

文章编号:1671-6833(2024)04-0038-08

基于去趋势多重互相关的深度回声状态网络剪枝算法

孙晓川^{1,2}, 王宇^{1,2}, 李莹琦^{1,2}, 黄天宇^{1,2}

(1. 华北理工大学 人工智能学院, 河北 唐山 063210; 2. 河北省工业智能感知重点实验室, 河北 唐山 063210)

摘要: 针对储备池中存在的冗余结构导致深度回声状态网络预测精度不佳的问题, 提出了一种基于去趋势多重互相关的深度回声状态网络剪枝算法。首先, 根据去趋势协方差函数和去趋势方差函数, 依次计算所选储备池中每 2 个神经元之间的去趋势互相关系数, 构建去趋势互相关矩阵, 基于该矩阵评估该储备池中所选神经元与所有剩余神经元之间的去趋势多重互相关性。其次, 依次删除每个储备池中高相关性神经元到输出层的连接, 从而去除网络中的冗余结构。最后, 通过最小二乘回归重新训练剪枝后的网络, 以获得最优的深度回声状态网络拓扑结构。仿真结果表明: 经过所提算法优化后的深度回声状态网络在 Mackey-Glass 时间序列上的预测精度和记忆能力分别提高了 89.80% 和 30.93%, 在 Call 时间序列上的预测精度和记忆能力分别提高了 14.34% 和 0.10%。

关键词: 深度回声状态网络; 结构优化; 剪枝; 去趋势多重互相关; 时间序列预测

中图分类号: TP183

文献标志码: A

doi: 10.13705/j.issn.1671-6833.2024.04.005

深度回声状态网络 (deep echo state network, DESN) 作为一种深度学习背景下构建的多储备池计算框架, 能够在不断增加的抽象层次上开发时间信息的高级特征表示, 从而以自然的方式实现具有多时间尺度特征的时间序列建模任务。经典的 DESN 采用最小二乘回归训练输出权重, 这种训练方式避免了传统循环神经网络在采用反向传播算法训练时出现的收敛速度慢和极易陷入局部最优值等问题。简单的训练方式和强大的非线性特征映射能力使 DESN 在诸多领域中发挥了重要作用, 如时间序列预测、语音建模、信号处理和异常识别等。尽管有上述优势, 但当前对于 DESN 的结构设计仍然缺乏足够的理论指导, 尤其是如何为特定任务选择合适的储备池规模仍然面临着严峻的挑战^[1]。在这种情况下, 设计人员倾向于使用更大规模的储备池以期使网络获得令人满意的学习性能。然而, 这往往会导致储备池中存在大量的冗余神经元, 为储备池引入了更多的噪声, 进而影响网络的泛化性能。

剪枝是一种帮助剔除神经网络冗余组件的有效

策略^[2]。近年来, 已有学者将剪枝策略引入到 DESN 的建模中。例如, Shen 等^[3] 基于储备池神经元之间的相似性提出了迭代剪枝合并算法, 通过剪枝或合并一些相似的神元, 获得了具有良好适应性的网络结构。Gao 等^[4] 通过删除相邻储备池之间的连接, 阻止了输入信号的劣质特征表示由低隐层向高隐层的传递。然而, 上述剪枝算法很容易改变储备池的动力学演化轨迹, 进而影响网络的预测性能。因此, 如何设计合理的剪枝策略以避免在优化后储备池的动力学受到影响仍有待研究。

此外, 为从剪枝的角度优化 DESN 的结构, 精准评估神经元之间的相关性尤为关键。这种相关性可为剪枝过程提供重要的先验知识, 以准确地识别应该从储备池中移除的神经元。传统的皮尔逊相关系数常用于度量 2 个时间序列之间的相关性。然而, 此类相关性分析工具在处理非平稳数据时存在一定的局限性, 且仅适用于二元线性关联, 难以准确地评估储备池神经元之间的相关性。去趋势多重互相关 (detrended multiple cross-corre-

收稿日期: 2023-12-20; 修订日期: 2024-02-18

基金项目: 河北省海洋生态修复与智慧海洋工程研究中心开放基金项目 (HBMESO2315)

作者简介: 孙晓川 (1983—), 男, 山东烟台人, 华北理工大学副教授, 博士, 主要从事神经网络建模优化研究, E-mail: sunxiaochuan@ncst.edu.cn。

通信作者: 黄天宇 (1983—), 男, 河北秦皇岛人, 华北理工大学讲师, 博士, 主要从事神经网络、机器学习研究, E-mail: htyu@ncst.edu.cn。

引用本文: 孙晓川, 王宇, 李莹琦, 等. 基于去趋势多重互相关的深度回声状态网络剪枝算法[J]. 郑州大学学报(工学版), 2024, 45(4): 38-45. (SUN X C, WANG Y, LI Y Q, et al. Deep echo state network pruning algorithm based on detrended multiple cross-correlation[J]. Journal of Zhengzhou University (Engineering Science), 2024, 45(4): 38-45.)

lation, DMC^2)^[5]是一种有效的时间序列分析方法,提供了对多个时间序列之间非线性关联的鲁棒估计。此外,通过去除时间序列中的趋势成分, DMC^2 可以有效地处理非平稳数据。到目前为止, DMC^2 已被广泛应用于多个领域中。Wang 等^[6]利用 DMC^2 准确地识别出了 $PM_{2.5}$ 是影响北京市空气质量的主要因素,并证明其分析结果具备统计显著性。De Almeida Brito 等^[7]将 DMC^2 应用于生态环境数据的建模中,表明其能够提供比传统方法更好的统计结果。鉴于 DMC^2 在时序数据分析方面的优势,本文采用其评估储备池神经元之间的相关性,进而为后续的剪枝操作识别冗余神经元提供准确、可靠的先验知识。

综上,本文提出了一种基于 DMC^2 的 DESN 剪枝算法 DMCP。一方面,利用 DMC^2 量化储备池神经元之间的相关性,筛选出相关性较高的神经元;另一方面,仅删除所筛选神经元到网络输出层之间的连接以实现剪枝,以确保优化后的储备池动力学不受剪枝的影响,并最终通过最小二乘回归重新训练网络以完成 DESN 的结构优化任务。

1 深度回声状态网络

经典的 DESN 模型如图 1 所示。结构上,该模型主要由输入层、多个堆叠在一起的储备池和输出层组成。在每个时间步 t , 第一个储备池由外部输入 $u(t)$ 驱动,而每个后续层均由其前一层的输出驱动。这些储备池的状态转移函数表示为

$$\begin{cases} \mathbf{x}^l(t) = (1 - a^l)\mathbf{x}^l(t-1) + a^l \tanh(\mathbf{W}_{in}^l u(t) + \mathbf{W}_r^l \mathbf{x}^l(t-1)), l = 1; \\ \mathbf{x}^l(t) = (1 - a^l)\mathbf{x}^l(t-1) + a^l \tanh(\mathbf{W}_{in}^l \mathbf{x}^{(l-1)}(t) + \mathbf{W}_r^l \mathbf{x}^l(t-1)), 1 < l \leq L. \end{cases} \quad (1)$$

式中: $l \in [1, L]$ 为储备池编号; $\mathbf{x}^l(t)$ 为第 l 个储备池在时刻 t 的激活状态; $a^l \in [0, 1]$ 为第 l 个储备池的泄漏参数; \mathbf{W}_{in}^l 为第 l 个储备池的输入权重矩阵; \mathbf{W}_r^l 为第 l 个储备池的内部权重矩阵。

与标准 ESN 一样, DESN 的储备池结构在初始化后受到稳定性约束而不需要训练。此类约束可以用满足 DESN 回声状态属性^[8]的条件来表示,定义为

$$\max_{1 \leq l \leq L} r((1 - a^l)\mathbf{I} + a^l \mathbf{W}_r^l) = \max_{1 \leq l \leq L} r^l < 1. \quad (2)$$

式中: r^l 为第 l 个储备池的谱半径。

DESN 在 t 时刻的输出 $y(t)$ 为所有储备池激活状态的线性组合:

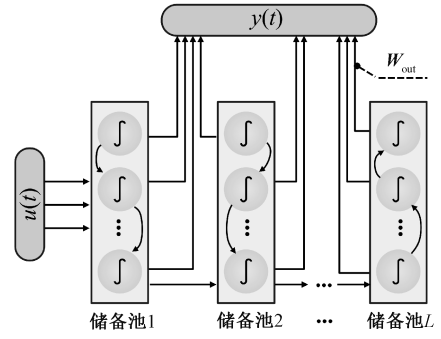


图1 深度回声状态网络的结构示意图

Figure 1 Schematic diagram of the DESN structure

$$y(t) = \mathbf{W}_{out} \mathbf{X}(t). \quad (3)$$

式中: $\mathbf{X}(t) = [\mathbf{x}^1(t); \mathbf{x}^2(t); \dots; \mathbf{x}^L(t)]$ 为网络在 t 时刻的全局状态; \mathbf{W}_{out} 为网络的输出权重矩阵。事实上, DESN 训练的主要目的是获得最优的 \mathbf{W}_{out} , 可通过求解以下的最小二乘问题实现:

$$\mathbf{W}_{out} = \underset{\mathbf{W}_{out}}{\operatorname{argmin}} \|\mathbf{Y}_{real} - \mathbf{M} \mathbf{W}_{out}\|. \quad (4)$$

式中: $\mathbf{M} = [\mathbf{X}(1), \mathbf{X}(2), \dots, \mathbf{X}(H)]^T$ 为网络 H 个时间步长的全局状态矩阵; $\mathbf{Y}_{real} = [y(1), y(2), \dots, y(H)]^T$ 表示网络的期望输出; $\|\cdot\|$ 为欧几里得范数。通常, 伪逆方法可用于对式(4)进行求解:

$$\mathbf{W}_{out} = (\mathbf{M}^+ \mathbf{Y}_{real})^T. \quad (5)$$

式中: \mathbf{M}^+ 为 \mathbf{M} 的广义逆矩阵。一旦完成 \mathbf{W}_{out} 的计算, 则标志着 DESN 的训练结束。

2 DMCP 算法描述

2.1 总体架构

图 2 显示了使用 DMCP 算法优化 DESN 结构的示意图。结构主要由储备池神经元的 DMC^2 测度模块和储备池剪枝模块组成, 前者可以根据定量的 DMC^2 度量评估储备池神经元之间的相关性, 为储备池剪枝提供先验知识; 后者通过将高相关性神经元到输出层的连接权重置零, 来实现 DESN 的剪枝优化。

2.2 储备池神经元的 DMC^2 测度

DMC^2 系数作为二元去趋势互相关分析 (detrended cross-correlation analysis, DCCA)^[9] 的推广, 提供了对多个 (2 个及 2 个以上) 时间序列之间非线性关联的鲁棒度量。鉴于此, 考虑利用 DMC^2 评估储备池神经元之间的相关性。首先获取选定储备池中各神经元的状态序列, 计算每 2 个神经元之间的 DCCA 系数 (定义为去趋势协方差函数和去趋势方差函数之间的比率), 最后通过获得的 DCCA 矩阵计算多个神经元之间的 DMC^2 系数。储备池神经元之间 DMC^2 系数的计算过程如下。

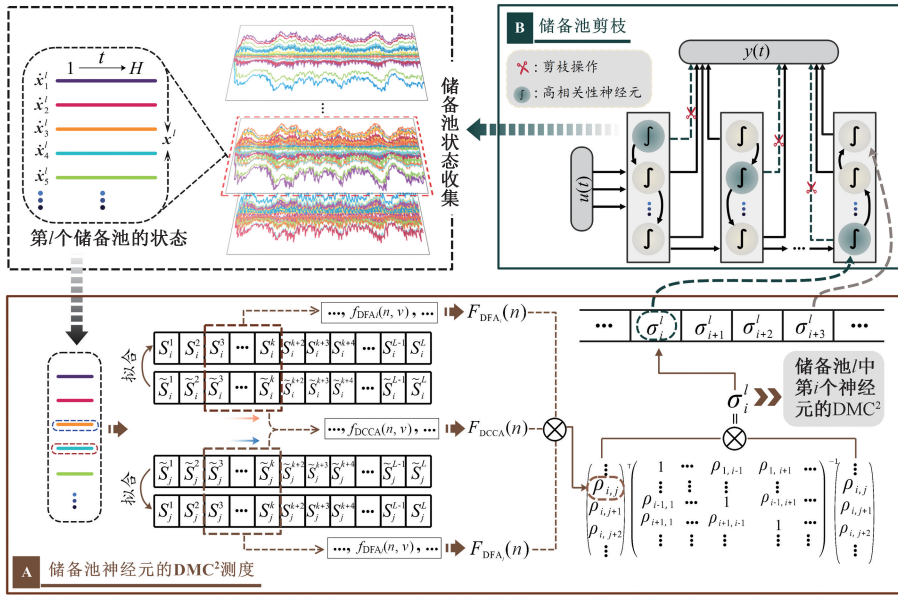


图2 基于DMCP优化DESN结构示意图

Figure 2 Schematic diagram of optimizing DESN structure based on DMCP

假设 \mathbf{x}^l 为 DESN 中具有 N^l 个神经元的第 l 个储备池共计 H 个时间步的状态矩阵, 即

$$\mathbf{x}^l = \begin{bmatrix} \mathbf{x}_1^l \\ \mathbf{x}_2^l \\ \vdots \\ \mathbf{x}_{N^l}^l \end{bmatrix} = \begin{bmatrix} x_1^l(1) & x_1^l(2) & \cdots & x_1^l(H) \\ x_2^l(1) & x_2^l(2) & \cdots & x_2^l(H) \\ \vdots & \vdots & \vdots & \vdots \\ x_{N^l}^l(1) & x_{N^l}^l(2) & \cdots & x_{N^l}^l(H) \end{bmatrix}. \quad (6)$$

对于第 i 个和第 j 个神经元的状态序列 \mathbf{x}_i^l 和 \mathbf{x}_j^l , 分别计算这 2 个序列的累计偏差以获得 2 个新序列, 即

$$\begin{cases} S_i^k = \sum_{t=1}^k (x_i^l(t) - m_i); \\ S_j^k = \sum_{t=1}^k (x_j^l(t) - m_j). \end{cases} \quad (7)$$

式中: m_i 和 m_j 分别为 \mathbf{x}_i^l 和 \mathbf{x}_j^l 的平均值; $k=1, 2, \dots, H$. 随后, 将整合后的新序列 $S_i = [S_i^1, S_i^2, \dots, S_i^H]$ 和 $S_j = [S_j^1, S_j^2, \dots, S_j^H]$ 分成 γ 个等长为 n 的窗口, 并通过最小二乘法计算每个窗口的局部趋势。基于获得的局部趋势, 可以计算每个窗口在去趋势后残差的协方差 f_{DCCA} 以及方差 f_{DFA_i} 和 f_{DFA_j} , 即

$$\begin{cases} f_{\text{DCCA}}(n, v) = \left[n^{-1} \sum_{k=1}^n (S_i^{(v-1)n+k} - \tilde{S}_i^{k,v}) (S_j^{(v-1)n+k} - \tilde{S}_j^{k,v}) \right]^{1/2}; \\ f_{\text{DFA}_i}(n, v) = \left[n^{-1} \sum_{k=1}^n (S_i^{(v-1)n+k} - \tilde{S}_i^{k,v})^2 \right]^{1/2}; \\ f_{\text{DFA}_j}(n, v) = \left[n^{-1} \sum_{k=1}^n (S_j^{(v-1)n+k} - \tilde{S}_j^{k,v})^2 \right]^{1/2}. \end{cases} \quad (8)$$

式中: $\tilde{S}_i^{k,v}$ 和 $\tilde{S}_j^{k,v}$ 分别为 S_i 和 S_j 在第 v 个窗口的局部趋势, $v=1, 2, \dots, \gamma$. 在此基础上计算相应的去趋势协方差函数 F_{DCCA} 以及去趋势方差函数 F_{DFA_i} 和 F_{DFA_j} :

$$\begin{cases} F_{\text{DCCA}}(n) = \left[\gamma^{-1} \sum_{v=1}^{\gamma} f_{\text{DCCA}}^2(n, v) \right]^{1/2}; \\ F_{\text{DFA}_i}(n) = \left[\gamma^{-1} \sum_{v=1}^{\gamma} f_{\text{DFA}_i}^2(n, v) \right]^{1/2}; \\ F_{\text{DFA}_j}(n) = \left[\gamma^{-1} \sum_{v=1}^{\gamma} f_{\text{DFA}_j}^2(n, v) \right]^{1/2}. \end{cases} \quad (9)$$

基于此, 第 i 个和第 j 个神经元之间的 DCCA 系数可定义为

$$\rho_{i,j}(n) = \frac{F_{\text{DCCA}}^2(n)}{F_{\text{DFA}_i}(n) F_{\text{DFA}_j}(n)}. \quad (10)$$

式中: $-1 \leq \rho_{i,j} \leq 1$. 至此, 此储备池中第 i 个神经元对剩余 N^l-1 个神经元的 DMC² 系数可表示为 $\sigma_i^l(n) =$

$$\begin{pmatrix} \vdots \\ \rho_{i,j} \\ \rho_{i,j+1} \\ \rho_{i,j+2} \\ \vdots \end{pmatrix}^T \begin{pmatrix} 1 & \cdots & \rho_{1,i-1} & \rho_{1,i+1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{i-1,1} & \cdots & 1 & \rho_{i-1,i+1} & \cdots \\ \rho_{i+1,1} & \cdots & \rho_{i+1,i-1} & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}^{-1} \begin{pmatrix} \vdots \\ \rho_{i,j} \\ \rho_{i,j+1} \\ \rho_{i,j+2} \\ \vdots \end{pmatrix}. \quad (11)$$

式中: $i \neq j$; $(\cdot)^{-1}$ 为逆矩阵; σ_i^l 为 $[0, 1]$.

2.3 储备池剪枝

本文所提剪枝算法的指导思想是逐层剪枝, 依次删除各储备池中高相关性神经元到输出层之间的

连接以降低网络冗余度,并最终通过最小二乘回归重训练网络以获得最优的模型结构。具体来说,构建具有 L 个储备池的 DESN,利用样本数据驱动网络,同时收集每个储备池的状态 $\mathbf{x}^l (l = 1, 2, \dots, L)$ 。从第 1 个储备池开始,基于 \mathbf{x}^1 ,通过式(11)计算神经元之间的 DMC² 系数 σ_i^1 ,并将具有高 σ_i^1 值神经元到输出层的权重置零,实现网络剪枝。此过程需要反复运行网络,比较误差大小,以确保是否继续执行剪枝,当网络误差超过可接受范围时停止对第 1 个储备池的剪枝。按照上述操作对剩余 $L-1$ 个储备池执行剪枝优化,当剪枝到第 L 个储备池且网络误差超过可接受范围时,意味着 DESN 的剪枝结束。算法 1 给出了基于 DMCP 算法优化 DESN 结构的整体流程,其时间复杂度为 $O(T^2 \times (N^l)^5 \times m \times L)$ 。

算法 1 DMCP 算法的伪代码。

输入:样本数据 $u(t)$,储备池数量 L ,谱半径 r^l ,遗忘时间 T_0 ,状态收集时间 T ,对每个储备池执行剪枝的次数 m ,每次剪枝过程中删除的神经元数量 k^l ;

输出:DMCP-DESN。

```

① #初始化 DESN
② 随机生成  $\mathbf{W}_{in}^l, \mathbf{W}_r^l$ 
③ 设置  $r^l \in (0, 1)$ 
④ 配置 DESN
⑤ for  $t = T_0$  to  $T$  do
⑥   通过式(1)更新储备池状态;
⑦   收集网络状态  $\mathbf{M}$ ;
⑧   通过式(5)计算  $\mathbf{W}_{out}$  并记录误差  $E$ ;
⑨ end
⑩ for  $l = 1$  to  $L$  do
⑪   for 1 to  $m$  do
⑫      $N^l \leftarrow$  通过  $\mathbf{x}^l$  获得神经元数量;
⑬     #储备池神经元的 DMC2 测度
⑭     for  $i = 1$  to  $N^l$  do
⑮       通过式(11)计算第  $i$  个神经元的  $\sigma_i^l$ ;
⑯       保存  $\sigma_i^l$  到  $\Gamma(i)$  中;
⑰     end
⑱     #储备池剪枝
⑲     for  $p = 1$  to  $k^l$  do
⑳        $j \leftarrow$  获得  $\Gamma$  中最大值的索引;
㉑       置零第  $j$  个神经元到输出层的权重;
㉒        $\Gamma(j) = 0$ ;
㉓     end
㉔     将未被剪枝的神经元的状态保存到  $\mathbf{x}^l$  中;
```

```

㉕     通过式(5)更新  $\mathbf{W}_{out}$  并记录误差  $E$ ;
㉖     if  $E$  超过可接受范围 then
㉗       break;
㉘     end
㉙   end
㉚ end
```

3 仿真结果与性能比较

3.1 数据集与度量指标

本文选取 Mackey-Glass (MG) 混沌时间序列和来自真实世界的蜂窝网络流量时间序列来验证 DMCP 算法的有效性。

MG 混沌系统是测试时间序列预测模型的经典基准,可由以下形式的时滞微分方程导出:

$$\frac{dy(t)}{dt} = \frac{0.2y(t-\tau)}{1+y(t-\tau)^{10}} - 0.1y(t). \quad (12)$$

当 $\tau > 16.8$ 时系统表现出混沌特性。因此,本实验通过式(12)在 $\tau = 17$ 时生成 1 000 个数据点,按 1:1 的比例划分为训练集和测试集。

蜂窝网络流量时间序列选择来自欧洲一家大型电信服务提供商(意大利电信)提供的开源数据集^[10]。该数据集覆盖整个意大利米兰市,包括 3 种业务流量类型:短信息业务(SMS)、通话业务(Call)和互联网业务(Internet)。每种业务流量的收集时间为 2013 年 11 月 1 日至 2014 年 1 月 1 日,每 10 min 采集 1 次,共计 3×10^6 条记录。实验选择 1 000 条 Call 记录作为分析对象,按 1:1 的比例划分为训练集和测试集。

采用归一化均方根误差 $NRMSE$ 、均方根误差 $RMSE$ 和决定系数 R^2 作为模型预测性能的量化标准,对应的数学表达式为

$$NRMSE = \sqrt{\frac{\sum_{k=1}^l (y_{\text{real}}(k) - \hat{y}_{\text{pre}}(k))^2}{l \cdot D(y_{\text{real}})}}; \quad (13)$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^l (y_{\text{real}}(k) - \hat{y}_{\text{pre}}(k))^2}{l}}; \quad (14)$$

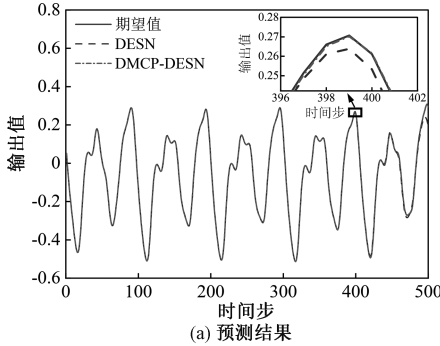
$$R^2 = 1 - \frac{\sum_{k=1}^l (y_{\text{real}}(k) - \hat{y}_{\text{pre}}(k))^2}{\sum_{k=1}^l (y_{\text{real}}(k) - \bar{y}_{\text{real}})^2}. \quad (15)$$

式中: l 为时间序列长度; $y_{\text{real}}(k)$ 和 $\hat{y}_{\text{pre}}(k)$ 分别为在时刻 k 的期望值和预测值; \bar{y}_{real} 为真实数据的平均值; $D(y_{\text{real}})$ 为 y_{real} 的方差。此外,采用记忆力

MC 作为 DESN 重构过去时间信息能力的度量标准:

$$\begin{cases} MC_{\tau} = \frac{\text{Cov}^2(u_{in}(k-\tau), \hat{y}_{pre}(k))}{\text{Var}(u_{in}(k)) \text{Var}(\hat{y}_{pre}(k))}; \\ MC = \sum_{\tau=1}^T MC_{\tau}. \end{cases} \quad (16)$$

式中: τ 为时间延迟; $u_{in}(k)$ 为在 k 时刻的模型输入; $\hat{y}_{pre}(k)$ 为模型对于 $u_{in}(k-\tau)$ 的预测输出; $\text{Cov}(\cdot)$ 和 $\text{Var}(\cdot)$ 分别为协方差函数和方差函数; 实验中设置 $T = 40$ 。



3.2 仿真结果

3.2.1 Mackey-Glass 混沌时间序列预测

图 3 显示了 DESN 和 DMCP-DES 对 MG 时间序列的预测结果和误差分布箱型图(网络初始规模为 1-50-50-50-50-1)。从图 3(a) 的局部放大图可知, DMCP-DES 预测曲线与期望曲线的变化趋势更相似, 且图 3(b) 中较扁且整体上更接近 0 的误差箱体意味着其在预测 MG 数据时的偏差更小。因此, 相比于原始的 DESN, DMCP-DES 具有更强的非线性逼近能力。

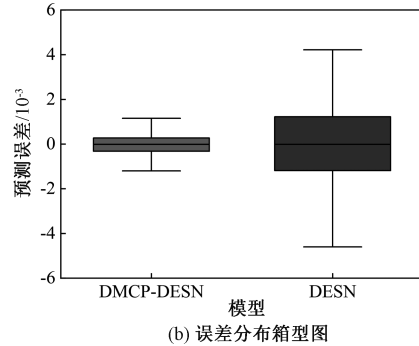


图 3 DMCP-DES 和 DESN 在 MG 预测任务中的预测结果和误差分布箱型图

Figure 3 Prediction results and error distribution box plots of DMCP-DES and DESN in the MG prediction task

图 4 显示了每次剪枝后 DMCP-DES 在 MG 时间序列预测任务中的预测误差曲线(网络初始规模为 1-50-50-50-50-1)。可以看出, 在结构优化初期, $RMSE$ 曲线随剪枝神经元数量的增加显著下降, 意味着模型预测性能的提升。当 95 个神经元被剪枝时, DMCP-DES 具有最佳的预测性能。然而, 当更多的神经元被剪枝时, 模型的预测性能下降, 导致 $RMSE$ 曲线急剧上升。实际上, 剪枝适当数量的冗余神经元等同于模型的特征选择过程。这个过程可以有效地剔除储备池中的冗余特征和无关特征, 降低数据中的噪声对网络的不利影响, 进而提高 DMCP-DES 的预测性能。但过多的剪枝会导致储备池中的重要神经元被剔除, 网络无法检测与输入信号有关的关键特征, 进而降低了 DMCP-DES 的预测能力。因此, 应对剪枝的神经元数量加以控制, 从而保证 DMCP-DES 具备良好的性能。

表 1 给出了 2 种网络规模下 DESN 和 DMCP-DES 的性能比较结果。显然, 在同一网络规模下, 更小的 $NRMSE$ 、 $RMSE$ 以及更大的 R^2 再一次证明了 DMCP-DES 在非线性逼近能力上的优势。同时, 注意到 DMCP-DES 具有更大的 MC , 表明本文所提算法能够增强网络的记忆能力, 使其更善于捕捉和利用过去信息处理序列数据。

实验也对比了其他现有剪枝策略优化后的

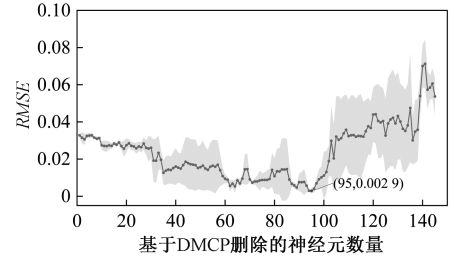


图 4 每次剪枝操作后 DMCP-DES 在 MG 预测任务中的 $RMSE$

Figure 4 $RMSE$ of DMCP-DES in the MG prediction task after each pruning operation

DESN, 包括基于贡献度剪枝的 C-DES^[11]、基于皮尔逊相似性剪枝的 PC-IPMA-DES^[3] 和基于斯皮尔曼相似性剪枝的 SC-IPMA-DES^[3]。同时, 也将其与时间序列预测建模中常见的深度学习模型进行了比较, 包括 TCN^[12]、GRU^[13]、LSTM^[14]、BiLSTM^[15]、DBN-SVM^[16] 和 DBN-ELM^[17]。

图 5 显示了在 MG 任务中各模型对选定时间步范围内数据的预测结果和误差分布箱型图。从图 5 可以看出, DMCP-DES 能够对目标信号进行最准确的拟合。表 2 给出了各模型预测性能的量化结果。从表 2 可以看出, DMCP-DES 与其他 3 种剪枝策略优化后的 DESN 相比, 预测精度具有较大优势, 这表明本文提出的剪枝算法能够更精准地识别并移除网络中的冗余成分, 从而最大限度地提高网络对

表 1 DMCP-DESN 在 MG 预测任务中的性能评价

Table 1 Performance evaluation of DMCP-DESN in the MG prediction task

模型	初始网络规模	<i>NRMSE</i>	<i>RMSE</i>	R^2	<i>MC</i>
DESN	1-50-50-50-50-1	0.137 3±0.021 5	0.029 7±0.004 7	0.980 7±0.005 8	8.053 1±2.196 6
DMCP-DESN		0.014 0±0.006 3	0.003 0±0.001 4	0.999 8±0.000 2	10.544 0±2.419 8
DESN	1-100-100-100-100-1	0.104 3±0.006 4	0.022 6±0.001 4	0.989 1±0.001 3	9.593 3±2.274 4
DMCP-DESN		0.004 8±0.002 5	0.001 0±0.000 5	0.999 9±0	11.624 3±2.040 4

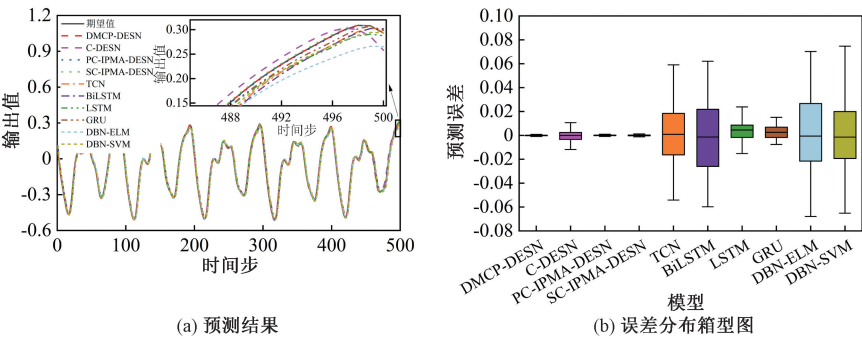


图 5 各模型在 MG 预测任务中的预测结果和误差分布箱型图

Figure 5 Prediction results and error distribution box plots of each model in the MG prediction task

表 2 各模型在 MG 预测任务中的性能评价

Table 2 Performance evaluation of each model in the MG prediction task

模型	<i>NRMSE</i>	<i>RMSE</i>	R^2
TCN	0.120 3±0.020 3	0.026 0±0.004 4	0.985 09±0.004 9
BiLSTM	0.144 0±0.000 5	0.031 2±0.000 1	0.979 20±0.000 1
DBN-ELM	0.151 3±0.001 2	0.032 8±0.000 3	0.977 07±0.000 4
DBN-SVM	0.148 5±0.000 5	0.032 2±0.000 1	0.977 90±0.000 1
GRU	0.060 5±0.003 4	0.013 1±0.000 7	0.996 32±0.000 4
LSTM	0.048 5±0.019 3	0.010 5±0.004 2	0.997 27±0.002 0
C-DESN	0.043 2±0.039 5	0.009 4±0.008 6	0.996 57±0.004 6
PC-IPMA-DESN	0.005 8±0.001 6	0.001 3±0.000 4	0.999 96±0.000 0
SC-IPMA-DESN	0.007 5±0.005 8	0.001 6±0.001 3	0.999 90±0.000 1
DMCP-DESN	0.004 8±0.002 5	0.001 0±0.000 5	0.999 97±0.000 0

MG 序列的预测精度。此外,与其他深度学习模型进行比较,DMCP-DESN 的优势更加突出。

3.2.2 蜂窝网络流量时间序列预测

图 6 显示了对于 Call 流量数据,DMCP-DESN 结构调整过程中 *RMSE* 的变化曲线(网络初始规模为 1-50-50-50-50-1)。可以看出,当 138 个储备池神经元被剪枝时,网络具有最佳的非线性逼近能力。表 3 给出了 DESN 和 DMCP-DESN 在 Call 流量预测任务中的性能对比结果。可以看出,DMCP-DESN 在预测精度和记忆能力方面均优于原始 DESN,说明 DMCP 能够有效优化网络结构,提高对复杂网络流量数据的建模能力。

图 7 显示了在 Call 流量预测任务中各模型对选定时间步范围内数据的预测结果分布脊线图和误差分布箱型图。从图 7(a)可知,DMCP-DESN 的预测输出与期望输出具有十分相似的密度分布;从图

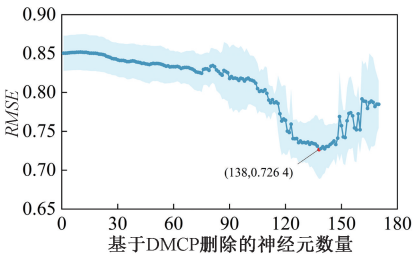


图 6 每次剪枝操作后 DMCP-DESN 在 Call 预测任务中的 *RMSE*

Figure 6 *RMSE* of DMCP-DESN in the Call prediction task after each pruning operation

7(b)可知,DMCP-DESN 的箱体最扁且整体上更接近 0,表明其在预测 Call 数据时的偏差最小。表 4 给出了各模型预测性能的量化结果。可以看出,DMCP-DESN 仍然以较大优势领先于对比模型,成为此类网络流量预测任务中的最佳架构。

表 3 DMCP-DESN 在 Call 预测任务中的性能评价

Table 3 Performance evaluation of DMCP-DESN in the Call prediction task

模型	初始网络规模	<i>NRMSE</i>	<i>RMSE</i>	<i>R</i> ²	<i>MC</i>
DESN	1-50-50-50-50-1	0.184 1±0.003 3	0.835 2±0.015 0	0.966 0±0.001 2	12.155 7±0.014 1
DMCP-DESN		0.157 7±0.006 6	0.715 4±0.030 1	0.975 0±0.002 1	12.168 0±0.000 9
DESN	1-100-100-100-100-1	0.173 0±0.001 5	0.785 0±0.006 8	0.970 0±0.000 5	12.103 2±0.027 7
DMCP-DESN		0.154 9±0.001 8	0.702 8±0.008 1	0.976 0±0.000 5	12.158 4±0.018 8

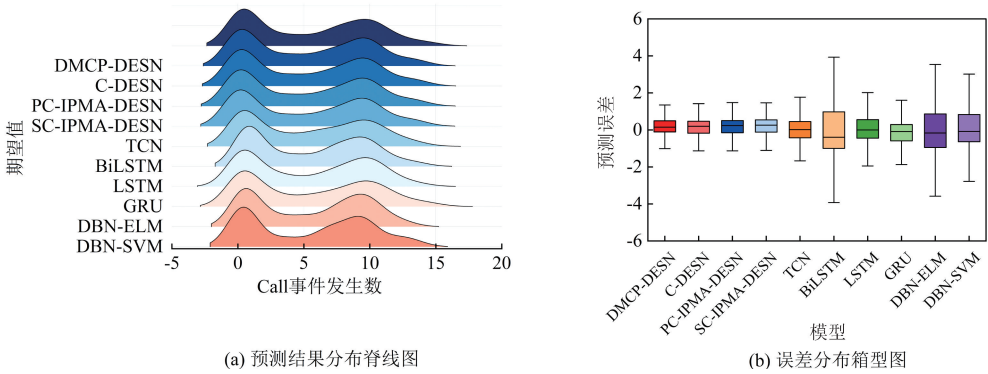


图 7 各模型在 Call 预测任务中的预测结果分布脊线图和误差分布箱型图

Figure 7 Prediction result distribution ridge plots and error distribution box plots of each model in the Call prediction task

表 4 各模型在 Call 预测任务中的性能评价

Table 4 Performance evaluation of each model in the Call prediction task

模型	<i>NRMSE</i>	<i>RMSE</i>	<i>R</i> ²
TCN	0.214 0±0.001 6	0.970 6±0.007 4	0.954 1±0.000 7
BiLSTM	0.392 6±0.000 0	1.781 0±0.000 0	0.845 6±0.000 0
DBN-ELM	0.389 0±0.000 0	1.764 6±0.000 1	0.848 4±0.000 0
DBN-SVM	0.388 0±0.002 5	1.760 1±0.011 3	0.849 2±0.001 9
GRU	0.195 7±0.002 6	0.887 6±0.011 6	0.961 6±0.001 0
LSTM	0.242 2±0.013 0	1.098 7±0.059 0	0.941 1±0.006 3
C-DESN	0.158 9±0.006 6	0.720 7±0.029 8	0.974 7±0.002 1
PC-IPMA-DESN	0.163 2±0.004 9	0.740 2±0.022 1	0.973 3±0.001 6
SC-IPMA-DESN	0.165 9±0.005 2	0.752 6±0.023 4	0.972 4±0.001 7
DMCP-DESN	0.154 9±0.001 8	0.702 8±0.008 1	0.976 0±0.000 5

4 结论

本文提出了一种基于去趋势多重互相关的深度回声状态网络剪枝算法 DMCP。该算法首先利用去趋势多重互相关准则评估储备池神经元之间的相关性,随后通过剪枝高相关性神经元到输出层的连接,有效地剔除了网络中的冗余成分,优化了 DESN 的拓扑结构。仿真结果表明:利用所提算法剔除适当数量的冗余神经元有助于提高网络的学习能力,使 DMCP-DESN 在预测性能和记忆能力方面的表现优于原始的 DESN。此外,与其他类型的预测模型相比,DMCP-DESN 预测器在预测精度方面仍具有显著优势。

参考文献:

[1] BARREDO ARRIETA A, GIL-LOPEZ S, LAÑA I, et

al. On the post-hoc explainability of deep echo state networks for time series forecasting, image and video classification[J]. Neural Computing and Applications, 2022, 34(13): 10257-10277.

[2] 李北明, 金荣璐, 徐召飞, 等. 基于特征蒸馏的改进 Ghost-YOLOv5 红外目标检测算法[J]. 郑州大学学报(工学版), 2022, 43(1): 20-26.

LI B M, JIN R L, XU Z F, et al. An improved Ghost-YOLOv5 infrared target detection algorithm based on feature distillation [J]. Journal of Zhengzhou University (Engineering Science), 2022, 43(1): 20-26.

[3] SHEN Q Y, ZHANG H W, MAO Y. Improving deep echo state network with neuronal similarity-based iterative pruning merging algorithm[J]. Applied Sciences, 2023, 13(5): 2918.

[4] GAO R B, LI R L, HU M H, et al. Dynamic ensemble deep echo state network for significant wave height fore-

casting[J]. *Applied Energy*, 2023, 329: 120261.

[5] ZEBENDE G F, DA SILVA FILHO A M. Detrended multiple cross-correlation coefficient[J]. *Physica A: Statistical Mechanics and its Applications*, 2018, 510: 91–97.

[6] WANG F, XU J, FAN Q J. Statistical properties of the detrended multiple cross-correlation coefficient[J]. *Communications in Nonlinear Science and Numerical Simulation*, 2021, 99: 105781.

[7] DE ALMEIDA BRITO A, DE ARAÚJO H A, ZEBENDE G F. Detrended multiple cross-correlation coefficient applied to solar radiation, air temperature and relative humidity[J]. *Scientific Reports*, 2019, 9: 19764.

[8] GALLICCHIO C, MICHELI A. Echo state property of deep reservoir computing networks[J]. *Cognitive Computation*, 2017, 9(3): 337–350.

[9] BEN-SALHA O, MOKNI K. Detrended cross-correlation analysis in quantiles between oil price and the US stock market[J]. *Energy*, 2022, 242: 122918.

[10] BARLACCHI G, DE NADAI M, LARCHER R, et al. A multi-source dataset of urban life in the city of Milan and the Province of Trentino [J]. *Scientific Data*, 2015, 2: 150055.

[11] LI D Y, LIU F, QIAO J F, et al. Structure optimization for echo state network based on contribution[J]. *Tsinghua Science and Technology*, 2019, 24(1): 97–105.

[12] LI D, JIANG F X, CHEN M, et al. Multi-step-ahead wind speed forecasting based on a hybrid decomposition method and temporal convolutional networks[J]. *Energy*, 2022, 238: 121981.

[13] 高金峰, 庞昊, 杜耀恒. 基于 GRU 网络的配电网故障数量等级预测方法[J]. *郑州大学学报(工学版)*, 2019, 40(5): 39–44.

GAO J F, PANG H, DU Y H. A method for predicting the number of faults in distribution network based on GRU neural network [J]. *Journal of Zhengzhou University (Engineering Science)*, 2019, 40(5): 39–44.

[14] CHEN Z L, YANG C L, QIAO J F. The optimal design and application of LSTM neural network based on the hybrid coding PSO algorithm[J]. *The Journal of Supercomputing*, 2022, 78(5): 7227–7259.

[15] CUI M S. District heating load prediction algorithm based on bidirectional long short-term memory network model [J]. *Energy*, 2022, 254: 124283.

[16] HUANG Y P, YEN M F. A new perspective of performance comparison among machine learning algorithms for financial distress prediction[J]. *Applied Soft Computing*, 2019, 83: 105663.

[17] LI Y M, PENG T, HUA L, et al. Research and application of an evolutionary deep learning model based on improved grey wolf optimization algorithm and DBN-ELM for AQI prediction [J]. *Sustainable Cities and Society*, 2022, 87: 104209.

Deep Echo State Network Pruning Algorithm Based on Detrended Multiple Cross-correlation

SUN Xiaochuan^{1, 2}, WANG Yu^{1, 2}, LI Yingqi^{1, 2}, HUANG Tianyu^{1, 2}

(1. College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China; 2. Hebei Provincial Key Laboratory of Industrial Intelligent Perception, Tangshan 063210, China)

Abstract: To address the problem of undesirable prediction accuracy of a deep echo state network caused by redundant structures in the reservoirs, a pruning algorithm for the deep echo state network based on detrended multiple cross-correlation was proposed. Firstly, according to the detrended covariance function and the detrended variance function, the detrended cross-correlation coefficient between each two neurons in the selected reservoirs in turn was calculated, and the detrended cross-correlation matrix was constructed. Based on this matrix, the detrended multiple cross-correlation between a selected neuron and all remaining neurons in this reservoir could be evaluated. Subsequently, the connections from the highly correlated neurons in each reservoir to the output layer were pruned sequentially, thus removing redundant components in the network. Finally, the network after pruning was retrained by least squares regression to obtain the optimal deep echo state network topology. Simulation results showed that the prediction accuracy and memory capacity of the deep echo state network optimized by the proposed algorithm on Mackey-Glass time series were improved by 89.80% and 30.93%, respectively, and on Call time series by 14.34% and 0.10%, respectively.

Keywords: deep echo state network; structure optimization; pruning; detrended multiple cross-correlation; time series prediction