

基于层次化可塑性回声状态网络的混沌时间序列预测

那晓栋¹, 王嘉宁¹, 刘墨燃¹, 任伟杰², 韩敏^{3,4†}

- (1. 大连理工大学 电子信息与电气工程学部, 辽宁 大连 116024;
2. 哈尔滨工程大学 智能科学与工程学院, 哈尔滨 150001;
3. 大连理工大学 教育部工业设备智能控制与优化重点实验室, 辽宁 大连 116024;
4. 大连理工大学 辽宁省工业装备分布式控制专业技术创新中心, 辽宁 大连 116024)

摘要: 为了提高回声状态网络对于混沌时间序列特征提取与预测的能力, 提出一种层次化可塑性回声状态网络模型. 该模型将多个储备池顺序连接, 通过逐层特征变换的方式增强对非线性多尺度动态特征的提取能力. 同时, 引入神经科学中的内在可塑性机制模拟真实生物神经元的放电率分布, 以最大化神经元的信息传递为目标对储备池进行预训练. 层次化可塑性回声状态网络不仅能够增加模型的容量, 降低随机投影所带来的不稳定性, 而且也理解储备池的表示、处理、记忆及储存操作提供一种新的思路. 仿真实验结果表明, 相比于其他 7 种改进的回声状态网络模型, 所提出的模型在人造数据和真实数据所构成的混沌时间序列预测任务中均能取得最优的预测精度.

关键词: 混沌时间序列预测; 回声状态网络; 层次化策略; 神经内在可塑性; 预训练; 神经网络

中图分类号: TP183

文献标志码: A

DOI: 10.13195/j.kzyjc.2021.0773

开放科学(资源服务)标识码(OSID):



引用格式: 那晓栋, 王嘉宁, 刘墨燃, 等. 基于层次化可塑性回声状态网络的混沌时间序列预测[J]. 控制与决策, 2023, 38(1): 133-142.

Hierarchical plasticity echo state network for chaotic time series prediction

NA Xiao-dong¹, WANG Jia-ning¹, LIU Mo-ran¹, REN Wei-jie², HAN Min^{3,4†}

- (1. Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China;
2. College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China;
3. Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of Ministry of Education, Dalian University of Technology, Dalian 116024, China;
4. Professional Technology Innovation Center of Distributed Control for Industrial Equipment of Liaoning Province, Dalian University of Technology, Dalian 116024, China)

Abstract: To improve the ability of echo state network for feature extraction and prediction on chaotic time series, a hierarchical plasticity echo state network (HPESN) model is proposed. In this model, multiple reservoirs are connected in sequence, and the ability of nonlinear multi-scale dynamic feature extraction is enhanced through layer-by-layer feature transformation. Meanwhile, the intrinsic plasticity mechanism in neuroscience is introduced to simulate the firing rate distribution of real biological neurons, and the reservoir is pre-trained with the goal of maximizing neuronal information transmission. The HPESN not only increases the capacity of the model and reduces the instability caused by random projection, but also provides a new idea for understanding the representation, processing, memory and storage operations of the reservoir. The simulation results show that compared with other seven modified echo state network models, the proposed HPESN model achieves the best prediction accuracy in the chaotic time series prediction task composed of synthetic data and real-world data.

Keywords: chaotic time series prediction; echo state network; hierarchical strategy; neuronal intrinsic plasticity; pre-training; neural network

0 引言

综观 20 世纪下半叶, 非线性科学得到了蓬勃的发展, 其中对混沌的研究就占据了极大的份额. 混沌

科学与其他科学相互渗透, 无论是在数学、物理学还是经济学、天文学等领域, 混沌受到了广泛的关注^[1-2]. 在混沌的应用方面, 对混沌系统中观测得到的非线性

收稿日期: 2021-05-02; 录用日期: 2021-08-18.

基金项目: 国家自然科学基金项目(62173063); 中央高校基本科研业务费专项资金项目(DUT20LAB114).

†通讯作者. E-mail: minhan@dlut.edu.cn.

时间序列进行分析与预测,一直都是科学研究和工程技术领域的研究热点^[3-4].

近些年来,基于数据驱动的混沌时间序列预测方法受到越来越多的关注.在众多机器学习方法中,回声状态网络(echo state network, ESN)作为一种结构简洁、训练高效的特殊循环神经网络而被广泛应用于混沌时间序列的预测^[5]. ESN的核心思想是把原始信号通过一个固定的输入层随机投影到一个稀疏高维的动态空间中,以获得更为丰富的高维动态表示^[6-7].值得注意的是,目前的研究已从理论分析和实际应用的角度全面验证了ESN模型的有效性.从非线性动力学的角度,文献[8]提出了一个能够解释ESN模型是如何学习到混沌系统中长期非线性动力学演化规律的理论框架;而文献[9]则证明了在输入有界的情况下,ESN具有通用的一致逼近性.从工程应用的角度,ESN已被成功地应用于风能预测^[10]、电力负荷预测^[11]、太阳辐射量预测^[12]和燃料电池剩余寿命估计^[13]等众多领域.然而,正如文献[14]所指出的那样:时间序列数据中普遍具有复杂的多尺度动态特性,在预测难度逐步增大的情况下,单隐层循环神经网络在提取和处理时序数据中复杂的动态依赖关系时仍面临严峻挑战.

随着计算机算力与数据获取能力的不断提升,网络结构的层次化逐渐成为一种提升模型多尺度特征提取与非线性拟合能力的有效手段^[15].近些年来,ESN模型的研究中也出现了对网络结构层次化的尝试.文献[16]和文献[17]分别提出了 φ -ESN和随机静态投影储备池(reservoir with random static projections, R²SP)模型, φ -ESN和R²SP的核心思想都是在动态储备池结构的基础上添加静态层,使改进后的模型具有更强的短期记忆与非线性映射能力.此后,多层回声状态机(multilayered echo state machine, ML-ESM)^[18]和深度回声状态网络(deep echo state network, DeepESN)^[19]被相继提出,从真正意义上实现了储备池结构的层次化.近期,文献[20]提出了异步深度储备池计算模型(asynchronous deep reservoir computing, ADRC),该模型通过异步处理方式将长时记忆任务转化为一组短时记忆任务,显著提高了模型的记忆容量.文献[21]提出一种多层投影编码的层次化储备池计算框架,称为深度投影编码回声状态网络(deep projection-encoding echo state network, DeePr-ESN),并通过实验验证了DeePr-ESN模型在捕捉多尺度动态特征方面的优势,进一步完善了对ESN模型层次化的研究和探索.

目前所提出的大多数层次化ESN模型都着重于网络整体结构的改进,忽略了储备池内部拓扑结构的设计.文献[22]指出,随机生成的储备池内部权重可能会产生不清晰的动态特性,难以形成与具体任务相匹配的最优结构.此外,随着储备池个数的增多以及网络规模的扩大,随机生成的储备池也会在一定程度上增加模型的不稳定性.因此,对储备池内部结构的约束和改进也是需要重点考虑的因素之一.对于储备池内部拓扑结构的改进,文献[23-24]提出了两种确定性的储备池拓扑结构,分别为简单环形储备池(simple cycle reservoir, SCR)和跳跃式环形储备池(cycle reservoir with jumps, CRJ).SCR和CRJ这两种模型均表现出在不影响网络学习性能的前提下同时具备最小化储备池计算复杂度的特点.随着脑科学研究的不断深入,神经科学的相关理论也在指导着储备池结构的设计.受大脑皮层神经元连接方式的启发,文献[25-26]提出了具有小世界网络拓扑结构的ESN模型,并验证了小世界回声状态网络(small world echo state network, SWESN)在鲁棒性和抗干扰性等方面均优于标准ESN.另外,文献[27]发现,神经科学中的可塑性机制能使输入数据在某种给定约束下完成最大化信息传递,基于可塑性理论的预训练策略能使储备池具有更强的稳定性和记忆能力.神经可塑性理论目前在生物实验和认知科学领域已取得了一系列成果,但作为一种生物学启发的无监督调节机制,其在计算科学领域的价值还有待于进一步的探索与研究.

基于层次化建模策略和神经可塑性机制,本文提出一种具有层次化结构的神经内在可塑性回声状态网络,称为层次化可塑性回声状态网络(hierarchical plasticity echo state network, HPESN).HPESN模型能通过逐层学习的方式挖掘和捕获隐藏在混沌系统中潜在的演化模式,并利用内在可塑性机制最大化神经元在储备池中的信息传递.最后采用直接预测的方式,实现混沌时间序列的单步和多步预测.

1 回声状态网络

标准ESN是由输入层、输出层和储备池构成的一种特殊循环神经网络^[5].储备池是ESN的核心,其本质上是一个随机生成的大规模稀疏连接的矩阵.储备池结构保证了ESN具有强大的非线性映射能力,同时也使得网络具有丰富的动态性、良好的自适应性和自组织性.标准ESN的状态更新方程可表示为

$$\mathbf{x}(t) = f(\mathbf{W}_{in}\mathbf{u}(t) + \mathbf{W}_{res}\mathbf{x}(t-1)). \quad (1)$$

其中: $f(\cdot)$ 表示非线性激活函数, $\mathbf{u}(t) \in \mathbf{R}^K$ 和 $\mathbf{x}(t) \in \mathbf{R}^{N_{\text{res}}}$ 分别表示储备池在 t 时刻的输入信号和内部状态信号, $\mathbf{W}_{\text{in}} \in \mathbf{R}^{N_{\text{res}} \times K}$ 和 $\mathbf{W}_{\text{res}} \in \mathbf{R}^{N_{\text{res}} \times N_{\text{res}}}$ 分别表示随机生成的输入权重矩阵和储备池内部权重矩阵, K 表示输入信号的维度, N_{res} 表示储备池内部神经元的个数. ESN 模型的输出可表示为

$$\mathbf{v}(t) = f_{\text{out}}(\mathbf{W}_{\text{out}}[\mathbf{u}(t); \mathbf{x}(t)]). \quad (2)$$

其中: $[\cdot; \cdot]$ 表示向量的垂直连接, $\mathbf{v}(t) \in \mathbf{R}^L$ 表示 t 时刻网络的输出, L 表示输出信号的维度. 在回归任务中, $f_{\text{out}}(\cdot)$ 选用恒等函数, 通常使用线性回归或岭回归方法对输出权重矩阵 $\mathbf{W}_{\text{out}} \in \mathbf{R}^{L \times (N_{\text{res}} + K)}$ 进行求解^[6-7].

2 层次化可塑性回声状态网络

本节首先介绍真实生物神经元中的放电率分布, 然后进一步解释如何利用生物神经内在可塑性机制对储备池进行预训练, 最后提出一种具有层次化结构的可塑性回声状态网络模型.

2.1 生物神经元的放电率分布

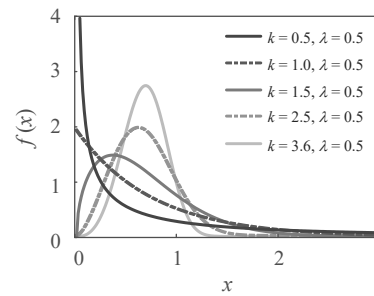
针对使用数学的概率分布描述生物神经元放电率的问题, 文献[28]和文献[29]分别使用指数分布和高斯分布进行尝试. 但是文献[30-31]的研究指出, 带有垂头的指数分布可以最大限度地将信息进行传递, 更适合描述真实生物元的放电率分布. 从概率分布的角度来看, 带有垂头的指数分布可以使用 Weibull 分布来近似表示.

Weibull 分布是由瑞典科学家 Waloddi Weibull 于 20 世纪中叶提出的一种连续性的概率分布, 其概率密度函数可以表示为

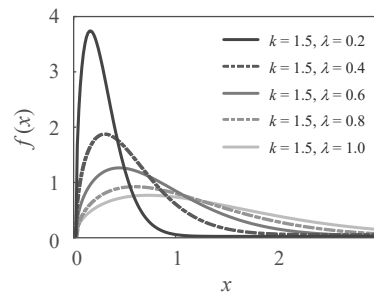
$$f(x; k, \lambda) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, & x \geq 0; \\ 0, & x < 0. \end{cases} \quad (3)$$

其中: x 为随机变量; $k > 0$ 为形状参数, 其决定 Weibull 分布概率密度函数的基本形状; $\lambda > 0$ 为尺度参数, 其起到缩放函数的作用. 形状参数 k 和尺度参数 λ 对 Weibull 分布概率密度函数曲线的影响规律如图 1 所示.

从图 1 中可以看出, 若尺度参数 λ 保持不变, 则 Weibull 分布概率密度函数曲线的形状会受到形状参数 k 的控制. 当 $0 < k \leq 1$ 时, Weibull 分布近似于指数分布; 当 $1 < k < 3.6$ 时, Weibull 分布近似于一个带有垂头的指数分布, 正是这种带有垂头的指数分布能有效地对真实生物元的放电率分布进行描述, 并最大化神经元的的信息传递. 而尺度参数 λ 的主要作用是对函数进行缩放, 当形状参数 k 保持不变时, 随着尺度



(a) 形状参数 k 对概率密度函数曲线的影响



(b) 尺度参数 λ 对概率密度函数曲线的影响

图 1 参数 k 和 λ 对 Weibull 概率密度函数曲线的影响

参数 λ 的增大或减小, Weibull 分布的概率密度函数曲线会随着横坐标伸张或收缩.

2.2 储备池的神经内在可塑性

从计算科学的角度进行分析, 神经内在可塑性规则可以被视为一种受生物启发的无监督学习. 内在可塑性使得单个神经元的活动水平维持在一定的稳态范围, 同时又能最大程度地将信息进行传递. 其中, 稳态特性可以维持系统稳定的代谢消耗, 而最大化信息传递能够最大化神经元输入与输出间的互信息, 使信息得到更为有效的利用. 基于信息最大化理论, 通过最小化实际输出与 Weibull 分布之间的 Kullback-Leiber(KL) 散度可以推导出内在可塑性规则. 假设神经元的激活函数 $f(\cdot)$ 为 sigmoid 函数, 则

$$y = \text{sigmoid}(x) = \frac{1}{1 + e^{-(ax+b)}}. \quad (4)$$

其中: 参数 a 和 b 分别表示激活函数的斜率和偏置; x 表示神经元的输入, 服从分布 $x \sim f_x(x)$. 于是, 神经元的输出 y 服从分布

$$y \sim f_y(y) = f_x(x) \frac{\partial x}{\partial y}. \quad (5)$$

对激活函数进行求导, 可得

$$\begin{aligned} \partial y / \partial x &= ay(1-y), \\ \partial y / \partial a &= xy(1-y), \\ \partial y / \partial b &= y(1-y). \end{aligned} \quad (6)$$

本文通过最小化实际输出分布 f_y 与期望输出分布 f_{Weibull} 的 KL 散度来推导可塑性规则, 具体过程为

$$\begin{aligned}
D &= d_{\text{KL}}(f_y \parallel f_{\text{Weibull}}) = \\
&\int f_y(y) \log\left(\frac{f_y(y)}{f_{\text{Weibull}}(y)}\right) dy = \\
&\int f_y(y) \log(f_y(y)) dy + \frac{1}{\lambda^k} \int f_y(y) y^k dy - \\
&(k-1) \int f_y(y) \log(y) dy - \log\left(\frac{k}{\lambda^k}\right) = \\
&-H(y) + \frac{1}{\lambda^k} E(y^k) - \\
&(k-1)E(\log(y)) - \log\left(\frac{k}{\lambda^k}\right). \quad (7)
\end{aligned}$$

其中: $E(\cdot)$ 表示数学期望; $H(y)$ 表示熵, 用来衡量一个随机事件的不确定性. 通过最小化 KL 散度 D 可以得到最大信息量, 熵 $H(y)$ 还可以表示为

$$\begin{aligned}
H(y) &= -\int f_y(y) \log(f_y(y)) dy = \\
&E\left(\log\frac{\partial y}{\partial x}\right) - E(\log f_x(x)). \quad (8)
\end{aligned}$$

KL 散度 D 对斜率 a 的计算过程可表示为

$$\begin{aligned}
\frac{\partial D}{\partial a} &= \frac{\partial D}{\partial y} \cdot \frac{\partial y}{\partial a} = \\
&-\frac{1}{a} + E\left(\left(-\frac{1}{y} + \frac{1}{1-y} + \frac{ky^{k-1}}{\lambda^k} - \frac{k-1}{y}\right) \frac{\partial y}{\partial a}\right) = \\
&-\frac{1}{a} + E\left(\left(\frac{1}{1-y} + \frac{ky^{k-1}}{\lambda^k} - \frac{k}{y}\right) xy(1-y)\right) = \\
&-\frac{1}{a} + E\left(\frac{kxy^k(1-y)}{\lambda^k} - kx + (k+1)xy\right). \quad (9)
\end{aligned}$$

KL 散度 D 对偏置 b 的计算过程可表示为

$$\begin{aligned}
\frac{\partial D}{\partial b} &= \frac{\partial D}{\partial y} \cdot \frac{\partial y}{\partial b} = \\
&E\left(\left(\frac{1}{1-y} + \frac{ky^{k-1}}{\lambda^k} - \frac{k}{y}\right) \frac{\partial y}{\partial b}\right) = \\
&E\left(\left(\frac{1}{1-y} + \frac{ky^{k-1}}{\lambda^k} - \frac{k}{y}\right) y(1-y)\right) = \\
&E\left(y + \frac{ky^k(1-y)}{\lambda^k} - k(1-y)\right). \quad (10)
\end{aligned}$$

由此得到神经内在可塑性的学习规则为

$$\begin{aligned}
a &= a + \Delta a = a + \eta_a \left(-\frac{\partial D}{\partial a}\right) = \\
&a + \eta_a \left(\frac{1}{a} + kx - (k+1)xy - \frac{kxy^k(1-y)}{\lambda^k}\right), \quad (11)
\end{aligned}$$

$$\begin{aligned}
b &= b + \Delta b = b + \eta_b \left(-\frac{\partial D}{\partial b}\right) = \\
&b + \eta_b \left(-y + k(1-y) - \frac{ky^k(1-y)}{\lambda^k}\right), \quad (12)
\end{aligned}$$

其中 η_a 和 η_b 表示内在可塑性规则的学习率.

通过以上描述的无监督内在可塑性学习规则对储备池进行预训练, 即可得到满足可塑性特性的储备池结构.

2.3 HPESN 模型结构

如图 2 所示, 使用层次化策略和可塑性规则的 HPESN 模型由多个顺序连接的储备池构成.

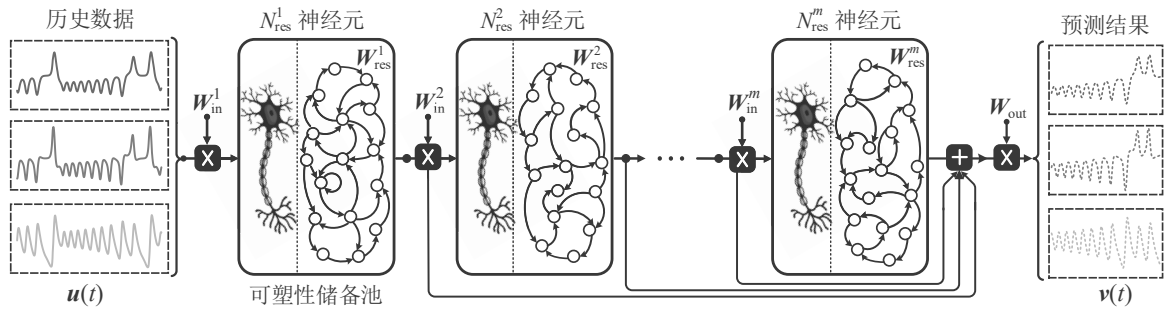


图 2 HPESN 结构

从整体结构上来看, 一个具有 m 个储备池的 HPESN 模型, 第 i 个储备池中神经元的个数记为 N^i_{res} , $i = 1, 2, \dots, m$; t 时刻的储备池内部状态记为 $\mathbf{x}^i_{\text{res}}(t)$; 输入 $\mathbf{u}(t)$ 和输出 $\mathbf{v}(t)$ 分别为 K 维和 L 维的向量; \mathbf{W}^i_{in} 为第 i 个随机生成的储备池输入权重矩阵, 当 $i = 1$ 时, $\mathbf{W}^i_{\text{in}} \in \mathbf{R}^{N^i_{\text{res}} \times D}$, 当 $i = 2, \dots, m$ 时, $\mathbf{W}^i_{\text{in}} \in \mathbf{R}^{N^i_{\text{res}} \times N^{i-1}_{\text{res}}}$; $\mathbf{W}^i_{\text{res}} \in \mathbf{R}^{N^i_{\text{res}} \times N^i_{\text{res}}}$ 为储备池内部权重矩阵, 其性质主要受两个超参数的影响, 即稀疏度 η_i 和谱半径 ρ_i , 稀疏度 η_i 决定 $\mathbf{W}^i_{\text{res}}$ 中非零元素的比例, 而 $\mathbf{W}^i_{\text{res}}$ 的生成则受谱半径 ρ_i 的控制, 具体可表示为

$$\mathbf{W}^i_{\text{res}} = \rho_i \cdot \frac{\mathbf{W}^i}{\lambda_{\max}(\mathbf{W}^i)}. \quad (13)$$

这里: 矩阵 \mathbf{W}^i 中的元素在 $[-1, 1]$ 的范围内随机生成; $\lambda_{\max}(\mathbf{W}^i)$ 表示矩阵 \mathbf{W}^i 的最大特征值; 一般情况下, 储备池谱半径 $\rho_i < 1$ 时可使模型满足回声状态特性^[16,32]. HPESN 的状态更新方程可表示为

$$\mathbf{x}^i_{\text{res}}(t) = f(\mathbf{W}^i_{\text{in}} \mathbf{x}^i_{\text{in}}(t) + \mathbf{W}^i_{\text{res}} \mathbf{x}^i_{\text{res}}(t-1)). \quad (14)$$

其中: $f(\cdot)$ 表示激活函数, 一般选用 sigmoid 函数; $\mathbf{x}^i_{\text{in}}(t)$ 为 t 时刻第 i 个储备池的输入信号. 当 $i = 1$ 时, 储备池的输入为原始信号, 即 $\mathbf{x}^1_{\text{in}}(t) = \mathbf{u}(t)$; 当 $i > 1$ 时, 当前储备池的输入为前一个储备池的输出, 即 $\mathbf{x}^i_{\text{in}}(t) = \mathbf{x}^{i-1}_{\text{res}}(t)$. 在多个储备池进行逐层计算完成后, 将各层的状态信号进行整合, 得到 t 时刻的采集

矩阵 $\mathbf{X}(t)$ 为

$$\mathbf{X}(t) = [\mathbf{u}(t); \mathbf{x}_{\text{res}}^1(t); \mathbf{x}_{\text{res}}^2(t); \dots; \mathbf{x}_{\text{res}}^m(t)], \quad (15)$$

其中 $[\cdot; \cdot]$ 表示向量的垂直连接. 假设序列长度为 T , 即 $t = 1, 2, \dots, T$, 采集矩阵可表示为 $\mathbf{X} = [\mathbf{X}(1), \mathbf{X}(2), \dots, \mathbf{X}(T)]$, 则 HPESN 输出的矩阵 $\mathbf{V} = [\mathbf{V}(1), \mathbf{V}(2), \dots, \mathbf{V}(T)]$ 可表示为

$$\mathbf{V} = \mathbf{W}_{\text{out}} \mathbf{X}. \quad (16)$$

假设已知的目标输出矩阵为 $\mathbf{Y} = [\mathbf{Y}(1), \mathbf{Y}(2), \dots, \mathbf{Y}(T)]$, 则目标函数可构造为

$$\min F(\mathbf{W}_{\text{out}}) = \|\mathbf{V} - \mathbf{Y}\|_2^2 + \beta \|\mathbf{W}_{\text{out}}\|_2^2. \quad (17)$$

其中, 所要求解的输出权重矩阵 \mathbf{W}_{out} 是一个 $L \times (K + \sum_{i=1}^m N_{\text{res}}^i)$ 维度的矩阵, 利用岭回归方法求解可得

$$\hat{\mathbf{W}}_{\text{out}} = \mathbf{Y} \mathbf{X}^\dagger = \mathbf{Y} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T + \beta \mathbf{I})^{-1}. \quad (18)$$

β 是 L_2 正则化的约束系数, 一般取值为很小的正数.

综上所述, HPESN 模型的具体建模步骤如下:

step 1: 网络参数的初始化. 建立具有 m 个储备池的 HPESN 模型, 设置每个储备池的神经元个数为 N_{res}^i , 随机生成元素范围在 $[-1, 1]$ 之间的输入权重 \mathbf{W}_{in}^i 和稀疏的储备池内部权重 $\mathbf{W}_{\text{res}}^i$, 其中 $\mathbf{W}_{\text{res}}^i$ 的谱半径和稀疏度分别为 ρ_i 和 η_i , $i = 1, 2, \dots, m$.

step 2: 回声状态信号的采集. 根据式(14)逐层计算并更新储备池的内部状态信号, 并以式(15)的方式构建包含各层状态信息的采集矩阵 \mathbf{X} .

step 3: 神经可塑性储备池的构建. 根据式(11)和(12), 通过神经内在可塑性规则预训练储备池神经元激活函数的斜率 a 和偏置 b , 并对采集矩阵 \mathbf{X} 进行更新.

step 4: 输出权重 \mathbf{W}_{out} 的求解. 以式(17)为目标函数, 利用式(18)训练输出层的权重 \mathbf{W}_{out} .

2.4 计算复杂度分析

假设 HPESN 模型具有 m 个储备池, 每个储备池中神经元个数均为 N , 给定长度为 T 的 K 维时间序列, 则使用神经内在可塑性机制对第 i 个储备池进行预训练的计算复杂度 C_{pre}^i 可近似表示为

$$C_{\text{pre}}^i \approx O(STN), \quad (19)$$

其中 S 表示求取式(11)和(12)中斜率 a 和偏置 b 的迭代次数. HPESN 模型中第 i 个储备池的计算复杂度 C_{res}^i 可表示为

$$C_{\text{res}}^i = \begin{cases} O(KTN + TN^2), & i = 1; \\ O(2TN^2), & i = 2, 3, \dots, m. \end{cases} \quad (20)$$

经过逐层计算, 输入变量和储备池的所有状态变量组成维度为 $(mN + K) \times T$ 的采集矩阵 \mathbf{X} , 因此, HPESN 模型求取 \mathbf{W}_{out} 的回归过程计算复杂度 C_{reg} 可表示为

$$C_{\text{reg}} = O(LTP + TP^2 + LP^2 + P^3) = O(LTP + (L + T + P)P^2), \quad (21)$$

其中 $P = mN + K$, 由于储备池中神经元数量 N 通常远大于输入变量的维度 K , 即 $N \gg K$, 因此可得 $P \approx mN$. 同时, 序列的长度 T 远大于输出变量的维度 L , 即 $T \gg L$, 因此 C_{reg} 可近似表示为

$$C_{\text{reg}} \approx O((T + P)P^2) \approx (Om^2TN^2 + m^3N^3). \quad (22)$$

HPESN 的计算复杂度 C_{HPESN} 可整理为

$$\begin{aligned} C_{\text{HPESN}} &= \sum_{i=1}^m (C_{\text{pre}}^i + C_{\text{res}}^i) + C_{\text{reg}} \approx \\ &O(mSTN + (TN^2 + KTN) + \\ &(m - 1)(2TN^2) + m^2TN^2 + m^3N^3) = \\ &O((mS + K)TN + \\ &(m^2 + 2m - 1)TN^2 + m^3N^3) \approx \\ &O(m^2TN^2 + m^3N^3). \end{aligned} \quad (23)$$

因此, HPESN 模型的计算复杂度可近似表示为 $O(m^2TN^2 + m^3N^3)$. 值得注意的是, HPESN 模型的计算复杂度 C_{HPESN} 与储备池的个数 m 以及储备池中神经元数量 N 成三次方关系. 因此, 若要在发挥 HPESN 逐层特征提取特点的同时保证储备池计算高效的优点, 储备池的数量和规模需要被适当限制.

3 仿真实验

本节选用 Lorenz 混沌时间序列和北京空气质量指数时间序列来验证本文所提出 HPESN 模型对于时间序列预测任务的有效性. 同时选用单储备池结构的 ESN^[5]、漏积分回声状态网络 (leaky integrator echo state network, LIESN)^[6]、SCR^[23]、CRJ^[24] 和 SWESN^[26] 与多储备池结构的 DeepESN^[19]、ADRC^[20] 和 DeePr-ESN^[21] 作为对比模型进行实验. 在对比模型中, SCR、CRJ 和 SWESN 模型均对储备池的内部拓扑结构进行了改进, DeepESN、ADRC 和 DeePr-ESN 则属于较为新颖的多储备池模型. 另外, 选用均方根误差 (root mean square error, RMSE) 作为定量评价模型预测性能指标, RMSE 的计算公式为

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2}. \quad (24)$$

其中: y_t 为真实值, \hat{y}_t 为预测值, T 为样本个数. 为了减小随机性对实验结果的影响, 本文仿真实验结果均取 10 次重复实验的平均值.

3.1 Lorenz 时间序列

1963 年, Edward Lorenz 在研究对流不稳定性时首次发现了气象系统中特有的“蝴蝶效应”, 并提出了经典的 Lorenz 系统. Lorenz 系统可以描述为

$$\begin{aligned} dx/dt &= a(y - x), \\ dy/dt &= cx - y - xz, \\ dz/dt &= xy - bz. \end{aligned} \quad (25)$$

其中, 当 $a = 10, b = 8/3, c = 28$ 时, 系统表现出混沌特性. Lorenz 系统的混沌吸引子和观测得到的时间序列如图 3 所示.

本文使用四阶龙格库塔法从初始条件 $(x_0, y_0, z_0) = (1, 1, 1)$ 以 0.01 的步长生成 10 000 个样本点. 将前 80% 的数据划分为训练集, 后 20% 的数据划分为测试集. 丢弃训练集的前 100 个样本用于克服储备池初始瞬态的影响. 为了让本文所提出的 HPESN 与各

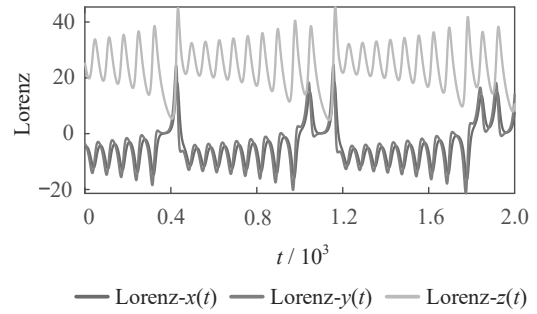


图 3 Lorenz 混沌时间序列

个对比模型进行公平比较, 所有模型中储备池的谱半径均设置为 0.9, 稀疏度设置为 0.1, 具有单储备池结构的 ESN、LIESN、SCR、CRJ 和 SWESN 模型神经元个数设置为 600, DeepESN、ADRC、DecPr-ESN 和 HPESN 均设置为具有 3 个储备池, 且每个储备池神经元个数为 200 的结构. 表 1 给出了 9 种不同模型对于 Lorenz- $x(t)$ 、Lorenz- $y(t)$ 和 Lorenz- $z(t)$ 序列在测试集中的多步预测结果. 可以看出, HPESN 模型在 1 步、5 步和 10 步预测中均表现出了最好的预测性能.

为了更直观地比较不同模型的多步预测结果,

表 1 不同模型对于 Lorenz- $x(t)$ 、Lorenz- $y(t)$ 和 Lorenz- $z(t)$ 多步预测的 RMSE 结果

模型 (10^{-3})	1 步预测			5 步预测			10 步预测			平均用时 /s
	$x(t)$	$y(t)$	$z(t)$	$x(t)$	$y(t)$	$z(t)$	$x(t)$	$y(t)$	$z(t)$	
ESN ^[5]	0.449	0.942	0.902	2.021	5.071	5.061	3.379	7.794	8.366	4.079
LIESN ^[6]	0.507	1.107	1.048	1.923	5.288	5.053	3.734	10.128	10.587	4.333
SCR ^[23]	0.401	0.819	0.801	1.687	4.802	4.255	3.369	7.514	7.311	4.012
CRJ ^[24]	0.398	0.815	0.801	1.711	4.824	4.352	3.338	7.511	7.097	5.215
SWESN ^[26]	0.401	0.819	0.812	1.757	4.745	4.337	3.268	7.227	6.921	5.564
DeepESN ^[19]	0.452	0.933	0.898	1.801	4.836	4.429	3.216	6.823	6.872	9.843
ADRC ^[20]	0.453	0.813	0.791	1.596	4.138	3.226	2.711	4.993	3.867	11.356
DecPr-ESN ^[21]	0.417	0.826	0.838	1.651	4.253	3.892	3.019	5.651	5.117	36.654
HPESN	0.382	0.756	0.779	1.450	3.850	3.024	2.561	4.882	3.757	62.216

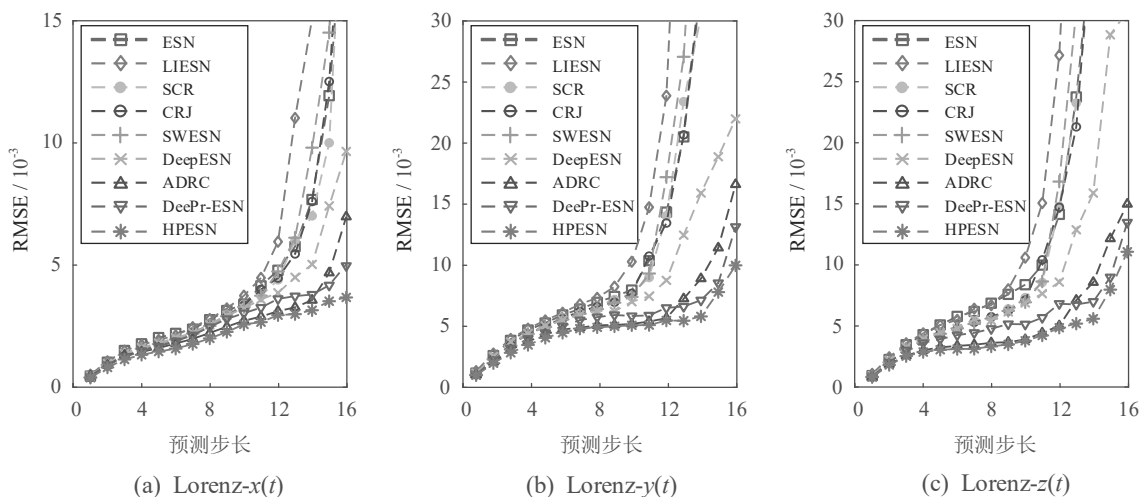
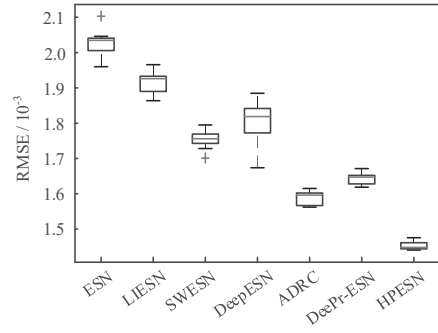


图 4 Lorenz- $x(t)$ 、Lorenz- $y(t)$ 和 Lorenz- $z(t)$ 的多步预测误差趋势

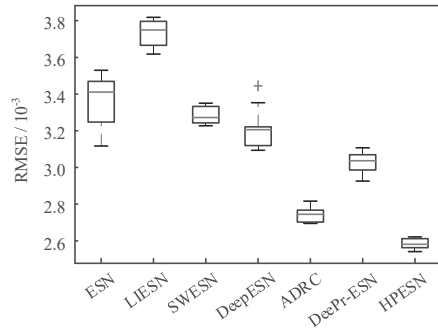
图4给出了9种模型1步预测~16步预测的误差趋势。从图4中可以看出:HPESN在面对短期预测任务时,预测性能的优势并不明显;当预测步长增大到一定程度时,HPESN逐步显示出其预测的优势。值得注意的是,当预测步长超出一定范围时,预测误差会出现一个陡然的上升。以Lorenz- $x(t)$ 序列为例,当预测步长 h 大于12时,ESN、LIESN、SCR、CRJ和SWESN等单储备池模型的预测误差会急剧增加。因为随着预测步长的增加,对预测模型的非线性特征提取与记忆能力会有更高的要求,尽管单个储备池拥有更多的神经元,但是,在逐层挖掘混沌系统中潜在的演化信息方面的能力还是弱于层次化储备池模型。相对而言,隶属于层次化储备池结构的ADRC、DeePr-ESN和HPESN表现出了较好的多步预测性能。因此,层次化策略有助于提高模型的非线性拟合和捕获隐藏在混沌系统中潜在的演化模式的能力。

由于以上多步预测的结果为10次相同实验的均值,为了进一步分析各个模型预测结果的稳定性,本文使用箱式图对Lorenz- $x(t)$ 序列的5步预测和10步预测的结果进行分析,结果如图5所示(因SCR和CRJ使用具有确定性环形结构的储备池,故不参与比较)。从图5中可以看出:在层次化储备池模型中,DeepESN由于没有特殊的设计与约束,其在Lorenz- $x(t)$ 序列的5步预测和10步预测中表现出了较大的

方差,预测结果呈现出较大的不稳定性;本文所提出的HPESN由于使用神经可塑性机制,最大化储备池中神经元的信息传递,使信息得到有效利用,因此,不仅得到了最好的预测性能,同时也表现出了良好的预测稳定性。



(a) 5步预测结果



(b) 10步预测结果

图5 Lorenz- $x(t)$ 序列的5步和10步预测结果的箱式图

表2 北京PM_{2.5}浓度多步预测的RMSE结果

模型	1步 预测	4步 预测	8步 预测	12步 预测	16步 预测	平均 用时/s
ESN ^[5]	1.129	5.583	11.844	18.739	24.035	9.641
LIESN ^[6]	1.102	5.554	11.802	18.838	22.963	10.354
SCR ^[23]	1.089	5.165	11.099	16.863	21.142	8.755
CRJ ^[24]	1.223	6.254	13.808	20.857	24.504	9.321
SWESN ^[26]	1.135	5.781	12.008	18.690	22.849	10.665
DeepESN ^[19]	1.125	5.349	11.680	18.098	22.103	26.541
ADRC ^[20]	1.062	4.960	10.781	16.510	19.991	31.012
DeePr-ESN ^[21]	1.078	5.242	11.322	17.414	20.973	92.554
HPESN	1.001	4.755	10.225	15.573	19.707	173.156

3.2 北京空气质量指数时间序列

北京空气质量指数数据集记录了2018年1月1日~2020年12月31日北京地区的主要空气污染物浓度,具体包括PM_{2.5}、PM₁₀、NO₂、CO、O₃和SO₂六种有害污染物的浓度。污染物浓度数据每小时采集一次,共采集样本26304个。使用全部样本的前80%作为训练集,后20%为测试集,预测模型关键超参数的设置与Lorenz系统的仿真实验保持一致。以PM_{2.5}浓度为预测目标,单步和多步预测的RMSE误差如表2

和图6所示。

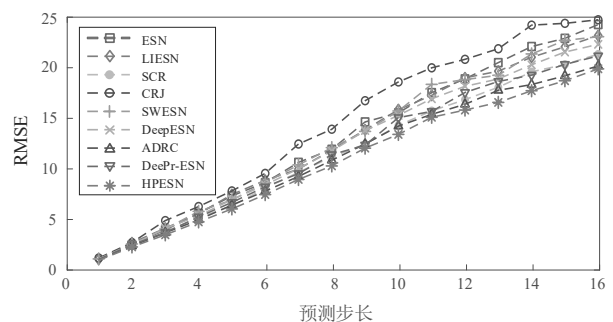
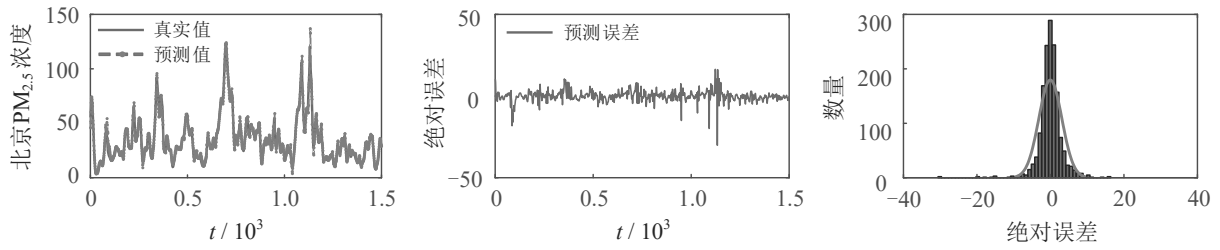
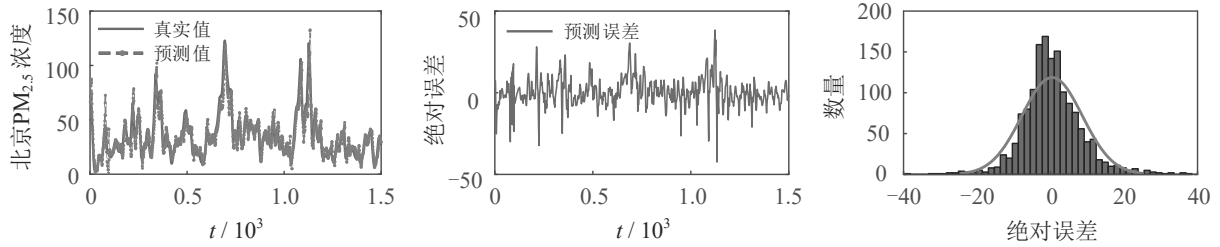


图6 北京PM_{2.5}浓度的多步预测误差趋势



(a) 4步预测结果



(b) 8步预测结果

图7 北京PM_{2.5}浓度的4步预测和8步预测效果

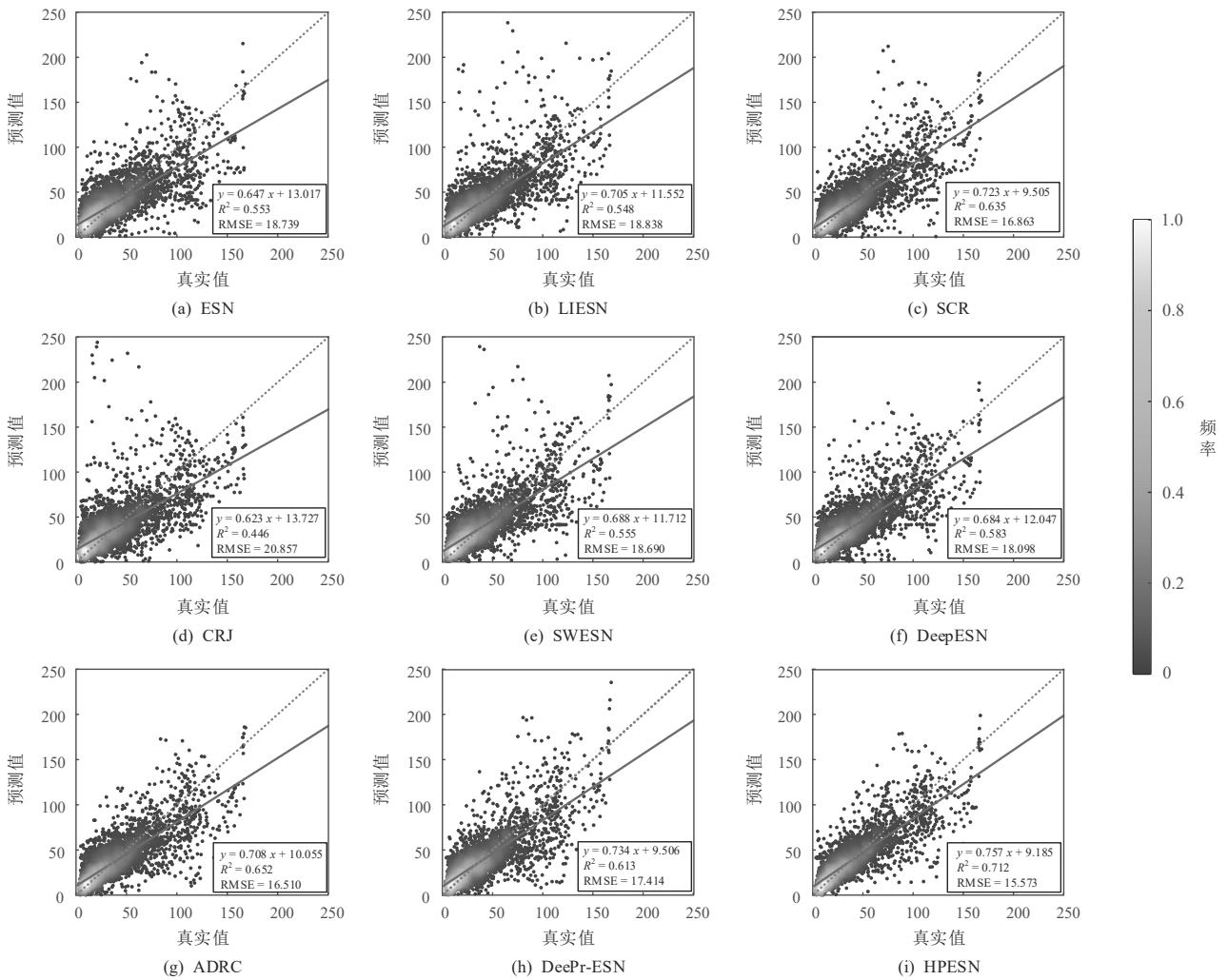


图8 北京PM_{2.5}浓度的不同模型12步预测的性能

从表2中可以看出,在真实数据的预测任务中,HPESN模型依旧表现出良好的预测性能.值得注意

的是,HPESN的用时明显高于其他模型,其原因是基于神经内在可塑性规则的预训练过程增加了计算成

本. 但是, 总体上HPESN建模所耗费的时间仍保持在可接受的范围之内. 另外, 从图6的误差变化趋势可以看出, 随着预测步长从1步增加到16步, HPESN一直保持着最小的预测误差.

为了更直观地验证HPESN模型的预测效果, 图7给出了HPESN模型4步预测和8步预测的效果. 从图7中可以看出, HPESN在进行4步预测时, 预测曲线能较好地真实曲线进行拟合, 预测绝对误差的范围基本控制在 $[-20, 20]$ 之间, 并且误差的分布近似于正态分布, 从而验证了HPESN模型在短期预测中的有效性. 当预测步长增加到8步时, 尽管预测曲线的拟合程度有所降低, 但是预测曲线依旧能反映出真实 $PM_{2.5}$ 浓度的变化趋势. 8步预测时的绝对误差仍呈现出正态分布, 同样说明了HPESN模型对于真实混沌时间序列多步预测的有效性.

最后, 为了对不同模型的预测效果进行比较与分析, 以真实值为横坐标, 预测值为纵坐标对测试集 $PM_{2.5}$ 浓度的12步预测结果进行可视化. 如图8所示, HPESN模型所表现出的散点更多地聚集在直线 $y = x$ 附近, 说明预测值与真实值的差异更小. HPESN模型12步预测结果得到了最大的决定系数 $R^2 = 0.712$, 同时得到了最小的 $RMSE = 15.573$, 从而表明HPESN模型相对于其他对比模型具有最优的预测效果.

4 结论

本文提出了一种层次化可塑性回声状态网络模型以进行混沌时间序列的单步和多步预测. 所提出的模型通过层次化策略逐层学习时序数据中多尺度的动态特征, 并通过神经可塑性机制最大化储备池中神经元的信息传递, 有效地提高了混沌时间序列的预测精度. 本文通过Lorenz时间序列和北京空气质量指数时间序列的仿真实验, 验证了所提出模型的有效性和优越性.

参考文献(References)

[1] Tang Y, Kurths J, Lin W, et al. Introduction to focus issue: When machine learning meets complex systems: Networks, chaos, and nonlinear dynamics[J]. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2020, 30(6): 063151.

[2] Mudelsee M. Trend analysis of climate time series: A review of methods[J]. *Earth-Science Reviews*, 2019, 190: 310-322.

[3] 王迪, 王萍, 石君志. 一种错误率可控的混沌时间序列区间预测算法[J]. *控制与决策*, 2019, 34(5): 956-964.

(Wang D, Wang P, Shi J Z. An interval prediction algorithm for chaos time series with controllable error rate[J]. *Control and Decision*, 2019, 34(5): 956-964.)

[4] Li D C, Han M, Wang J. Chaotic time series prediction based on a novel robust echo state network[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, 23(5): 787-799.

[5] Jaeger H, Haas H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication[J]. *Science*, 2004, 304(5667): 78-80.

[6] Jaeger H, Lukoševičius M, Popovici D, et al. Optimization and applications of echo state networks with leaky-integrator neurons[J]. *Neural Networks*, 2007, 20(3): 335-352.

[7] Lukoševičius M, Jaeger H. Reservoir computing approaches to recurrent neural network training[J]. *Computer Science Review*, 2009, 3(3): 127-149.

[8] Lu Z X, Hunt B R, Ott E. Attractor reconstruction by machine learning[J]. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2018, 28(6): 061104.

[9] Grigoryeva L, Ortega J P. Echo state networks are universal[J]. *Neural Networks*, 2018, 108: 495-508.

[10] Wang J Z, Niu T, Lu H Y, et al. A novel framework of reservoir computing for deterministic and probabilistic wind power forecasting[J]. *IEEE Transactions on Sustainable Energy*, 2020, 11(1): 337-349.

[11] Peng L, Lv S X, Wang L, et al. Effective electricity load forecasting using enhanced double-reservoir echo state network[J]. *Engineering Applications of Artificial Intelligence*, 2021, 99: 104132.

[12] Wu Z, Li Q, Xia X H. Multi-timescale forecast of solar irradiance based on multi-task learning and echo state network approaches[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(1): 300-310.

[13] Hua Z G, Zheng Z X, Péra M C, et al. Remaining useful life prediction of PEMFC systems based on the multi-input echo state network[J]. *Applied Energy*, 2020, 265: 114791.

[14] Chung J, Ahn S, Bengio Y. Hierarchical multiscale recurrent neural networks[C]. *The 5th International Conference on Learning Representations*. Toulon, 2017: 24-26.

[15] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.

[16] Gallicchio C, Micheli A. Architectural and Markovian factors of echo state networks[J]. *Neural Networks*, 2011, 24(5): 440-456.

[17] Butcher J B, Verstraeten D, Schrauwen B, et al. Reservoir computing and extreme learning machines for non-linear time-series data analysis[J]. *Neural Networks*, 2013, 38:

- 76-89.
- [18] Malik Z K, Hussain A, Wu Q J. Multilayered echo state machine: A novel architecture and algorithm[J]. IEEE Transactions on Cybernetics, 2017, 47(4): 946-959.
- [19] Gallicchio C, Micheli A, Pedrelli L. Design of deep echo state networks[J]. Neural Networks, 2018, 108: 33-47.
- [20] Bo Y C, Wang P, Zhang X. An asynchronously deep reservoir computing for predicting chaotic time series[J]. Applied Soft Computing, 2020, 95: 106530.
- [21] Ma Q L, Shen L F, Cottrell G W. DeePr-ESN: A deep projection-encoding echo-state network[J]. Information Sciences, 2020, 511: 152-171.
- [22] 王磊, 乔俊飞, 杨翠丽, 等. 基于灵敏度分析的模块化回声状态网络修剪算法[J]. 自动化学报, 2019, 45(6): 1136-1145.
(Wang L, Qiao J F, Yang C L, et al. Pruning algorithm for modular echo state network based on sensitivity analysis[J]. Acta Anatomica Sinica, 2019, 45(6): 1136-1145.)
- [23] Rodan A, Tio P. Minimum complexity echo state network[J]. IEEE Transactions on Neural Networks, 2011, 22(1): 131-144.
- [24] Rodan A, Tiño P. Simple deterministically constructed cycle reservoirs with regular jumps[J]. Neural Computation, 2012, 24(7): 1822-1852.
- [25] 薄迎春, 乔俊飞, 张昭昭. 一种具有 small world 特性的 ESN 结构分析与设计[J]. 控制与决策, 2012, 27(3): 383-388.
(Bo Y C, Qiao J F, Zhang Z Z. Analysis and design on structure of small world property ESN[J]. Control and Decision, 2012, 27(3): 383-388.)
- [26] Kawai Y J, Park J, Asada M. A small-world topology enhances the echo state property and signal propagation in reservoir computing[J]. Neural Networks, 2019, 112: 15-23.
- [27] Bozhkov L, Koprinkova-Hristova P, Georgieva P. Learning to decode human emotions with Echo State Networks[J]. Neural Networks, 2016, 78: 112-119.
- [28] Steil J J. Online reservoir adaptation by intrinsic plasticity for backpropagation-decorrelation and echo state learning[J]. Neural Networks, 2007, 20(3): 353-364.
- [29] Schrauwen B, Wardermann M, Verstraeten D, et al. Improving reservoirs using intrinsic plasticity[J]. Neurocomputing, 2008, 71(7/8/9): 1159-1171.
- [30] Li C G. A model of neuronal intrinsic plasticity[J]. IEEE Transactions on Autonomous Mental Development, 2011, 3(4): 277-284.
- [31] Wang X J, Jin Y C, Hao K R. Echo state networks regulated by local intrinsic plasticity rules for regression[J]. Neurocomputing, 2019, 351: 111-122.
- [32] Yildiz I B, Jaeger H, Kiebel S J. Re-visiting the echo state property[J]. Neural Networks, 2012, 35: 1-9.

作者简介

那晓栋(1992—), 男, 博士生, 从事储备池计算与时间序列预测的研究, E-mail: naxiaodong@mail.dlut.edu.cn;

王嘉宁(2000—), 男, 硕士生, 从事时间序列分析和特征选择的研究, E-mail: wjndllg@mail.dlut.edu.cn;

刘墨燃(2000—), 男, 硕士生, 从事时间序列分析和特征选择的研究, E-mail: 460884060@mail.dlut.edu.cn;

任伟杰(1990—), 男, 讲师, 博士, 从事神经网络及其在复杂系统建模中的应用的研究, E-mail: renweijie90@126.com;

韩敏(1959—), 女, 教授, 博士生导师, 从事复杂工业系统建模与控制、智能技术及优化算法等研究, E-mail: minhan@dlut.edu.cn.

(责任编辑: 李君玲)

著作权转让声明

论文作者须保证所投论文为原创作品且不存在涉密和一稿多投问题,若发生侵权或泄密问题,一切责任由论文作者承担. 论文作者保证所投论文的署名无争议,若发生署名争议,责任由论文作者承担.

本刊已许可中国知网以数字化方式复制、汇编、

发行、信息网络传播本刊全文. 本刊支付的稿酬已包含中国知网著作权使用费. 所有署名作者向本刊提交文章发表之行为视为同意上述声明. 如有异议,请在投稿时说明.