



## 一种深度回声状态网络的输入尺度自适应算法

刘鹏<sup>1</sup>, 叶润<sup>1</sup>, 闫斌<sup>1</sup>, 谢茜<sup>2</sup>, 刘睿<sup>2</sup>

(1.电子科技大学自动化工程学院, 成都 611731; 2.国网四川省电力公司电力科学研究院, 成都 610041)

**摘要:** 深度回声状态网络是回声状态网络与深度学习思想的结合, 合理选取不同谱半径的内部状态矩阵和弱积分参数能有效增强深度回声状态网络的多尺度时域特性。利用数据可视化分析输出矩阵在不同网络层中的分布关系, 发现高层网络中部分神经元处于饱和工作状态且该状态抑制了网络动态预测能力。提出一种深度回声状态网络的输入矩阵自适应算法, 在对网络内部状态的均值和方差进行递推估计的基础上判断神经元饱和状态, 通过自适应调整各层输入权重的值来增强神经元动态性。数值计算结果表明, 基于输入尺度自适应算法的深度回声状态网络相对同等规模的单层回声状态网络对于动态系统的预测精度有成倍提升。

**关键词:** 回声状态网络; 动态系统; 广义逆算法; 多尺度时域特性; 机器学习

开放科学(资源服务)标志码(OSID):



中文引用格式: 刘鹏, 叶润, 闫斌, 等. 一种深度回声状态网络的输入尺度自适应算法[J]. 计算机工程, 2022, 48(2): 92-98, 105.

英文引用格式: LIU P, YE R, YAN B, et al. An adaptive algorithm of input scale for deep echo state networks[J]. Computer Engineering, 2022, 48(2): 92-98, 105.

## An Adaptive Algorithm of Input Scale for Deep Echo State Networks

LIU Peng<sup>1</sup>, YE Run<sup>1</sup>, YAN Bin<sup>1</sup>, XIE Qian<sup>2</sup>, LIU Rui<sup>2</sup>

(1. School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China;

2. Electric Power Research Institute of State Grid Sichuan Electric Power Company, Chengdu 610041, China)

**[Abstract]** Deep Echo State Networks (DESN) is a combination of Echo State Networks (ESN) and the idea of deep learning. A reasonable selection of internal state matrices and weak integration parameters with different spectral radius can effectively enhance the multi-scale time domain characteristics of the DESN. By analyzing the distribution of output matrix in different network layers through data visualization, it is found that part of the neurons in higher network layers are partially working in a saturated state, which weakens the dynamic prediction of the network. An adaptive algorithm of input matrix for DESN is proposed, based on the recursive estimation of the mean and variance of the internal network state, whether a neuron is saturated is judged. Then the output weight of each layer is adjusted adaptively to improve neuron dynamics. The numeric analysis results show that the DESN based on the input scale adaptive algorithm has doubled the prediction accuracy of the dynamic system compared with the single-layer ESN of the same scale.

**[Key words]** Echo State Networks (ESN); dynamic system; generalized inverse algorithm; multi-scale time domain characteristic; machine learning

DOI: 10.19678/j.issn.1000-3428.0060532

### 0 概述

深度神经网络在图像处理<sup>[1]</sup>、语音识别<sup>[2-3]</sup>、数据建模<sup>[4]</sup>等领域得到广泛应用并取得重要研究成果, 显示出层级神经网络结构在特征学习与迁移中的性能优势。在自然语言处理(Natural Language Processing, NLP)、基于视觉的同时定位与地图构建(Vision-based Simultaneous Localization And Mapping,

VSLAM)<sup>[5]</sup>等实际工程应用中, 研究的系统一般为与时间相关的动态系统。对于动态系统, 典型特征是系统的状态随时间动态变化, 输入与输出之间是一个暂态过程。递归神经网络(Recurrent Neural Networks, RNN)<sup>[6]</sup>由于神经元在时域上递归连接, 使其具有丰富的时域行为, 因此在动态系统理论研究和实际应用中受到广泛关注。深度卷积神经网络(Deep Convolutional Neural Networks, DCNN)<sup>[7]</sup>的

基金项目: 国家自然科学基金(61703060, 61973055); 四川省科技计划项目(2019YJ0165); 中央高校基本科研业务费专项资金(ZYGX2020J011)。

作者简介: 刘鹏(1995—), 男, 硕士, 主研方向为机器学习及其应用; 叶润, 助理研究员; 闫斌, 讲师; 谢茜, 高级工程师; 刘睿, 教授级高级工程师。

收稿日期: 2021-01-08 修回日期: 2021-02-21 E-mail: pengliustd@163.com

网络层从低层到高层,所提取的特征由简单到复杂,正是因为这种层级结构,使得网络更容易训练,学习到的特征更容易泛化与迁移。DCNN主要使用前向连接,如果将其运用到动态系统学习中,可以使用一个滑动窗口输入一个时间段内的信息。一般需要学习的动态系统延时特性是未知的<sup>[4]</sup>,也就难以确定最优滑动窗口的长度,而且不同动态系统的延时特性不同,可能会使神经网络产生振荡、性能下降,甚至失去稳定性<sup>[8]</sup>。

研究人员通常利用RNN提高动态系统的学习精度,其中长短期记忆(Long Short Term Memory, LSTM)网络<sup>[9]</sup>是应用最为广泛的RNN之一,由HOCHREITER等在1997年提出,目前已经集成在TensorFlow等AI开发平台中。LSTM通常采用梯度下降方法进行学习<sup>[10]</sup>,在一次权值调整过程中需要使用过去时刻的部分或全部数据,学习难度相对于前向网络成倍增加。JAEGER<sup>[11]</sup>在2005年提出一种新型RNN,因输入信号在神经元之间来回衰减振荡类似于回声而被称作回声状态网络(Echo State Networks, ESN)。ESN只需训练输出矩阵,训练过程等价于线性规划问题,相对于LSTM而言训练难度显著降低,因训练简单,已在时间序列预测<sup>[12]</sup>、混沌系统学习<sup>[13-14]</sup>等任务中取得成功应用。SCHRAUWEN等<sup>[15]</sup>利用ESN的内部可塑性(Intrinsic Plasticity, IP)使得在特定输入下每一个神经元的输出都服从指数分布。IP方法能在很大程度上提升ESN的表达力,但要每一个神经元的输入进行调整,大幅增加了学习过程的计算代价。ESN应用过程中需要选择内部状态矩阵的谱半径、输入尺度缩放系数等参数。为了最大限度发挥ESN的学习能力,BIANCHI等<sup>[16]</sup>提出均值递归分析方法,使用图形定性分析和定量分析判据,将ESN推向临界稳定状态。

GALLICCHIO等<sup>[17-18]</sup>提出一种基于ESN的深度回声状态网络(Deep ESN, DESN)。文献[17]通过严格的实验分析多种输入方式对DESN不同层之间的延时特性的影响。实验结果表明,堆栈形式的DESN表现出更好的多尺度时域特性,而且合理选用不同谱半径的内部状态矩阵和弱积分参数能增强这种特性。文献[18]对如何选用DESN层数进行了深入研究,提出一种基于频谱分析的层数选用方法。虽然文献[19-20]已将DESN应用到一些实际工程中,但目前关于DESN网络结构特性的研究尚少。文献[17-18]因为侧重点是在分析时域特性而不是预测精度,所以并未与传统ESN进行预测精度对比,而在实际工程应用中,所需学习模型的时域特性本身是未知的,不能有效表明DESN对不同延时系统预测精度的影响。为更好地利用DESN的多尺度时域特性,本文将DESN内部状态和输出矩阵进行可

视化分析提出一种输入尺度自适应算法,通过学习两种数学模型完全已知的动态系统,分析改进后DESN的性能优势,并与ESN进行预测精度对比。

## 1 DESN结构与训练算法

DESN的基本单元是浅层ESN。本节首先介绍浅层ESN结构和数学形式,然后介绍DESN结构和基于广义逆的训练算法。

### 1.1 浅层回声状态网络

浅层ESN的基本结构如图1所示,其中, $\mathbf{u}$ 表示输入向量, $\mathbf{W}_i$ 表示输入权重, $\mathbf{W}$ 表示内部状态矩阵, $\mathbf{W}_{o1}$ 表示内部状态到输出的权重, $\mathbf{W}_{o2}$ 表示输入直连到输出的权重, $\mathbf{W}_b$ 表示输出反馈矩阵。ESN对参数选择的唯一限制就是需要满足回声状态特性<sup>[11]</sup>,等价于状态遗忘或输入遗忘特性,也就是说内部状态的初值以及某个时刻的输入在经历足够长的时间后,对系统输出产生的影响可以忽略不计。

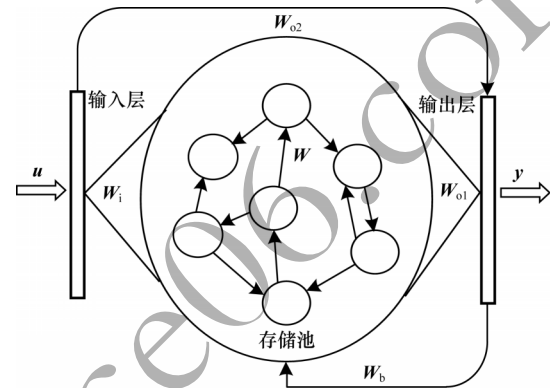


图1 浅层ESN基本结构

Fig.1 Basic structure of shallow ESN

输出反馈的引入能加快ESN初始化过程,使内部状态初值的影响快速消失,但由于输出反馈会使回声状态特性受输出矩阵的影响,ESN的鲁棒性会降低。一般在进行理论分析时,不考虑输出反馈的引入,而在实际工程应用中则可以通过引入输出反馈优化学习精度。

ESN的数学表达形式可以分为内部状态 $\mathbf{x}$ 和输出向量 $\mathbf{y}$ 两部分,如式(1)所示:

$$\begin{cases} \mathbf{x}(n+1) = \alpha \mathbf{x}(n) + (1-\alpha) f(\mathbf{W}_i \mathbf{u}(n+1) + \mathbf{W} \mathbf{x}(n)) \\ \mathbf{y}(n+1) = [\mathbf{W}_{o1} \quad \mathbf{W}_{o2}] \begin{bmatrix} \mathbf{x}(n+1) \\ \mathbf{u}(n+1) \end{bmatrix} = \mathbf{W}_o \begin{bmatrix} \mathbf{x}(n+1) \\ \mathbf{u}(n+1) \end{bmatrix} \end{cases} \quad (1)$$

其中:弱积分参数 $\alpha$ 表示当前状态与上一时刻状态的关联程度,满足 $0 \leq \alpha < 1$ ,通常而言参数 $\alpha$ 取决于所需学习的动态系统和采样时间间隔;激活函数 $f$ 选取tanh双曲正切函数,单个神经元的输出范围为 $-1 \sim 1$ 。设 $\rho$ 和 $\sigma$ 分别表示 $\mathbf{W}$ 的谱半径和奇异值。若 $\sigma < 1$ ,则回波状态特性一定能得到保证,若 $\rho < 1$ ,则回波状态特性在大多数情况下能得到保证。

## 1.2 深度回声状态网络

如图2所示,DESN结构是一种堆栈形式,外部输入只作用于第一层神经元,其他每一层都以前一层输出为输入。如果将外部输入引入到每一层,则DESN的多尺度时域特性会被削弱。

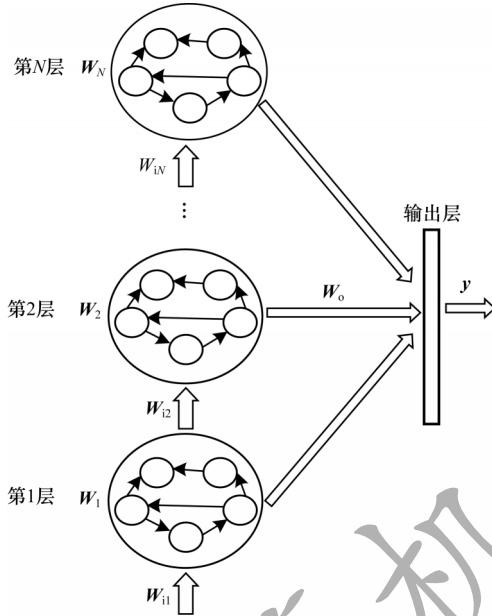


图2 DESN结构

Fig.2 Structure of DESN

每一层的状态如式(2)所示,形式上与式(1)保持一致。在DESN中,不同层之间可以选择不同的弱积分参数和不同的谱半径。

$$\mathbf{x}_k(n+1) = \alpha_k \mathbf{x}_k(n) + (1 - \alpha_k) f(\mathbf{W}_{ik} \mathbf{u}_k(n+1) + \mathbf{W}_k \mathbf{x}_k(n)) \quad (2)$$

其中: $k=1, 2, \dots, N$ ;  $\mathbf{u}_1$ 表示外部输入,  $\mathbf{u}_2 \sim \mathbf{u}_N$ 表示前一层的输出。

$$\mathbf{y}(n+1) = \mathbf{W}_o \begin{bmatrix} \mathbf{x}_1(n+1) \\ \mathbf{x}_2(n+1) \\ \vdots \\ \mathbf{x}_N(n+1) \\ \mathbf{u}_1(n+1) \end{bmatrix} \quad (3)$$

DESN和ESN只需训练输出矩阵 $\mathbf{W}_o$ 。输入矩阵 $\mathbf{W}_i$ 选为元素值在-1~1的随机矩阵, $\mathbf{W}$ 的选取需保证满足回声状态特性,参数 $\alpha$ 可以在不同层之间选择不同的值。

**算法1** 基于广义逆算法的输出矩阵训练

**步骤1** 初始化各层输入矩阵 $\mathbf{W}_{ik}$ 为-1~1的随机矩阵,随机生成各层状态矩阵 $\mathbf{W}_k$ ,计算 $\mathbf{W}_k$ 的奇异值 $\sigma_k$ ,将 $\mathbf{W}_k$ 除以奇异值 $\sigma_k$ 。

**步骤2** 选取训练集 $(\mathbf{u}(t), \mathbf{Y}_d(t))$ ,其中 $0 < t \leq T$ ,根据式(2)收集数据,如式(4)所示:

$$\mathbf{H} = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_N(1) & \mathbf{u}(1) \\ x_1(2) & x_2(2) & \cdots & x_N(2) & \mathbf{u}(2) \\ \vdots & \vdots & & \vdots & \vdots \\ x_1(T) & x_2(T) & \cdots & x_N(T) & \mathbf{u}(T) \end{bmatrix} \quad (4)$$

**步骤3** 使用广义逆算法计算输出矩阵,如式(5)所示:

$$\mathbf{W}_o = \mathbf{H}^+ \mathbf{Y}_d \quad (5)$$

## 2 DESN可视化分析

为研究DESN多尺度时域特性,选用具有不同延时特性的动态系统作为被学习对象。通过对比DESN与普通ESN的预测均方根误差(Root Mean Square Error, RMSE),分析DESN多尺度时域特性在预测精度上的优势和限制因素。

### 2.1 数值计算模型设定

参考滑动平均模型NARX<sup>[21]</sup>,并添加延时项使动态系统有更加丰富的时域特性,如式(6)所示:

$$\mathbf{y}(n) = 0.7\mathbf{u}(n) + 0.3\mathbf{u}^2(n-d) + \mathbf{y}(n) - \mathbf{y}^2(n) \quad (6)$$

其中: $d$ 表示延时, $d=0, 1, \dots, 19$ ;输入 $\mathbf{u}$ 为0~1的均匀分布。

ESN选取弱积分参数为0.7、神经元个数为500,DESN选取弱积分参数为0.0, 0.1,  $\dots$ , 0.9,共有10层,每层的神经元个数为50。ESN和DESN的激活函数均为tanh,训练算法使用广义逆算法。训练样本和测试样本均取自式(6)所示的动态系统,第1次实验延时设置为0,第2次实验延时设置为1,依此类推直至延时为19。在每次数值分析过程中,训练集和测试集都采样1000组数据。同时,为了避免初始状态设置对动态系统的影响,前50步迭代的数据被丢弃。

### 2.2 可视化分析

图3给出了DESN与ESN的预测RMSE随延迟时间的变化曲线,大致可分为三个阶段:第一阶段,延时弱影响阶段,DESN和ESN预测RMSE相近;第二阶段,延时较影响阶段,DESN预测RMSE明显小于ESN;第三阶段,延时强影响阶段,DESN预测RMSE大于ESN预测RMSE。在第一阶段中,由于延时特性不明显,因此DESN和ESN预测性能相近,此时DESN多尺度时域特性的优势并未显现。在第二阶段中,当 $d=5$ 时,DESN相对于ESN预测RMSE降低超过50%。由于DESN更高层的网络记忆输入的时间更长<sup>[17-18]</sup>,如果在第二阶段中连接高层网络的输出矩阵绝对值相对增大,那么就证明DESN多尺度时域特性提升了其预测能力。在第三阶段中,DESN预测性能劣于ESN,只有进一步分析,才能解释该阶段的反常现象。

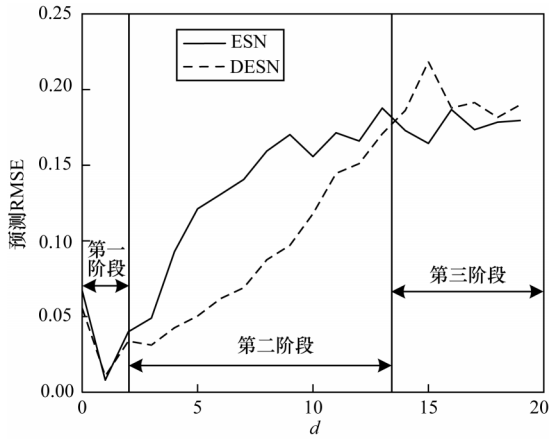


图3 DESN与ESN学习不同延时NARX模型的预测RMSE对比  
Fig.3 Comparison of forecast RMSE between DESN and ESN when learn NARX model with different delays

图4给出了DESN的 $d$ 取0、5、15时的输出矩阵 $W$ 在各层中的分布情况。从图4(a)可以看出,输出矩阵绝对值大的元素全都分布在第1层和输入直连输出中。图4(b)为图4(a)的局部放大图,从图4(b)中并未发现输出矩阵值会随着学习系统的延时增加而发生明显移动,由此得出图3中第二阶段DESN预测能力的提升不能归结为多尺度时域特性的影响。

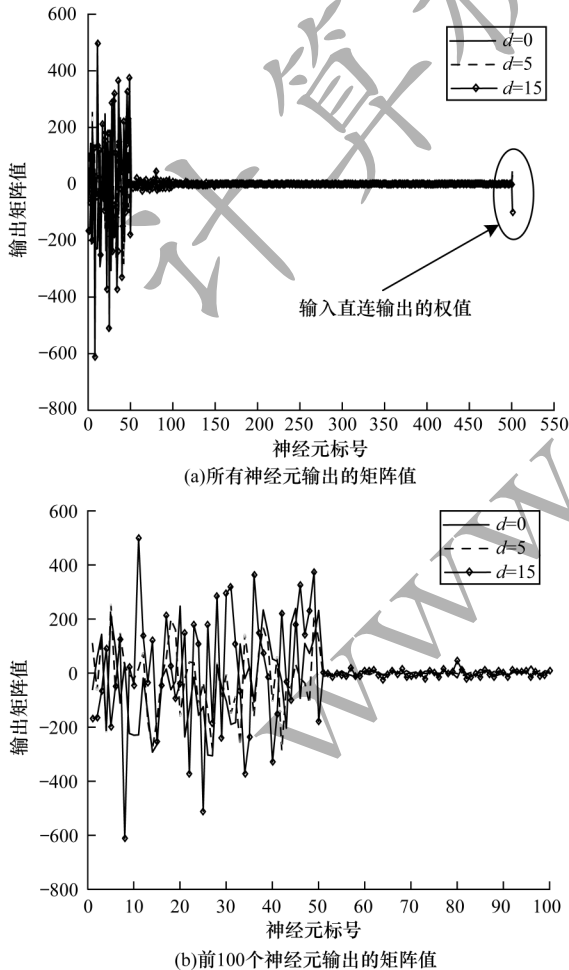


图4 不同延时的DESN输出矩阵对比  
Fig.4 Comparison of DESN output matrices with different delays

图5给出了 $d=15$ 时DESN第1、5、10层部分神经元的内部状态,其中横坐标 $t_{iteration}$ 表示迭代时间。在第5和10层中可以看出,超过一半神经元处于饱和和工作状态,即绝对值在大部分时间都维持在一个较高水平。这类神经元不仅自身动态性能差,而且会使得其他具有丰富动态的神经元、输出权值极小。实际上,从第3层开始,在后续层中大量神经元处于这种饱和和工作状态。从内部状态可视化分析可知,在DESN高层网络中容易出现处于饱和和工作状态的神经元,如果能够抑制这种饱和状态,则有可能更好地发挥DESN多尺度时域特性的优势。

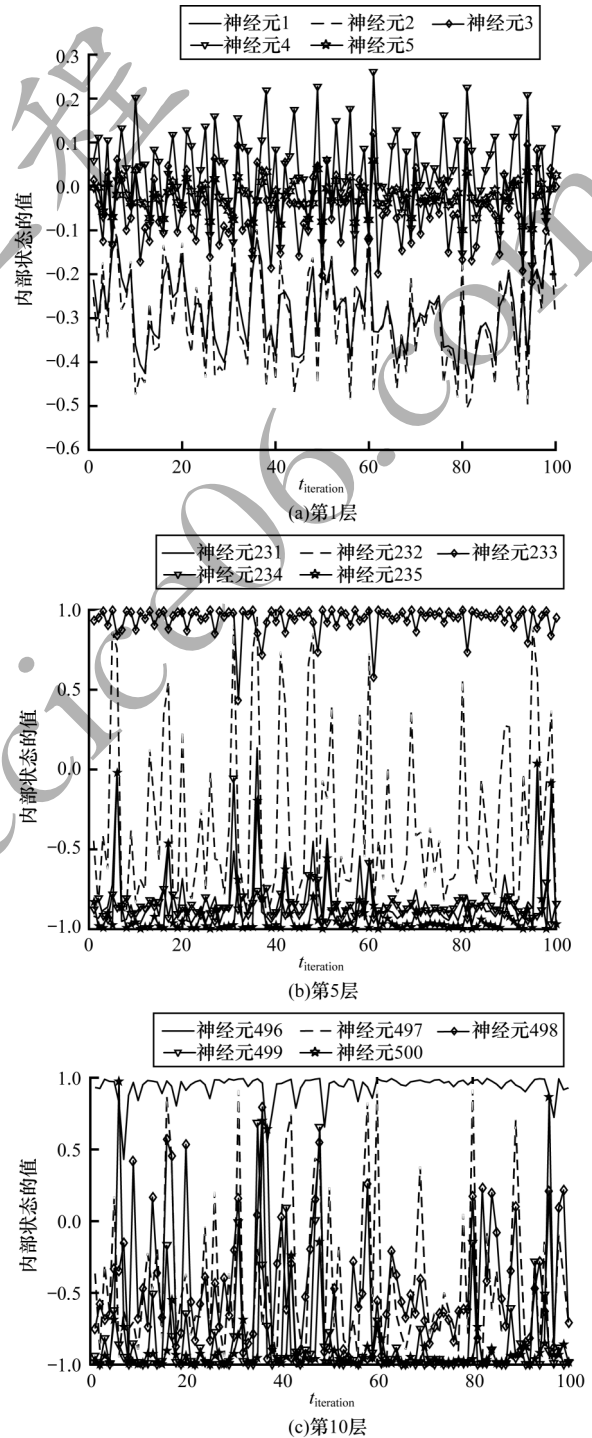


图5 DESN内部状态  
Fig.5 Internal state of DESN

### 3 DESN输入尺度自适应算法

由于DESN每一层的网络都满足回声状态特性,也就是说只要时间足够长,输入信号会被逐渐遗忘。然而,当输入信号足够强时,遗忘的速度小于输入的速度,使得部分神经元处于饱和和工作状态。为了保持DESN多尺度时域特性,需要避免神经元处于饱和和工作状态。

判断神经元是否处于饱和和工作状态主要包括状态保持恒定及维持较高值两个条件。状态保持恒定意味着方差小。在方差小的情况下维持高值意味着绝对均值相对较大。通过设定合适的方差与均值的阈值,综合判断神经元是否处于饱和状态。如果处于饱和状态,则意味着输入速度要大于遗忘速度,需减弱输入。同一层的神经元通过内部状态矩阵 $W$ 彼此互联,对一个神经元的输入会立即影响其他神经元。因此,对于具有饱和和工作状态的神经元的层,应该整体减小输入才能有效抑制饱和和工作状态。另外,每一层处于饱和和工作状态的神经元个数不同,处于饱和状态的神经元越多,输入的衰减程度就越大。

#### 算法2 饱和状态抑制

**步骤1** 对于内部状态序列 $x(t)$ ,  $0 < t \leq T$ ,其均值和方差分别可按式(7)和式(8)递推估计,初始化过程中的数据不能表达网络的稳定工作状态,不纳入均值和方差估计。

$$m(t+1) = m(t) + 1/(t+1) [x(t+1) - m(t)] \quad (7)$$

$$D(t+1) = (1-1/t)D(t) + (1/t)(m(t+1) - m(t))^2 \quad (8)$$

**步骤2** 判断当前时刻是否需要进行一次权值调整,判断依据是与上一次权值调整的时间间隔是否达到阈值 $\Delta T$ 。若需要调整,则进入步骤3,否则返回步骤1继续递推估计均值和方差。

**步骤3** 逐层分析各个神经元的均值和方差,若 $|m| > m_{\min}$ 且 $D < D_{\max}$ ,则判断该神经元处于饱和状态, $D_{\max}$ 和 $m_{\min}$ 都为可选的阈值。

**步骤4** 分层统计处于饱和状态的神经元个数 $n$ ,第 $k$ 层输入权重按式(9)动态调整:

$$W_{ik}(t+1) = W_{ik}(t)(1 - \eta n) \quad (9)$$

在算法2中,衰减系数 $\eta$ 是关键参数,若值过大会使内部状态工作在零值附近,若值太小会达不到调整输入权重的效果。使用算法2改进的DESN称为输入尺度自适应深度回声状态网络(Adaptive DESN, ADESN)。

#### 算法3 ADESN训练

**步骤1** 选定一个衰减系数 $\eta$ ,在训练集上执行算法2,调整输入矩阵 $W_{ik}$ ,  $k=1, 2, \dots, N$ 。

**步骤2** 在训练集上训练执行算法1,得到输出矩阵。

**步骤3** 在验证集上测试DESN预测性能,计算预测RMSE。

**步骤4** 重复进行步骤1~步骤3,选取验证集上

预测RMSE最小的衰减系数 $\eta$ 作为最终参数。

### 4 数值计算结果与分析

通过学习NARX模型和单输入单输出系统来验证ADESN的预测精度提升情况,并使用一个多输入多输出(Multiple Input Multiple Output, MIMO)系统<sup>[21]</sup>测试ADESN的适用性。

数据集共分为训练集、验证集和测试集,每一个样本采样数量为1000。DESN和ESN参数设置参考2.1节。ADESN衰减参数 $\eta$ 、方差阈值 $D_{\max}$ 和均值阈值 $m_{\min}$ 需设定,将在4.1节中进行讨论。

#### 4.1 ADESN参数选择

$D_{\max}$ 和 $m_{\min}$ 的选择会对神经元是否处于饱和状态的判断产生影响。如果 $D_{\max}$ 值选择过大,条件 $D < D_{\max}$ 容易满足,则大量神经元将被判断为饱和状态。 $m_{\min}$ 用于排除均值较小的神经元,如果值过大,大部分神经元都会判断为非饱和状态。对于ADESN而言,内部有成百上千的神经元。 $D_{\max}$ 和 $m_{\min}$ 的选择会影响各层判断为饱和状态的神经元数量。由算法2的描述可知,输入权值的调整是逐层调整而并不是逐个神经元调整,这就意味着算法对于参数 $D_{\max}$ 和 $m_{\min}$ 并不敏感,在每次调整权值的过程中都依赖衰减系数 $\eta$ 的选取,衰减系数 $\eta$ 的微小调整都可能由于累积效应而对ADESN产生较大影响,也就是说ADESN对参数 $\eta$ 较敏感。

为了避免参数之间的耦合,设定 $D_{\max} = 1 \times 10^{-4}$ 、 $m_{\min} = 0.1$ ,分析关键参数 $\eta$ 对ADESN的影响。选取ADESN衰减系数 $\eta$ 区间为0.001~0.01,衰减系数 $\eta$ 对不同延时的NARX模型预测RMSE的影响如图6所示。从图6可以看出:当延时 $d=0$ 时, RMSE随 $\eta$ 的变化略有波动;当延时 $d=5$ 时, RMSE随 $\eta$ 的增大先减小后增大;当延时 $d=15$ 时, RMSE随 $\eta$ 的增大逐渐减小并趋于稳定。由此可见, $\eta$ 对系统的影响与系统的时延特性有关,系统时延越大调整 $\eta$ 产生的影响越大。因此,参数 $\eta$ 需按算法3中描述针对不同时延的动态系统进行优化处理。

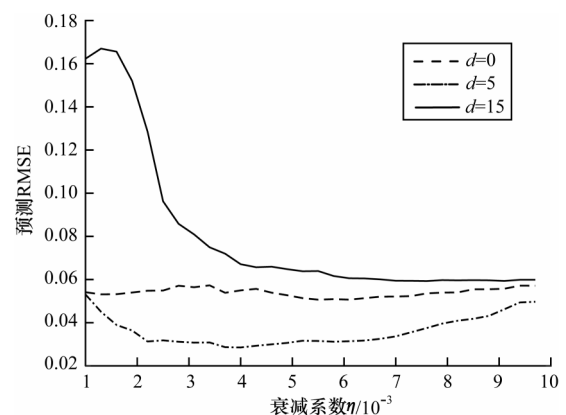


图6 不同延时测试集上衰减系数 $\eta$ 对预测RMSE的影响  
Fig.6 Effect of attenuation coefficient  $\eta$  to forecast RMSE on test set with different delays

### 4.2 NARX 模型预测

ESN、DESN 和 ADESN 在不同延时的 NARX 模型上的预测 RMSE 对比如图 7 所示,具体数据如表 1 所示。由图 7 和表 1 可以看出:ADESN 在全过程中都具有最佳预测精度,当  $d$  超过 5 时预测 RMSE 约为 ESN 的 30%,极大地提升了高延时系统的预测精度。ADESN 和 ESN 具有的神经元数目相同,但 ADESN 对于高延时系统能显著提升预测精度,其原因为充分发挥了多层结构的多尺度时域特性的优势。

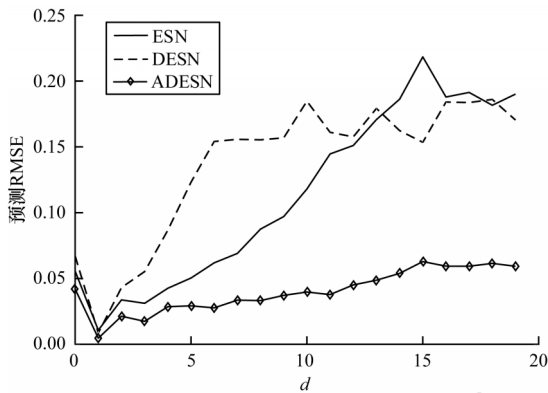


图 7 不同延时的 ESN、DESN、ADESN 预测 RMSE 对比曲线  
Fig.7 Comparison curves of forecast RMSE among ESN, DESN and ADESN under different delays

表 1 不同延时的 ESN、DESN、ADESN 预测 RMSE 对比数据  
Table 1 Comparison data of forecast RMSE among ESN, DESN and ADESN under different delays

$d$	ADESN	DESN	ESN
0	0.042	0.056	0.067
5	0.029	0.050	0.121
15	0.063	0.218	0.164

从图 8 可以看出,随着延时的增加,权值逐步向高层网络移动。高层网络记忆输入的时间更长,如果高层网络对应的权值更大,则意味着系统的延时特性更强。这种随着网络层次增加对输入的记忆时间增强的特性就是多层结构的多尺度时域特性。

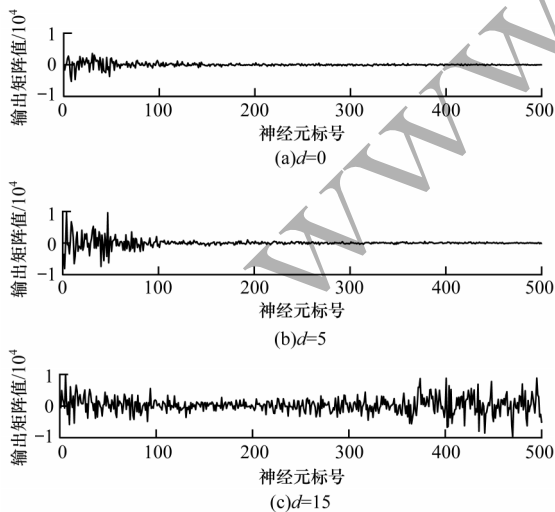


图 8 不同延时下 ADESN 输出矩阵在各层神经元之间的分布  
Fig.8 Distribution of ADESN output matrixes among neurons in each layer under different delays

图 9 给出了  $d=15$  时 ADESN 第 1、5、10 层部分神经元的内部状态。相对于图 5 中 DESN 内部状态, ADESN 在各层网络中不再有处于饱和工作状态的神经元,各层饱和状态神经元的消除一方面使得网络动态性更强,另一方面也使本来处于正常状态的神经元获得较大的输出权值。综上所述, ADESN 能够消除处于饱和状态的神经元,充分发挥多尺度时域特性的优势。

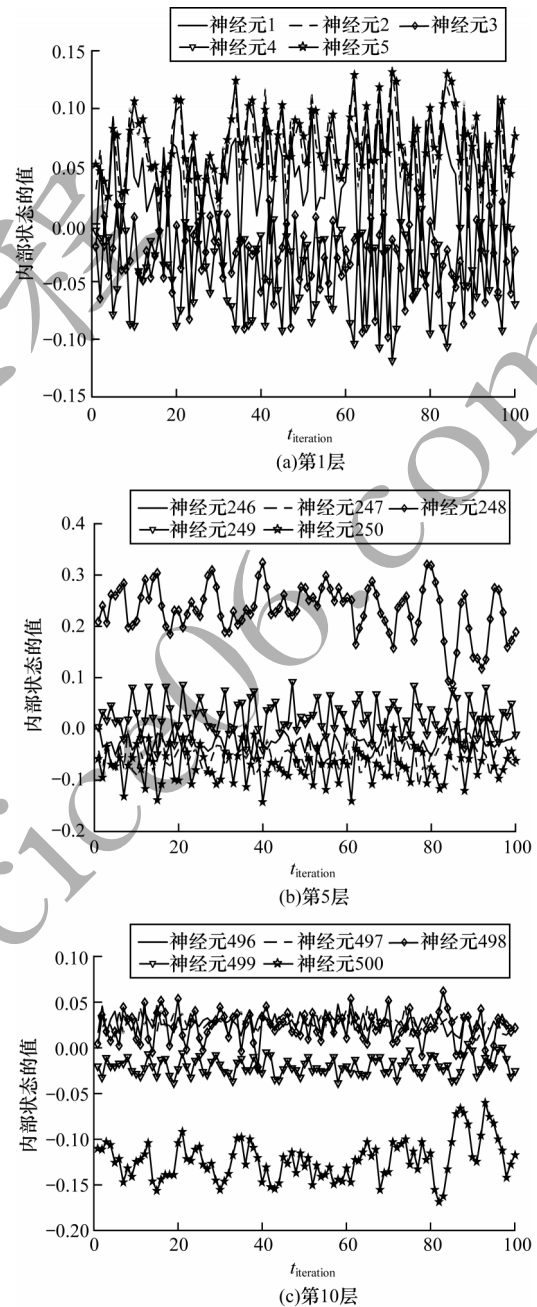


图 9 ADESN 内部状态

Fig.9 Internal state of ADESN

### 4.3 多输入多输出系统预测

应用多输入多输出系统(如式(10)、式(11)所示)进一步测试 ADESN 多尺度时域特性优势。

$$y_1(t) = \frac{0.75y_1(t-1)}{1+y_2^2(t-1)} + u_1(t-2)u_2(t-1) + 0.1y_2(t-1)u_1(t-1) + 0.5v_1(t-1) + 0.2y_1(t-2)v_1(t-1) + v_1(t) \quad (10)$$

$$y_2(t) = \frac{0.75y_2(t-1)}{1+y_2^2(t-1)} + \frac{0.85u_2^2(t-1)}{2+u_1^2(t-1)} + 0.2y_2(t-1)u_2(t-2) + 0.5v_2(t-1) + 0.1y_2(t-1)v_2(t-1) + v_2(t) \quad (11)$$

其中： $u_1$ 和 $u_2$ 表示值在0~1的均匀分布； $v_1$ 和 $v_2$ 表示均值为0、方差为0.00~0.02的高斯噪声。由于方差不同，因此以0.001为间隔共测试20种不同工况。

图10给出了ESN、DESN和ADESN在不同噪声条件下对多输入多输出系统的预测RMSE，具体数据如表2所示。由图10和表2可以看出，ADESN依然具有最佳的预测能力，但是随着噪声的增加，3种网络都严重劣化，在后期研究工作中需考虑提升网络对噪声的鲁棒性。

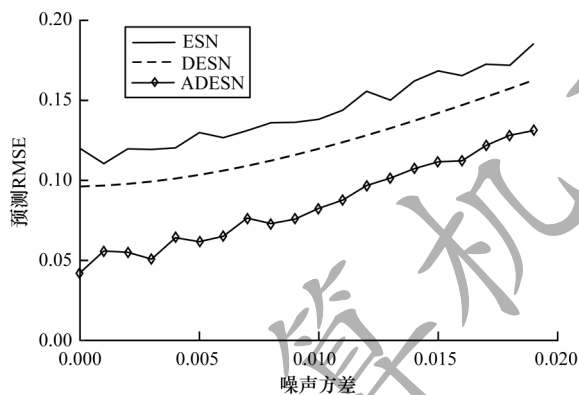


图10 不同噪声方差的ESN、DESN、ADESN预测RMSE对比曲线

Fig.10 Comparison curves of forecast RMSE among ESN, DESN and ADESN under different noise variances

表2 不同噪声方差的ESN、DESN、ADESN预测RMSE对比数据

Table 2 Comparison data of forecast RMSE among ESN, DESN and ADESN under different noise variances

噪声方差	ADESN	DESN	ESN
0.00	0.042	0.096	0.120
0.01	0.062	0.103	0.130
0.02	0.112	0.142	0.168

## 5 结束语

针对传统DESN内部状态容易陷入饱和和工作状态，不能重复利用多尺度时域特性的问题，本文通过自适应地调整输入矩阵构建一种深度回声状态网络ADESN，使用网络内部状态的均值和方差判断神经元是否处于饱和状态，加强网络动态预测能力。对两个不同动态系统的学习结果表明，ADESN相对DESN和ESN预测精度成倍提升。在后续研究中将改进ADESN的抗噪能力，进一步提升其在实际噪声环境中的预测精度。

## 参考文献

- [1] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 770-778.
- [2] CAMBRIA E, WHITE B. Jumping NLP curves: a review of natural language processing research[J]. IEEE Computational Intelligence Magazine, 2014, 9(2): 48-57.
- [3] 张征. 基于深度神经网络的汉语语音合成的研究[D]. 北京: 北京理工大学, 2014.  
ZHANG Z. Research on deep neural network based Chinese speech synthesis [D]. Beijing: Beijing Institute of Technology, 2014. (in Chinese)
- [4] LIN Y Y, CHANG J Y, LIN C T. Identification and prediction of dynamic systems using an interactively recurrent self-evolving fuzzy neural network[J]. IEEE Transactions on Neural Networks and Learning Systems, 2013, 24(2): 310-321.
- [5] WEISS S, ACHELNIK M W, LYNNEN S, et al. Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environments[C]//Proceedings of 2012 IEEE International Conference on Robotics and Automation. Washington D. C., USA: IEEE Press, 2012: 957-964.
- [6] 鲍刚. 基于忆阻递归神经网络的联想记忆分析与设计[D]. 武汉: 华中科技大学, 2012.  
BAO G. Analysis and design for associative memory based on delayed recurrent neural network with memristor[D]. Wuhan: Huazhong University of Science and Technology, 2012. (in Chinese)
- [7] KHAN A, SOHAIL A, ZAHOORA U, et al. A survey of the recent architectures of deep convolutional neural networks[J]. Artificial Intelligence Review, 2020, 53(8): 5455-5516.
- [8] SHIK B, LIU X Z, TANG Y Y, et al. Some novel approaches on state estimation of delayed neural networks[J]. Information Sciences, 2016, 372: 313-331.
- [9] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [10] HE H, YI S, LIU W W. Intelligent learning model of financial spoken English teaching based on BPTT algorithm and LSTM network model[J]. Journal of Intelligent & Fuzzy Systems, 2020, 39(4): 4835-4846.
- [11] JAEGER H. Reservoir riddles: suggestions for echo state network research [C]//Proceedings of 2005 IEEE International Joint Conference on Neural Networks. Washington D. C., USA: IEEE Press, 2005: 1460-1462.
- [12] XU M L, HAN M. Adaptive elastic echo state network for multivariate time series prediction[J]. IEEE Transactions on Cybernetics, 2016, 46(10): 2173-2183.
- [13] 张杰烁, 刘明, 李鑫, 等. 基于递归最小二乘法的回声状态网络算法用于心电信号降噪[J]. 生物医学工程学杂志, 2018, 35(4): 539-549.  
ZHANG J S, LIU M, LI X, et al. An echo state network algorithm based on recursive least square for electrocardiogram denoising[J]. Journal of Biomedical Engineering, 2018, 35(4): 539-549. (in Chinese)

(下转第105页)

(上接第 98 页)

- [14] 赵宸. 基于回声状态网络的卫星信道在线盲均衡算法研究[D]. 兰州:兰州大学, 2020.  
ZHAO B. Study on online blind equalization algorithm for satellite channel based on echo state network[D]. Lanzhou: Lanzhou University, 2020. (in Chinese)
- [15] SCHRAUWEN B, WARDERMANN M, VERSTRAETEN D, et al. Improving reservoirs using intrinsic plasticity[J]. *Neurocomputing*, 2008, 71(7/8/9): 1159-1171.
- [16] BIANCHI F M, LIVI L, ALIPPI C. Investigating echo-state networks dynamics by means of recurrence analysis[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(2): 427-439.
- [17] GALLICCHIO C, MICHELI A, PEDRELLI L. Deep reservoir computing: a critical experimental analysis[J]. *Neurocomputing*, 2017, 268: 87-99.
- [18] GALLICCHIO C, MICHELI A, PEDRELLI L. Design of deep echo state networks[J]. *Neural Networks*, 2018, 108: 33-47.
- [19] MCDERMOTT P L, WIKLE C K. Deep echo state networks with uncertainty quantification for spatio-temporal forecasting[J]. *Environmetrics*, 2019, 30(3): 1-23.
- [20] SUN X C, LI T, LI Q, et al. Deep belief echo-state network and its application to time series prediction[J]. *Knowledge-Based Systems*, 2017, 130: 17-29.
- [21] YAO X S, WANG Z S, ZHANG H G. Prediction and identification of discrete-time dynamic nonlinear systems based on adaptive echo state network[J]. *Neural Networks*, 2019, 113: 11-19.

编辑 陆燕菲