

控制与决策

Control and Decision

基于子空间映射的多任务储层计算方法

张昭昭, 陈豪, 朱应钦, 余文

引用本文:

张昭昭, 陈豪, 朱应钦, 等. 基于子空间映射的多任务储层计算方法[J]. *控制与决策*, 2026, 41(1): 221-233.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2025.0465>

您可能感兴趣的其他文章

Articles you may be interested in

基于观测器的网络化多智能体预测控制

Observer-based networked multi-agent predictive control

控制与决策. 2021, 36(9): 2290-2296 <https://doi.org/10.13195/j.kzyjc.2019.1801>

一种基于池计算的宽度学习系统

A broad learning system based on reservoir computing

控制与决策. 2021, 36(9): 2203-2210 <https://doi.org/10.13195/j.kzyjc.2019.1729>

一种基于深度学习的时间序列预测方法

A time series prediction method based on deep learning

控制与决策. 2021, 36(3): 645-652 <https://doi.org/10.13195/j.kzyjc.2019.0809>

基于多维泰勒网的超前d步预测模型

d-step-ahead predictive model based on multi-dimensional Taylor network

控制与决策. 2021, 36(2): 345-354 <https://doi.org/10.13195/j.kzyjc.2019.0722>

随机变批次长度的反馈辅助PD型量化迭代学习控制

Feedback-assisted PD-type quantized iterative learning control with randomly iteration varying lengths

控制与决策. 2021, 36(10): 2569-2576 <https://doi.org/10.13195/j.kzyjc.2020.0273>

基于子空间映射的多任务储层计算方法

张昭昭¹, 陈豪^{1†}, 朱应钦², 余文²

(1. 西安科技大学 人工智能与计算机学院, 西安 710054;

2. 墨西哥国立理工学院高级研究中心 控制科学与工程系, 墨西哥 07360)

摘要: 储层计算 (RC) 作为一种高效的循环神经网络训练范式, 在处理单个时序任务时表现出色. 但是在多任务场景下, 不同任务引起的储层状态易发生混叠, 限制了其应用. 鉴于此, 提出一种基于子空间映射的多任务储层计算框架, 在传统的回声状态网络 (ESN) 的基础上, 设计并实现一种多任务回声状态网络 (MT-ESN). 该方法为每个任务分配唯一的二元映射向量, 在每个时间步, 任务对应的原始储层状态与其映射向量进行 Hadamard 积运算, 将原始的高维储层状态选择性地投影至由其映射向量所定义的低维子空间内, 从而实现不同任务储层状态轨迹在共享储层内部的结构化分离, 从根本上抑制状态混叠现象, 进而有效降低不同任务状态间的重叠度. 通过对多个混沌吸引子短期预测和真实世界多个时序任务预测的实验验证, 与标准 ESN 相比, MT-ESN 能够在单一储层网络中显著提升多任务处理的准确性和稳定性, 尤其是在长时预测中能够有效避免状态崩溃; t 分布随机邻域嵌入 (t-SNE) 可视化也验证了其储层状态分离能力, 研究还发现映射向量存在最优稀疏度. 所提出方法为在资源受限设备上实现多任务储层计算提供了有效途径.

关键词: 回声状态网络; 储层计算; 时间序列预测; 动态系统建模; 多任务学习

中图分类号: TP183

文献标志码: A

DOI: 10.13195/j.kzyjc.2025.0465

引用格式: 张昭昭, 陈豪, 朱应钦, 等. 基于子空间映射的多任务储层计算方法 [J]. 控制与决策, 2026, 41(1): 221-233.

A multi-task reservoir computing method based on subspace mapping

ZHANG Zhao-zhao¹, CHEN Hao^{1†}, ZHU Ying-qin², YU Wen²

(1. College of Artificial Intelligence and Computer Science, Xi'an University of Science and Technology, Xi'an 710054, China; 2. Department of Control Automatic, Center for Research and Advanced Studies of the National Polytechnic Institute, Mexico City 07360, Mexico)

Abstract: Reservoir computing (RC), as an efficient training paradigm for recurrent neural networks (RNNs), performs excellently on single time series tasks. However, in multi-task scenarios, reservoir states induced by different tasks tend to overlap, limiting its application. To address this issue, this paper proposes a multi-task reservoir computing framework based on subspace mapping, and building upon the traditional echo state network (ESN), designs and implements a multi-task echo state network (MT-ESN). This method assigns a unique binary mapping vector to each task. At each time step, the original reservoir state corresponding to a task undergoes a Hadamard product operation with its mapping vector. This selectively projects the original high-dimensional reservoir state into a low-dimensional subspace defined by its mapping vector, thereby achieving structured separation of different task's reservoir state trajectories within the shared reservoir. This fundamentally suppresses the phenomenon of state overlap and effectively reduces the degree of overlap between states of different tasks. Through experimental validation on short-term prediction of multiple chaotic attractors and multiple real-world time series predictions, compared to the standard ESN, the MT-ESN significantly enhances the accuracy and stability of multi-task processing within a single reservoir network. Particularly in long-term prediction, it effectively avoids state collapse. t-distributed stochastic neighbor embedding (t-SNE) visualization also confirms its reservoir state separation capability. The study further reveals that there exists an optimal sparsity for the mapping vectors. The proposed method provides an effective approach for implementing multi-functional reservoir computing on resource-constrained devices.

Keywords: echo state network; reservoir computing; time series prediction; dynamic system modeling; multi-task learning

收稿日期: 2025-05-05; 录用日期: 2025-09-29.

责任编辑: 蒲志强.

†通信作者. E-mail: chenhao@stu.xust.edu.cn.

0 引言

储层计算 (RC) 的概念源于 2001 年提出的回声状态网络 (ESN)^[1] 和 2002 年提出的液体状态机 (LSM)^[2], 是早期为了规避循环神经网络 (RNN) 在训练过程中面临的梯度消失或爆炸等难题^[3] 而提出的一种计算范式. 其核心思想是通过大规模随机连接的非线性动力学系统 (称为储层), 将低维的输入投影至某个高维空间中, 以获得更丰富的高维动态表达^[4]. 然后, 在映射后的高维空间上, 利用简单的机器学习算法 (如线性回归方法), 只训练输出层的权重, 从而大大简化了传统循环神经网络的训练过程^[5].

深度学习的发展可认为已经解决了 RNN 面临的长距离依赖建模问题, 在许多大规模序列处理任务上的性能表现已显著超越了传统的 RC 模型^[6-7]. 但是, 这并不意味着 RC 的优势已完全消失. RC 固有的训练简洁性、低推理能耗等特点, 在需要快速部署、计算资源极其受限的场景时仍然具有其价值^[8]. 同时, 通过引入参数感知等机制, 储层计算在无需模型的情况下预测带有参数变化的混沌动态方面也展现了独特的潜力^[9]. 值得关注的是, 物理储层计算 (PRC)^[10] 正成为一个活跃的研究方向. PRC 是利用各种物理载体 (如光子^[11]、忆阻器^[12]、自旋电子^[13]、软体机器人动力学^[14] 等) 的非线性响应直接实现储层功能, 这为在物联网 (IoT) 设备、可穿戴设备、嵌入式系统等资源极度受限的场景中部署智能计算提供了新的可能性^[15], 如在无人机载系统中实现状态估计和轨迹预测^[16]. 然而, 这些物理或边缘设备往往需要同时处理来自不同传感器或执行不同控制逻辑的多样化任务. 如何在单一、紧凑的储层上高效地实现多任务处理, 成为充分释放 RC 潜力, 特别是 PRC 潜力的关键技术瓶颈.

当前, 针对储层计算 (RC) 的研究主要集中于储层结构创新和超参数优化等方面. 在参数优化方面, Zhang 等^[17] 提出了一种基于行为空间的参数优化方法, 该方法不仅优化效率高, 且学习性能优于增长回声状态网络 (GESN) 等方法, 测试精度提升了 1% ~ 20%; Chouikhi 等^[18] 提出了一种基于粒子群优化 (PSO) 的回声状态网络 (ESN) 参数优化方法, 该方法在 ESN 正常训练前, 使用 PSO 对网络的部分权重 (包括输入权重、储层权重和反馈权重) 进行预训练优化; 在储层结构设计层面, 王磊等^[19] 提出了增量式正则化回声状态网络 (IRESN), 该网络由相互独立的子储备池模块构成, 利用奇异值分解方法生成子储备池, 根据问题复杂度或残差逐一添加子储备

池, 采用正则化方法训练输出权重, 具有结构紧凑、预测精度高的优势; 郭伟等^[20] 提出了一种基于强化学习的储层神经元筛选优化方法 (SC-ESN), 通过集成学习构建多个初始储备池, 利用互信息度量储层池中每个神经元对网络性能的贡献, 并结合强化学习的决策机制筛选出对网络输出有效的神经元, 从而实现了网络结构优化和性能提升. 此外, 将 RC 与其他机器学习方法融合也是一个重要的研究方向. Zhu 等^[21] 提出了一种基于误差补偿迁移学习回声状态网络 (ETL-ESN) 模型, 该模型结合了基于 ESN 的计算层和基于迁移学习的补偿层, 通过将误差自相关识别为 ESN 预测中增加方差的关键因素, 并利用误差补偿层来降低系统误差, 同时, 利用迁移学习防止误差域过拟合.

在此基础上, 近期研究进一步拓展了 RC 在生物启发学习、深度架构以及光子硬件实现等方面的创新. Zhang 等^[22] 提出了一种生物启发的数字液体状态机 (LSM), 采用基于脉冲的在线学习机制, 避免了传统岭回归所需的中转数据存储, 实现了本地突触权重更新, 并在 TI46 语音识别任务上达到了与 Sphinx-4 相当的识别精度, 为低功耗 VLSI 部署提供了新思路; Manna 等^[23] 则提出了并行化 LSM (PLSM), 引入时空读出层和语义约束, 构建了 GPU 兼容的脉冲神经网络算法, 在 Oops 数据集上优于传统深度学习与自监督模型, 显著提升了 RC 在视频理解任务中的实用性; 面对混沌时间序列预测中短期记忆容量不足的问题, Gallicchio 等^[24] 提出了深度回声状态网络 (DeepESN), 该模型通过将多个储层以分层方式堆叠, 形成了一种高效训练的深度神经网络架构, 每层储层可以捕获不同时间尺度的动态特征, 从而实现了时序数据的多尺度表示; 此外, Picco 等^[25] 探索了光子深度储层计算在语音识别中的应用, 提出了一种可物理实现的光子深度 RC 架构, 在保持低功耗特性的同时实现了高维音频信号的高速处理, 为构建高性能神经形态硬件奠定了基础.

相较于上述针对单一任务建模的研究, 如何利用单个储层高效地处理多个任务, 在储层计算领域的研究相对较为有限. 当前存在的一些研究是从动态系统的角度出发, 如 Kong 等^[26] 提出了一种基于索引的 RC 存储模型, 能够存储和检索多个复杂的动态吸引子, 包括混沌和周期性吸引子, 该模型在基于索引的存储和基于提示信号的存储下实现了多个吸引子的存储和切换; Flynn 等^[27] 利用网络的多功能性, 通过调节超参数实现在不改变 RC 网络的结构下, 根据初始条件重建了多个混沌吸引子. 与此同

时, 在物理储层计算, 特别是光子储层计算领域, 研究人员也在积极探索多任务处理方案, 如利用波分复用技术或受生物启发的功能分区来实现多个任务的并行计算^[28-29]. 这些研究未考虑因不同任务输入导致共享储层状态发生混叠或重叠的关键问题. 这种状态重叠会严重限制输出层区分任务信息的能力, 是实现单储层多任务学习必须克服的障碍.

鉴于此, 本文提出一种基于子空间映射的多任务储层计算方法, 摒弃传统储层计算采用的单一、任务无关的动态演化模式, 转而引入一种任务条件化的状态空间调控机制. 通过为不同任务分配互不相同的映射向量, 即使这些子空间共享同一个物理或虚拟的储层载体, 也能够确保不同任务的储层状态演化轨迹被有效隔离在各自专属的子空间内. 从根本上避免因储层状态在高维空间中过于接近或路径交叉而引发的混叠问题. 使得单一的储层网络能够更清晰稳定地同时对多个任务进行并行建模和预测, 为资源受限场景下的多功能智能计算提供了新的解决方案.

为便于理解, 本文所使用的主要符号及其说明如表1所示.

表1 主要符号及其说明

符号	说明
W_{in}	输入权重矩阵
W_{res}	储层内部连接权重矩阵
W_{out}	输出权重矩阵
s	任务映射向量
ρ_{ij}	任务 <i>i</i> 与任务 <i>j</i> 间的相似度
\mathcal{O}	重叠度
$\tilde{\mathcal{O}}$	映射后重叠度
C_{ij}	原始内积
\tilde{C}_{ij}	映射后内积
$C_{ij,n}$	原始内积在第 <i>n</i> 个维度上的贡献

1 相关理论

1.1 储层计算

一个典型的 RC 由输入层、储层以及输出层 3 部分组成, 如图 1 所示. 储层由大量随机连接的神经元构成, 提供丰富的动态特性和短期记忆能力. 储层的内部权重通常保持固定, 仅通过调整输出层的权重进行训练, 从而实现学习功能.

假设 $u(t)$ 为从某个未知动态系统中测量得到的时序数据, 该系统的状态为 $s(t)$, 测量函数为 h , 则 $u(t)$ 可表示为

$$u(t) = h(s(t)), \quad (1)$$

其中 h 和 $s(t)$ 均是未知的.

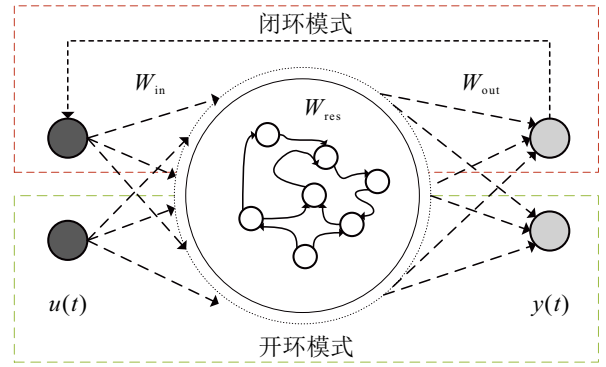


图1 储层计算及其两种模式

储层计算一般可分为监听、训练以及预测 3 个阶段. 在监听阶段, 储层计算的目标是通过外部输入驱动储层, 使其内部状态能够稳定地映射至真实动力系统的状态. 在外部输入 $u(t)$ 的驱动下, 储层的状态 $x(t)$ 可表示为

$$x(t + \tau) = g(x(t), u(t)). \quad (2)$$

其中: τ 为时间增量; g 为储层的演化函数, 通常是未知的. 储层的状态更新不仅取决于当前状态 $x(t)$, 还取决于 $u(t)$ 的驱动. 若储层的设计符合回声状态属性 (ESP), 则在监听阶段结束后, 储层计算机已与未知动力系统间发生了广义同步^[30], 使得其内部状态 $x(t)$ 与系统状态 $s(t)$ 间建立了稳定的映射关系^[31], 即存在一个函数 ϕ , 使得

$$x(t) = \phi(s(t)). \quad (3)$$

在训练阶段, 储层计算的任务是构建一个后处理函数, 使得储层状态 $x(t)$ 能够有效映射至观测空间, 从而恢复系统的测量数据 $u(t)$. 然而, 通常情况下无法直接获得该映射的解析表达式, 只能通过训练数据来学习一个近似映射 $\hat{\psi}$, 使得

$$\hat{\psi}(x(t)) \approx u(t). \quad (4)$$

在预测阶段, 储层计算的目标是利用训练阶段获得的后处理映射 ϕ 预测系统未来的动态演化. 这个阶段储层可分为开环预测和闭环预测两种模式. 开环预测适用于单步预测, 储层使用真实的输入 $u(t)$ 来更新状态并预测, 如下所示:

$$\begin{aligned} \hat{x}(t + \tau) &= g(\hat{x}(t), u(t)), \\ \hat{u}(t + \tau) &= \hat{\psi}(\hat{x}(t + \tau)). \end{aligned} \quad (5)$$

在闭环模式下, 储层使用预测的输出 $\hat{u}(t)$ 作为未来的输入来更新其状态. 这种模式适用于长期预测, 但是, 误差会随着时间的推移而累积, 其过程为

$$\begin{aligned} \hat{x}(t + \tau) &= g(\hat{x}(t), \hat{\psi}(\hat{x}(t))), \\ \hat{u}(t + \tau) &= \hat{\psi}(\hat{x}(t + \tau)). \end{aligned} \quad (6)$$

1.2 回声状态网络

回声状态网络 (ESN) 属于储层计算范式. 假设输入层、储层和输出层的神经元数量分别为 D 、 N 和 Q , 则在时间步 t 的输入信号用向量 $u(t) \in \mathbb{R}^D$ 表示, 储层的状态为 $x(t) \in \mathbb{R}^N$, 输出为 $y(t) \in \mathbb{R}^Q$.

ESN 的储层状态通过以下公式更新, 利用输入信号以及上一时间步的储层状态进行计算:

$$\begin{aligned}\tilde{x}(t) &= f(W_{\text{in}}u(t) + W_{\text{res}}x(t-1)), \\ x(t) &= (1 - \alpha)x(t-1) + \alpha\tilde{x}(t).\end{aligned}\quad (7)$$

其中: $\tilde{x}(t)$ 为储层的临时状态更新; $f(\cdot)$ 为储层的激活函数, 通常采用双曲正切函数 \tanh , 用于引入非线性动态特性; $\alpha \in (0, 1]$ 为泄露率, 用于控制储层状态的更新速度; $W_{\text{in}} \in \mathbb{R}^{N \times D}$ 和 $W_{\text{res}} \in \mathbb{R}^{N \times N}$ 分别为输入权重矩阵和储层内部的连接权重矩阵.

在 ESN 中, 通常只对输出层的权重 W_{out} 进行训练. 给定目标输出 Y_{target} , 可通过最小化预测误差计算 W_{out} , 常见的优化方法是使用岭回归^[32], 以减小共线性对于预测精度的影响. 岭回归公式为

$$W_{\text{out}} = Y_{\text{target}}X^T(XX^T + \beta I)^{-1}, \quad (8)$$

其中 β 为正则化参数. ESN 的预测输出公式为

$$y(t) = W_{\text{out}}x(t). \quad (9)$$

2 基于子空间映射的多任务储层计算

2.1 多任务储层计算

所提出多任务储层计算是用单个储层建模并预测多个任务, 考虑一个具有 N 个节点的储层计算系统, 给定 K 个不同的时序任务输入, 每个时序任务的长度为 L , 其在储层中诱导产生的状态轨迹分别表示为 $X_k = [\mathbf{x}_k(1), \mathbf{x}_k(2), \dots, \mathbf{x}_k(L)] \in \mathbb{R}^{N \times L}$, 其中 $\mathbf{x}_k(t) \in \mathbb{R}^N$ 为第 k 个时序任务在 t 时刻的储层状态向量. 将所有时序任务的储层状态矩阵拼接为 $X_{\text{all}} = [X_1, X_2, \dots, X_k] \in \mathbb{R}^{N \times KL}$.

在传统的 ESN 框架下, 输出权重矩阵 $W_{\text{out}} \in \mathbb{R}^{M \times N}$ 的训练通常采用岭回归方法. 给定与输入时序任务对应的目标输出 $Y_k \in \mathbb{R}^{M \times L}$, 将所有目标输出拼接为 $Y_{\text{all}} = [Y_1, Y_2, \dots, Y_k] \in \mathbb{R}^{M \times KL}$, 岭回归的目标是最小化以下正则化损失函数:

$$\mathcal{L}(W_{\text{out}}) = \|Y_{\text{all}} - W_{\text{out}}X_{\text{all}}\|_F^2 + \beta \|W_{\text{out}}\|_F^2. \quad (10)$$

其中: $\beta > 0$ 为正则化参数, 该优化问题的闭式解为

$$W_{\text{out}} = Y_{\text{all}}X_{\text{all}}^T(X_{\text{all}}X_{\text{all}}^T + \beta I)^{-1}. \quad (11)$$

2.2 ESN 在多任务中的储层状态重叠

当模型用于处理多个不同任务时, 不同任务在同一储层中产生的状态向量可能会在高维空间中发生重叠, 从而引发“储层状态重叠干扰”现象. 该现象

表现为不同任务的储层状态在分布上趋于相似, 导致任务间的特征表达不具备足够的区分性, 从而影响 ESN 在多任务学习中的泛化性能和预测精度. 为了定量描述不同任务储层状态间的相似程度, 引入重叠度的概念. 首先, 定义平均成对归一化重叠度, 如下所示:

$$\mathcal{O} = \frac{2}{K(K-1)} \sum_{i < j} |\rho(X_i, X_j)|. \quad (12)$$

其中: $\rho(X_i, X_j)$ 为矩阵 X_i 与矩阵 X_j 间的归一化相似度, 定义为它们 Frobenius 内积^[33] 的余弦值, 有

$$\rho(X_i, X_j) = \frac{\langle X_i, X_j \rangle_F}{\|X_i\|_F \|X_j\|_F}. \quad (13)$$

重叠度 \mathcal{O} 的取值范围为 $[0, 1]$, 值越接近 1, 不同时序任务的储层状态在 Frobenius 空间中越趋于共线, 即高度重叠; 反之, 值越接近 0, 则它们之间的差异越显著.

对于闭式解中的 $X_{\text{all}}X_{\text{all}}^T$ 部分, 对其展开, 有

$$X_{\text{all}}X_{\text{all}}^T = \sum_{i=1}^K X_iX_i^T + \sum_{i \neq j} X_iX_j^T. \quad (14)$$

当储层状态重叠度 \mathcal{O} 较高时, 意味着不同时序任务的储层状态矩阵 X_i 与 X_j 间存在显著的线性相关性, 这将直接导致交叉项 $\sum_{i \neq j} X_iX_j^T$ 在 $X_{\text{all}}X_{\text{all}}^T$ 中占更大的比重, 由于 X_i 与 X_j 的列向量趋于共线, 矩阵乘积 $X_iX_j^T$ 的秩较低. 多个低秩矩阵叠加将导致 $X_{\text{all}}X_{\text{all}}^T$ 矩阵的特征值分布不均匀, 部分特征值接近于 0, 使得 $X_{\text{all}}X_{\text{all}}^T$ 接近奇异.

对于接近奇异的矩阵, 即使加入小的正则化项 βI_N , 矩阵 $(X_{\text{all}}X_{\text{all}}^T + \beta I_N)$ 条件数也会较大, 将导致求逆运算对于微小扰动非常敏感, 这意味着通过岭回归求解得到的 W_{out} 对于训练数据中的噪声和细微变化具有较高的敏感性, 从而降低了解的稳定性.

在闭环模式下, 模型的输出将被反馈回储层作为下一步的输入. 若 W_{out} 是在高度重叠的储层状态上训练得到的, 其区分不同时序任务状态的能力较弱, 则在预测过程中产生的误差易将累积和放大.

2.3 子空间映射分离储层状态理论分析

为增强储层计算对多个时序任务的区分能力, 降低储层状态的重叠度 \mathcal{O} , 本文在传统储层网络中引入一个特定的向量 $s \in \{0, 1\}^N$, 称为映射向量 (projection vector), 如图 2 所示. 在一个 N 维的储层状态空间中, 对于时序任务 i 所产生的储层状态, 均有一个与其对应的映射向量 s_i , 储层状态通过与映射向量 s_i 做 Hadamard 乘积, 产生一个新的储层状态

$$\tilde{x}_i(t) = s_i \odot x_i(t).$$

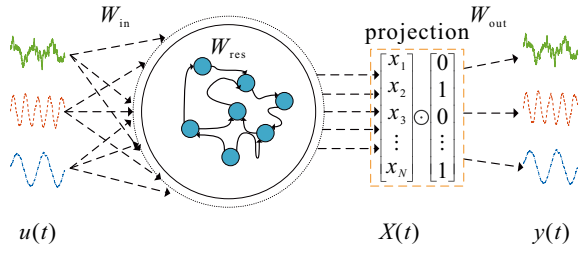


图2 多任务储层计算框架

对于任意一对不同的任务 ($k = 2$), 在引入随机映射向量后, 其重叠度 \mathcal{O} 为

$$\mathcal{O} = \frac{2}{K(K-1)} \sum_{i < j} |\rho(X_i, X_j)| = |\rho(X_i, X_j)|. \quad (15)$$

将 $\rho(X_i, X_j)$ 简写为 ρ_{ij} , 则原始重叠度定义为

$$|\rho_{ij}| = \frac{|\langle X_i, X_j \rangle_F|}{\|X_i\|_F \|X_j\|_F} = \frac{|C_{ij}|}{\|X_i\|_F \|X_j\|_F}, \quad (16)$$

其中 C_{ij} 为原始内积. 引入映射向量 $s_i, s_j \in \{0, 1\}^N$ 后, 新的状态矩阵为 $\tilde{X}_i = X_i \odot s_i$, 新的重叠度为

$$\tilde{\rho}_{ij} = \frac{|\langle \tilde{X}_i, \tilde{X}_j \rangle_F|}{\|\tilde{X}_i\|_F \|\tilde{X}_j\|_F} = \frac{|\tilde{C}_{ij}|}{\|\tilde{X}_i\|_F \|\tilde{X}_j\|_F}. \quad (17)$$

假设每个映射向量的元素 $s_{in} (n = 1, 2, \dots, N)$ 为独立同分布的随机变量, 遵循伯努利分布, 则

$$P(s_{in} = 1) = 1 - p \text{ and } P(s_{in} = 0) = p, \quad (18)$$

其中 p 为映射向量的稀疏度 (0 元素的比例). 不同任务的映射向量 s_i 与 s_j 也是相互独立的. 对于重叠度的分子项, 映射后内积 \tilde{C}_{ij} 的期望值为

$$\begin{aligned} \tilde{C}_{ij} &= \langle \tilde{X}_i, \tilde{X}_j \rangle_F = \\ &= \sum_{n=1}^N \sum_{l=1}^L (s_{in} x_{in}(l)) (s_{jn} x_{jn}(l)) = \\ &= \sum_{n=1}^N s_{in} s_{jn} C_{ij,n}, \end{aligned} \quad (19)$$

这里 $C_{ij,n}$ 为原始内积在第 n 个维度上的贡献, 取期望值为

$$\mathbb{E}[\tilde{C}_{ij}] = \mathbb{E} \left[\sum_{n=1}^N s_{in} s_{jn} C_{ij,n} \right] = \sum_{n=1}^N \mathbb{E}[s_{in} s_{jn}] C_{ij,n}. \quad (20)$$

由于 s_i 与 s_j 相互独立, 其元素也相互独立, 有

$$\mathbb{E}[s_{in} s_{jn}] = \mathbb{E}[s_{in}] \mathbb{E}[s_{jn}]. \quad (21)$$

对于伯努利变量 $s_{in} \sim \text{Bernoulli}(1-p)$, 其期望为 $\mathbb{E}[s_{in}] = 1-p$. 因此, 有

$$\mathbb{E}[s_{in} s_{jn}] = (1-p)(1-p) = (1-p)^2. \quad (22)$$

随机映射是对原始内积贡献的子采样, 它不会

改变各分量的符号. 对于大规模储层, 根据大数定律, \tilde{C}_{ij} 将集中在其期望值 $\mathbb{E}[\tilde{C}_{ij}]$, 基于此, 分子绝对值的期望为

$$\mathbb{E}[|\tilde{C}_{ij}|] \approx |\mathbb{E}[\tilde{C}_{ij}]| = (1-p)^2 |C_{ij}|. \quad (23)$$

因此, 分子项绝对值的期望值以 $(1-p)^2$ 的比例因子缩减. 对于分母项, 范数的平方的期望值为

$$\begin{aligned} \|\tilde{X}_i\|_F^2 &= \sum_{n=1}^N \sum_{l=1}^L (s_{in} x_{in}(l))^2 = \\ &= \sum_{n=1}^N s_{in}^2 \left(\sum_{l=1}^L x_{in}(l)^2 \right) = \\ &= \sum_{n=1}^N s_{in}^2 \|X_{i,n}\|_2^2, \end{aligned} \quad (24)$$

其中 $\|X_{i,n}\|_2^2$ 为原始状态矩阵第 n 列, 取期望值为

$$\mathbb{E}[\|\tilde{X}_i\|_F^2] = \sum_{n=1}^N \mathbb{E}[s_{in}^2] \|X_{i,n}\|_2^2. \quad (25)$$

对于一个取值为 0 或 1 的伯努利变量, $s_{in}^2 = s_{in}$. 因此, 有

$$\mathbb{E}[s_{in}^2] = \mathbb{E}[s_{in}] = 1-p. \quad (26)$$

将式 (26) 代入 (25), 可得到

$$\mathbb{E}[\|\tilde{X}_i\|_F^2] = \sum_{n=1}^N (1-p) \|X_{i,n}\|_2^2 = (1-p) \|X_i\|_F^2. \quad (27)$$

因此, 分母项范数的平方项期望值以 $(1-p)$ 的比例因子缩减. 映射后重叠度绝对值的期望值 $\mathbb{E}[|\tilde{\rho}_{ij}|]$ 可用期望的比率近似, 即

$$\mathbb{E}[|\tilde{\rho}_{ij}|] \approx \frac{\mathbb{E}[|\tilde{C}_{ij}|]}{\mathbb{E}[\|\tilde{X}_i\|_F] \mathbb{E}[\|\tilde{X}_j\|_F]}. \quad (28)$$

由于范数的平方项 $\|\tilde{X}_i\|_F^2 = \sum_{n=1}^N s_{in}^2 \|X_{i,n}\|_F^2$ 是 N 个独立的、有界的随机变量的和, 而储层的规模通常是一个较大的数值, 根据大数定律和中心极限定理, 当 N 很大时, 这个和的分布会高度集中于其期望值附近, 其方差相对于期望值的平方会很小, 因此, 利用琴生不等式的近似 $\mathbb{E}[\sqrt{Y}] \approx \sqrt{\mathbb{E}[Y]}$, 有

$$\begin{aligned} \mathbb{E}[\|\tilde{X}_i\|_F] &\approx \sqrt{\mathbb{E}[\|\tilde{X}_i\|_F^2]} = \sqrt{(1-p) \|X_i\|_F^2} = \\ &= \sqrt{1-p} \cdot \|X_i\|_F. \end{aligned} \quad (29)$$

将所有结果代入式 (28), 可得到

$$\begin{aligned} \mathbb{E}[|\tilde{\rho}_{ij}|] &\approx \frac{(1-p)^2 |C_{ij}|}{(\sqrt{1-p} \|X_i\|_F) (\sqrt{1-p} \|X_j\|_F)} = \\ &= \frac{(1-p)^2}{1-p} \cdot \frac{|C_{ij}|}{\|X_i\|_F \|X_j\|_F} = (1-p) |\rho_{ij}|. \end{aligned} \quad (30)$$

因此,映射后成对重叠度 ($k=2$) 绝对值的期望值降低为原始值的 $(1-p)$ 倍. 当 $k>2$ 时, 整体重叠度为

$$\tilde{\mathcal{O}}' = \frac{2}{K(K-1)} \sum_{i<j}^K |\tilde{\rho}_{ij}|. \quad (31)$$

其期望为

$$E[\tilde{\mathcal{O}}'] = E\left[\frac{2}{K(K-1)} \sum_{i<j}^K |\tilde{\rho}_{ij}|\right]. \quad (32)$$

期望算子 $E[\cdot]$ 为一个线性算子. 这意味着对于常数 a 和随机变量 X 、 Y , 有 $E[aX] = aE[X]$ 和 $E[X+Y] = E[X] + E[Y]$. 因此, 有

$$\begin{aligned} E[\tilde{\mathcal{O}}'] &= E\left[\frac{2}{K(K-1)} \sum_{i<j}^K |\tilde{\rho}_{ij}|\right] = \\ &= \frac{2}{K(K-1)} \sum_{i<j}^K E[|\tilde{\rho}_{ij}|] \approx \\ &= \frac{2}{K(K-1)} \sum_{i<j}^K (1-p) |\rho_{ij}| = \\ &= (1-p) \left[\frac{2}{K(K-1)} \sum_{i<j}^K |\rho_{ij}|\right] = \\ &= (1-p) \mathcal{O}'. \end{aligned} \quad (33)$$

因此, 映射后整体重叠度的期望值降低为原始值的 $(1-p)$ 倍.

从上述推导可以看出, 映射向量的实质是将时序任务 i 的储层状态投影至一个由 s_i 确定的子空间中. 若 s_i 中有 $N_i^{(1)}$ 个 1, 则 $\tilde{x}_i(t)$ 本质上是原始 N 维储层状态在 $N_i^{(1)}$ 维子空间上的表示. 虽然映射后的子空间维度略低于原始空间, 但是, 只要储层规模 N 足够大, 这种维度的微小降低并不会显著影响模型的表达能力. 通过引入子空间映射机制, 将不同任务对应的储层状态投影至特定的储层子空间.

2.4 回声状态属性与稳定性分析

对于储层计算, 所引入的子空间映射操作是否会破坏储层计算赖以成立的回声状态属性 (ESP) 是一个关键问题. ESP 要求: 在相同输入序列的驱动下, 系统的状态最终会“遗忘”初始状态, 收敛至由输入唯一决定的轨迹上, 即对于任意两个不同的初始状态 $x_a(0)$ 和 $x_b(0)$, 有

$$\lim_{t \rightarrow \infty} \|x_a(t) - x_b(t)\| = 0. \quad (34)$$

在 MT-ESN 模型中, 需要区分两个过程: 1) 决定系统动力学特性的储层状态演化过程; 2) 状态生成后的静态映射过程. 子空间映射 $\tilde{x}(t) = x(t) \odot s$ 是一个无记忆的逐元素操作, 它不改变决定 ESP 的动力学核心, 即储层权重矩阵 W_{res} 及其谱半径. 只要

底层 ESN 的设计满足 ESP (通常要求 $\rho(W_{\text{res}}) < 1$), 该属性便能够被完整保留.

假设底层 ESN 满足 ESP, 即 $\lim_{t \rightarrow \infty} \|x_a(t) - x_b(t)\| = 0$. 考虑映射后状态 $\tilde{x}_a(t) = x_a(t) \odot s$ 和 $\tilde{x}_b(t) = x_b(t) \odot s$, 它们差值的范数满足

$$\|\tilde{x}_a(t) - \tilde{x}_b(t)\| = \|(x_a(t) - x_b(t)) \odot s\|. \quad (35)$$

根据 Hadamard 积的性质以及映射向量 s 的元素 $s_k \in \{0, 1\}$, 对于任意向量 v 恒有 $\|v \odot s\| \leq \|v\|$. 因此, 可得到

$$0 \leq \|\tilde{x}_a(t) - \tilde{x}_b(t)\| \leq \|x_a(t) - x_b(t)\|. \quad (36)$$

由于已知 $\lim_{t \rightarrow \infty} \|x_a(t) - x_b(t)\| = 0$, 根据夹逼定理, 可得到

$$\lim_{t \rightarrow \infty} \|\tilde{x}_a(t) - \tilde{x}_b(t)\| = 0. \quad (37)$$

上述推导表明, 只要底层的 ESN 满足 ESP, 经子空间映射后的状态同样具备 ESP. 因此, 该框架在保持储层计算基本动力学特性的前提下, 实现了对多任务状态的有效分离.

为从理论上保证所提出随机映射方法的可靠性, 使用集中不等式来验证其稳定性. 为简化验证, 在引入随机映射向量后, 假设只有两个不同任务, 即两个不同任务状态矩阵的 Frobenius 内积.

内积的降低量可表示为

$$\begin{aligned} \Delta C_{ij} &= C_{ij} - \hat{C}_{ij} = \\ &= \sum_{n=1}^N (1 - s_{in} s_{jn}) \sum_{l=1}^L x_{in}(l) x_{jn}(l). \end{aligned} \quad (38)$$

假设映射向量 s_i 是通过随机过程生成的, 其中每个元素 s_{in} 以概率 p 被设置为 0, 以概率 $1-p$ 被设置为 1. 这个 p 对应实验中的映射向量稀疏度 (即 0 元素比例). 为简化分析, 假设 s_{in} 和 s_{jn} 的选择是独立的.

一个维度 n 的贡献被屏蔽 (即 $s_{in} s_{jn} = 0$) 的概率为

$$\begin{aligned} P(s_{in} s_{jn} = 0) &= \\ &= 1 - P(s_{in} = 1 \text{ and } s_{jn} = 1) = \\ &= 1 - (1-p)^2 = 2p - p^2. \end{aligned} \quad (39)$$

这是一个非 0 的概率, 只要稀疏度 $p > 0$.

令 $Z_n = (1 - s_{in} s_{jn}) \sum_{l=1}^L x_{in}(l) x_{jn}(l)$ 为第 n 个维度对内积降低量的贡献, 这是一个独立的随机变量,

$$\Delta C_{ij} = \sum_{n=1}^N Z_n.$$

首先, 计算期望降低量, 如下所示:

$$\begin{aligned}
E[\Delta C_{ij}] &= E\left[\sum_{n=1}^N Z_n\right] = \\
&\sum_{n=1}^N E[1 - s_{in}s_{jn}]\left(\sum_{l=1}^L x_{in}(l)x_{jn}(l)\right) = \\
(2p - p^2) &\sum_{n=1}^N \left(\sum_{l=1}^L x_{in}(l)x_{jn}(l)\right) = \\
(2p - p^2) &C_{ij}. \quad (40)
\end{aligned}$$

这个结果表明, 内积的降低量与原始内积 C_{ij} 和稀疏度 p 直接相关.

为了验证这种降低效果是高度可靠的, 使用霍夫丁不等式. 该不等式为 一组有界独立随机变量的和偏离其期望值的概率提供了一个上界.

随机变量 Z_n 是有界的. 假设储层状态 $x(t)$ 的范数有界, 对于平均每个维度的降低量 $\frac{\Delta C_{ij}}{N}$, 有

$$|C_{ij,n}| = \left| \sum_{l=1}^L x_{in}(l)x_{jn}(l) \right| \leq C_{\max}, \quad (41)$$

对于所有 n 均成立. 因此, Z_n 的取值范围长度小于或等于 C_{\max} .

根据霍夫丁不等式, 对于一组有界独立随机变量的均值, 其偏离期望值的概率满足以下上界 (对于任意 $\epsilon' > 0$):

$$P\left(\left|\frac{\Delta C_{ij}}{N} - E\left[\frac{\Delta C_{ij}}{N}\right]\right| \geq \epsilon\right) \leq 2 \exp\left(-\frac{2N(\epsilon')^2}{C_{\max}^2}\right), \quad (42)$$

其中 ϵ 是为平均降低量设定的一个很小的容忍偏差. 从结果可以看出, 不等式 (42) 的右侧随着 N 的增加呈指数级下降. 这意味着对于一个典型规模的储层 (即 N 足够大), 随机映射导致的重叠度降低效果偏离其期望值的概率极小. 稀疏度 p 主要决定期望降低量的幅度, 并不直接影响概率尾界的指数衰减速率; 而序列长度 L 通过影响上界 C_{\max} 间接影响尾界. 更长的序列将导致 C_{\max} 增大, 从而使得概率上界变得相对宽松. 因此, 在处理更长的时间序列时, 需要更大的储层规模 N 来维持同等水平的稳定性.

2.5 算法流程

为了更加清晰地阐述多任务储层计算方法, 本节将详细描述其具体算法流程. 假设共有 K 个时序任务. 每个时序任务的长度为 L , 储层的节点设置为 N , 子空间的维度设置为 M .

step 1: 随机构建一个回声状态网络 (ESN), 储层的节点数为 N .

step 2: 对于每个时序任务, 随机生成一个对应的映射向量 $s_i (i = 1, 2, \dots, K)$, 每个映射向量 s_i 包

含 M 个 1 和 $N - M$ 个 0, 且所有映射向量互不相同.

step 3: 将 K 个时序任务依次输入至储层. 对于每个任务, 首先, 执行一个独立的 washout 阶段 (前 100 步), 即使用其初始的一段数据驱动储层来消除初始状态的影响, 此阶段的状态不用于训练; 然后, 对于有效数据部分, 由式 (7) 得到的储层状态向量 $x_i(t)$ 与对应的映射向量 s_i 进行 Hadamard 乘积, 可得到投影后的储层状态 $\tilde{x}_i(t) = x_i(t) \odot s_i$. 在处理下一个任务前, 储层状态将被重置.

step 4: 将所有时序任务的投影储层状态按照输入顺序拼接为矩阵 X_{all} , 并将其对应的目标输出拼接为矩阵 Y_{all} .

step 5: 使用拼接后的储层状态矩阵 X_{all} 和目标输出矩阵 Y_{all} , 由式 (8) 求解输出权重矩阵 W_{out} .

step 6: 对于第 i 个任务, 预测公式为 $y_i(t+1) = W_{\text{out}}\tilde{x}_i(t)$.

在 step 2 生成映射向量时, 已生成的向量被存储于一个哈希集合中, 其唯一性检查的平均时间复杂度接近 $O(N)$, 且由于映射向量的总可能空间为 $C(N, M)$, 随机生成时发生碰撞的概率极低. 因此, 为全部 K 个任务生成唯一的映射向量, 其总时间复杂度近似为 $O(K \cdot N)$.

上述流程描述的是一种离线的批处理实验模式. 在需要处理交替或并发任务流的真实在线部署中, 频繁重置和 washout 储层状态是低效的. 更优的策略是为每个任务 i 维护一个状态缓存 $x_i(t)$. 当属于任务 i 的新数据到达时, 系统加载其缓存状态, 执行一步更新, 然后将新状态存回缓存. 这样, 单一的储层便可通过分时复用和状态缓存, 高效地并行追踪多个任务的动态演化.

3 实验分析

多任务储层计算是在 ESN 的基础上实现的, 为方便后续描述, 将该实现称为 MT-ESN (multi-task echo state networks). 所有数据集的总长度为 $L_{\text{washout}} + L_{\text{train}} + L_{\text{test}}$ 的样本. 其中: 前 L_{washout} 步用于洗出初始储层状态以减小初始条件影响, L_{train} 步作为训练集, L_{test} 步作为测试集. 所有实验均采用闭环多步预测方式, 并通过均方根误差 (RMSE) 来检验模型性能, 其计算公式如下所示:

$$\text{RMSE} = \sqrt{\frac{1}{L_{\text{test}}} \sum_{l=1}^{L_{\text{test}}} (y - y_{\text{pred}})^2}. \quad (43)$$

3.1 基准实验

本研究所用的基准数据集由 6 个典型非线性动力学吸引子系统生成的时序任务构成: 包括

Lissajous 系统和 Sprott 系统 (周期性), 以及 Lorenz 系统、Rössler 系统、食物链系统和 Hindmarsh-Rose 神经元系统 (混沌). 所有序列均在 Matlab 2024 中以

表 2 中各自指定的时间步长 Δt 进行仿真. 为保证评估结果的可靠性, 每个系统在相同参数配置下重复实验 10 次取其平均值.

表2 Lissajous 系统、Sprott 系统、Lorenz 系统、Rössler 系统、食物链系统以及神经元系统

Lissajous系统	Sprott系统	Lorenz系统
$u_x(t) = \sin\left(\frac{\pi t}{100}\right)$	$\dot{u}_x = 0.3u_x + u_z$	$\dot{u}_x = 10(u_y - u_x)$
$u_y(t) = \sin\left(\frac{3\pi t}{100} + \frac{\pi}{2}\right)$	$\dot{u}_y = u_x u_z - u_y$	$\dot{u}_y = u_x(28 - u_z) - u_y$
$u_z(t) = \sin\left(\frac{\pi t}{20}\right)$	$\dot{u}_z = -u_x + u_y$	$\dot{u}_z = u_x u_y - \frac{8}{3}u_z$
$\Delta t = 1$	$\Delta t = 0.015$	$\Delta t = 0.03$
Rössler系统	食物链系统	神经元系统
$\dot{u}_x = -u_y - u_z$	$\dot{u}_x = u_x - \frac{0.2u_x u_y}{1 + 0.05u_x}$	$\dot{u}_x = u_y - u_x^3 + 3u_x^2 - u_z + 3.25$
$\dot{u}_y = u_x + 0.2u_y$	$\dot{u}_y = -u_y + \frac{0.2u_x u_y}{1 + 0.05u_x} - u_y u_z$	$\dot{u}_y = 1 - 5u_x^2 - u_y$
$\dot{u}_z = u_z(u_x - 5.7) + 0.2$	$\dot{u}_z = -10(u_z - 0.006) + u_y u_z$	$\dot{u}_z = 0.006\left(4\left(u_x + \frac{8}{5}\right) - u_z\right)$
$\Delta t = 0.1$	$\Delta t = 0.002$	$\Delta t = 0.06$

3.1.1 6个吸引子的短期预测

为验证在单一储层结构下仅调整输出层的 MT-ESN 模型能否有效学习并预测多种具有不同动力学特性的复杂系统, 本文选取上述 6 个代表性吸引子任务. 结果如图 3(a) 所示, MT-ESN 能够在同一储层网络中精确重构所有系统的全局动力学轨迹, 并有效捕捉各系统的局部动态特征, 展现出优异的多任务学习和预测能力; 相比之下, 传统 ESN (如图 3(b) 所示) 在相同设置下性能显著退化, 其预测轨迹随着时间步迭代不断偏离真实轨迹, 累积误差放大最终导致储层状态崩溃, 验证了第 2.2 节中所述的状态重叠会削弱传统 ESN 多任务预测稳定性的理论推导.

为进一步探究 MT-ESN 与传统 ESN 性能差异的内在机制, 本实验利用 t-分布随机邻域嵌入 (t-SNE) 对两种模型在处理 6 个吸引子任务时所激活的储层状态进行降维可视化. 如图 4(a) 所示: 引入映射向量机制后, MT-ESN 能够将不同动态系统的储层状态映射至相对分离的子空间, 形成清晰可辨的簇状结构, 直观验证了所提出子空间映射方法在区分任务状态方面的显著有效性, 为输出层权重求解提供了更具区分度的内部表征; 而传统 ESN (如图 4(b) 所示), 降维后各任务的储层状态相互混杂, 难以区分, 佐证了状态重叠是其多任务性能不佳的关键因素.

3.1.2 与其他模型对比

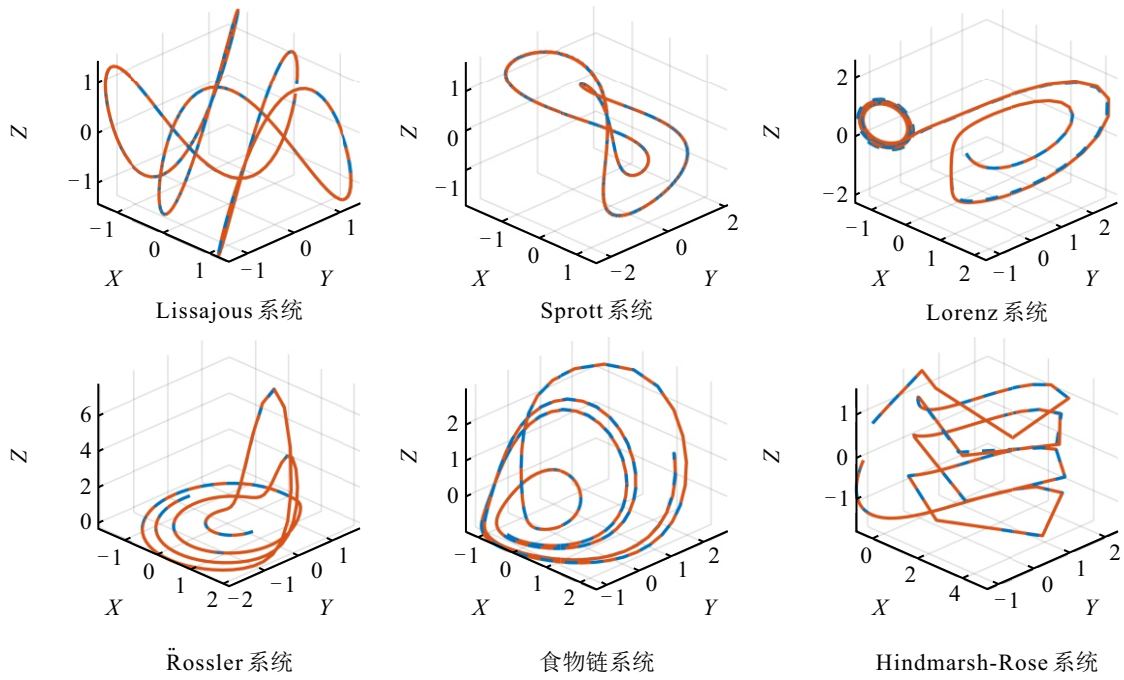
本节旨在验证所提出 MT-ESN 模型的性能. 现有两种基于单一储层的多任务 RC 方法: Kong 等^[26]

提出的基于索引的 IB-ESN (index based echo state network) 以及 Flynn 等^[27] 提出的基于多功能性的 MF-ESN (multifunctional echo state network). 首先, 将 MT-ESN 与 IB-ESN 在 6 个混沌吸引子重构任务上的表现进行比较, 结果如表 3 所示. 由表 3 可见: MT-ESN 在所有 6 个系统上的预测误差均显著低于 IB-ESN, 尤其是在 Lorenz 系统和神经元系统上, RMSE 分别降低了约 52.8% 和 82.2%, 充分表明 MT-ESN 能够在单一储层中更高效地处理多任务并保持优异的预测精度.

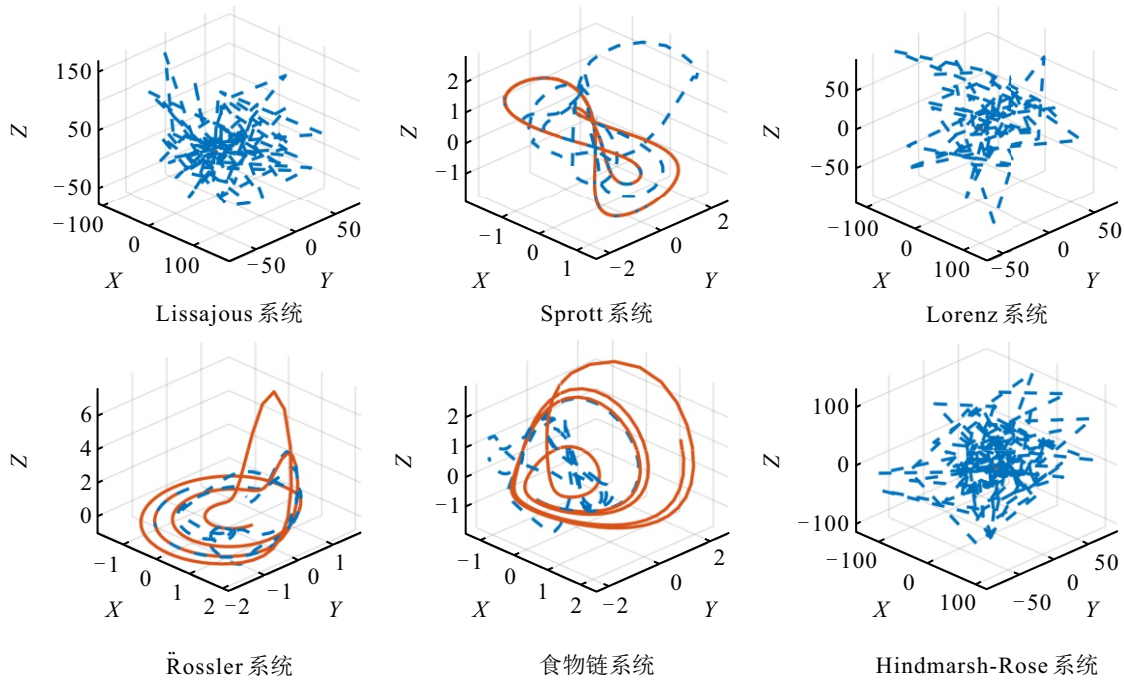
如表 4 所示, 进一步将 MT-ESN 与基于多功能性的 MF-ESN 在仅包含 Lorenz 系统和 Halvorsen 系统两项任务的多任务重构场景下进行对比. MF-ESN 通过将不同吸引子的输入信号按照预设权重混合后一并输入储层以实现多功能性, 然而, 这种混合机制在任务间干扰较小时仍然能够保持一定性能, 但是, 随着任务数量增加干扰效应会明显加剧. 为保证对比公平, MF-ESN 的实验设置采用其原始文献中的最优配置, 且两种模型均使用相同规模的储层节点. 结果表明, 即便在 MF-ESN 相对擅长的双任务场景下, MT-ESN 的预测误差依然更低, 在 Lorenz 系统上 RMSE 降低约 10.22%, 在 Halvorsen 系统上 RMSE 降低约 29.89%. 该结果进一步表明, MT-ESN 所依赖的子空间映射机制能够在多任务环境中有效抑制不同动态系统间的干扰, 实现更精确的多任务预测.

3.2 映射向量对模型的影响

映射向量是多任务储层计算方法的核心, 为了



(a) MT-ESN对6个吸引子的预测轨迹



(b) ESN对6个吸引子的预测轨迹

— 真实轨迹 - - - 预测轨迹

图3 6个吸引子的短期预测

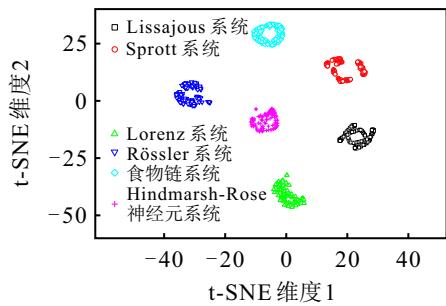
探究映射向量对所提出模型的影响, 本节在第 3.1 节实验的基础上, 从两个方面对映射向量进行研究.

3.2.1 映射向量稀疏度的影响

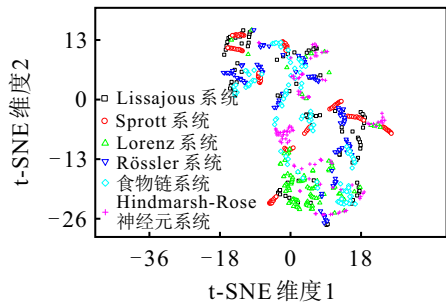
映射向量的稀疏度, 即其中包含的 0 元素比例, 是决定子空间投影特性的一个关键参数, 0 元素占比越大, 映射向量的稀疏度越高. 本节对 MT-ESN 中映射向量的稀疏度进行系统性评估. 第 3.1 节的实验将高维储层状态投影至一个维度为其 90% 的不同子空间中. 映射向量指明了原始储层状态向量中哪些

分量被保留以构成目标子空间的基准, 保留的分量占总数的 90%, 而被丢弃的分量占 10%. 本实验延续第 3.1 节实验的设置, 但是, 将被忽略分量的比例从 0% 系统地增加至 50%. 当此比例为 0% 时, 意味着所有原始分量均被保留, 模型退化为标准 ESN.

图 5 为映射向量稀疏度的影响. 如图 5(a) 所示, 随着映射向量中稀疏度的增大, MT-ESN 的总体误差呈现先下降再上升的趋势. 当 0 比例从 0% 开始增加, 预测误差持续减小, 性能不断提升, 到达



(a) MT-ESN 储层状态t-SNE 可视化



(b) ESNN 储层状态t-SNE 可视化

图4 储层状态的 t-SNE 可视化

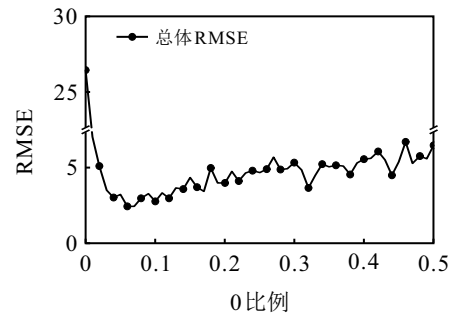
表3 MT-ESN 与 IB-ESN 对多个吸引子的短期预测

吸引子名称	IB-ESN	MT-ESN
Lissajous系统	0.013 103	0.002 228
Sprott系统	0.003 322	0.000 478
Lorenz系统	0.300 384	0.141 685
Rössler系统	0.007 779	0.001 873
食物链系统	0.017 771	0.004 886
神经元系统	0.203 733	0.036 349

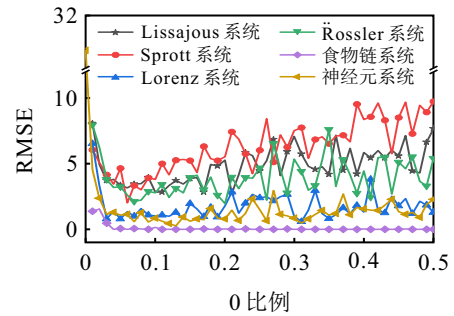
表4 MF-ESN 和 MT-ESN 重构两个吸引子

吸引子名称	MF-ESN	MT-ESN
Lorenz系统	0.464 667	0.417 179
Halvorsen系统	0.594 747	0.416 996

5%左右时误差达到最小值,继续提升 θ 比例,误差逐渐增大,各任务的误差也与总体误差趋势相似(如图5(b)所示).这表明引入一定稀疏度,可以过滤掉部分冗余的储层维度,使得投影后的子空间更具有任务特异性.为验证该结论的普适性,在不同任务数量(2个、4个以及6个任务)下均进行测试,并观察到了高度一致的“U”型误差曲线,这表明所发现的最优区间具有良好的鲁棒性.虽然高稀疏度可能会进一步增强子空间的分离性,但是,过高的稀疏度会导致在子空间投影中丢失过多的有效信息.因此,映射向量的稀疏度并非越高越好,存在一个最优的 θ 元素比例区间(在本实验条件下约为 2% ~ 10%),在该比例下,模型能够在信息保留与子空间有效划分之间达到最佳平衡,从而实现最优的多任务学习和预测性能.



(a) 总体RMSE随 θ 比例的变化



(b) 不同系统RMSE随 θ 比例的变化

图5 映射向量稀疏度的影响

综合上述实验结果和理论分析,映射向量稀疏度的选取,其本质是在状态分离度与信息保真度间的一种权衡.根据式(33)的理论推导,提高 θ 比例能够有效降低任务间储层状态的期望重叠度,增强子空间的分离性,从而抑制任务间干扰.但是,过高的 θ 比例意味着从原始储层状态中舍弃了更多的信息维度,可能会削弱模型对单一任务复杂动态的表征能力,导致信息损失.因此,在实际应用中, θ 比例应被视为一个关键的超参数.其最佳选取准则是通过在独立的验证集上进行评估,寻找能够使得模型泛化误差最小化的值.

3.2.2 映射向量中 θ 元素分布模式的影响

除稀疏度的影响, θ 元素在映射向量中的分布也是一个重要因素.在先前的实验中,映射向量中 θ 元素的位置被设定为随机选取,这与储层内部随机连接生成的理念一致.但是,随机分布未必是最优解,特定优化算法往往会引入额外的计算复杂度,因此,本节提出4种简单的分布模式.具体如下:随机分布,为基准参照;连续块状分布,将所有 θ 元素均安排在同一连续的索引区间;交替分布,各任务 θ 元素间交替分布;任务特定分布,为每个任务分配特定的区间, θ 元素在区间内随机排列.

实验设置与第3.2.1节的实验保持一致, θ 比例设置为映射向量的2%.图6为 θ 元素分布模式的影响.由图6实验结果可见, θ 元素的分布模式对于模型预测精度和有一定的影响.特定分布模式导致模型性能显著下降且不稳定.随机分布、连续分布与交

替分布的平均性能相近且较优,但是,其可能会产生少量性能较差的离群结果,其中交替分布的离群现象相对更突出.因此,在无特定先验知识指导时,除特定分布模式,其余3种分布均可作为选择.

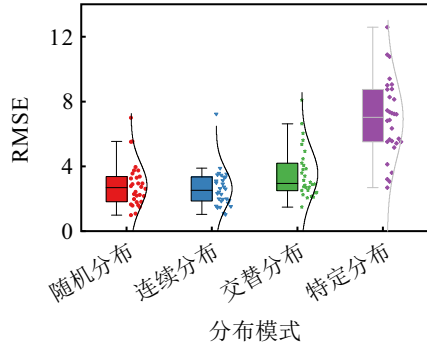


图6 0元素分布模式的影响

3.3 4个真实时序任务实验

为验证 MT-ESN 的实际效果,本实验使用一个包含4个真实时序任务的数据集.其中:前两个时序

任务为某市两个路口每小时通过的车辆数;后两个时序任务为西安市2019年和2020年气体污染物CO的质量浓度^[34],数据采集频率为每5min一次.数据采集点是西安市雁塔区科技一路15号关中平原生态环境变化与综合治理野外科学观测研究站高新子观测场内(经纬度:108.894°N,34.234°E).对于其中的缺失值,采用样条插值法进行填补,以确保数据的完整性和连续性.为避免储层的激活状态在动态过程中出现过度放大或过度收缩的现象,影响网络的稳定性,将所有数据归一化至[0,1]之间.

实验取 $L_{train} = 6000$, $L_{test} = 800$,考虑到真实世界的时序数据往往充满了噪声和不确定性,直接进行一次性的长期预测是非常困难的.因此,本实验采用递归多步预测的策略.首先,模型基于训练数据预测未来 h 步(本实验中分别设置为 $h = 5$ 和 $h = 10$) 的输出;然后,利用上一轮预测的终点作为新的起点,更新储层状态,并继续预测接下来的步,如此

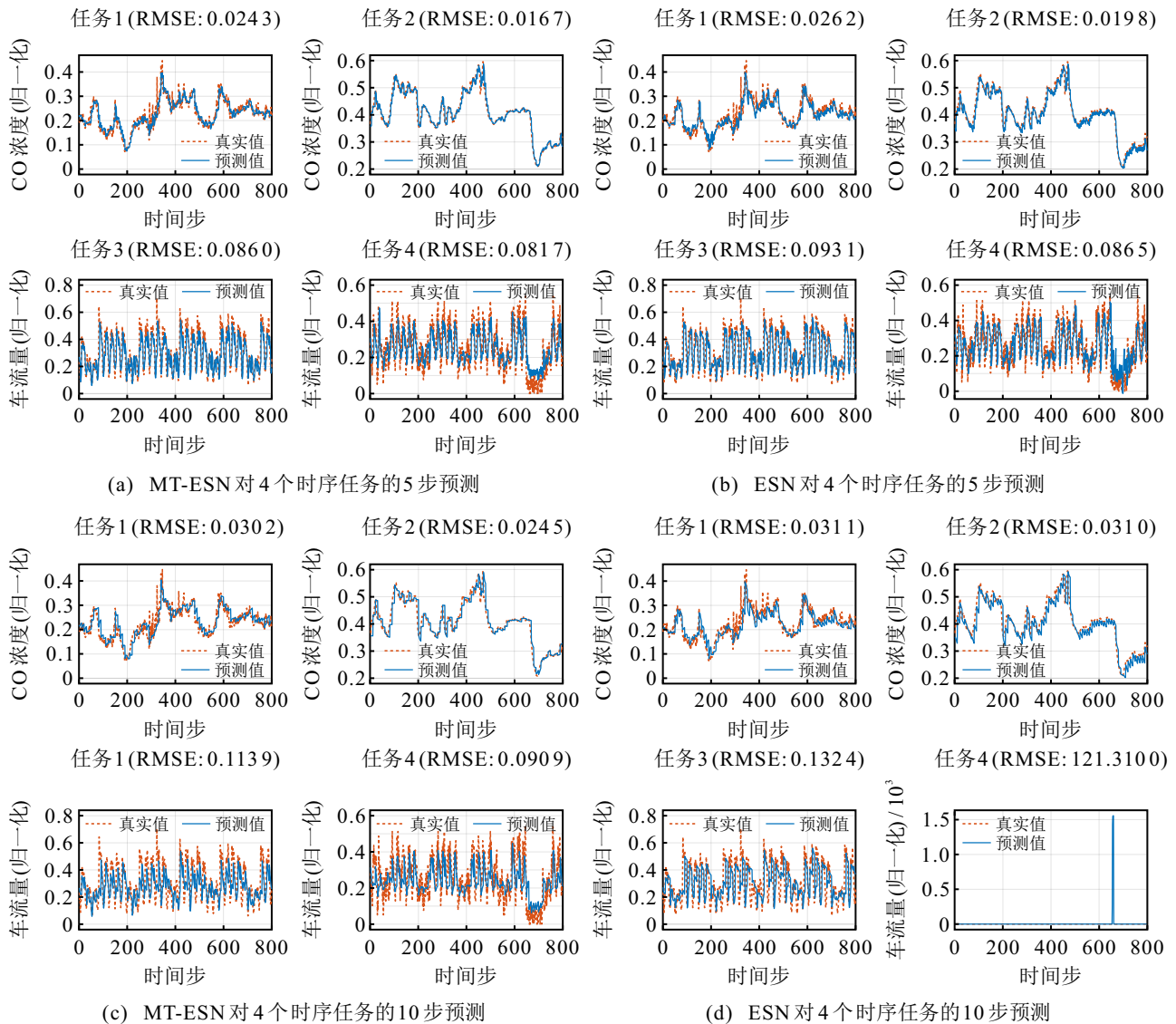


图7 4个真实时序任务的多步预测

循环往复,直至预测完所需步;最后,将各预测片段拼接起来,与真实的测试数据进行对比。

将预测步长设置为 5 时(如图 7(a)和图 7(b)所示),无论是 MT-ESN 还是 ESN,均表现出了良好的预测性能。但是,通过比较二者的均方根误差(RMSE)可以发现,MT-ESN 的误差略低于 ESN,这表明引入映射向量后,MT-ESN 的多任务预测效果得到了一定程度的提升。进一步仔细观察两种模型在任务 2 上的预测情况,会发现 ESN 的预测值出现了明显的锯齿状波动。这种剧烈的波动表明模型的输出值不够平稳,预测精度受到了影响,表明储层的状态已不再稳定。相比之下,MT-ESN 在序列 2 的预测中表现出了更加平滑稳定的输出。

图 7 为 4 个真实时序任务的多步预测。将预测步长设置为 10 时,两种模型在稳定性方面的差异变得更为显著。如图 7(d)所示:传统 ESN 对任务 2 的预测中,锯齿状波动变得更加剧烈和频繁,这种不稳定的输出表明,在长期预测任务中,传统的 ESN 可能会面临储层状态崩溃的风险。为进一步验证这一问题,对未引入映射向量的储层计算机进行多次重复实验。传统 ESN 对任务 4 的预测中,出现了明显的崩溃现象,完全偏离了实际的时序任务走势。储层状态的崩溃不仅影响了预测精度,也使得网络无法继续有效地学习和更新。相比之下,如图 7(c)所示,MT-ESN 在步长为 10 的预测中表现出了良好的稳定性。即使在长期预测任务中,其输出的预测值依然保持平稳,没有出现剧烈的波动或崩溃现象。这表明引入映射向量有效地增强了储层计算的鲁棒性,使其能够在复杂的多时序任务预测任务中维持稳定的性能。

4 结论

本文提出并验证了一种基于子空间映射的多任务储层计算方法,旨在克服标准储层计算在处理多任务时因内部状态重叠而导致的性能瓶颈。研究表明,通过为不同任务引入独特的映射向量并将储层状态投影至相应子空间,能够显著增强单一储层网络区分和处理多个异构时序任务的能力。实验结果表明了该方法在复杂动态系统短期预测和真实世界多变量时序任务预测方面的优越性,相较于传统 ESN,MT-ESN 不仅提高了预测精度,更重要的是大幅度提升了模型的稳定性和鲁棒性,避免了在多任务干扰下的状态崩溃。此外,对映射向量稀疏度和结构模式的初步探索揭示了其对于模型性能存在的影响,并指出了优化的可能性。所提出子空间映射策略

为储层计算的多任务扩展提供了简洁而有效的解决方案,对推动其在需要集成多种功能的智能传感、边缘计算等领域的实际应用具有重要意义。未来的工作将集中于映射向量的自适应优化、任务间关联性利用以及物理储层实现等方面。

参考文献 (References)

- [1] Jaeger H. The “echo state” approach to analysing and training recurrent neural networks-with an erratum note[R]. Bonn: German National Research Center for Information Technology, 2001.
- [2] Maass W, Natschläger T, Markram H. Real-time computing without stable states: A new framework for neural computation based on perturbations[J]. *Neural Computation*, 2002, 14(11): 2531-2560.
- [3] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. *IEEE Transactions on Neural Networks*, 1994, 5(2): 157-166.
- [4] Lukoševičius M, Jaeger H. Reservoir computing approaches to recurrent neural network training[J]. *Computer Science Review*, 2009, 3(3): 127-149.
- [5] Jaeger H, Haas H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication[J]. *Science*, 2004, 304(5667): 78-80.
- [6] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]. *Advances in Neural Information Processing Systems 30*. Long Beach, 2017: 5998-6008.
- [7] Schmidhuber J. Deep learning in neural networks: An overview[J]. *Neural Networks*, 2015, 61: 85-117.
- [8] Antonik P, Dupont F, Hermans M, et al. Online training of an opto-electronic reservoir computer applied to real-time channel equalization[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(11): 2686-2698.
- [9] Guo J M, Du Y, Luo H B, et al. Model-free prediction of chaotic dynamics with parameter-aware reservoir computing[J]. *Chinese Physics B*, 2025, 34(4): 040505.
- [10] Tanaka G, Yamane T, Héroux J B, et al. Recent advances in physical reservoir computing: A review[J]. *Neural Networks*, 2019, 115: 100-123.
- [11] Vandoorne K, Mechet P, Van V T, et al. Experimental demonstration of reservoir computing on a silicon photonics chip[J]. *Nature Communications*, 2014, 5: 3541.
- [12] Du C, Cai F X, Zidan M A, et al. Reservoir computing using dynamic memristors for temporal information processing[J]. *Nature Communications*, 2017, 8(1): 2204.
- [13] Torrejon J, Riou M, Araujo F A, et al. Neuromorphic computing with nanoscale spintronic oscillators[J]. *Nature*, 2017, 547(7664): 428-431.
- [14] Nakajima K, Hauser H, Li T, et al. Exploiting the dynamics of soft materials for machine learning[J]. *Soft Robotics*, 2018, 5(3): 339-347.
- [15] Wu X S, Wang S C, Huang W, et al. Wearable in-sensor reservoir computing using optoelectronic polymers with through-space charge-transport characteristics for multi-task learning[J]. *Nature Communications*, 2023, 14: 468.

- [16] Souli N, Kardaras P, Grigoriou Y, et al. Onboard UAV state estimation and trajectory prediction using multi-task reservoir computing[C]. International Conference on Unmanned Aircraft Systems. Charlotte, 2025: 75-82.
- [17] Zhang Z Z, Zhu Y Q, Wang X H, et al. Optimal echo state network parameters based on behavioural spaces[J]. *Neurocomputing*, 2022, 503: 299-313.
- [18] Chouikhi N, Ammar B, Rokbani N, et al. PSO-based analysis of echo state network parameters for time series forecasting[J]. *Applied Soft Computing*, 2017, 55: 211-225.
- [19] 王磊, 苏中, 乔俊飞, 等. 基于增量式学习的正则化回声状态网络[J]. 控制与决策, 2022, 37(3): 661-668. (Wang L, Su Z, Qiao J F, et al. Design of incremental regularized echo state network[J]. *Control and Decision*, 2022, 37(3): 661-668.)
- [20] 郭伟, 姚欢, 张昭昭, 等. 基于强化学习的储层神经元筛选优化方法[J]. 控制与决策, 2024, 39(9): 2876-2884. (Guo W, Yao H, Zhang Z Z, et al. Optimization method for reservoir neuron selection based on reinforcement learning[J]. *Control and Decision*, 2024, 39(9): 2876-2884.)
- [21] Zhu Y Q, Liu Y, Wang N, et al. Real-time error compensation transfer learning with echo state networks for enhanced wind power prediction[J]. *Applied Energy*, 2025, 379: 124893.
- [22] Zhang Y, Li P, Jin Y, et al. A digital liquid state machine with biologically inspired learning and its application to speech recognition[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 26(11): 2635-2649.
- [23] Manna S, Das D, Bhattacharya S, et al. PLSM: A parallelized liquid state machine for unintentional action detection[J]. *IEEE Transactions on Emerging Topics in Computing*, 2023, 11(2): 474-484.
- [24] Gallicchio C, Micheli A, Pedrelli L. Deep reservoir computing: A critical experimental analysis[J]. *Neurocomputing*, 2017, 268: 87-99.
- [25] Picco E, Lupo A, Massar S. Deep photonic reservoir computer for speech recognition[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2025, 36(4): 7606-7614.
- [26] Kong L W, Brewer G A, Lai Y C. Reservoir-computing based associative memory and itinerancy for complex dynamical attractors[J]. *Nature Communications*, 2024, 15(1): 4840.
- [27] Flynn A, Tsachouridis V A, Amann A. Multifunctionality in a reservoir computer[J]. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2021, 31(1): 013125.
- [28] Castro B J G, Peucheret C, Zibar D, et al. Multi-task photonic reservoir computing: Wavelength division multiplexing for parallel computing with a silicon microring resonator[J]. *Advanced Optical Technologies*, 2024, 13: 1471239.
- [29] Zhang L Y, Hong W J, Li S S, et al. Multifunctional photonic reservoir computing based on semiconductor laser with optical feedback[J]. *IEEE Photonics Journal*, 2025, 17(1): 1500407.
- [30] Kocarev L, Parlitz U. Generalized synchronization, predictability, and equivalence of unidirectionally coupled dynamical systems[J]. *Physical Review Letters*, 1996, 76(11): 1816-1819.
- [31] 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.
- [32] 史志伟, 韩敏. ESN岭回归学习算法及混沌时间序列预测[J]. 控制与决策, 2007, 22(3): 258-261. (Shi Z W, Han M. Ridge regression learning in ESN for chaotic time series prediction[J]. *Control and Decision*, 2007, 22(3): 258-261.)
- [33] 张贤达. 矩阵分析与应用[M]. 第2版. 北京: 清华大学出版社, 2013. (Zhang X D. Matrix analysis and applications[M]. The 2nd edition. Beijing: Tsinghua University Press, 2013.)
- [34] 冉伟康, 周家茂, 方焱, 等. 2019-2020年西安城市PM_{2.5}、气体污染物及气象数据集[J]. 中国科学数据, 2024, 9(1): 110-121. (Ran W K, Zhou J M, Fang Y, et al. A dataset of PM_{2.5}, gaseous pollutants, and meteorological data in Xi'an City from 2019 to 2020. *China Scientific Data*, 2024, 9(1): 110-121.)

作者简介

张昭昭 (1973-), 男, 副教授, 博士, 主要研究方向为智能信息处理、神经网络结构优化设计, E-mail: zzzhao123@126.com;

陈豪 (1997-), 男, 硕士生, 主要研究方向为复杂系统建模、神经网络结构设计, E-mail: chen hao@stu.xust.edu.cn;

朱应钦 (1997-), 男, 博士生, 主要研究方向为池网络结构优化设计, E-mail: yzhu@ctrl.cinvestav.mx;

余文 (1965-), 男, 教授, 博士, 博士生导师, 主要研究方向为神经网络、自适应控制、模糊系统, E-mail: yuw@ctrl.cinvestav.mx.