

DOI:10.16356/j.1005-2615.2019.05.017

## 抗离群值的鲁棒正则化贯序超限学习机

郭威 汤克明 于建江

(盐城师范学院信息工程学院,盐城,224002)

**摘要:** 针对离群值环境下的在线学习问题,提出一种鲁棒正则化贯序超限学习机(Robust regularized online sequential extreme learning machine, RR-OSELM)。RR-OSELM 以增量学习新样本的方式实现在线学习,并在学习过程中基于样本的先验误差进行逆向加权计算以降低学习模型对于离群值的敏感性;同时 RR-OSELM 通过融合使用 Tikhonov 正则化技术进一步增强了其在实际应用中的稳定性。实验结果表明,RR-OSELM 具有较同类算法更好的鲁棒性和实用性,对于离群值环境下的在线建模与预测问题是积极有效的。

**关键词:** 在线贯序超限学习机;离群值;鲁棒性;Tikhonov 正则化;在线学习

**中图分类号:** TP181      **文献标志码:** A      **文章编号:** 1005-2615(2019)05-0704-07

## Robust Regularized Online Sequential Extreme Learning Machine for Outliers Restraining

GUO Wei, TANG Keming, YU Jianjiang

(College of Information Engineering, Yancheng Teachers University, Yancheng, 224002, China)

**Abstract:** Aiming at the online learning with outliers, this paper proposes a robust regularized online sequential extreme learning machine (RR-OSELM). The proposed RR-OSELM is able to learn the newly arrived samples incrementally by a recursive fashion, and assign inverse weights for each example based on the priori error so as to reduce its sensibility to outliers. The Tikhonov regularization technique is incorporated in the RR-OSELM to further enhance the stability of the algorithm in real applications. Experimental results show that the proposed RR-OSELM is more robust than its counterparts, and it can be applied to the online modeling and prediction of data streams with outliers.

**Key words:** online sequential extreme learning machine; outlier; robustness; Tikhonov regularization; online learning

在线学习是当前机器学习研究领域的一个重要方向,其基本思想是随着数据样本的不断到达,学习器能够对当前样本进行“局部”学习并保持模型的同步更新。与采用批量计算模式的离线学习方法相比,在线学习方法能以增量的方式实现对数据样本的持续快速学习,为解决目前广泛存在的数据流在线建模与预测问题提供了天然的解决方案<sup>[1-3]</sup>。此外,在现实的在线应用中,由于机械故

障、设备退化、人为错误等原因,实际采集的在线数据中不可避免会存在少量的离群值,离群值的存在会严重影响在线建模的准确性以及后续的预测效果。因此,如何增强在线学习算法对于离群值的抗干扰能力和鲁棒性是一个非常重要而现实的研究课题。

在线贯序超限学习机(Online sequential extreme learning machine, OSELM)是近年来广受关

**基金项目:** 国家自然科学基金(61603326, 61379064, 61273106)资助项目。

**收稿日期:** 2018-06-17; **修订日期:** 2018-10-12

**通信作者:** 郭威,男,副教授, E-mail: weiguo031@163.com。

**引用格式:** 郭威,汤克明,于建江. 抗离群值的鲁棒正则化贯序超限学习机[J]. 南京航空航天大学学报, 2019, 51(5): 704-710. GUO Wei, TANG Keming, YU Jianjiang. Robust Regularized Online Sequential Extreme Learning Machine for Outliers Restraining[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(5): 704-710.

注的一种新颖而实用的在线学习算法<sup>[4]</sup>,该算法以超限学习机(Extreme learning machine, ELM)<sup>[5]</sup>的逼近理论为基础,将单隐层前馈神经网络(Single layer feedforward networks, SLFNs)的训练问题转化为线性方程组的求解问题,并采用递归最小二乘(Recursive least squares, RLS)方法递推计算输出权值以实现算法的在线更新。相比其他常见的在线学习算法,OSELM不仅具有更好的泛化性能,而且在学习速度上优势明显,目前已在数据流在线建模与预测研究领域获得了成功应用,如降雨量预测<sup>[6]</sup>,风速预测<sup>[7]</sup>,时间序列预测<sup>[8-9]</sup>,时变系统预测<sup>[10-11]</sup>等。然而,RLS方法对于噪声数据的天然敏感性<sup>[12]</sup>也使得基于RLS的OSELM算法在噪声或离群值环境下的泛化性能将严重下降。为了增强OSELM对于噪声数据的处理能力,文献[13]提出了一种基于双目标优化的正则化OSELM算法(Regularized OSELM, R-OSELM),该算法在学习过程中同步追求训练残差和输出权值模的最小化,有效提高了模型的泛化性能和稳定性,并在一定程度上增强了对噪声和离群值的抗干扰能力,但该算法并非解决离群值问题的针对性方法,其抗离群值能力相对现实需求仍然较差。结合噪声补偿机制,文献[14]提出了一种基于最小均方差的在线超限学习机ESEP-ELM,并将其应用于含离群值混沌时间序列的在线建模与预测。实验结果显示,ESEP-ELM具有较原始OSELM更好的稳定性和更高的计算效率,并且对离群值有较强的抵抗能力,但其学习精度不如OSELM。近年来,学者们以超限学习理论为基础并结合不同的优化学习方法,提出了多种抗离群值的鲁棒ELM算法<sup>[15-19]</sup>,但这些算法均基于离线学习方式,在在线情形下并不适用。

针对离群值环境下的在线学习问题,本文提出一种具有抗离群值能力的鲁棒正则化贯序超限学习机(Robust regularized online sequential extreme learning machine, RR-OSELM)。RR-OSELM继承了OSELM贯序学习的基本思想,并在在线学习过程中基于样本的先验误差进行逆向加权计算,以有效抑制离群值对学习模型的影响;同时,RR-OSELM通过引入Tikhonov正则化技术使得算法具有良好的稳定性。仿真实验验证了RR-OSELM的鲁棒性和实用性。

## 1 OSELM算法

OSELM<sup>[4]</sup>是ELM<sup>[5]</sup>的在线学习版本。对于 $N$ 个任意的不同样本 $(x_i, y_i) \in \mathbf{R}^d \times \mathbf{R}^m$ ,具有 $n$ 个隐

层节点并且隐层激活函数为 $g(x)$ 的SLFNs可表示为

$$\sum_{j=1}^n \beta_j g_j(x_i) = \sum_{j=1}^n \beta_j G(a_j, b_j, x_i) \quad i = 1, 2, \dots, N \quad (1)$$

式中 $a_j$ 和 $b_j$ 分别是第 $j$ 个隐层节点的输入权值和偏移, $\beta_j$ 是连接第 $j$ 个隐层节点与输出层的输出权值, $g_j(x_i) = G(a_j, b_j, x_i)$ 表示第 $j$ 个隐层节点关于输入向量 $x_i$ 的输出。

该SLFNs能以零误差逼近这 $N$ 个样本,即存在 $(a_j, b_j)$ 及 $\beta_j$ 使得

$$\sum_{j=1}^n \beta_j G(a_j, b_j, x_i) = y_i \quad i = 1, 2, \dots, N \quad (2)$$

式(2)可写成如下矩阵形式

$$H\beta = Y \quad (3)$$

$$H = \begin{bmatrix} h_1 \\ \vdots \\ h_N \end{bmatrix} = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_n, b_n, x_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & \cdots & G(a_n, b_n, x_N) \end{bmatrix}_{N \times n} \quad (4)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_n^T \end{bmatrix}_{n \times m}, \quad Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times m} \quad (5)$$

根据文献[5]中的插值理论,只要 $g(x)$ 无限可微,SLFNs的隐层节点参数 $(a_j, b_j)$ 可随机生成并保持不变,则 $H$ 为一常数矩阵,这样就将SLFNs的训练问题等价转化为求解线性系统 $H\beta = Y$ 的最小二乘解 $\hat{\beta}$

$$\|H\hat{\beta} - Y\| = \min_{\beta} \|H\beta - Y\| \quad (6)$$

ELM采用最小模最小二乘解作为网络的输出权值,即

$$\hat{\beta} = H^+ Y \quad (7)$$

式中, $H^+$ 为 $H$ 的Moore-Penrose广义逆。若 $H^T H$ 非奇异,式(7)可进一步表示为

$$\hat{\beta} = H^+ Y = (H^T H)^{-1} H^T Y \quad (8)$$

为了满足在线学习的需要,Liang等<sup>[4]</sup>将贯序学习的思想用于ELM并提出一种可增量学习训练样本的在线学习算法OSELM,其学习过程描述如下。

在初始化阶段,对于给定的初始训练子集 $\Omega_{k-1} = \{(x_i, y_i) | i = 1, \dots, k-1\}$ ,根据式(8)初始的输出权值可表示为

$$\beta_{k-1} = P_{k-1} H_{k-1}^T Y_{k-1} \quad (9)$$

式中 $P_{k-1} = (H_{k-1}^T H_{k-1})^{-1}$ ,  $H_{k-1} = [h_1^T \ h_2^T \ \cdots \ h_{k-1}^T]^T$ ,  $Y_{k-1} = [y_1 \ y_2 \ \cdots \ y_{k-1}]^T$ 。

在在线学习阶段,每当获取到一个新的训练样本  $(\mathbf{x}_k, y_k)$ , 则按照下式递推计算输出权值

$$\mathbf{P}_k = \mathbf{P}_{k-1} - \frac{\mathbf{P}_{k-1} \mathbf{h}_k^T \mathbf{h}_k \mathbf{P}_{k-1}}{1 + \mathbf{h}_k \mathbf{P}_{k-1} \mathbf{h}_k^T} \quad (10)$$

$$\boldsymbol{\beta}_k = \boldsymbol{\beta}_{k-1} + \mathbf{P}_k \mathbf{h}_k^T (y_k - \mathbf{h}_k \boldsymbol{\beta}_{k-1})$$

式中  $\mathbf{h}_k = [G(a_1, b_1, x_k) \cdots G(a_n, b_n, x_k)]$ 。

## 2 鲁棒正则化贯序超限学习机

### 2.1 RR-OSELM 学习模型

根据 OSELM 算法<sup>[4]</sup>的实现过程可知,原始 OSELM 算法以最小二乘(Least squares, LS)作为学习准则,追求误差平方和的最小化,即其代价函数可表示为

$$J(\boldsymbol{\beta}_k) = \|\mathbf{H}\boldsymbol{\beta}_k - \mathbf{Y}\|^2 = \sum_{i=1}^k |y_i - \mathbf{h}_i \boldsymbol{\beta}_k|^2 \quad (11)$$

然而,LS 学习准则对于离群值非常敏感<sup>[12]</sup>并易于生成过拟合的学习模型,从而导致 OSELM 在离群值环境下的泛化性能将严重下降。作为改进算法,R-OSELM<sup>[13]</sup>借鉴统计学习理论中的结构风险最小化原理,在学习过程中同时追求经验风险(训练残差)和结构风险(输出权值的模)的最小化,即其对应的代价函数为

$$J_R(\boldsymbol{\beta}_k) = \|\mathbf{H}\boldsymbol{\beta}_k - \mathbf{Y}\|^2 + \delta \|\boldsymbol{\beta}_k\|^2 = \sum_{i=1}^k |y_i - \mathbf{h}_i \boldsymbol{\beta}_k|^2 + \delta \|\boldsymbol{\beta}_k\|^2 \quad (12)$$

式中  $\delta$  为权衡这两种风险的正则化参数。尽管 R-OSELM 通过使用正则化方法有效避免了迭代学习过程中潜在的病态矩阵求逆问题,从而增强了算法的稳定性,并在一定程度上减轻了对于离群值的敏感性,但其抗离群值能力仍然非常有限。直观上分析,若当前训练样本  $(\mathbf{x}_k, y_k)$  为离群值样本,则其对应的训练残差  $y_k - \mathbf{h}_k \boldsymbol{\beta}_k$  将严重偏离正常范围,此时代价函数  $J_R(\boldsymbol{\beta}_k)$  在整体上是偏的,显然在此情形下训练得到的学习模型也将是不准确的。

为进一步增强 R-OSELM 对于离群值的鲁棒性,本文引入 Tikhonov 正则化技术和样本加权的思想,提出一种新的鲁棒正则化 OSELM 学习模型(RR-OSELM)。该模型对应的代价函数如下

$$J_{RR}(\boldsymbol{\beta}_k) = \sum_{i=1}^k \varphi_i |y_i - \mathbf{h}_i \boldsymbol{\beta}_k|^2 + \delta \|\boldsymbol{\beta}_k\|^2 \quad (13)$$

式中  $\varphi_i$  为一个非负的权值系数,用来对不同的学习样本进行加权。本文借鉴文献<sup>[20]</sup>中的方法,基于当前样本的先验误差进行逆向加权计算,以对离群值样本所产生的较大扰动误差进行反向抑制,从而减轻离群值对学习模型所带来的不利影响,即对于当前学习样本  $(\mathbf{x}_k, y_k)$ , 其对应的权值为

$$\varphi_k = \frac{1}{\|\mathbf{h}_k e_k\|_1} \quad (14)$$

式中  $e_k = y_k - \mathbf{h}_k \boldsymbol{\beta}_{k-1}$  为当前样本的先验误差。这里注意到,  $\varphi_k$  的计算结果可能大于 1,而这会影响算法的收敛性,为解决此问题,本文使用

$$\varphi_k = \min\left(1, \frac{1}{\|\mathbf{h}_k e_k\|_1}\right) \quad (15)$$

### 2.2 RR-OSELM 模型的递归求解

下面将推导出 RR-OSELM 模型中输出权值  $\boldsymbol{\beta}_k$  的递推计算公式以满足在线学习的要求。让代价函数  $J_{RR}(\boldsymbol{\beta}_k)$  对  $\boldsymbol{\beta}_k$  求微分并令结果为零,可得法方程

$$\mathbf{R}_k \boldsymbol{\beta}_k = \boldsymbol{\Phi}_k \quad (16)$$

式中

$$\mathbf{R}_k = \sum_{i=1}^k \varphi_i \mathbf{h}_i^T \mathbf{h}_i + \delta \mathbf{I} = \mathbf{R}_{k-1} + \varphi_k \mathbf{h}_k^T \mathbf{h}_k \quad (17)$$

$$\boldsymbol{\Phi}_k = \sum_{i=1}^k \varphi_i \mathbf{h}_i^T y_i = \boldsymbol{\Phi}_{k-1} + \varphi_k \mathbf{h}_k^T y_k \quad (18)$$

对式(17)应用 Sherman-Morrison 公式<sup>[21]</sup>可得

$$\mathbf{R}_k^{-1} = (\mathbf{R}_{k-1} + \varphi_k \mathbf{h}_k^T \mathbf{h}_k)^{-1} = \mathbf{R}_{k-1}^{-1} - \frac{\varphi_k \mathbf{R}_{k-1}^{-1} \mathbf{h}_k^T \mathbf{h}_k \mathbf{R}_{k-1}^{-1}}{1 + \varphi_k \mathbf{h}_k \mathbf{R}_{k-1}^{-1} \mathbf{h}_k^T} \quad (19)$$

令  $\mathbf{P}_k = \mathbf{R}_k^{-1}$ , 则式(19)可重写为

$$\mathbf{P}_k = \mathbf{P}_{k-1} - \frac{\varphi_k \mathbf{P}_{k-1} \mathbf{h}_k^T \mathbf{h}_k \mathbf{P}_{k-1}}{1 + \varphi_k \mathbf{h}_k \mathbf{P}_{k-1} \mathbf{h}_k^T} \quad (20)$$

综合式(16)和式(18),  $\boldsymbol{\beta}_k$  可表示为

$$\boldsymbol{\beta}_k = \mathbf{R}_k^{-1} \boldsymbol{\Phi}_k = \mathbf{P}_k (\boldsymbol{\Phi}_{k-1} + \varphi_k \mathbf{h}_k^T y_k) = \mathbf{P}_k (\mathbf{R}_{k-1} \boldsymbol{\beta}_{k-1} + \varphi_k \mathbf{h}_k^T y_k) \quad (21)$$

将式(17)代入式(21)中,有

$$\begin{aligned} \boldsymbol{\beta}_k &= \mathbf{P}_k ((\mathbf{R}_k - \varphi_k \mathbf{h}_k^T \mathbf{h}_k) \boldsymbol{\beta}_{k-1} + \varphi_k \mathbf{h}_k^T y_k) = \\ &= \boldsymbol{\beta}_{k-1} - \varphi_k \mathbf{P}_k \mathbf{h}_k^T \mathbf{h}_k \boldsymbol{\beta}_{k-1} + \varphi_k \mathbf{P}_k \mathbf{h}_k^T y_k = \\ &= \boldsymbol{\beta}_{k-1} + \varphi_k \mathbf{P}_k \mathbf{h}_k^T (y_k - \mathbf{h}_k \boldsymbol{\beta}_{k-1}) = \\ &= \boldsymbol{\beta}_{k-1} + \varphi_k \mathbf{P}_k \mathbf{h}_k^T e_k \end{aligned} \quad (22)$$

综上,RR-OSELM 模型中输出权值  $\boldsymbol{\beta}_k$  的递推计算公式可最终表示为

$$\begin{aligned} \mathbf{P}_k &= \mathbf{P}_{k-1} - \frac{\varphi_k \mathbf{P}_{k-1} \mathbf{h}_k^T \mathbf{h}_k \mathbf{P}_{k-1}}{1 + \varphi_k \mathbf{h}_k \mathbf{P}_{k-1} \mathbf{h}_k^T} \\ \boldsymbol{\beta}_k &= \boldsymbol{\beta}_{k-1} + \varphi_k \mathbf{P}_k \mathbf{h}_k^T e_k \end{aligned} \quad (23)$$

式中  $\varphi_k$  由式(15)计算得到,并且当  $\varphi_k \equiv 1$  时,RR-OSELM 退化 R-OSELM。

### 2.3 RR-OSELM 的鲁棒性分析与比较

根据式(10),OSELM(及 R-OSELM)算法中输出权值的微分向量可表示为

$$\Delta \boldsymbol{\beta}_k = \boldsymbol{\beta}_k - \boldsymbol{\beta}_{k-1} = \mathbf{P}_k \mathbf{h}_k^T e_k \quad (24)$$

则其  $L_1$  范数为

$$\|\Delta\beta_k\|_1 = \|e_k\| \|P_k h_k^T\|_1 \quad (25)$$

由式(25)容易看出,受离群值样本扰动误差  $e_k$  的影响,OSELM(及 R-OSELM)算法中输出权值的微分向量的  $L_1$  范数将快速增加,导致学习模型不够准确。

类似地,根据式(23)可以得到 RR-OSELM 算法中输出权值的微分向量的  $L_1$  范数为

$$\|\Delta\beta_k\|_1 = \|\varphi_k P_k h_k^T e_k\|_1 = \frac{\|P_k h_k^T\|_1}{\|h_k\|_1} \|e_k\|_1 \quad (26)$$

可以看到离群值样本所带来的扰动误差  $e_k$  并没有对  $\|\Delta\beta_k\|_1$  造成直接影响。使用著名的向量范数不等式

$$\frac{1}{\sqrt{n}} \|\Delta\beta_k\|_1 \leq \|\Delta\beta_k\|_2 \leq \|\Delta\beta_k\|_1 \quad (27)$$

可进一步得到 RR-OSELM 算法中  $\Delta\beta_k$  的  $L_2$  范数也是有界的,即 RR-OSELM 算法对离群值具有良好的抗干扰能力和鲁棒性。

### 3 实 验

为了评估 RR-OSELM 算法的鲁棒性,本文通过 1 个人工函数逼近问题和 7 个标杆混沌时间序列预测问题对其进行了实验验证,并将实验结果与 ELM<sup>[5]</sup>, R-ELM<sup>[19]</sup>, OSELM<sup>[4]</sup>, R-OSELM<sup>[13]</sup> 进行比较。对于每个测试实例,报告的实验结果为 20 次独立实验的平均值,算法性能评价标准采用均方根误差(RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (28)$$

式中:  $y_i, \hat{y}_i$  分别为真实值和预测值,  $N$  为测试样本的数目。

#### 3.1 “SinC”函数逼近

“SinC”函数逼近是评估学习模型学习拟合能

力的常用实例<sup>[5]</sup>,其函数表达式为

$$y(x) = \begin{cases} (\sin x)/x & x \neq 0 \\ 1 & x = 0 \end{cases} \quad (29)$$

根据式(29)分别生成各包含 5 000 个数据样本  $(x_i, y_i)$  的训练集和测试集,其中自变量  $x_i$  为区间  $(-10, 10)$  内的随机数。为了检验各种算法对于离群值的鲁棒性,训练集中 10% 的正常样本被离群值样本所替换,即将这些样本的输出  $y_i$  用范围在  $(-2, 2)$  之间的随机数来代替。

本实例的实验设置如下。参照文献[5],所有学习算法的隐层节点个数统一设置为 20;对于 OSELM, R-OSELM 和 RR-OSELM,初始训练样本个数设置为 100;此外,对于 R-ELM, R-OSELM 和 RR-OSELM,笔者验证了其在不同正则化参数条件下的泛化性能,并给出了最优条件下的实验结果。为了便于比较,每种算法均在训练集包含离群值和不包含离群值两种情形下分别进行对照实验,并收集各自情形下的测试 RMSE 和训练时间。

根据 OSELM 算法的稳定性分析<sup>[22]</sup>,该算法的泛化性能对隐层节点个数较为敏感,而在当前实验条件下 OSELM 的性能极不稳定,其 20 次实验的平均测试误差超出了合理的范围,故本实例中略去了 OSELM 的实验结果而仅给出其他 4 种算法的实验结果,如表 1 所示。比较 4 种算法在非离群值和离群值环境下的测试结果可得,当训练集中包含离群值时,ELM, R-ELM, R-OSELM 的预测误差增加了数十倍,而 RR-OSELM 的测试 RMSE 仅为原来的 1.7 倍。另外,在训练时间方面,ELM 和 R-ELM 这两种离线算法的训练时间明显少于 R-OSELM 和 RR-OSELM 这两种在线算法的训练时间;尽管 RR-OSELM 在在线学习过程中需要进行额外的加权计算,但其在线学习时间仅比 R-OSELM 稍长。总之,与 R-OSELM 相比,RR-OSELM 能在少量增加学习时间的情况下显著增强算法对于离群值的鲁棒性。

表 1 “SinC”函数逼近问题的性能比较

Tab. 1 Performance comparisons for learning “SinC”

算法	测试 RMSE		训练时间/s	
	无离群值	有离群值	无离群值	有离群值
ELM	2.62E-04	2.05E-02	0.010 93	0.012 50
R-ELM	8.96E-04	1.74E-02	0.005 46	0.003 12
R-OSELM	1.78E-03	1.72E-02	0.224 37	0.232 18
RR-OSELM	1.74 E-03	2.91E-03	0.246 09	0.259 06

为了更直观地比较 R-OSELM 和 RR-OSELM 对于离群值的抵抗能力,给出了这两种学习算法在训练集不包含离群值和包含离群值两种情形下对

“SinC”函数进行逼近的典型输出曲线图,分别如图 1,2 所示。由图 1 容易看出,R-OSELM 在非离群值环境下的实际输出与期望输出拟合较好,但在

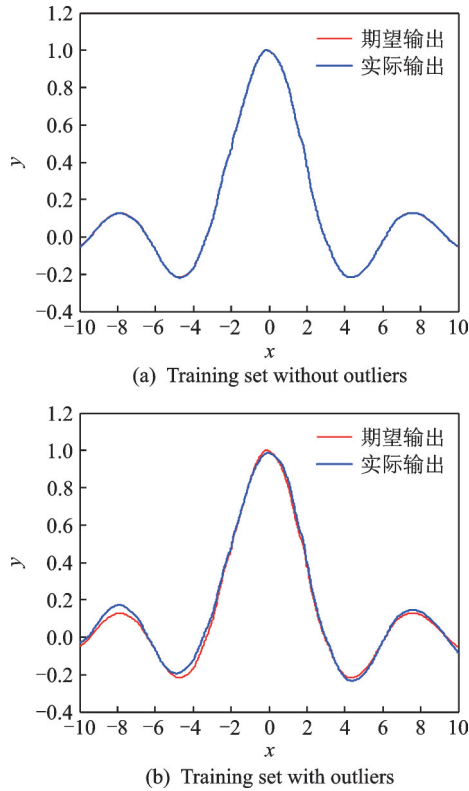


图1 R-OSELM的函数逼近结果

Fig. 1 Function approximation results of R-OSELM

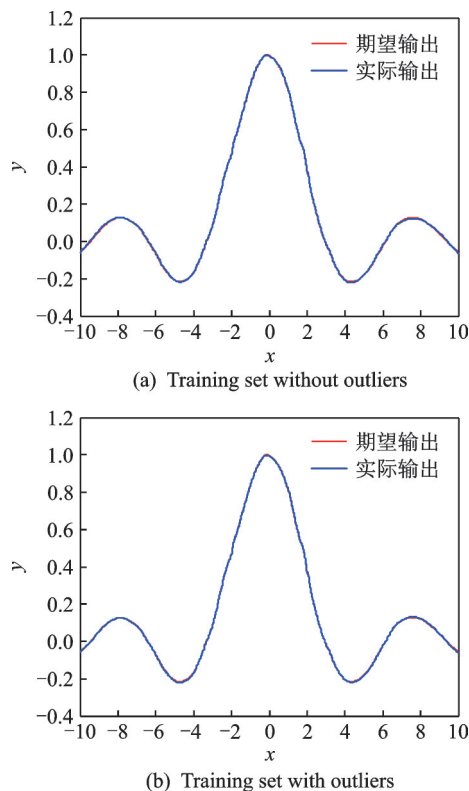


图2 RR-OSELM的函数逼近结果

Fig. 2 Function approximation results of RR-OSELM

训练集包含离群值时,其实际输出为了拟合无效的离群值数据而严重偏离期望输出;相对地,图2所示的拟合曲线表明RR-OSELM在两种情形下都

具有良好的学习拟合效果,这进一步验证了本文提出的RR-OSELM算法对于离群值具有较好的抗干扰能力。

### 3.2 混沌时间序列预测

本小节选择 Mackey-Glass, Rossler, Logistic, Henon, Lorenz, Ikeda 和 Quadratic 这7个经典的混沌时间序列作为实验数据集,时间序列数据采用四阶 Runge-Kutta 算法生成得到。为评估各种算法对于离群值的鲁棒性,在各时间序列的训练集中根据其大小不同添加了一定数量的离群值,这里离群值由随机数发生器随机生成,生成范围为各训练集中的最小值与最大值之间的开区间。表2给出了7个时间序列数据集的完整描述,其中  $d, \tau$  分别表示对时间序列进行相空间重构的嵌入维数和时间延迟。

表2 数据集描述

Tab. 2 Dataset description

时间序列	训练样本	测试样本	离群值个数	$(d, \tau)$
Mackey-Glass	1 000	500	8	(17, 1)
Rossler	1 500	500	13	(1, 5)
Logistic	1 500	500	13	(4, 1)
Henon	1 500	500	13	(4, 1)
Lorenz	1 500	500	13	(9, 1)
Ikeda	1 500	500	13	(5, 1)
Quadratic	1 500	500	13	(5, 1)

对于上述包含离群值的时间序列数据集,分别使用 ELM, R-ELM, OSELM, R-OSELM 和 RR-OSELM 5种算法对训练样本进行学习建模,并在测试集上对其预测性能进行比较。各算法的实验参数设置如下:对于 ELM 和 OSELM,需要确定的模型参数仅有隐层节点个数  $n$ ,为此,笔者在  $[10, 200]$  区间内以 10 为间隔逐渐增加  $n$ ,同时记录算法在不同  $n$  下的预测 RMSE,并报告 20 次实验中平均预测误差最小的实验结果。对于 R-ELM, R-OSELM 和 RR-OSELM,本文采用网格遍历法来评估算法在不同隐层节点个数  $n$  和正则化参数  $\delta$  的二元参数组合下的预测性能:其中  $n = \{10, 20, \dots, 190, 200\}$ ,  $\delta = \{10^{-10}, 10^{-9}, \dots, 10^4, 10^5\}$ ,平均预测误差最小的实验结果及其对应的模型参数将被报告。另外,对于 OSELM, R-OSELM, RR-OSELM 这3种在线学习算法,初始训练样本个数统一设置为 200。各种算法最终选择得到的最优模型参数如表3所示。

与表3对应,表4给出了各种算法在测试集上

的最优预测 RMSE。由表 4 可以看出,在所有 7 个包含离群值的时间序列数据集上,ELM、OSELM 的预测结果最差,采用正则化技术的 R-ELM、R-

OSELM 的预测结果稍好,而本文提出的 RR-OS-ELM 总能够取得最好的预测性能,表现出更强的鲁棒性。

表 3 各种算法的模型参数

Tab. 3 Parameter values of each algorithm

时间序列	ELM $n$	R-ELM ( $n, \delta$ )	OSELM $n$	R-OSELM ( $n, \delta$ )	RR-OSELM( $n, \delta$ )
Mackey-Glass	100	(150, $10^{-4}$ )	140	(160, $10^{-4}$ )	(200, $10^{-8}$ )
Rossler	120	(100, $10^{-8}$ )	110	(80, $10^{-8}$ )	(200, $10^{-7}$ )
Logistic	90	(200, $10^{-8}$ )	110	(190, $10^{-8}$ )	(200, $10^{-8}$ )
Henon	110	(160, $10^{-6}$ )	170	(200, $10^{-6}$ )	(200, $10^{-8}$ )
Lorenz	150	(200, $10^{-1}$ )	160	(200, $10^{-1}$ )	(200, $10^{-4}$ )
Ikeda	190	(200, $10^{-8}$ )	190	(200, $10^{-7}$ )	(190, $10^{-8}$ )
Quadratic	140	(190, $10^{-4}$ )	200	(170, $10^{-4}$ )	(200, $10^{-8}$ )

表 4 各种算法的预测 RMSE 比较

Tab. 4 Prediction RMSE comparisons of each algorithm

时间序列	学习算法				
	ELM	R-ELM	OSELM	R-OSELM	RR-OSELM
Mackey-Glass	6.63E-03	5.64E-03	6.95E-03	5.69E-03	8.45E-04
Rossler	2.79E-02	4.70E-03	7.03E-03	4.79E-03	2.66E-03
Logistic	8.52E-03	8.21E-03	1.22E-02	8.26E-03	3.74E-04
Henon	1.98E-02	1.72E-02	2.64E-02	1.72E-02	8.03E-04
Lorenz	5.80E-01	5.30E-01	6.42E-01	5.25E-01	2.51E-01
Ikeda	1.70E-01	1.64E-01	1.80E-01	1.65E-01	1.54E-01
Quadratic	4.28E-02	3.80E-02	5.24E-02	3.86E-02	4.82E-04

## 4 结 论

为增强 OSELM 在离群值环境下的鲁棒性,本文提出一种具有抗离群值能力的鲁棒正则化贯序超限学习机 RR-OSELM。与原始 OSELM 算法完全使用最小二乘作为代价函数不同,RR-OSELM 在代价函数中引入权值系数并基于样本的先验误差进行逆向加权计算,以抑制离群值残差对模型更新所造成的不利影响,从而有效降低了在线学习模型对于离群值的敏感性;同时 RR-OSELM 内置的正则化技术也保证了算法的持续稳定性。RR-OS-ELM 的鲁棒性通过实验进行了验证,实验结果表明 RR-OSELM 确实具有较对比算法更强的离群值抵抗能力。

与目前常见的基于离线批处理学习方式的鲁棒学习模型<sup>[15-19]</sup>相比,RR-OSELM 最突出的优势在于其在保持 OSELM 优越的在线学习能力的同时还兼具良好的抗离群值能力,特别适用于训练样本连续到达且包含离群值的在线学习问题。

### 参考文献:

[1] OLORUNNIMBE M K, VIKTOR H L, PAQUET E. Dynamic adaptation of online ensembles for drifting

data streams[J]. Journal of Intelligent Information Systems, 2018, 50(2): 291-313.

[2] MEJRI D, LIMAM M, WEIHS C. A new dynamic weighted majority control chart for data streams [J]. Soft Computing, 2018, 22(2): 511-522.

[3] LI C, WEI F, DONG W, et al. Dynamic structure embedded online multiple-output regression for streaming data[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2019, 41(2): 323-336.

[4] LIANG N Y, HUANG G B, SARATCHANDRAN P, et al. A fast and accurate online sequential learning algorithm for feedforward networks[J]. IEEE Transactions on Neural Networks, 2006, 17(6): 1411-1423.

[5] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2006, 70(1): 489-501.

[6] ALI M, DEO R C, DOWNS N J, et al. Multi-stage hybridized online sequential extreme learning machine integrated with Markov chain Monte Carlo copula-Bat algorithm for rainfall forecasting[J]. Atmospheric Research, 2018, 213: 450-464.

[7] ZHANG D, PENG X, PAN K, et al. A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust extreme learning

- machine[J]. *Energy Conversion and Management*, 2019, 180: 338-357.
- [8] NOBREGA J P, OLIVERRA A L. A sequential learning method with Kalman filter and extreme learning machine for regression and time series forecasting[J]. *Neurocomputing*, 2019, 337: 235-250.
- [9] GUO W, XU T, TANG K. M-estimator-based online sequential extreme learning machine for predicting chaotic time series with outliers[J]. *Neural Computing & Applications*, 2017, 28(12): 4093-4110.
- [10] GUO W, XU T, TANG K, et al. Online sequential extreme learning machine with generalized regularization and adaptive forgetting factor for time-varying system prediction[J]. *Mathematical Problems in Engineering*, 2018(7): 1-22.
- [11] 郭威, 徐涛, 于建江, 等. 基于M-estimator与可变遗忘因子的在线贯序超限学习机[J]. *电子与信息学报*, 2018, 40(6): 1360-1367.  
GUO Wei, XU Tao, YU Jianjiang, et al. Online sequential extreme learning machine based on m-estimator and variable forgetting factor[J]. *Journal of Electronics & Information Technology*, 2018, 40(6): 1360-1367.
- [12] LIAO B, ZHANG Z G, CHAN S C. A new robust Kalman filter-based subspace tracking algorithm in an impulsive noise environment[J]. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2010, 57(9): 740-744.
- [13] HUYNH HT, WON Y. Regularized online sequential learning algorithm for single-hidden layer feedforward neural networks[J]. *Pattern Recognition Letters*, 2011, 32(14): 1930-1935.
- [14] SUN L, CHEN B, TOH K A, et al. Sequential extreme learning machine incorporating survival error potential[J]. *Neurocomputing*, 2015, 155(5): 194-204.
- [15] ZHANG K, LUO M. Outlier-robust extreme learning machine for regression problems[J]. *Neurocomputing*, 2015, 151: 1519-1527.
- [16] BARRETO G A, BARROS A. A robust extreme learning machine for pattern classification with outliers[J]. *Neurocomputing*, 2016, 176: 3-13.
- [17] FRENAY B, VERLEYSSEN M. Reinforced extreme learning machines for fast robust regression in the presence of outliers[J]. *IEEE Transactions on Cybernetics*, 2016, 46(12): 3351-3363.
- [18] CHEN K, LV Q, LU Y, et al. Robust regularized extreme learning machine for regression using iteratively reweighted least squares[J]. *Neurocomputing*, 2017, 230: 345-358.
- [19] LI G, NIU P. An enhanced extreme learning machine based on ridge regression for regression[J]. *Neural Computing and Applications*, 2013, 22(3/4): 803-810.
- [20] BHOTTO M Z A, ANTONIOU A. Robust recursive least-squares adaptive-filtering algorithm for impulsive-noise environments[J]. *IEEE Signal Processing Letters*, 2011, 18(3): 185-188.
- [21] GOLUB G H, VAN LOAN C F. *Matrix computations* [M]. Baltimore: JHU Press, 2012: 65.
- [22] HORATA P, CHIEWCHANWATTANA S, SUNAT K. Enhancement of online sequential extreme learning machine based on the householder block exact inverse QRD recursive least squares[J]. *Neurocomputing*, 2015, 149: 239-252.

(编辑:张蓓)