证 delay-ESN 模型的有效性. 实验中, 输入数据和教师数据都被归一化为具有均值 0 和标准差 1. 储备池层的节点数设为 500,  $\mathbf{W}_{in}$  和  $\mathbf{W}_{res}$  中的元素由  $(-0.5, 0.5)$  内随机均匀分布的随机数组成. 用均方根误差(RMSE)衡量模型的性能:

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (\mathbf{y}_*(t) - \mathbf{y}(t))^2}, \tag{7}$$

其中  $\mathbf{y}_*(t)$  表示输出的教师信号,  $\mathbf{y}(t)$  表示预测值,  $T$  为预测的时间步长.

考虑 Rössler 混沌系统, 定义为

$$\begin{cases} \dot{x} = -(y + z), \\ \dot{y} = x + \alpha y, \\ \dot{z} = \beta + z(x - \theta), \end{cases} \quad (8)$$

其中  $\alpha=0.15, \beta=0.2, \theta=10$ . 则系统展现出混沌特性.

Rössler 系统  $z$  分量的轨迹表现出不规则的跳跃行为, 许多 ESN 方法很难准确预测该系统. 选择初始条件  $(x_0, y_0, z_0) = (0, 1, 0)$ , 使用 Runge-Kutta 方法生成 6 000 个数据点的三维时间序列, 时间步长为 0.01. 其中前 1 000 个样本用于预热, 2 500 个数据作为训练储层的输入, 其余的 2 500 个数据用于测试 ESN 的预测性能, 实验中参数的设定满足定理 1 条件, 并且  $r_{\max} = 0.9, \eta = 0.1, \xi = 10^{-8}$ .

选择参数  $\gamma = -0.22, m = 6$ , 图 2 为 delay-ESN 和 ESN 的预测曲线和 RMSE, 其中阴影部分表示 delay-ESN 的 RMSE. 由于 Rössler 混沌系统对初始值的敏感性, 因此由图 2 可见, 当离散时间步长约小于 153 时, ESN 有较好的预测结构, 其中  $RMSE < 10^{-2}$ . 但随着时间的延长, ESN 的预测逐渐偏离了实际轨迹, 当离散时间大于 900 时,  $RMSE > 10^{-1}$ , 预测性能不佳. 但当离散时间步长达 2 000 时, delay-ESN 的预测轨迹仍可以准确地跟踪实际的混沌曲线, 预测  $RMSE < 10^{-2}$ . 实验结果表明, 本文提出的 delay-ESN 可以实现混沌系统的长期预测.

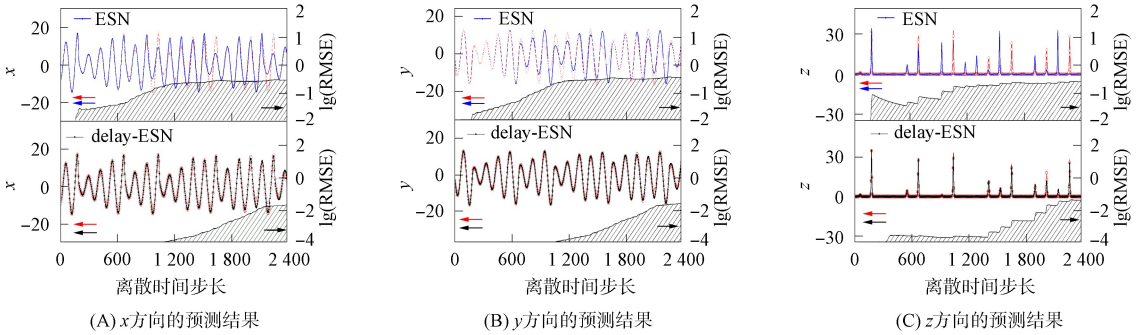


图 2 ESN 和 delay-ESN 预测结果比较

Fig. 2 Comparison of predication results of ESN and delay-ESN

对参数  $(\gamma, m)$  进行网络划分, 在参数的交叉点上计算 RMSE, 则可以得到 RMSE 的区域颜色图, 其中颜色深浅表示了 RMSE 的大小. 当参数位于绿色和蓝色区域时, 预测误差非常小 ( $RMSE < 0.1$ ).

图 3 为 delay-ESN 在  $\gamma$ - $m$  平面内的预测误差颜色图. 由图 3 可见:

1) delay-ESN 可以在很大的  $(\gamma, m)$  参数范围内给出更好的预测结果, 因此更容易选择参数  $\gamma$  和  $m$ , 以获得更好的预测结果, 降低了经典 ESN 方法选择最优参数的困难;

2) 与经典 ESN 相比, 负反馈增益可实现良好的收敛性, 这与许多现有的带有参数的 ESN 方法 (如 Leaky ESN) 非常不同, 这些模型中参数通常是正的;

3) 随着延迟的增加, 具有更好预测精度的参数逐渐减小, 呈现出近似周期性变化模式;

4) 预测精度较高的参数通常位于具有小延迟和小反馈增益的区域, 在实际问题中, 可以从这些区域搜索参数以获得更好的精度.

综上所述, 作为一种新型的回归神经网络模型, 回声状态网络有效克服了传统 RNN 建模中遇到的梯度爆炸和梯度消失问题, 在混沌时间序列预测中取得了优异的效果. 但由于 ESN 的短期记忆能

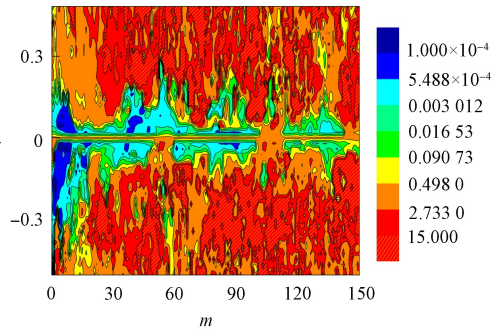


图 3 delay-ESN 在  $\gamma$ - $m$  平面上的预测误差颜色图  
Fig. 3 Colormaps for predicted errors of delay-ESN on  $\gamma$ - $m$  plane

力, 其动态性能略有不足, 难以实现复杂系统的长期行为分析和预测. 本文考虑了神经元信号有限的传播速度, 将储备池层神经元的延迟反馈引入到储备池状态的更新过程中, 提高了传统 ESN 方法记忆能力不足的问题. 并且对模型的回声状态特性进行了分析, 给出了确保 delay-ESN 稳定收敛的充分条件. 虽然提出的模型引入了两个额外参数, 但易通过对这些参数的搜索获得更优的计算性能. 数值实验结果表明, 本文提出的方法在混沌时间序列数据的分析和预测中效果很好. 本文的数值实验虽然只与经典的 ESN 模型进行了对比, 但可以在一些改进的模型中引入延迟反馈获得更好的记忆性能, 从而更有利于在实际问题中应用.

### 参 考 文 献

- [1] JAEGER H, HASS H. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communications [J]. *Science*, 2004, 304: 78-80.
- [2] GAUTHIER D J, BOLLT E, GRIFFITH A, et al. Next Generation Reservoir Computing [J]. *Nature Communications*, 2021, 12: 5564-1-5564-7.
- [3] GONG Z C, CHEN H H, YUAN B, et al. Multiobjective Learning in the Model Space for Time Series Classification [J]. *IEEE Transactions on Cybernetics*, 2019, 49(3): 918-932.
- [4] BIANCHI F M, SCARDAPANE S, LØKSE S, et al. Reservoir Computing Approaches for Representation and Classification of Multivariate Time Series [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(5): 2169-2179.
- [5] GALLICCHIO C, MICHELI A, PEDRELLI L. Deep Reservoir Computing: A Critical Experimental Analysis [J]. *Neurocomputing*, 2017, 268: 87-99.
- [6] GALLICCHIO C, MICHELI A, PEDRELLI L. Design of Deep Echo State Networks [J]. *Neural Networks*, 2018, 108: 33-47.
- [7] OTTE S, BUTZ M V, KORYAKIN D, et al. Optimizing Recurrent Reservoirs with Neuro-Evolution [J]. *Neurocomputing*, 2016, 192: 128-138.
- [8] QIAO J F, LI F J, HAN H G, et al. Growing Echo-State Network with Multiple Subreservoirs [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, 28(2): 391-404.
- [9] WANG H S, LIU Y X, WANG D S, et al. Discriminative and Regularized Echo State Network for Time Series Classification [J]. *Pattern Recognition*, 2022, 130: 108811-1-108811-14.
- [10] ROY M, MANDAL S, HENS C, et al. Model-Free Prediction of Multistability Using Echo State Network [J]. *Chaos*, 2022, 32(10): 101104-1-101104-8.
- [11] PATHAK J, HUNT B, GIRVAN M, et al. Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach [J]. *Physical Review Letters*, 2018, 120(2): 024102-1-024102-5.
- [12] ZHANG H, FAN H W, WANG L, et al. Learning Hamiltonian Dynamics with Reservoir Computing [J]. *Physical Review E*, 2021, 104(2): 024205-1-024205-12.

(责任编辑: 赵立芹)