

控制与决策

Control and Decision

基于增量式学习的正则化回声状态网络

王磊, 苏中, 乔俊飞, 赵静

引用本文:

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.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.1311>

您可能感兴趣的其他文章

Articles you may be interested in

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>

基于矩阵的双论域模糊概率粗糙集增量更新算法

Incremental updating of fuzzy probability rough sets over two universes based on matrix method

控制与决策. 2021, 36(3): 553–564 <https://doi.org/10.13195/j.kzyjc.2019.0692>

基于FWADE-ELM的短时交通流预测方法

Short-term traffic flow forecasting based on hybrid FWADE-ELM

控制与决策. 2021, 36(4): 925–932 <https://doi.org/10.13195/j.kzyjc.2019.1103>

基于FWADE-ELM的短时交通流预测方法

Short-term traffic flow forecasting based on hybrid FWADE-ELM

控制与决策. 2021, 36(4): 925–932 <https://doi.org/10.13195/j.kzyjc.2019.1103>

基于生成对抗网络学习被遮挡特征的目标检测方法

Object detection via learning occluded features based on generative adversarial networks

控制与决策. 2021, 36(5): 1199–1205 <https://doi.org/10.13195/j.kzyjc.2019.1319>

基于增量式学习的正则化回声状态网络

王磊^{1,2,3†}, 苏中¹, 乔俊飞³, 赵静⁴

- (1. 北京信息科技大学 高动态导航技术北京市重点实验室, 北京 100192;
2. 北京信科高端信息产业技术研究院有限公司, 北京 100192;
3. 北京工业大学 信息学部, 北京 100124; 4. 中国标准化研究院, 北京 100191)

摘要: 针对回声状态网络(ESN)的结构设计问题,提出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,首先,子储备池根据奇异值分解方法生成,且可以保证每个子储备池权值矩阵的奇异值都小于 1;其次,利用问题复杂度或者残差,将网络中逐一添加子储备池,直至满足预设的终止条件,在生成 IRESN 的过程中,回声状态特性能得以保证,并不需要缩放储备池权值矩阵;然后,为了解决不适定问题,在网络增量式学习过程中,利用正则化方法训练输出权值,并利用留一交叉验证方法选择正则化参数. 仿真结果表明,与其他 ESNs 相比较,所得网络具有紧凑的结构和较高的预测精度.

关键词: 回声状态网络; 增量式学习; 奇异值分解; 正则化; 留一交叉验证

中图分类号: TP183

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.1311

开放科学(资源服务)标识码(OSID):



引用格式: 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.

Design of incremental regularized echo state network

WANG Lei^{1,2,3†}, SU Zhong¹, QIAO Jun-fei³, ZHAO Jing⁴

- (1. Beijing Key Laboratory of High Dynamic Navigation Technology, Beijing Information Science and Technology University, Beijing 100192, China; 2. Beijing Jingxinke High-end Information Industry Technology Research Institute Co., Ltd., Beijing 100192, China; 3. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China; 4. China National Institute of Standardization, Beijing 100191, China)

Abstract: Aiming at the structure design of an echo state network(ESN), an incremental regularized echo state network(IRESN) is proposed in this paper. The reservoir of the IRESN is composed of independent sub-reservoir modular networks. Firstly, the sub-reservoirs are obtained using the singular value decomposition method, and the singular values of the weight matrix of each sub-reservoir can be guaranteed to be less than one. Then, depending on the problem, complexity or residual error, the sub-reservoirs are added to the network one after another until the preset termination conditions are met. In the process of generating the IRESN, the echo state property can be guaranteed without scaling the reservoir weight matrix. Furthermore, in order to tackle the ill-posed problem, in the process of incremental learning, the output weights are trained using the regularization method, and the leave-one-out cross-validation method is used to select the regularization parameter. The simulation results show that the IRESN has compact structure and high prediction accuracy compared with other ESNs.

Keywords: echo state network; incremental learning; singular value decomposition; regularization; leave-one-out cross-validation

0 引言

回声状态网络(echo state network, ESN)是一种新型递归神经网络,关键部分是随机生成的稀疏储备池^[1]. 对于经典 ESN 的训练,只需要利用线性回归方法训练输出权值,输入权值和储备池权值根据特定的要求随机生成. ESN 的这种训练方式能够保证权值

的全局最优,克服了基于梯度的递归神经网络计算效率低、训练方法复杂以及容易陷入局部最优等问题,具有较高的泛化能力^[2]. ESN 已经成功应用于时间序列预测^[3-5]、非线性系统控制^[6-7]、故障诊断^[8]等领域.

ESN 的核心是储备池结构设计,设计与具体任务相匹配的储备池仍然是一个亟待解决的问题^[9]. ESN

收稿日期: 2020-09-19; 录用日期: 2020-12-25.

基金项目: 国家重点研发计划项目(2020YFC1511702); 国家自然科学基金项目(61771059).

责任编委: 牛玉刚.

†通讯作者. E-mail: leiw@bistu.edu.cn.

储备池结构设计的主要方法有确定型方法^[10-11]、进化算法^[12-13]、修剪算法^[5,14]和增长算法^[3]。

确定型方法克服了传统ESN储备池的随机性。Xue等^[10]给出了具有侧抑制的解耦回声状态网络,预测精度高、鲁棒性好、实验设计成功率高。为了构建最小复杂度储备池,Rodan等^[11]给出了简单环形储备池,具有较少储备池神经元。进化算法具有较好的全局搜索能力。Duan等^[12]利用鸽群优化算法对稀疏度、储备池规模、谱半径等参数进行优化,改进的网络较好地解决了图像复原问题。Otte等^[13]利用微分进化算法训练储备池权值矩阵,实验结果表明优化的ESN能有效预测MSO问题。利用进化算法训练ESN存在问题,如收敛速度慢且容易陷入局部最优。修剪算法一般是从一个较大储备池结构开始,通过某种机制修剪掉冗余神经元。王磊等^[5]利用灵敏度分析给出了一种修剪型模块化回声状态网络,将灵敏度低的子模块进行删除。Dutoit等^[14]利用正则化技术给出了一种修剪算法来修剪储备池与输出层之间的权值连接,提高了泛化能力。修剪算法初始的神经元个数一般较多,具有较大的计算复杂度。增长算法在网络的初始阶段结构简单,但需要给出合适的停止准则。Qiao等^[3]利用SVD方法提出了增长型回声状态网络,该网络存在的问题是输出权值的计算利用广义逆求得,会导致不适定问题的出现。为了给出与任务相匹配的网络结构,自动确定储备池结构,Yang等^[15]给出了动态正则化回声状态网络,当新的神经元被加入到网络后,当前的网络可以被新生成的具有更紧凑网络结构的储备池替代。

针对以上分析,本文提出增量式正则化回声状态网络(incremental regularized ESN, IRESN)。将子储备池逐一添加至IRESN中,若满足预定的条件,则迭代结束。在网络结构调整的过程中,根据正则化方法对输出权值进行训练,利用留一交叉验证方法选择正则化参数。

1 回声状态网络

1.1 经典回声状态网络

经典ESN由输入层、储备池层和输出层构成,假设神经元节点数量分别为 K 、 N 、 L ,拓扑结构如图1所示。对于含有 J 个样本的训练样本集 $\{(u(n), t(n)) | u(n) \in R^K, t(n) \in R^L, n = 1, 2, \dots, J\}$,ESN的基本方程如下:

$$x(n+1) = f(W^{\text{in}}u(n+1) + Wx(n)), \quad (1)$$

$$y(n+1) = W^{\text{out}}(u(n+1), x(n+1)). \quad (2)$$

其中: $u(n) \in R^K, x(n) \in R^N, y(n) \in R^L$ 分别为ESN的输入变量、状态变量和输出变量; $W^{\text{in}}, W, W^{\text{out}}$ 分别表示输入权值矩阵、储备池权值矩阵、输出权值矩阵; f 为储备池神经元的激活函数,一般取为sigmoid函数。基本ESN的 W^{in} 和 W 均随机初始化,在训练和测试过程中固定不变。

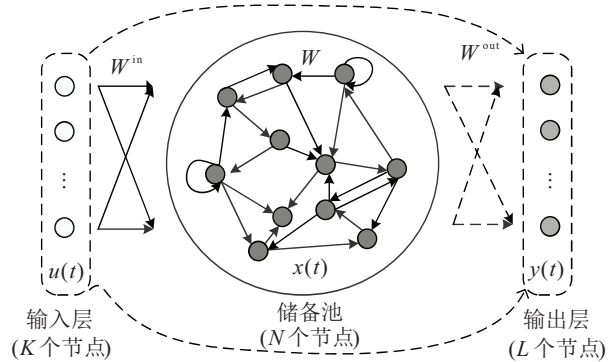


图1 ESN拓扑结构

记内部状态阵 $H = [X(1), X(2), \dots, X(J)]^T$,其中 $X(n) = [u(n)^T, x(n)^T]^T$,对应的期望输出矩阵为 $T = [t(1), t(2), \dots, t(J)]^T$,输出权值矩阵满足

$$HW^{\text{out}} = T. \quad (3)$$

式(3)的求解等价于求解如下的优化问题:

$$\min_{W^{\text{out}}} \|HW^{\text{out}} - T\|_2^2. \quad (4)$$

据最小二乘原理可得

$$W^{\text{out}} = H^+T = (H^T H)^{-1} H^T T,$$

其中 H^+ 表示内部状态矩阵 H 的Moore-Penrose逆。

1.2 正则化回声状态网络

为了得到输出权值的稳定解,正则化方法被广泛应用,采用比较多的是lasso正则化(也称 l_1 正则化)^[16]和Tikhonov正则化(也称 l_2 正则化或者岭回归)^[14]。lasso正则化的优点是可以产生稀疏解,缺点是比较难以获得正则化参数。与传统的最小二乘法相比较,根据方差-偏差的权衡,Tikhonov正则化能够具有较好的预测能力。

式(4)的求解可能产生不适定问题,作为最小二乘代价函数的改进,本文考虑用Tikhonov正则化方法训练输出权值,输出权值的求解等价于寻求下式的最优解:

$$\min_{W^{\text{out}}} (\|HW^{\text{out}} - T\|_2^2 + c\|W^{\text{out}}\|_2^2). \quad (5)$$

其中:添加的惩罚项 $c\|W^{\text{out}}\|_2^2$ 用于解决不适定问题, $c > 0$ 为Tikhonov正则化参数,其取值方法参见2.3小节。求解式(5)的过程如下:记

$$F(W^{\text{out}}) = (HW^{\text{out}} - T)^T (HW^{\text{out}} - T) +$$

$$c(W^{\text{out}})^T W^{\text{out}}, \quad (6)$$

式(6)关于 W^{out} 的共轭梯度

$$\frac{\partial F(W^{\text{out}})}{\partial W^{\text{out}}} = H^T H W^{\text{out}} - H^T T + c W^{\text{out}}. \quad (7)$$

令 $\partial F(W^{\text{out}})/\partial W^{\text{out}} = 0$, 可以得到

$$(cI + H^T H)W^{\text{out}} = H^T T, \quad (8)$$

其中 I 为单位阵. 由式(8)可得

$$W^{\text{out}} = (cI + H^T H)^{-1} H^T T. \quad (9)$$

显然, 当 $c = 0$ 时, 式(9)即是经典的回声状态网络输出权值, 在本文中输出权值利用 Tikhonov 正则化方法进行训练.

2 增量式正则化回声状态网络

为了保证 ESN 能正常工作, 需要具有回声状态特性 (echo state property, ESP) 以保持网络的稳定性. ESN 具有 ESP 的充分条件是储备池权值矩阵的最大奇异值小于 1, 必要条件是储备池权值矩阵的谱半径小于 1. 具有 ESP 的必要条件比较容易满足, 可以通过缩放因子产生储备池, 但是具有 ESP 的充分条件比较难以满足, 也容易被忽略^[9]. 为了保证所构造的网络具有 ESP 的充分条件, 考虑采用 SVD 方法构造储备池.

为了使得构造模型的网络结构较为紧凑, 同时具有较好的泛化能力, 下面给出 IRESN 拓扑结构如图 2 所示.

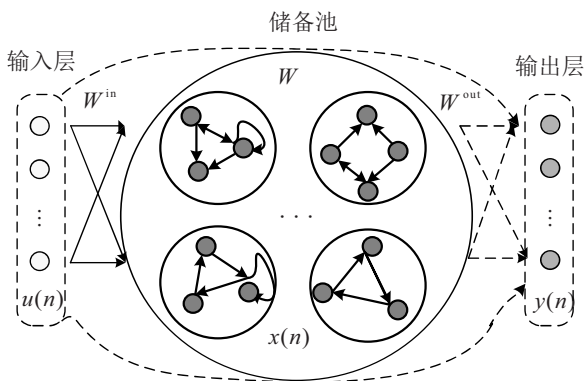


图 2 IRESN 拓扑结构

2.1 IRESN 算法分析

给定 ESN 模型 Ψ , 其残差函数定义如下:

$$E(\Psi) = \|HW^{\text{out}} - T\|_2^2 + c\|W^{\text{out}}\|_2^2. \quad (10)$$

其中: H 和 W^{out} 分别表示网络 Ψ 的内部状态矩阵和输出权值矩阵, $E(\Psi)$ 的值越小, 说明网络的训练效果越好.

IRESN 的设计理念是根据分块矩阵理论, 构造增量迭代学习的正则化回声状态网络. 若干子储备池构成 IRESN, 利用问题复杂度或者残差, 在网络中

逐一增加子储备池, 直至满足预设的终止条件. 根据 SVD 方法构造子储备池权值矩阵, 使得储备池权值矩阵的奇异值均于小 1, 意味着最大奇异值和稀疏度能够自适应设定, 详细描述如下:

首先, 生成具有 N_1 个储备池节点的子储备池 Ψ_1 , 随机生成对角阵 $\Lambda_1 = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_{N_1})$ 和两个正交阵 $B_1 = (b_{ij})_{N_1 \times N_1}$, $V_1 = (v_{ij})_{N_1 \times N_1}$, 其中 $0 < \sigma_i < 1 (i = 1, 2, \dots, N_1)$, $b_{ij}, v_{ij} \in (-1, 1)$. 假设子储备池 Ψ_1 对应的权值矩阵 $W_1 = B_1 \Lambda_1 V_1$, 依据 SVD 原理, W_1 和 Λ_1 的奇异值相同. 求得内部状态 x_1 , 得到内部状态矩阵 H_1 , 利用式(9)得到输出权值 W_1^{out} .

其次, 构造类似的子储备池 Ψ_2 , 对应的储备池权值矩阵 $W_2 = B_2 \Lambda_2 V_2$, 这里 B_2 和 V_2 为正交矩阵, 内部状态、内部状态矩阵以及输入权值矩阵分别为 x_1 、 H_2 和 W_2^{in} , 将 Ψ_2 增添至网络中.

按照下式更新内部状态和模型输出:

$$\bar{x}_2(n+1) = \begin{bmatrix} x_1(n+1) \\ x_2(n+1) \end{bmatrix} = f \begin{bmatrix} W_1 & \\ & W_2 \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} + \begin{bmatrix} W_1^{\text{in}} \\ W_2^{\text{in}} \end{bmatrix} u(n+1), \quad (11)$$

$$y_2(n) = (W_2^{\text{out}})^T X_2(n), \quad (12)$$

其中 $X_2(n) = [\bar{x}_2(n)^T, u(n)^T]^T$. 更新子储备池之后的内部状态矩阵为 $\bar{H}_2 = [H_1, H_2]$, 按照下式计算输出权值矩阵:

$$W_2^{\text{out}} = (cI + \bar{H}_2^T \bar{H}_2)^{-1} \bar{H}_2^T T. \quad (13)$$

下面利用待定系数方法给出式(13)的求解推导过程, 其关键是计算 $(cI + \bar{H}_2^T \bar{H}_2)^{-1}$. 不妨假设

$$(cI + \bar{H}_2^T \bar{H}_2)^{-1} = \begin{bmatrix} Q_1 & Q_2 \\ Q_2^T & Q_3 \end{bmatrix},$$

其中 Q_1, Q_2, Q_3, Q_4 待定. 因为 $\bar{H}_2 = [H_1, H_2]$, 有

$$(cI + \bar{H}_2^T \bar{H}_2)^{-1} = \left(cI + \begin{bmatrix} H_1^T \\ H_2^T \end{bmatrix} [H_1, H_2] \right)^{-1} = \begin{bmatrix} H_1^T H_1 + cI & H_1^T H_2 \\ H_2^T H_1 & H_2^T H_2 + cI \end{bmatrix}^{-1} = \begin{bmatrix} A_1 & A_2 \\ A_2^T & A_3 \end{bmatrix}^{-1}. \quad (14)$$

其中

$$\begin{aligned} A_1 &= H_1^T H_1 + cI, \\ A_2 &= H_1^T H_2, \\ A_3 &= H_2^T H_2 + cI. \end{aligned} \quad (15)$$

由式(14)及待定假设, 有 $\begin{bmatrix} A_1 & A_2 \\ A_2^T & A_3 \end{bmatrix}^{-1} = \begin{bmatrix} Q_1 & Q_2 \\ Q_2^T & Q_3 \end{bmatrix}$,

则

$$I = \begin{bmatrix} A_1 & A_2 \\ A_2^T & A_3 \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 \\ Q_2^T & Q_3 \end{bmatrix} = \begin{bmatrix} A_1 Q_1 + A_2 Q_2^T & A_1 Q_2 + A_2 Q_3 \\ A_2^T Q_1 + A_3 Q_2^T & A_2^T Q_2 + A_3 Q_3 \end{bmatrix}. \quad (16)$$

由分块矩阵求逆矩阵的公式有

$$\begin{aligned} Q_1 &= A_1^{-1} + A_1^{-1} A_2 (A_3 - A_2^T A_1^{-1} A_2)^{-1} (A_1^{-1} A_2)^T, \\ Q_2 &= -A_1^{-1} A_2 (A_3 - A_2^T A_1^{-1} A_2)^{-1}, \\ Q_3 &= (A_3 - A_2^T A_1^{-1} A_2)^{-1}. \end{aligned} \quad (17)$$

则

$$(cI + \bar{H}_2^T \bar{H}_2)^{-1} \bar{H}_2^T = \begin{bmatrix} Q_1 & Q_2 \\ Q_2^T & Q_3 \end{bmatrix} \begin{bmatrix} H_1^T \\ H_2^T \end{bmatrix} = \begin{bmatrix} Q_1 H_1^T & Q_2 H_2^T \\ Q_2^T H_1^T & Q_3 H_2^T \end{bmatrix} = \begin{bmatrix} M \\ P \end{bmatrix}. \quad (18)$$

其中记 $M = Q_1 H_1^T + Q_2 H_2^T$, $P = Q_2^T H_1^T + Q_3 H_2^T$. 根据式(15)、(17)和(18),有

$$\begin{aligned} P &= - (A_3 - A_2^T A_1^{-1} A_2)^{-1} (A_1^{-1} A_2)^T H_1^T + \\ & (A_3 - A_2^T A_1^{-1} A_2)^{-1} H_2^T = \\ & (A_3 - A_2^T A_1^{-1} A_2)^{-1} (H_2^T - (H_1^T H_2)^T) \times \\ & (H_1^T H_1 + cI)^{-1} H_1^T = \\ & (cI + H_2^T (I - H_1 D) H_2)^{-1} H_2^T (I - H_1 D), \end{aligned} \quad (19)$$

$$\begin{aligned} M &= A_1^{-1} A_2 (A_3 - A_2^T A_1^{-1} A_2)^{-1} (A_1^{-1} A_2)^T H_1^T + \\ & A_1^{-1} H_1^T - A_1^{-1} A_2 (A_3 - A_2^T A_1^{-1} A_2)^{-1} H_2^T = \\ & D + D H_2 (A_3 - A_2^T A_1^{-1} A_2)^{-1} H_2^T (H_1 D - I) = \\ & D (I - H_2 P), \end{aligned} \quad (20)$$

其中记 $D = (cI + H_1^T H_1)^{-1} H_1^T$.

综上,可以得到

$$W_2^{\text{out}} = (cI + \bar{H}_2^T \bar{H}_2)^{-1} \bar{H}_2^T T = \begin{bmatrix} M \\ P \end{bmatrix} T. \quad (21)$$

这说明根据式(19)~(21), IRESN 的输出权值 W_2^{out} 可以通过迭代增量的方法计算,使得计算复杂度在一定程度上得到了降低.

2.2 IRESN 算法设计

设 $\{(u(n), t(n)) | n = 1, 2, \dots, J\}$ 为训练样本集, 储备池的激活函数为 sigmoid 函数, k_{\max} 为最大迭代次数, 则 IRESN 的主要步骤如下.

step 1: 初始化子储备池数 $k = 0$.

step 2: 对子储备池数 k 进行更新, 即 $k = k + 1$,

对含有 N_k 个储备池节点的子储备池 Ψ_k 进行初始化. 随机生成对角阵 $A_k = \text{diag}(\sigma_k^1, \sigma_k^2, \dots, \sigma_k^{N_k})$ 和正交阵 $B_k = (b_{ij})_{N_k \times N_k}$, $V_k = (v_{ij})_{N_k \times N_k}$, 其中 $0 < \sigma_k^i < 1$ ($i = 1, 2, \dots, N_k$), $b_{ij}, v_{ij} \in (-1, 1)$. 取 $W_k = B_k A_k V_k$ 为第 k 个子储备池的权值矩阵, 并随机生成对应的输入权值矩阵 W_k^{in} . 储备池状态 x_k 可以由式(1)计算, 并从 n_{\min} 时刻构建第 k 个子储备池内部状态矩阵 $H_k = [X_k(n_{\min}), X_k(n_{\min} + 1), \dots, X_k(J)]^T$, $X_k(n) = [x_k(n)^T, u(n)^T]^T$, $n = n_{\min}, \dots, J$, $\bar{H}_k = [\bar{H}_{k-1}, H_k]$.

step 3: 根据下式得到输出权值矩阵 W_k^{out} :

$$D_k = (cI + \bar{H}_{k-1}^T \bar{H}_{k-1})^{-1} \bar{H}_{k-1}^T, \quad (22)$$

$$P_k = (cI + H_k^T (I - \bar{H}_{k-1} D) H_k)^{-1} H_k^T \times (I - \bar{H}_{k-1} D_k), \quad (23)$$

$$M_k = D_k (I - H_k P), \quad (24)$$

$$W_k^{\text{out}} = \begin{bmatrix} M_k \\ P_k \end{bmatrix} T. \quad (25)$$

step 4: 训练误差由下式计算(若达到预设的终止条件转到 step 6, 否则转到 step 5, 至最大迭代次数 k_{\max} 停止):

$$E(\Psi_k) = \|\bar{H}_k W_k^{\text{out}} - T\|_2^2 + c \|W_k^{\text{out}}\|_2^2. \quad (26)$$

step 5: 权值矩阵和内部状态矩阵均利用下式进行更新:

$$\begin{aligned} \bar{W}_k^{\text{in}} &= \begin{bmatrix} \bar{W}_{k-1}^{\text{in}} \\ W_k^{\text{in}} \end{bmatrix}, \\ \bar{W}_k &= \text{diag}(\bar{W}_{k-1}, W_k), \\ \bar{H}_k &= [\bar{H}_{k-1}, H_k]. \end{aligned} \quad (27)$$

step 6: 对训练好的 IRESN 进行测试.

2.3 正则化参数的选取准则

在 IRESN 输出权值的学习过程中, 需要选择恰当的正则化参数, 以便得到最优的网络模型. 为了能够得到可靠稳定的模型, 本文用交叉验证来选取正则化参数. 交叉验证是一种模型选择方法, 根据数据集切分的方法不同, 可以分为 Holdout 验证、 k -重交叉验证和留一交叉验证 (leave-one-out cross validation, LOO-CV) 等^[17-18]. 与前两者相比, 留一交叉验证的优点是最接近原始样本的分布, 评估所得的结果比较可靠, 实验过程中没有随机因素会影响实验数据, 能够确保实验过程是可以被复制的.

综上, 利用 LOO-CV 选择正则化参数. LOO-CV 的主要思想是, 如果设原始数据有 J 个样本, 则 LOO-CV 就是 J -CV, 即每个样本单独作为验证集, 其余的

$J - 1$ 个样本作为训练集, 所以 LOO-CV 会得到 J 个模型. LOO-CV 的缺点是计算成本高, 为了克服这个问题, 利用 PRESS 统计公式^[19] 计算 LOO-CV 误差, 即

$$E_{\text{LOO-CV}} = \frac{1}{J} \sum_{i=1}^J \left(\frac{t(i) - y(i)}{1 - g(i)} \right)^2, \quad (28)$$

$$G = H(H^T H + cI)^{-1} H^T. \quad (29)$$

其中: $t(i)$ 为第 i 个样本的期望, $y(i)$ 为对应的预测值, $g(i)$ 为 G 的对角线的第 i 个元素. 正则化参数 c 的选取目标是 minimized LOO-CV 误差.

3 仿真实验

为了验证 IRESN 的性能, 利用如下 3 个实验对比研究 IRESN 的网络性能, 包括 NARMA 系统辨识, Lorenz 时间序列预测, 污水处理过程出水总磷浓度预测. 对比的网络模型取 OESN^[3], RESN^[14], GESN^[3], ESN-LAR^[14] 和 ESN-DE^[13]. 所有实验的软硬件参数如下: i7-4790, CPU 3.60 GHz, 8.0 GB RAM, Windows7 操作系统, 2013b 版 Matlab.

测试标准均方根误差 (NRMSE) 用于检验 IRESN 的性能^[3], 即

$$\text{NRMSE} = \sqrt{\frac{\sum_{n=1}^J \frac{(t(n) - y(n))^2}{J\sigma^2}}{J}}. \quad (30)$$

其中: $t(n)$ 表示期望输出, $y(n)$ 表示预测输出, J 表示样本数量, σ^2 表示期望输出的方差. NRMSE 的值越小意味着具有较好的预测性能.

3.1 NARMA 系统辨识

因具备长时记忆及非线性等特点, NARMA 系统建模难度大, 其相应方程可由下式给出^[11]:

$$\begin{aligned} x(n+1) = & 0.3x(n) + 0.05x(n) \left(\sum_{i=0}^9 x(n-i) \right) + \\ & 1.5u(n-9)u(n) + 0.1. \end{aligned} \quad (31)$$

其中: $u(n)$ 于 $[0, 1]$ 上随机生成, 初始值 $x(n) (n = 1, 2, \dots, 10) = 0$. 利用式 (31) 生成的样本量为 2500, 训练

样本的长度为 1500, 测试样本的长度为 1000. 为了克服初态瞬态的影响, 训练样本中的丢弃点数为 500.

针对 $x(n)$, 图 3 给出了测试输出, 图 4 给出了测试误差, 为了方便查看, 给出了测试样本的前 200 个点的图形. 由图 3 和图 4 可知, IRESN 测试精度高, 意味着 IRESN 的泛化能力较好.

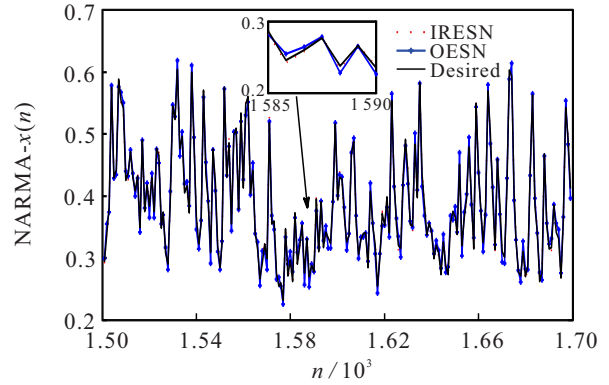


图 3 NARMA 系统测试输出

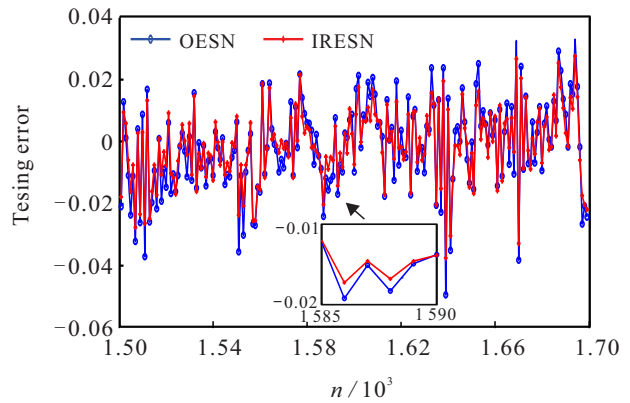


图 4 NARMA 系统测试误差

为了详细核验 IRESN 算法的性能, 进行 100 次独立实验, 详细的参数设置与实验结果如表 1 所示. 由表 1 可知, IRESN 的最大奇异值和稀疏度能够自适应设定, 且最大奇异值小于 1. 在学习时间方面, IRESN 并不是最小的, 但是在精度方面, IRESN 具有最小的训练/测试 NRMSE, 在结构方面, IRESN 的储备池层比较紧凑.

表 1 10 阶 NARMA 系统预测辨识实验结果对比

网络模型	储备池规模	正则化参数 c	谱半径	最大奇异值	稀疏度	训练时间/s	训练 NRMSE		测试 NRMSE	
							平均值	标准差	平均值	标准差
IRESN	210	10^{-3}	0.945 8	0.992 1	0.015 2	492.86	0.024 9	2.18×10^{-3}	0.062 9	3.32×10^{-3}
OESN	600	-	0.900 0	1.769 1	0.025 0	273.09	0.043 1	4.48×10^{-3}	0.102 1	3.41×10^{-2}
RESN	600	10^{-4}	0.900 0	1.373 2	0.025 0	595.56	0.034 9	3.86×10^{-3}	0.089 3	4.61×10^{-2}
GESN	400	-	0.935 6	0.991 2	0.019 5	436.29	0.031 8	4.57×10^{-3}	0.076 1	4.85×10^{-2}
ESN-LAR	900	-	0.900 0	0.979 2	0.035 0	602.76	0.035 1	4.69×10^{-3}	0.095 1	6.77×10^{-2}
ESN-DE	500	-	0.900 0	0.913 8	0.035 0	1 213.95	0.041 9	7.11×10^{-4}	0.098 6	4.61×10^{-3}

3.2 Lorenz时间序列预测

混沌是非线性动力系统的固有特性,Lorenz发现第1个混沌吸引子——Lorenz系统,其状态方程^[20-21]如下:

$$\begin{cases} \frac{dx}{dt} = a(y - x), \\ \frac{dy}{dt} = -xz + bx - y, \\ \frac{dz}{dt} = xy - cz. \end{cases} \quad (32)$$

当 $a = 10, b = 28, c = 8/3$ 时呈现混沌状态. 在实验中,用四阶龙格-库塔方法产生样本量为5000. 为了验证改进模型的性能,将样本叠加均匀分布于 $[-0.01, 0.01]$ 的随机噪声对 $y(t)$ 进行预测. 在生成的样本中,训练样本的长度为3000,测试样本的长度为2000,训练数据中的丢弃点数为1000. 步长为0.01,初值 $x(0) = 1, y(0) = 1, z(0) = 0$.

为了验证 IRESN 的预测性能,针对 IRESN 和 OESN 模型,图5和图6分别给出了测试输出及测试误差. 由图6可知, IRESN 比 OESN 具有更小的测试误差,位于 $[-0.01, 0.015]$.

为了进一步说明 IRESN 模型的性能,进行100次独立实验,详细结果如表2所示. 由表2可知, IRESN 的训练/测试NRMSE值较小,说明具有很好的泛化能力,但是训练时间不是最短的,储备池神经元数为

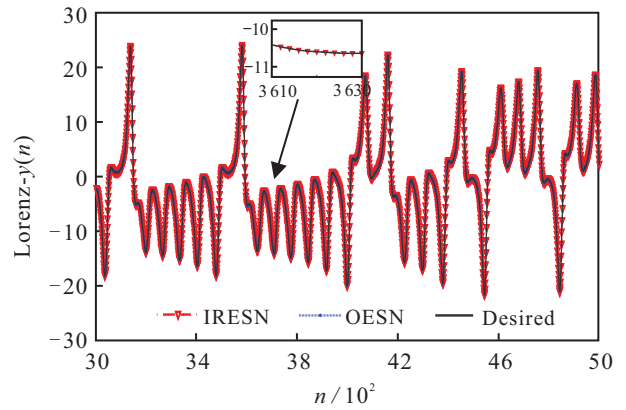


图5 Lorenz系统测试输出

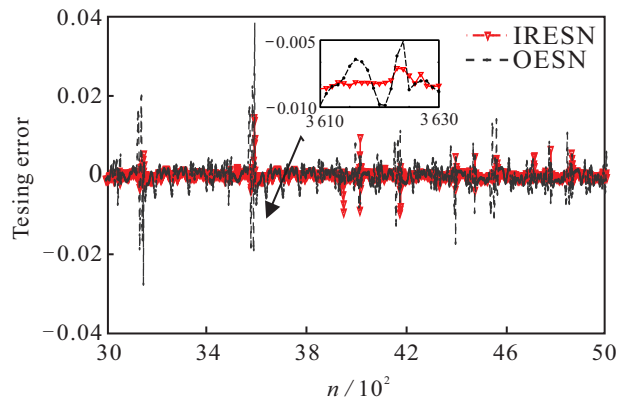


图6 Lorenz系统测试误差

285,表明所提出的模型较好地解决了结构设计问题,最大奇异值和稀疏度能够自适应地得到.

表2 Lorenz系统实验结果对比

网络模型	储备池规模	正则化参数 c	谱半径	最大奇异值	稀疏度	训练时间/s	训练NRMSE		测试NRMSE	
							平均值	标准差	平均值	标准差
IRESN	285	10^{-5}	0.9536	0.9920	0.0155	362.39	1.28×10^{-3}	2.79×10^{-4}	8.12×10^{-3}	8.29×10^{-4}
OESN	600	-	0.9000	0.9771	0.0300	221.45	9.21×10^{-3}	9.55×10^{-4}	9.29×10^{-2}	3.49×10^{-3}
RESN	600	10^{-4}	0.9000	0.9641	0.0300	411.28	4.29×10^{-3}	5.49×10^{-4}	6.11×10^{-2}	5.78×10^{-3}
GESN	350	-	0.9250	0.9910	0.0250	325.76	1.65×10^{-3}	1.95×10^{-4}	3.59×10^{-2}	6.32×10^{-4}
ESN-LAR	700	-	0.8000	0.9486	0.0400	369.53	3.31×10^{-3}	7.12×10^{-4}	4.86×10^{-2}	5.79×10^{-4}
ESN-DE	500	-	0.8000	0.9795	0.0400	852.35	3.75×10^{-3}	7.89×10^{-4}	5.12×10^{-2}	3.81×10^{-3}

3.3 污水处理过程出水总磷浓度预测

污水处理过程出水总磷是水质评价的重要指标,总磷超标会引起水体富营养化,发生水华或赤潮,造成水环境污染^[22-23]. 在实际污水处理过程中,总磷的值难以测量,现存的总磷检测仪存在价格昂贵、维护费用高等问题,同时难以实时在线检测^[24]. 在本实验中,利用 IRESN 模型预测污水处理过程出水总磷浓度值,实验数据来源于北京市某污水处理厂2018-06-01~2019-05-31期间的数据,选取进水总磷、温度、氧化还原电位、溶解氧、总固体悬浮物和pH值6个易测

变量作为模型输入辅助变量. 经过去噪、剔除异常值得到333组样本数据并进行归一化处理,前270组样本数据用于训练,其中训练集中丢弃点的样本数量为20,后63组数据样本用于模型测试.

图7和图8分别给出了 IRESN 和 OESN 的出水总磷测试输出及测试误差曲线,可以看出, IRESN 具有较高的精度,其预测性能优于传统的 OESN, IRESN 的测试误差为 $[-0.02, 0.02]$, OESN 的测试误差为 $[-0.08, 0.08]$.

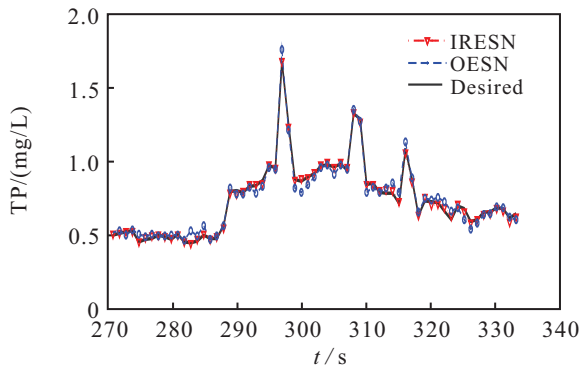


图7 出水总磷测试输出

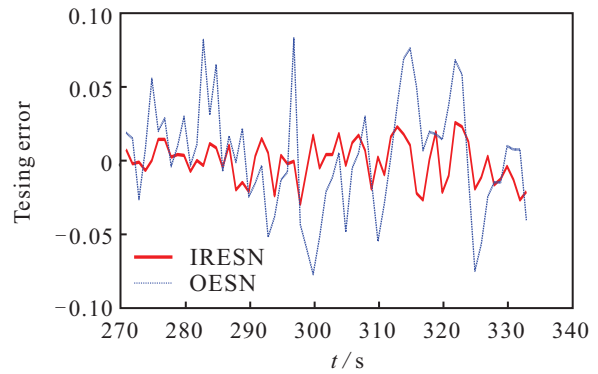


图8 出水总磷测试误差

表3 出水总磷实验结果对比

网络模型	储备池规模	正则化参数 c	谱半径	最大奇异值	稀疏度	训练时间/s	训练NRMSE		测试NRMSE	
							平均值	标准差	平均值	标准差
IRESN	280	10^{-3}	0.938 2	0.990 0	0.012 5	453.31	0.008 6	9.15×10^{-4}	0.024 5	2.56×10^{-3}
OESN	500	—	0.850 0	1.873 2	0.025 0	256.36	0.069 5	6.52×10^{-3}	0.123 5	2.02×10^{-2}
RESN	500	10^{-4}	0.850 0	1.679 8	0.025 0	581.21	0.045 3	4.61×10^{-3}	0.089 8	8.67×10^{-2}
GESN	300	—	0.945 2	0.990 0	0.016 5	412.56	0.011 6	2.78×10^{-3}	0.031 8	3.35×10^{-2}
ESN-LAR	700	—	0.850 0	0.987 5	0.025 0	587.56	0.042 1	4.31×10^{-3}	0.085 6	8.62×10^{-3}
ESN-DE	500	—	0.850 0	0.926 5	0.025 0	1 125.31	0.048 6	5.01×10^{-3}	0.081 2	8.45×10^{-3}

为了详细给出 IRESN 的测试性能,进行 100 次独立实验,详细的参数设置与对比结果如表 3 所示. 由表 3 可以看出, IRESN 的最大奇异值为 0.99,稀疏度为 0.0125,较好地达到了与具体任务相匹配的参数自适应. IRESN 的训练时间没有优势,但是 IRESN 具有较小的储备池规模,训练/测试 NRMSE 的平均值和标准差在对比模型中最小,说明 IRESN 具有较好的模型精度和较高的泛化能力,同时 IRESN 的最大奇异值小于 1,说明 IRESN 具有很好的稳定性.

4 结论

为了较好地回声状态网络进行结构设计,本文利用奇异值分解方法提出了基于增量式学习的正则化回声状态网络 IRESN,其优点是能够得到与具体任务相匹配的回声状态网络,且能够自适应设定储备池及其相关参数. 在结构调整过程中,能够添加子模块至新的网络中,并采取迭代式方法学习输出权值,同时正则化参数根据留一交叉验证方法进行选择. 仿真结果表明,所提出的方法具有与具体任务相匹配的网络结构和较好的泛化能力. 不足之处在于, IRESN 的子模块神经元数量仍是凭经验设定,正则化参数的选择方法缺少一定的理论依据. 下一步,可以对回声状态网络进行结构实时动态在线调整,实现储备池自组织结构,掌握储备池的丰富动力学特性,依然是研究的重点和热点.

参考文献(References)

- [1] Jaeger H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication[J]. Science, 2004, 304(5667): 78-80.
- [2] Jaeger H. The echo state approach to analysing and training recurrent neural networks[R]. Bonn: German National Research Center for Information Technology GMD, 2001: 148.
- [3] 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, 2017, 28(2): 391-404.
- [4] Qiao J F, Wang L, Yang C L. Adaptive lasso echo state network based on modified Bayesian information criterion for nonlinear system modeling[J]. Neural Computing and Applications, 2019, 31(10): 6163-6177.
- [5] 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 Automatica Sinica, 2019, 45(6): 1136-1145.)
- [6] Pan Y P, Wang J. Model predictive control of unknown nonlinear dynamical systems based on recurrent neural networks[J]. IEEE Transactions on Industrial Electronics,

- 2012, 59(8): 3089-3101.
- [7] Xia Y L, Jelfs B, van Hulle M M, et al. An augmented echo state network for nonlinear adaptive filtering of complex noncircular signals[J]. IEEE Transactions on Neural Networks, 2011, 22(1): 74-83.
- [8] Liu L, Wang Z S, Yao X S, et al. Echo state networks based data-driven adaptive fault tolerant control with its application to electromechanical system[J]. IEEE/ASME Transactions on Mechatronics, 2018, 23(3): 1372-1382.
- [9] Strauss T, Wustlich W, Labahn R. Design strategies for weight matrices of echo state networks[J]. Neural Computation, 2012, 24(12): 3246-3276.
- [10] Xue Y B, Yang L, Haykin S. Decoupled echo state networks with lateral inhibition[J]. Neural Networks, 2007, 20(3): 365-376.
- [11] Rodan A, Tino P. Minimum complexity echo state network[J]. IEEE Transactions on Neural Networks, 2011, 22(1): 131-144.
- [12] Duan H B, Wang X H. Echo state networks with orthogonal pigeon-inspired optimization for image restoration[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(11): 2413-2425.
- [13] Otte S, Butz M V, Koryakin D, et al. Optimizing recurrent reservoirs with neuro-evolution[J]. Neurocomputing, 2016, 192: 128-138.
- [14] Dutoit X, Schrauwen B, van Campenhout J, et al. Pruning and regularization in reservoir computing[J]. Neurocomputing, 2009, 72(7/8/9): 1534-1546.
- [15] Yang C L, Qiao J F, Wang L, et al. Dynamical regularized echo state network for time series prediction[J]. Neural Computing and Applications, 2019, 31(10): 6781-6794.
- [16] 韩敏, 任伟杰, 许美玲. 一种基于 L_1 范数正则化的回声状态网络[J]. 自动化学报, 2014, 40(11): 2428-2435. (Han M, Ren W J, Xu M L. An improved echo state network via L_1 -norm regularization[J]. Acta Automatica Sinica, 2014, 40(11): 2428-2435.)
- [17] Stone M. Cross-validators choice and assessment of statistical predictions[J]. Journal of the Royal Statistical Society—Series B: Methodological, 1974, 36(2): 111-133.
- [18] Rivals I, Personnaz L. On cross validation for model selection[J]. Neural Computation, 1999, 11(4): 863-870.
- [19] Cornell J A. Classical and modern regression with applications[J]. Technometrics, 1987, 29(3): 377-378.
- [20] Lorenz E N. Deterministic nonperiodic flow[J]. Journal of the Atmospheric Sciences, 1963, 20(2): 130-141.
- [21] Qiao J F, Wang L, Yang C L, et al. Adaptive levenberg-marquardt algorithm based echo state network for chaotic time series prediction[J]. IEEE Access, 2018, 6: 10720-10732.
- [22] Tang Q H, Peng L, Yang Y, et al. Total phosphorus-precipitation and Chlorophyll a-phosphorus relationships of lakes and reservoirs mediated by soil iron at regional scale[J]. Water Research, 2019, 154: 136-143.
- [23] Wu Y, Luo J Y, Zhang Q, et al. Potentials and challenges of phosphorus recovery as vivianite from wastewater: A review[J]. Chemosphere, 2019, 226: 246-258.
- [24] Li F J, Qiao J F, Han H G, et al. A self-organizing cascade neural network with random weights for nonlinear system modeling[J]. Applied Soft Computing, 2016, 42: 184-193.

作者简介

王磊(1980—), 男, 讲师, 博士, 从事神经网络结构设计与优化、复杂过程建模与优化控制等研究, E-mail: leiw@bistu.edu.cn;

苏中(1962—), 男, 教授, 博士生导师, 从事高动态导航与控制、盲环境导航等研究, E-mail: sz@bistu.edu.cn;

乔俊飞(1968—), 男, 教授, 博士生导师, 从事智能控制与智能信息处理、复杂过程建模与优化控制等研究, E-mail: junfeiq@bjut.edu.cn;

赵静(1982—), 女, 副研究员, 博士, 从事统计推理、假设检验等研究, E-mail: zhaoj@cnis.ac.cn.

(责任编辑: 孙艺红)