

## 自适应权重的级联增强节点的宽度学习算法

蔡欣雨, 冯翔, 虞慧群

### 引用本文

蔡欣雨, 冯翔, 虞慧群. 自适应权重的级联增强节点的宽度学习算法[J]. 计算机科学, 2022, 49(6): 134-141.

CAI Xin-yu, FENG Xiang, YU Hui-qun. Adaptive Weight Based Broad Learning Algorithm for Cascaded Enhanced Nodes[J]. Computer Science, 2022, 49(6): 134-141.

---

### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于混合样本自动数据增强技术的半监督学习方法](#)

Semi-supervised Learning Method Based on Automated Mixed Sample Data Augmentation Techniques  
计算机科学, 2022, 49(3): 288-293. <https://doi.org/10.11896/jsjcx.210100156>

#### [融合多策略数据增强的低资源依存句法分析方法](#)

Improving Low-resource Dependency Parsing Using Multi-strategy Data Augmentation  
计算机科学, 2022, 49(1): 73-79. <https://doi.org/10.11896/jsjcx.210900036>

#### [基于 i\\_ResNet34 模型和数据增强的深度伪造视频检测方法](#)

Deepfake Videos Detection Method Based on i\_ResNet34 Model and Data Augmentation  
计算机科学, 2021, 48(7): 77-85. <https://doi.org/10.11896/jsjcx.210300258>

#### [基于智能数据增强和改进 YOLOv3 算法的接触网吊弦及支架检测研究](#)

Study on Catenary Dropper and Support Detection Based on Intelligent Data Augmentation and Improved YOLOv3  
计算机科学, 2020, 47(11A): 178-182. <https://doi.org/10.11896/jsjcx.200200053>

#### [流式数据处理的动态自适应缓存策略研究](#)

Study on Dynamic Adaptive Caching Strategy for Streaming Data Processing  
计算机科学, 2020, 47(11): 122-127. <https://doi.org/10.11896/jsjcx.190800093>

# 自适应权重的级联增强节点的宽度学习算法

蔡欣雨 冯翔 虞慧群

华东理工大学计算机科学与工程系 上海 200237

上海智慧能源工程技术研究中心 上海 200237

(787997078@163.com)

**摘要** 进入智能化时代,需要在大数据平台上进行持续自主学习和优化,而持续自主学习的第一步就是进行数据增强。文中提出基于级联增强节点的宽度学习方法,为大数据平台上的持续自主学习提供了新的数据增强方法,也为后续在学习架构基础上的演化优化提供了可能。以时序预测问题为依托,但由于经典宽度学习是典型的前馈神经网络,并不适合建模动态时间序列,因此在传统的宽度学习系统中引入反馈结构,将增强节点层顺序连接,使得增强节点具有记忆性,能够保留部分历史信息。在进行特征提取时,采用了相空间重构来提取数据更本质的特征;同时,引入了权重因子,在训练时依据每个样本对模型的贡献度,为其独立分配不同的权重,从而消除噪声和离群点对学习过程的干扰,提高算法的预测准确率以及鲁棒性。实验结果表明所提算法是有效的。

**关键词**:宽度学习;时序预测;权重因子;数据增强

**中图分类号** TP183

## Adaptive Weight Based Broad Learning Algorithm for Cascaded Enhanced Nodes

CAI Xin-yu, FENG Xiang and YU Hui-qun

Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

Shanghai Smart Energy Engineering Technology Research Center, Shanghai 200237, China

**Abstract** In the era of intelligence, continuous autonomous learning and optimization need to be carried out on the big data platform, and the first step of continuous autonomous learning is data enhancement. This paper proposes a broad learning method based on cascaded enhancement nodes, which provides a new data enhancement method for continuous autonomous learning on big data platform, and makes it possible for subsequent evolutionary optimization on the basis of learning architecture. Classical broad learning is a typical feedforward neural network, which is not suitable for modeling dynamic time series. In this paper, the feedback structure is introduced into the traditional broad learning system, which makes the enhancement nodes have memory and retains part of the historical information. In feature extraction, phase space reconstruction is used to extract more essential features of the data. At the same time, a weight factor is introduced to assign different weights to each sample according to its contribution to model during training, so as to eliminate the interference of noise and outliers to the learning process and improve the robustness of the algorithm. Experimental results show that the proposed algorithm is effective.

**Keywords** Broad learning, Time series prediction, Weight factor, Data enhancement

### 1 引言

在智能化时代,大数据平台上的可持续自主学习和优化极为必要。进行持续自主学习和优化的几个关键步骤包括:

1)进行数据增强学习,研究区分时间、区分空间和特征的数据增强学习;2)进行机器学习;3)进行从知识表示、稀疏矩阵研究解空间的分层搜索,从特征状态快速确定子解空间,加速

搜索研究策略学习、迁移学习,从构建奖惩函数结合数据训练研究深度演化优化算法的强化学习等;4)在完成以上步骤的前提下在学习架构基础上进行演化优化,研究基于数据动力学模型的深度演化、分布并行智能处理理论与方法。

作为持续自主学习的第一步,也是关键的一步,数据增强学习主要是解决神经网络训练时没有足够的数据来最大化深层神经网络的泛化能力的问题<sup>[1-2]</sup>。Lemley 等于 2017 年

到稿日期:2021-05-17 返修日期:2021-10-18

基金项目:国家自然科学基金(61772200,61772201,61602175);上海市浦江人才计划(17PJ1401900);上海市经信委“信息化发展专项资金”(201602008)

This work was supported by the National Natural Science Foundation of China(61772200,61772201,61602175),Shanghai Pujiang Talent Program(17PJ1401900) and Shanghai Economic and Information Commission “Special Fund for Information Development”(201602008).

通信作者:冯翔(xfeng@ecust.edu.cn)

提出了一种智能增强学习方法,其将用于训练的数据按一种最优的数据增强策略进行扩维<sup>[2]</sup>。增强学习一般用于图像数据扩维,针对时空数据的扩维方法目前并不多见。在时序预测领域,也常常面临着没有足够的数据来进行学习的窘境,因此本文将数据增强引入时空数据来提升时序预测的精度。

传统的时间序列预测模型主要是基于统计学方法,如自回归移动平均模型(Autoregressive Integrated Moving Average Model, ARIMA)<sup>[3]</sup>、马尔可夫链模型<sup>[4]</sup>和贝叶斯模型<sup>[5]</sup>等。如文献[6]将贝叶斯模型作为基回归器进行集成学习以提高对时序模型的预测精度,此过程计算量较小且精度较高。随后,一些判别模型被用于时间序列的预测,如支持向量回归模型(Support Vector Regression, SVR)<sup>[7]</sup>、随机森林(Random Forest, RF)<sup>[8]</sup>等。文献[9]提出的一种数据集成方法通过随机重组将原始的一维预测变量重组为高维变量,保留了随机森林的收敛性质。近年来,神经网络(Artificial Neural Network, ANN)被越来越多地应用到时空数据预测当中,如循环神经网络以及RNN的两个变体:长短期记忆网络(Long Short-Term Memory, LSTM)<sup>[10]</sup>和门控单元(Gated Recurrent Unit, GRU)<sup>[11]</sup>。文献[12]提出一种预测方法选择策略,依据时间序列的特征,利用神经网络来学习每个序列和其最佳预测方法的决策规则,同时通过周期性重构来把多步预测转为单步预测。该模型克服了周期项多步预测的缺点,但同时由于提取的序列特征和选用预测方法不够全面,该系统的预测性能不够稳定。文献[13]结合了循环神经网络和自注意力机制,利用自注意力模块来建模多个序列之间的依赖关系,进一步提高了模型鲁棒性。以上预测方法虽然有一些较为成熟,但都属于传统批量学习方法(Batch Learning),不适合数据不断新增的增量学习场景,也并未对数据量较大的情况进行优化。因此,传统的基于批量学习的时间序列预测技术将不再适用于需要随数据新增而快速响应的实时预测场景;同时,传统深度结构存在运行时间长、难以收敛等问题。对此,本文期望寻求一种简单、高效的机器学习技术,在缩短模型训练时间、节约成本的基础上,对于新增的数据能够用增量的方式来更新模型,从而提高系统的实时预测准确率。

最近,一种新兴的基于平面网络结构的高效的随机神经网络<sup>[14-16]</sup>——宽度学习系统(Broad Learning System, BLS)<sup>[17]</sup>,引起了学术界和工业界的广泛关注。该结构只有连接到输出的权重才可通过线性回归进行调整,其他权重是随机初始化的。在BLS中,有两个部分非常关键,即特征节点和增强节点。在文献[18]中,使用相空间重构将原始输入转换成特征节点中的高维特征,且增强节点在宽度上扩展了特征节点。BLS的主要优点是,当需要新的节点或新数据进入网络时,网络会横向地扩展神经元并逐步更新权重,因此BLS在大数据环境中表现良好。文献[19]针对宽度学习模型提取不平衡数据特征不足、识别结果不佳的问题,提出一种基于AdaBoost集成加权宽度学习的不平衡数据分类方法,通过迭代实现权重的动态更新,提升集成模型对少数类的识别能力。但宽度学习模型是典型的前馈神经网络,不适合建模动态时间序列<sup>[20]</sup>。受回声状态网络思想的启发,本文通过在增强节点中加入反馈连接,使得网络能够记忆历史信息,使其适用于时间序列预测。

## 2 问题定义

时间序列模型是机器学习中常用来研究系统运行规律的工具,建立在系统输出数据基础上的时间序列模型是分析研究系统状态和预测未来值的有效手段。本研究基于负荷受多种自变量影响的特点,采用了多元时序分析。一个时间序列 $X^{(i)} = (X_1^{(i)}, X_2^{(i)}, \dots, X_T^{(i)})$ 是完全观测到的时间有序的测量序列,其中 $X_T^{(i)}$ 测量是在时间戳 $T$ 记录的。通常,两次连续测量之间的时间间隔是恒定的。多变量时间序列表示为: $X = (X^{(1)}, X^{(2)}, \dots, X^{(D)})$ ,其中 $X$ 中的时间序列相互关联,测量值 $X_T \in \mathbb{R}^D$ 且在时间戳 $T$ 被记录。

给定一组多元时间序列 $X = (X^{(1)}, X^{(2)}, \dots, X^{(D)})$ ,其中 $D$ 是单变量时间序列的数量, $X^{(D)} \in \mathbb{R}^T$ 是输入窗口大小的长度,我们旨在以滑动窗口的方式来进行预测,即根据已知的 $(X_1, X_2, \dots, X_T)$ 来预测 $X_{T+h}$ ,其中 $h$ 是当前时间戳之前的理想水平。同样,假设窗口内的信息足以进行预测,并且窗口大小是固定的,我们根据 $(X_{1+k}, X_{2+k}, \dots, X_{T+k})$ , $k \in \mathbb{R}^+$ 来预测 $X_{T+h+k}$ 的值。因此,对于预测目标 $X_{T+h} \in \mathbb{R}^D$ ,时间戳 $T$ 处的输入矩阵是 $\mathbf{X} = (X^{(1)}, X^{(2)}, \dots, X^{(D)}) \in \mathbb{R}^{D \times T}$ 。

## 3 宽度学习模型

### 3.1 标准宽度学习系统

宽度学习系统是由澳门大学陈俊龙教授提出的一种基于随机向量函数连接(Random Vector Functional Link Network, RVFL)平面网络结构的增量学习算法,模型结构如图1所示。宽度学习方法的输入样本经过一次线性变换后,将特征表达映射在特征平面上形成特征节点,得到的特征节点再经过激活函数非线性变换生成增强节点。特征节点和增强节点共同连接作为系统的实际输入信号,经连接矩阵线性输出。与RVFL相同,考虑到经典BP算法存在高时间成本和容易陷入局部最优等缺点,宽度学习方法采用岭回归广义逆直接求取输出连接矩阵。

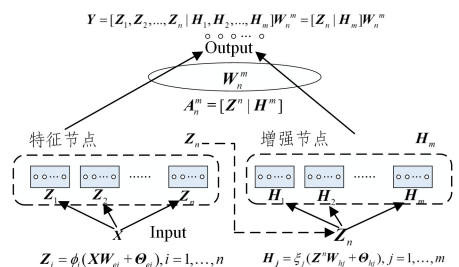


图1 基本宽度学习结构

Fig. 1 Structure of basic broad learning system

首先对输入数据进行线性映射产生映射特征:

$$Z_i = \Phi_i(\mathbf{X}\mathbf{W}_{ei} + \Theta_{ei}), i = 1, \dots, n \quad (1)$$

得到映射特征之后,通过非线性映射生成增强特征:

$$H_j = \xi_j(\mathbf{Z}^m \mathbf{W}_{hj} + \Theta_{hj}), j = 1, \dots, m \quad (2)$$

其中, $\mathbf{X}$ 表示原始数据, $\mathbf{W}_{ei}$ 表示呈高斯分布的随机权重矩阵, $\Theta_{ei}$ 表示偏置常量, $\Phi_i$ 表示稀疏化和正则化处理函数, $\mathbf{Z}^m = [Z_1, \dots, Z_n]$ 为所有的映射特征, $\mathbf{W}_{hj}$ 是经过正交规范化后的随机矩阵, $\Theta_{hj}$ 为偏置常量, $\xi_j$ 为非线性函数。

通过合并映射特征和增强特征,得到最终网络的输入 $\mathbf{A}$ :

$$\mathbf{A} = [\mathbf{Z}^n | \mathbf{H}^m] \quad (3)$$

将  $\mathbf{A}$  重新表示为式(4):

$$\mathbf{A} = [a(x_1)^T, \dots, a(x_i)^T, \dots, a(x_N)^T]^T \quad (4)$$

那么每一个样本在网络中的输出就可以表示为:

$$f(x_i) = a(x_i)\beta \quad (5)$$

预测值和真实值之间的差值就可以表示为:

$$e_i^T = a(x_i)\beta - y_i^T \quad (6)$$

$\beta$  可以通过最小化以下目标函数得到:

$$\arg \min \frac{\theta}{2} \|\beta\|^2 + \frac{1}{2} \sum_{i=1}^N \|e_i\|^2 \quad (7)$$

其中, 第一项是正则化项, 用来避免过拟合,  $\theta$  表示对  $\beta$  和的进一步限制, 第二项表示损失值向量。式(7)可以进一步表达为岭回归形式:

$$\arg \min L_{\text{BLS}} \frac{\theta}{2} \|\beta\|^2 + \frac{1}{2} \sum_{i=1}^N \|A\beta - \mathbf{Y}\|^2 \quad (8)$$

对式(8)求梯度, 并令其等于 0:

$$\nabla L_{\text{BLS}} \theta \beta + \mathbf{A}^T (\mathbf{A}\beta - \mathbf{Y}) = 0 \quad (9)$$

移项后可得  $\beta$  的表达式为:

$$\beta = (\mathbf{A}^T \mathbf{A} + \theta \mathbf{I}_{n_h})^{-1} \mathbf{A}^T \mathbf{Y} \quad (10)$$

其中,  $\mathbf{I}_{n_h}$  是单位阵,  $n_h$  是 BLS 模型中所有节点的个数。

如果训练样本的数量少于神经节点的数量, 那么  $\mathbf{A}$  的行数就会少于列数。在这种情况下,  $\beta$  就可能有无穷多个解。BLS 通过计算  $\mathbf{A}$  的伪逆来解决这个问题。给出  $\mathbf{A}$  的近似伪逆, 得到:

$$\beta = \mathbf{A}^+ \mathbf{Y} \quad (11)$$

进而求得参数矩阵, 得到模型。

### 3.2 宽度学习模型的增量学习过程

BLS 的主要优点之一就是其结构灵活。如果模型的学习效果不能令人满意, 那么我们可以通过增量学习的方式插入一些增强节点和特征节点。

增加增强节点的增量学习算法流程如下: 假设当前网络包含  $n$  组特征节点和  $m$  组增强节点, 那么状态矩阵表示为  $\mathbf{A}_n^m = [\mathbf{Z}^n | \mathbf{H}^m]$ 。现在通过添加第  $m+1$  组增强节点来扩展网络, 添加节点后的新的状态矩阵为:

$$\mathbf{A}_n^{m+1} = [\mathbf{A}_n^m | \xi(\mathbf{Z}^n \mathbf{W}_{h(m+1)} + \Theta_{h(m+1)})]$$

其中,  $\mathbf{W}_{h(m+1)}$  和  $\Theta_{h(m+1)}$  是随机初始化的。 $\mathbf{A}_n^{m+1}$  的伪逆可以表示为:

$$[\mathbf{A}_n^{m+1}]^+ = \begin{bmatrix} [\mathbf{A}_n^m]^+ - \mathbf{D}\mathbf{B}^T \\ \mathbf{B}^T \end{bmatrix}$$

其中:

$$\mathbf{D} = [\mathbf{A}_n^m]^+ \mathbf{H}_{m+1}$$

$$\mathbf{B}^T = \begin{cases} \mathbf{C}^+, & \text{if } \mathbf{C} \neq 0 \\ (\mathbf{I} + \mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T [\mathbf{A}_n^m]^+, & \text{if } \mathbf{C} = 0 \end{cases}$$

$$\mathbf{C} = \mathbf{H}_{m+1} - \mathbf{A}_n^m \mathbf{D}$$

因此, 新的输出权重矩阵表示为:

$$\mathbf{W}_n^{m+1} = \begin{bmatrix} \mathbf{W}_n^m - \mathbf{D}\mathbf{B}^T \mathbf{Y} \\ \mathbf{B}^T \mathbf{Y} \end{bmatrix}$$

可以看出, 只有与第  $m+1$  组增强节点相关的权重需要重新计算, 这一特点保证了 BLS 的快速学习特性。

增加特征节点的增量学习算法流程如下: 同样假设当前网络包含  $n$  组特征节点和  $m$  组增强节点, 并且我们可以通过

将第  $n+1$  组特征节点添加到新的状态矩阵来扩展网络, 添加节点之后的新的状态矩阵为:

$$\mathbf{A}_{n+1}^m = [\mathbf{A}_n^m | \mathbf{Z}_{n+1} | \mathbf{H}_{ex}^1 | \mathbf{H}_{ex}^2 | \dots | \mathbf{H}_{ex}^m]$$

其中,  $\mathbf{H}_{ex}^1 | \mathbf{H}_{ex}^2 | \dots | \mathbf{H}_{ex}^m$  是对应于第  $n+1$  组特征节点的增强节点的附加输出。附加状态表示为:

$$\mathbf{A}^a = [\mathbf{Z}_{n+1} | \mathbf{H}_{ex}^1 | \mathbf{H}_{ex}^2 | \dots | \mathbf{H}_{ex}^m]$$

此时, 新的状态矩阵表示为:

$$[\mathbf{A}_{n+1}^m]^+ = \begin{bmatrix} [\mathbf{A}_n^m]^+ - \mathbf{D}\mathbf{B}^T \\ \mathbf{B}^T \end{bmatrix}$$

其中:

$$\mathbf{D} = [\mathbf{A}_n^m]^+ + \mathbf{A}^a$$

$$\mathbf{B}^T = \begin{cases} \mathbf{C}^+, & \text{if } \mathbf{C} \neq 0 \\ (\mathbf{I} + \mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T [\mathbf{A}_n^m]^+, & \text{if } \mathbf{C} = 0 \end{cases}$$

$$\mathbf{C} = \mathbf{A}^a - \mathbf{A}_n^m \mathbf{D} \quad (19)$$

因此, 新的输出权重矩阵表示为:

$$\mathbf{W}_{n+1}^m = \begin{bmatrix} \mathbf{W}_n^m - \mathbf{D}\mathbf{B}^T \mathbf{Y} \\ \mathbf{B}^T \mathbf{Y} \end{bmatrix}$$

## 4 级联增强节点的宽度学习模型

宽度学习模型是典型的前馈神经网络, 不适合建模动态的时间序列。受到回声状态网络的启发, 本文在宽度学习结构中的增强节点层中加入反馈连接, 使得网络能够记忆历史信息, 使其适用于时间序列预测; 同时, 为了增加算法的鲁棒性, 引入了权重因子, 提出了加权级联增强节点宽度学习系统 (Weight Based Cascaded Enhanced Nodes Broad Learning System, WCEBLS)。

在级联宽度学习的基础上, 为了更好地提取数据的内在特征, 本文引入了相空间重构, 相空间可以反映包含所有可能状态的确定性动力系统; 同时, 本文引入权重因子对不同的数据点分配不同的权重, 以此来减少随机事件、人为干扰等因素对系统产生的噪声数据。通过给不同的样本分配不同的权重因子, 来控制样本对建模的贡献度。

WCEBLS 将相空间重构、权重因子与级联宽度学习系统相结合, 其结构简单, 模型训练效率高, 网络参数较少, 很好地解决了许多传统时序预测方法中存在的无法提取数据深层次特征, 以及深度学习方法中大量超参数优化、容易陷入局部最优以及训练时间过长等问题, 因此其对于时序预测具备可行性。

### 4.1 级联增强节点的宽度学习系统

为了获取时间序列的动态特征, 需要对宽度学习结构进行调整, 将每组增强节点顺序连接, 使增强节点中存在反馈结构, 将特征节点层和增强节点层的输出作为最后输出层的输入进行计算, 其结构如图 2 所示。

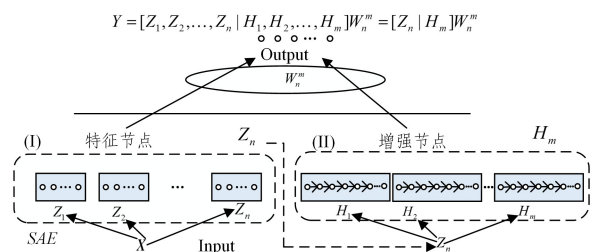


图 2 WCEBLS 结构图

Fig. 2 Structure of WCEBLS

该 BLS 模型通过级联函数合成重构增强节点。同样,对于输入数据,前  $n$  组特征节点可以通过式(12)生成:

$$\mathbf{Z}_i \triangleq \phi(\mathbf{X}\mathbf{W}_{e_i} + \Theta_{e_i}), i=1, \dots, n \quad (12)$$

其中,  $\mathbf{W}_{e_i}$  和  $\Theta_{e_i}$  是符合固定分布的随机值,  $\phi(\cdot)$  是线性映射函数。通过  $\xi(\cdot)$  函数对特征节点进行投影映射,得到第一组增强节点:

$$\mathbf{H}_1 \triangleq \xi(\mathbf{Z}^n \mathbf{W}_{h_1} + \Phi_{h_1}) \triangleq \xi(\mathbf{Z}^n; \{\mathbf{W}_{h_1}, \Phi_{h_1}\}) \quad (13)$$

其中,相关权重都是随机采样得到的。第二组增强节点按照如下方式复合生成:

$$\begin{aligned} \mathbf{H}_2 &= \xi(\mathbf{H}_1 \mathbf{W}_{h_2} + \Phi_{h_2}) \\ &= \xi(\xi(\mathbf{Z}^n \mathbf{W}_{h_1} + \Phi_{h_1}) + \Phi_{h_2}) \\ &= \xi^2(\mathbf{Z}^n; \{\mathbf{W}_{h_i}, \Phi_{h_i}\}_{i=1,2}) \end{aligned} \quad (14)$$

同理,前  $m$  组增强节点形式如下:

$$\mathbf{H}_u = \xi^u(\mathbf{Z}^n; \{\mathbf{W}_{h_i}, \Phi_{h_i}\}_{i=1}^u), u=1, \dots, m \quad (15)$$

因此,特征节点  $\mathbf{Z}^n$  和增强节点  $\mathbf{H}^m = [\mathbf{H}_1, \dots, \mathbf{H}_m]$  与输出直接相连,最后的输出形式如下:

$$\mathbf{Y} = [\mathbf{Z}^n | \mathbf{H}^m] \mathbf{W}_n^m$$

其中,  $\mathbf{W}_n^m$  通过  $[\mathbf{Z}^n | \mathbf{H}^m]$  的伪逆计算得到。

级联增强节点的宽度学习过程如算法 1 所示。

#### 算法 1 级联增强节点的宽度学习

输入:训练样本  $\mathbf{X}, \mathbf{Y}$

输出:权重  $\mathbf{W}$

1. for  $i=0; i \leq m$ ;
2. 随机生成  $\mathbf{W}_{e_i}, \beta_{e_i}$ ;
3. 计算  $\mathbf{Z}_i = \varphi(\mathbf{X}\mathbf{W}_{e_i} + \beta_{e_i})$ ;
4. end;
5. 特征映射节点组  $\mathbf{Z}^n = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n]$ ;
6. for  $u=1, u \leq m$ ;
7. 随机生成  $\mathbf{W}_{h_i}, \beta_{h_i}$ ;
8. 计算  $\mathbf{H}_u = \xi^u(\mathbf{Z}^n; \{\mathbf{W}_{h_i}, \Phi_{h_i}\}_{i=1}^u)$ ;
9. end;
10. 增强节点组  $\mathbf{H}^m = [\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_m]$ ;
11. 表示  $\mathbf{A}^m$  并计算  $\mathbf{A}^m$  的伪逆  $(\mathbf{A}^m)^+$ ;
12. 利用  $(\mathbf{A}^m)^+$  计算  $\mathbf{W}_m$ ;
13. 得到  $\mathbf{W} = \mathbf{W}_m$ .

接下来详细介绍级联宽度学习的增量学习过程。假设现在增加了第  $n+1$  组特征节点:  $\mathbf{Z}_{n+1} \triangleq \phi(\mathbf{X}\mathbf{W}_{e_{n+1}} + \Theta_{e_{n+1}})$ , 那么第  $u$  组增强节点应该修正为  $\xi^u(\mathbf{Z}_{n+1}; \{\mathbf{W}_{e_{i1}}, \Phi_{e_{i1}}\}_{i=1}^u)$ , 其中,  $u=(1, \dots, m)$ 。相应的前  $m$  组增强节点矩阵为  $\mathbf{H}_{e_{r_m}} \triangleq [\xi(\mathbf{Z}_{n+1}; \{\mathbf{W}_{e_{r_1}}, \Phi_{e_{r_1}}\}), \dots, \xi^m(\mathbf{Z}_{n+1}; \{\mathbf{W}_{e_{r_m}}, \Phi_{e_{r_m}}\}_{i=1}^m)]$

接下来,第  $m+1$  组增强节点被表示为:

$$\mathbf{H}_{m+1} \triangleq \xi^{m+1}(\mathbf{Z}_{n+1}; \{\mathbf{W}_{h_i}, \Phi_{h_i}\}_{i=1}^{m+1})$$

其中,  $\mathbf{Z}^{n+1} \triangleq [\mathbf{Z}_1, \dots, \mathbf{Z}_{n+1}]$ ,  $\mathbf{W}_{h_{m+1}}$  和  $\Phi_{h_{m+1}}$  是随机生成的。因此,矩阵  $\mathbf{A}_n^m \triangleq [\mathbf{Z}^n | \mathbf{H}^m]$  可以更新为  $\mathbf{A}_{n+1}^{m+1} \triangleq [\mathbf{A}_n^m | \mathbf{Z}_{n+1} | \mathbf{H}_{e_{r_m}} | \mathbf{H}_{m+1}]$ , 这个级联宽度学习网络的更新的伪逆和新的权重为:

$$(\mathbf{A}_{n+1}^{m+1})^+ = \begin{bmatrix} (\mathbf{A}_n^m)^+ - \mathbf{D}\mathbf{B}^T \\ \mathbf{B}^T \end{bmatrix} \quad (16)$$

$$\mathbf{W}_{n+1}^{m+1} = \begin{bmatrix} \mathbf{W}_n^m - \mathbf{D}\mathbf{B}^T\mathbf{Y} \\ \mathbf{B}^T\mathbf{Y} \end{bmatrix} \quad (17)$$

其中:

$$\begin{aligned} \mathbf{D} &= (\mathbf{A}_n^m)^+ [\mathbf{Z}_{n+1} | \mathbf{H}_{e_{r_m}} | \mathbf{H}_{m+1}] \\ \mathbf{B}^T &= \begin{cases} (\mathbf{C})^+, & \text{if } \mathbf{C} \neq 0 \\ (1 + \mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T (\mathbf{A}_n^m)^+, & \text{if } \mathbf{C} = 0 \end{cases} \\ \mathbf{C} &= [\mathbf{Z}_{n+1} | \mathbf{H}_{e_{r_m}} | \mathbf{H}_{m+1}] - \mathbf{A}_n^m \mathbf{D} \end{aligned} \quad (18)$$

## 4.2 权重因子

本节将介绍基于 Huber 权重函数的计算惩罚加权因子的方法。

Huber M 估计是由 Huber 提出的,其借鉴了迭代加权最小二乘法的思想, M 估计采用了 Huber 权函数来处理样本残差。Huber 权重函数表示为:

$$\theta_i = \begin{cases} 1, & |u_i| \leq b \\ b/|u_i|, & |u_i| > b \end{cases} \quad (19)$$

其中,  $\theta_i$  表示第  $i$  组权重惩罚因子;  $b$  表示一个可调正参数,  $u_i$  代表第  $i$  组样本的标准化残差,其表示为:

$$u_i = \frac{r_i}{\hat{z}}$$

其中,  $r_i$  表示第  $i$  个样本的预测值和测量值之间的残差。所有的样本残差构成了残差矩阵  $\mathbf{r} = [r_1, \dots, r_N]^T$ 。  $\hat{z}$  是鲁棒尺度估计,  $\hat{z} = \text{MAR}/0.6745$ , 其中 MAR 表示绝对剩余中位数, 可以通过下式计算得到:

$$\text{MAR} = \text{med}(|r - \text{med}(|r_i|)|)$$

其中,  $\text{med}(\cdot)$  是中位数函数。

添加权重因子的级联宽度学习系统的基础是标准宽度学习系统。级联宽度学习的连接权重可以通过式(20)加权岭回归算法计算得到:

$$\arg \min_{\theta} \|\theta \mathbf{A} \mathbf{W} - \theta \mathbf{Y}\|_2^2 + \lambda \|\mathbf{W}\|_2^2 \quad (20)$$

其中,  $\theta$  是权重惩罚因子, 表示样本权重,  $\theta = [\theta_1, \theta_2, \dots, \theta_N]^T$ 。 WCEBLS 的连接权重计算公式如下:

$$\mathbf{W} = (\lambda \mathbf{I} + \mathbf{A}^T \theta^2 \mathbf{A})^{-1} \mathbf{A}^T \theta^2 \mathbf{Y} \quad (21)$$

$\mathbf{A}$  的伪逆的具体推导公式如下:

$$\mathbf{A}^+ = \lim_{\lambda \rightarrow 0} (\lambda \mathbf{I} + \mathbf{A}^T \theta^2 \mathbf{A})^{-1} \mathbf{A}^T \theta \quad (22)$$

该算法可以通过迭代计算连接权重和加权惩罚因子获得更好的精度, 迭代过程如图 3 所示。连续解之间差值的最大绝对值如下式所示:

$$\text{MAX} = \max(|\mathbf{W}_m^{(l)} - \mathbf{W}_m^{(l-1)}|)$$

其中,  $l$  表示迭代次数,  $l=1, 2, 3, \dots$ ;  $\mathbf{W}_m^{(l)}$  表示第  $l$  次迭代得到的参数矩阵。若 MAX 小于设定的阈值或者  $l$  达到最大值, 则迭代过程终止; 否则, 该过程继续。

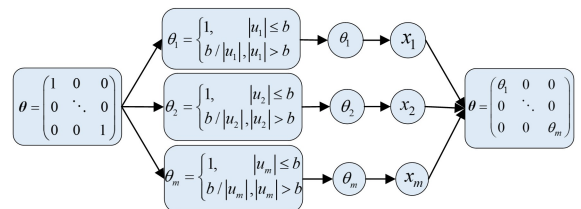


图 3 WCEBLS 中的权重更新规则

Fig. 3 Weight updating rules in WCEBLS

算法 2 给出了加权宽度学习的过程。

#### 算法 2 加权宽度学习

输入:训练样本  $\mathbf{X}, \mathbf{Y}$

输出:权重  $\mathbf{W}$

1. for  $j=1, j \leq m$ ;

2. 随机生成  $W_{h_i}, \beta_{h_i}$
3. 计算  $H_j = \xi(XW_{h_i} + \beta_{h_i})$
4. end;
5.  $H^m = [H_1, H_2, \dots, H_m]$ ;
6. 表示  $A^m$  并计算剩余误差;
7. 初始化  $\theta$  为单位矩阵;
8. if 不满足停止迭代的条件:
9. 更新  $\theta$ ;
10. 用公式计算  $W_m, A_m^+$ ;
11. end;
12. 得到  $W = W_m$ ;

## 5 仿真实验及参数分析

在这一节中,本文使用了两个典型的混沌系统和一个真实的上海市燃气使用量数据集来评估提出的级联增强节点的宽度学习系统的性能;同时,为了验证所提出的加权方式的有效性,通过在数据集中添加一定的噪声来测试鲁棒性。

算法的性能表现通过均方根误差(RMSE)和平均绝对百分比误差(MAPE)来评估。RMSE 测量预测值和目标值之间的差异,定义如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (\tilde{y}(t) - y(t))^2}$$

其中,  $\tilde{y}(t)$  是目标值,  $y(t)$  是预测值,  $N$  表示测试数据集的大小。

MAPE 测量预测值和目标值之间的百分比误差,定义如下:

$$MAPE = \frac{100\%}{N} \sum_{t=1}^N \left| \frac{\tilde{y}(t) - y(t)}{y(t)} \right|$$

为了对不同模型进行客观的对比,下文实验结果在 20 个不同的随机初始化中进行了平均。

### 5.1 数据集描述

Lorenz 时间序列由式(23)生成:

$$\begin{cases} \frac{dx}{dt} = \alpha(y-x) \\ \frac{dy}{dt} = (\beta-z) - y \\ \frac{dz}{dt} = xy - \gamma z \end{cases} \quad (23)$$

当  $\alpha=10, \gamma=8/3, \beta=28$  时,整个序列呈现出混沌特性。Lorenz 系统是一个非线性、非周期、三维的确定性混沌系统。本文采用的是四阶龙格-库塔法来生成 Lorenz 序列。将初始状态设置为  $(x, y, z) = (1, 1, 1)$ , 时间步长为 0.01。实验中一共有 40000 个样本组,每组有 3 个变量。首先选择 35000 个样本作为训练集,5000 个样本作为测试集;然后通过相空间重构将 Lorenz 序列重构到高维空间中,嵌入维数分别设置为  $[20, 20, 20]$ , 将  $x, y, z$  序列的时滞分别设置为  $[1, 1, 1]$ 。

另外,实验中使用了上海市某区 2013—2016 年间每一天的燃气使用量数据集。该数据集包含 1460 组数据,每组数据皆由当天采集的燃气使用量组成。在重构过程中,将变量的嵌入维数设置为 20,延迟时间设置为 1,取前 1176 组数据作为训练集,后 284 组数据作为测试集。

Rossler 时间序列定义如下:

$$\begin{cases} \frac{dx}{dt} = -y - z \\ \frac{dy}{dt} = x + ay \\ \frac{dz}{dt} = b + z(x - c) \end{cases} \quad (24)$$

其中,  $a=0.15, b=0.2, c=10.0$ 。当  $a=0.15, b=0.2, c=10.0$  时,整个序列呈现出混沌特性。Rossler 系统是一个非线性、非周期、三维的确定性混沌系统。本文采用的是四阶龙格-库塔法来生成 Rossler 序列。与 Lorenz 序列相同,我们将 Rossler 序列的初始状态设置为  $(x, y, z) = (1, 1, 1)$ , 时间步长为 0.01。通过相空间重构将 Rossler 序列重构到高维空间中,嵌入维数分别设置为  $[20, 20, 20]$ , 将  $x, y, z$  序列的时滞分别设置为  $[1, 1, 1]$ 。

表 1 列出了在 Rossler 和 Lorenz 两个混沌时间序列上进行实验仿真时,不同算法的参数设置。极限学习机的隐层节点设置为 1000;对于 BLS,设置 10 个特征节点窗口,每个窗口包含 50 个特征节点,增强节点的个数为 100,正则化参数设置为  $2^{-30}$ ;Huber 函数中的非负调节参数  $b$  设置为 1.345。

表 1 实验参数设置

Table 1 Setting of experimental parameters

Series	Methods	Parameters
Rossler series	ELM	Number of hidden nodes: 500, 1000
	CFBLS	$n=5, k=10, h=60, s=0.8, c=2 \times 10^{-30}$
	BLS	$n=5, k=10, h=60, s=0.8, c=2 \times 10^{-30}$
	WCEBLS	$n=5, k=10, h=60, s=0.8, c=2 \times 10^{-30}$
Lorenz series	ELM	Number of hidden nodes: 200, 1000
	CFBLS	$n=50, k=10, h=100, s=0.8, c=2 \times 10^{-30}$
	BLS	$n=50, k=10, h=100, s=0.8, c=2 \times 10^{-30}$
	WCEBLS	$n=50, k=10, h=100, s=0.8, c=2 \times 10^{-28}$

### 5.2 实验结果与分析

为了保证实验高速且平稳地运行,我们在 Lenovo Shengteng 6800 HPCC 机架式优化高性能服务器平台上运行算例,它含有 1 个控制节点和 8 个计算节点,每个计算节点都配置了两个 2.4GHz 的四核处理器以及 24GB 的内存空间。服务器中运行的操作系统为 Red Hat Enterprise Linux 7,实验的软件计算平台为 Python 3.7。在两个数据集上验证了所提方法,对重构的典型混沌时间序列即 Lorenz 和 Rossler 时间序列以及上海市燃气日消耗量数据集进行了仿真,验证了 WCEBLS 模型的有效性。

由于源数据中缺乏离群点,因此我们在所有的训练数据的输出加上服从正态分布的噪声( $\epsilon \sim N(0, 2)$ ),使得函数数据更符合实际采集的数据。另外,将本文方法与极限学习机(Extreme Learning Machine, ELM)和 BLS 进行横向对比。预测的性能由 3 个指标来衡量:均方根误差(RMSE)、平均绝对百分比误差(MAPE)、平均绝对误差(MAE)。图 4 给出了由 WCEBLS 产生的 Rossler 和 Lorenz 时间序列的预测曲线和误差曲线。可以看出, Lorenz 序列预测的绝对误差遵循高斯分布,这意味着 WCEBLS 能够充分捕捉到 Lorenz 混沌系统的动态。然而,对于 Rossler 序列预测而言,某些时间步长的绝对误差很大,这是由 Rossler 序列的数值跳跃造成的。在所有被评估的模型中,所提 WCEBLS 获得了最好的结果。

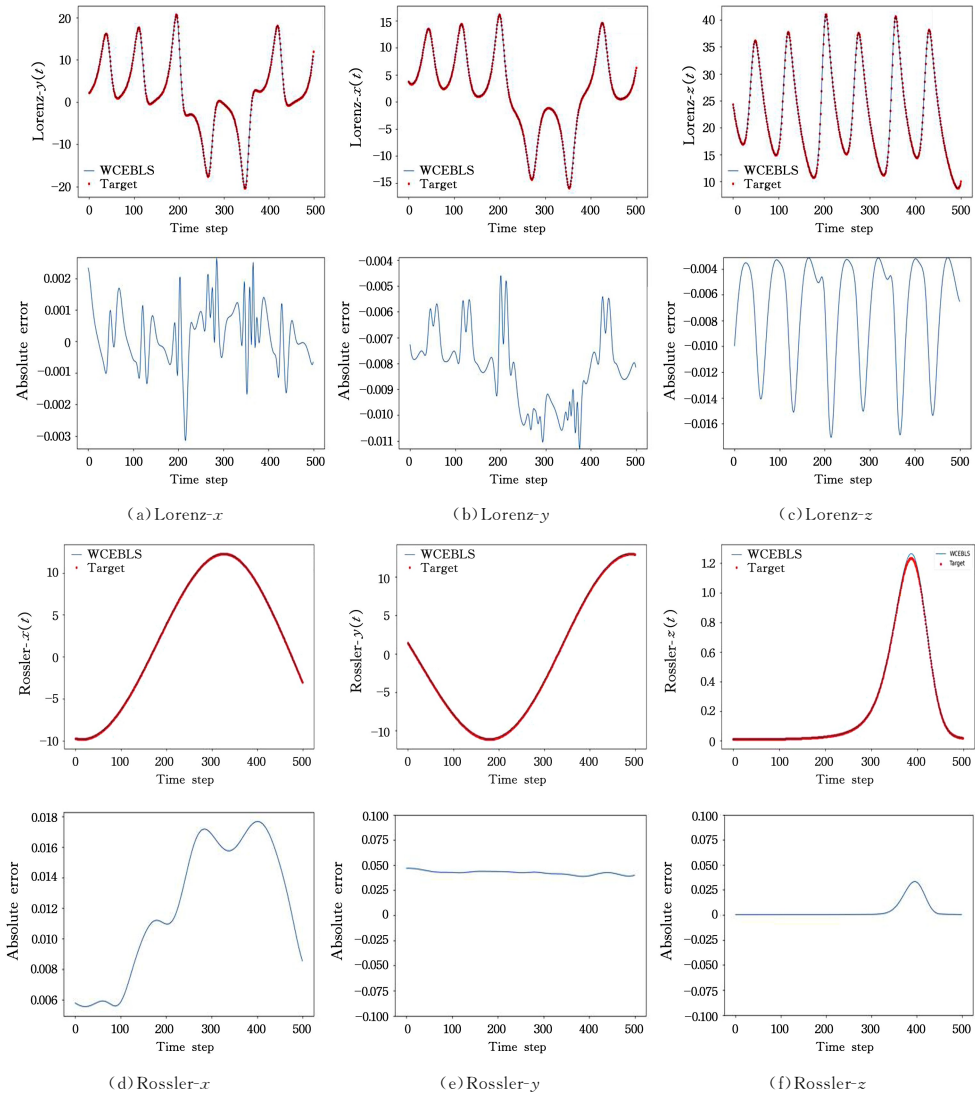


图4 WCEBLS分别为Rossler- $x(t)$ ,  $y(t)$ ,  $z(t)$ 和Lorenz- $x(t)$ ,  $y(t)$ ,  $z(t)$ 时生成的预测曲线和误差曲线

Fig. 4 Prediction curves and error curves generated by WCEBLS for Rossler- $x(t)$ ,  $y(t)$ ,  $z(t)$  and Lorenz- $x(t)$ ,  $y(t)$ ,  $z(t)$

表2列出了不同评估模型的预测结果,包括20次运行的平均值和标准差。从图4则可以明显看到,本文提出的WCEBLS在两个混沌基准时间序列上表现最好,拥有最低的RMSE, MAPE和MAE值。因为预测误差非常小,所以在这个实验中我们没有对WCEBLS的权重进行微调。可以看出,WCEBLS的性能优于ELM,因为连接的增强节点使得模型

能够存储与当前输出相关的历史信息。此外,在Rossler- $x(t)$ 序列上,BLS的性能比基于神经网络的模型性能差得多,BLS没有随机性,因为20次运行的标准偏差为0。这里还注意到,ELM在Rossler- $x(t)$ 时间序列中获得了良好的性能,但在Rossler- $y(t)$ 序列中表现不佳。以上结果说明,这两种方法都有一定的适用范围,而本文提出的WCEBLS则是普遍适用的。

表2 Rossler时间序列中第 $t+1$ 步的预测结果

Table 2 Prediction results of step  $t+1$  in Rossler series

Methods	RMSE			MAPE			MAE		
	$x$	$y$	$z$	$x$	$y$	$z$	$x$	$y$	$z$
ELM(200)	0.2378	3.1684	0.5572	0.3945	1.0568	0.1130	0.2134	3.1445	0.1418
ELM(1000)	0.0070	0.0069	0.0043	0.0091	0.0020	0.0060	0.1681	0.0233	0.0255
BLS	3.3601	0.6144	0.8085	1.0286	0.0551	0.0824	0.5565	0.1640	0.1034
WCEBLS	0.0043	0.0028	0.0017	0.0067	0.0019	0.0057	0.0019	0.0015	0.0017

图5给出了两个时间序列的原始分布以及训练样本在添加均值为0、方差为2时的噪声分布。其中,蓝线代表原始序列曲线,红点代表被噪声和异常值污染的训练样本。图6给出了测试样本在不同算法中的预测结果。由图5可以看出,

由于受到噪声以及异常值的影响,所有训练样本都分布在真实曲线周围。同时,由图7可知,与标准BLS相比,当训练样本被严重污染时,WCEBLS的预测曲线更接近于真实的测试曲线。

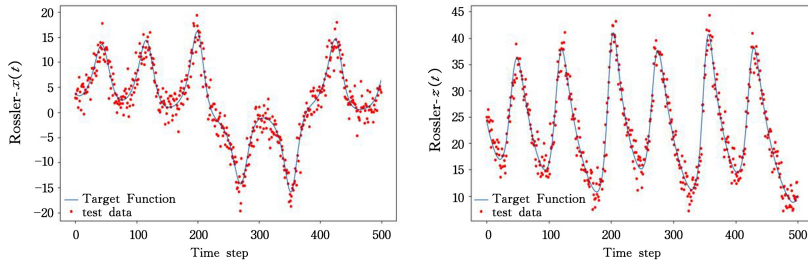


图5 带有噪声和异常值的训练样本和目标函数(电子版为彩图)

Fig. 5 Training samples and objective function with noise and outliers

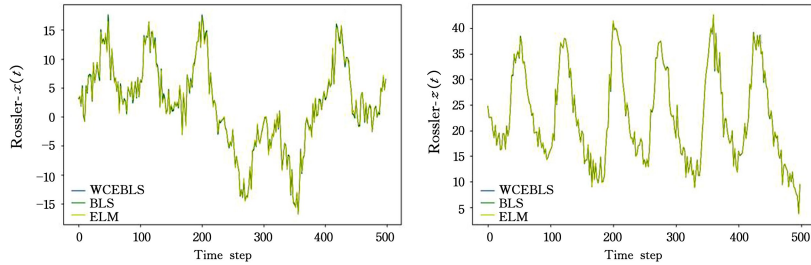


图6 测试数据的不同算法的测试结果

Fig. 6 Test results of testing data with different algorithms

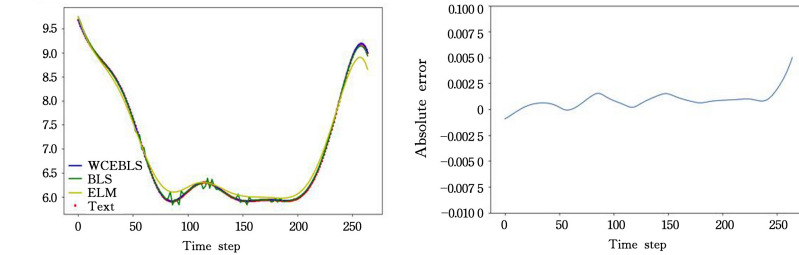


图7 燃气数据集上各算法的预测结果

Fig. 7 Prediction results of each algorithm on gas data set

真实世界的数据——上海市燃气日使用量数据集的预测对比结果如表3所列。在该表中,WCEBLS的不同衡量指标均达到了最优值。燃气使用量的 $t+1$ 步预测结果如图7所示。从图7可以看出,WCEBLS的预测值与燃气测试集拟合

程度最为相似,并且对曲线的上下起伏也能很好地拟合;而ELM对曲线波动趋势敏感度较低,在曲线的极值点处误差较大,且在大部分时间尺度下的预测误差均大于BLS和WCEBLS。

表3 Lorenz时间序列中第 $t+1$ 步的预测结果

Table 3 Prediction results of step  $t+1$  in Lorenz series

Methods	RMSE			MAPE			MAE		
	$x$	$y$	$z$	$x$	$y$	$z$	$x$	$y$	$z$
ELM(200)	0.0252	0.0076	0.0776	0.0432	0.0119	0.0032	0.0247	0.0075	0.0750
ELM(1000)	0.0072	0.0080	0.0047	0.0038	0.0048	0.0027	0.0074	0.0083	0.0053
BLS	0.2570	0.4784	0.6740	0.2370	0.4413	0.0117	0.1354	0.2800	0.2755
WCEBLS	0.0003	0.0005	0.0003	0.0003	0.00046	0.0002	0.0010	0.0015	0.0009

从预测误差曲线来看,尽管燃气数据随着日期不断上下波动,且有快速上升期和快速下降期,但从图7来看,WCEBLS的预测误差一直比较平稳。

在训练时长上,ELM和BLS的训练均属于求解伪逆,计算速度较快,不需要进行迭代。而ELM作为深度学习算法,在训练时要确定大量的网络参数,训练过程极为耗时,当数据集更新时难以对模型进行重新训练。

**结束语** 本文提出了一种自适应权重的级联增强节点的宽度学习算法。作为一种数据增强算法,其在大数据平台上进行持续自主学习提供了数据前提,并为后续的优化过程

提供了数据准备。首先,在特征提取阶段对输入数据进行相空间重构,并将其映射到高维空间中;然后,将高维数据输入到加权宽度学习系统中,通过级联增强节点,使得宽度结构存在记忆,保留了部分历史信息;最后,考虑到输入数据中存在一定的离群点或者噪声,在宽度结构中引入权重因子,自适应地为不同的样本分配不同的权重,以提高预测的性能以及鲁棒性。研究表明,无论是在基准数据集还是案例数据集上,WCEBLS都优于ELM和BLS,且能够提高时间序列的预测精度。可见在实际数据的预测应用中,WCEBLS的性能更加突出。

## 参 考 文 献

- [1] CAO B, WANG N, LI J, et al. Data Augmentation-Based Joint Learning for Heterogeneous Face Recognition[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 30(6):1731-1743.
- [2] LEMLEY J, BAZRAFKAN S, CORCORAN P. Smart Augmentation Learning an Optimal Data Augmentation Strategy[J]. IEEE Access, 2017, 5:5858-5869.
- [3] BOX G E P, PIERCE D A. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models[J]. Journal of the American Statistical Association, 1970, 65(332):1509-1526.
- [4] YU G, HU J, ZHANG C, et al. Short-Term Traffic Flow Forecasting Based on Markov Chain Model[C]//IEEE IV2003 Intelligent Vehicles Symposium. IEEE, 2003:208-212.
- [5] WANG J, DENG W, GUO Y. New Bayesian Combination Method for Short-Term Traffic Flow Forecasting[J]. Transportation Research Part C: Emerging Technologies, 2014, 43:79-94.
- [6] FENG L, ZHAO C, CHEN C L P, et al. BNGBS: An Efficient Network Boosting System With Triple Incremental Learning Capabilities for More Nodes, Samples, and Classes[J]. Neurocomputing, 2020, 412:486-501.
- [7] HOU J X, LI Q, ZHU Y J, et al. Real-Time Forecasting System of PM<sub>2.5</sub> Concentration Based on Spark Framework and Random Forest Model[J/OL]. Science of Surveying and Mapping, 2007. [http://en.cnki.com.cn/Article\\_en/CJFDTotal-CHKD201701001.htm](http://en.cnki.com.cn/Article_en/CJFDTotal-CHKD201701001.htm).
- [8] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [9] XIE K, RONG Y T, HU F P, et al. Random Forest Algorithm Based on Data Integration[J]. Computer Engineering, 2020, 46(12):290-298.
- [10] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[J]. arXiv:1412.3555, 2014.
- [11] CHEN T T, LEE S J. A Weighted LS-SVM Based Learning System for Time Series Forecasting[J]. Information Sciences, 2015, 299:99-116.
- [12] SONG X L, LIU Y Z, CHEN S F. Seasonal Time Series Forecasting Based on Seasonality Method Selection[J]. Computer Engineering, 2011, 37(21):131-132, 135.
- [13] HUANG S, WANG D, WU X, et al. Dsanet: Dual Self-attention Network for Multivariate Time Series Forecasting[C]// Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019:2129-2132.
- [14] CHEN C L P. A Rapid Supervised Learning Neural Network for Function Interpolation and Approximation[J]. IEEE Transactions on Neural Networks, 1996, 7(5):1220-1230.
- [15] CHEN C L P, LECLAIR S R, PAO Y H. An Incremental Adaptive Implementation of Functional-Link Processing for Function Approximation, Time-Series Prediction, and System Identification[J]. Neurocomputing, 1998, 18(1/2/3):11-31.
- [16] CHEN C L P, WAN J Z. A Rapid Learning and Dynamic Stepwise Updating Algorithm for Flat Neural Networks and the Application to Time-Series Prediction[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 1999, 29(1):62-72.
- [17] CHEN C L P, LIU Z. Broad Learning System: An Effective and Efficient Incremental Learning System without the Need for Deep Architecture[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 29(1):10-24.
- [18] HAN M, ZHANG R, QIU T, et al. Multivariate Chaotic Time Series Prediction Based on Improved Grey Relational Analysis[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017, 49(10):2144-2154.
- [19] WANG M D, XU X Y, YAN G W, et al. Ensemble Weighted Broad Learning System with AdaBoost for Imbalanced Classification[J]. Computer Engineering, 2020, 48(4):99-105, 112.
- [20] XU M, HAN M, CHEN C L P, et al. Recurrent Broad Learning Systems for Time Series Prediction[J]. IEEE Transactions on Cybernetics, 2018, 50(4):1405-1417.



**CAI Xin-yu**, born in 1998, postgraduate, is a member of China Computer Federation. His main research interests include swarm intelligence and broad learning.



**FENG Xiang**, born in 1977, Ph.D, professor, is a member of China Computer Federation. Her main research interests include artificial intelligence, swarm intelligence and evolutionary computing, and big data intelligence.

(责任编辑:李亚辉)