

基于 M-estimator 与可变遗忘因子的在线贯序超限学习机

郭 威^{*①②} 徐 涛^② 于建江^① 汤克明^①

^①(盐城师范学院信息工程学院 盐城 224002)

^②(南京航空航天大学计算机科学与技术学院 南京 210016)

摘 要: 该文针对时变离群值环境下的在线学习问题, 提出一种基于 M-estimator 与可变遗忘因子的在线贯序超限学习机算法(VFF-M-OSELM)。VFF-M-OSELM 以在线贯序超限学习机模型为基础, 通过引入一种更加鲁棒的 M-estimator 代价函数来替代传统的最小二乘代价函数, 以提高模型对于离群值的在线处理能力和鲁棒性。同时 VFF-M-OSELM 通过融合使用一种新的可变遗忘因子方法进一步增强了其在时变环境下的动态跟踪能力和自适应性。仿真实例验证了所提算法的有效性。

关键词: 在线贯序超限学习机; M-估计; 可变遗忘因子; 鲁棒性; 自适应性

中图分类号: TP183

文献标识码: A

文章编号: 1009-5896(2018)06-1360-08

DOI: 10.11999/JEIT170800

Online Sequential Extreme Learning Machine Based on M-estimator and Variable Forgetting Factor

GUO Wei^{①②} XU Tao^② YU Jianjiang^① TANG Keming^①

^①(College of Information Engineering, Yancheng Teachers University, Yancheng 224002, China)

^②(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: To solve the online learning problem under the scenario of time-varying and containing outliers, this paper proposes an M-estimator and Variable Forgetting Factor based Online Sequential Extreme Learning Machine (VFF-M-OSELM). The VFF-M-OSELM is developed from the online sequential extreme learning machine algorithm and retains the same excellent sequential learning ability as it, it replaces the conventional Least-Squares (LS) cost function with a robust M-estimator based cost function to enhance the robustness of the learning model to outliers. Meanwhile, a new variable forgetting factor method is designed and incorporated in the VFF-M-OSELM to enhance further the dynamic tracking ability and adaptivity of the algorithm to time-varying system. The simulation results verify the effectiveness of the proposed algorithm.

Key words: Online Sequential Extreme Learning Machine (OSELM); M-estimator; Variable Forgetting Factor (VFF); Robustness; Adaptivity

1 引言

在线学习是当前机器学习领域的一个研究热点, 对于解决流式数据环境下的在线建模与预测问题具有重要价值^[1,2]。在线贯序超限学习机(Online Sequential Extreme Learning Machine, OSELM)是近年来提出的一种新颖而实用的在线学习算法^[3], 该算法以超限学习机(Extreme Learning Machine, ELM)^[4]的插值理论和逼近定理为基础, 将单隐层前馈神经网络(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方法递推计算输出权值以实现模型的在线更新。相比其他流行的在线学习算法, OSELM 不仅具有更好的泛化能力, 而且在学习速度上优势明显, 目前已在非线性系统在线识别^[5], 时间序列在线预测^[6,7]等研究领域展现出良好的应用价值和发展潜力。

离群值的在线检测与处理是实际在线应用中需要特别关注的问题, 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。为了提高 OSELM 对于含噪数据的处理能力, 文献[8]提出一种基于双目标优化的正则化 OSELM 算法(R-OSELM), 该算法通过使用 Tikhonov 正则化技术有效提高了模型的稳定性和泛化性能, 并在一定程度上增强了对噪声和离群值数据的抗干扰能力, 但该

收稿日期: 2017-08-08; 改回日期: 2018-01-16; 网络出版: 2018-03-15

*通信作者: 郭威 weigu031@163.com

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

Foundation Items: The National Natural Science Foundation of China (61603326, 61379064, 61273106)

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很弱。最近, 文献[9]结合噪声补偿机制提出一种基于最小均方差的贯序超限学习机算法 ESEP-ELM, 并用于含离群值混沌时间序列的在线学习与预测, 相比原始 OSELM, ESEP-ELM 具有更高的计算效率和稳定性, 并且对离群值有较好的抵抗能力, 但其学习精度较 OSELM 有所下降。

此外, 在许多在线应用中, 在线数据的潜在分布和变化趋势随时间不断发生变化^[10], 要求在线模型能够不断学习新的数据样本以获取最新知识, 并及时消除旧的失效样本的影响, 以快速跟踪目标系统的实时状态。为了增强 OSELM 在时变环境下的学习和跟踪能力, 学者们引入遗忘因子(Forgetting Factor, FF)的概念, 提出了基于 FF 的改进 OSELM 算法(FF-OSELM)^[11,12]。与原始 OSELM 算法对新旧学习样本采取等权处理方式不同, FF-OSELM 在在线学习过程中更加重视新近样本的贡献, 并以遗忘旧样本的方式间接突出新近样本的作用, 从而能够更好地跟踪时变系统的实时变化。为进一步增强模型的跟踪性能, 学者们还提出了多种遗忘因子可在线调整的 OSELM 算法, 如 LAFF-OSELM^[13], DFF-OSELM^[14]等。尽管这些算法对于一般的渐变系统展现出良好的跟踪能力, 但对于突变系统其遗忘因子的收敛速度相对较慢, 实时跟踪能力较差。

针对上述问题, 本文提出一种新的基于 M-estimator 与可变遗忘因子的 OSELM 算法(OSELM based on M-estimator and Variable Forgetting Factor, VFF-M-OSELM)。VFF-M-OSELM 沿用了 OSELM 贯序学习的基本思想, 但在学习模型中使用一种更加鲁棒的 M-estimator 代价函数来替代传统的最小二乘代价函数, 以有效降低学习模型对于离群值的敏感性; 此外, 本文设计了一种新的基于相对误差的可变遗忘因子方法并整合到 VFF-M-OSELM, 使得遗忘因子在在线学习过程中能够根据环境的变化而进行自适应迭代调整, 从而进一步提高了在线模型对于时变系统的跟踪性能。最后通过两个典型仿真实例验证了 VFF-M-OSELM 算法的鲁棒性和自适应性。

2 VFF-M-OSELM 算法

2.1 VFF-M-OSELM 学习模型

根据 OSELM 算法^[3]的推导过程可知, 原始 OSELM 算法采用最小二乘(LS)作为学习准则, 追求训练残差平方和的最小化; 然而, LS 学习准则对于离群值的天然敏感性使其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。此外, OSELM 在在线学习过程中只是不断地将新的学习样本加入学习器并进行贯序学习, 而并未对旧样本作任何处理, 极大地限制了其对于时变系统的跟踪性能。

为了增强 OSELM 在时变离群值环境下的鲁棒性和自适应跟踪能力, 本文引入 M-estimator 技术和可变遗忘因子(VFF)方法, 提出一种新的鲁棒自适应 OSELM 学习模型 VFF-M-OSELM。该学习模型对应的代价函数表示为

$$J(\beta_k) \triangleq \sum_{i=1}^k \lambda^{k-i} \rho(y_i - \mathbf{h}_i \beta_k) + \frac{1}{2} \delta \|\beta_k\|^2 \quad (1)$$

式中, $e_i = y_i - \mathbf{h}_i \beta_k$ 为目标输出与实际输出之间的残差, $\rho(\cdot)$ 为 M-estimator 函数, λ 为遗忘因子, δ 为广义正则化参数。

与原始 OSELM 相比, VFF-M-OSELM 主要有以下两点改进之处: 其一, VFF-M-OSELM 在其代价函数中使用了一种更加鲁棒的 M-estimator 函数来替代传统 OSELM 中的 LS 学习准则, 以减轻或消除离群值对学习模型所造成的不利影响。本文选择如下一种形式简单且性能良好的改进 Huber 函数作为 $\rho(\cdot)$:

$$\rho(e) = \begin{cases} e^2/2, & |e| \leq \alpha \\ \alpha^2/2, & |e| > \alpha \end{cases} \quad (2)$$

式中, α 为阈值参数, $| \cdot |$ 表示取绝对值操作。容易看出, 若当前训练残差 e 的绝对值大于阈值 α , 则当前训练样本很可能是无效的离群值样本, 此时 $\rho(\cdot)$ 为一常数, 从而能够对噪声或离群值样本所产生的较大扰动误差起到很好的抑制作用。

其二, VFF-M-OSELM 采用了与 FF-OSELM^[11,12]类似的遗忘因子方法来对不同时刻点的数据样本进行加权, 以体现新旧样本的不同贡献; 在此基础上, 我们还设计了一种新的基于相对误差的可变遗忘因子方法(详见 2.4 节), 使得遗忘因子在在线学习过程中能够根据环境的变化而进行自适应迭代调整, 从而进一步增强在线模型的动态跟踪能力和自适应性。

2.2 模型的递归求解

VFF-M-OSELM 模型的递归求解过程如下。让代价函数 $J(\beta_k)$ 对 β_k 求微分并令结果为零, 可得

$$\mathbf{R}_k \beta_k = \Phi_k \quad (3)$$

式中

$$\begin{aligned} \mathbf{R}_k &= \sum_{i=1}^k \lambda^{k-i} \varphi(e_i) \mathbf{h}_i^T \mathbf{h}_i + \delta \mathbf{I} \\ &= \lambda \mathbf{R}_{k-1} + \delta(1-\lambda) \mathbf{I} + \varphi(e_k) \mathbf{h}_k^T \mathbf{h}_k \end{aligned} \quad (4)$$

$$\Phi_k = \sum_{i=1}^k \lambda^{k-i} \varphi(e_i) \mathbf{h}_i^T \mathbf{y}_i = \lambda \Phi_{k-1} + \varphi(e_k) \mathbf{h}_k^T \mathbf{y}_k \quad (5)$$

其中, $\varphi(e) \triangleq \zeta(e)/e$, $\zeta(e) \triangleq \partial \rho(e)/\partial e$ 。定义:

$$\mathbf{R}_k^* = \lambda \mathbf{R}_{k-1} + \delta(1-\lambda) \mathbf{I} \quad (6)$$

则式(4)变形为

$$\mathbf{R}_k = \mathbf{R}_k^* + \varphi(e_k) \mathbf{h}_k^T \mathbf{h}_k \quad (7)$$

将 Sherman-Morrison-Woodbury 公式和 Sherman-Morrison 公式^[15]分别应用于式(6), 式(7)可得

$$\begin{aligned} (\mathbf{R}_k^*)^{-1} &= \lambda^{-1} \mathbf{R}_{k-1}^{-1} - \delta(1-\lambda)/\lambda^2 \mathbf{R}_{k-1}^{-1} \\ &\quad \cdot (\mathbf{I} + \delta(1-\lambda)/\lambda \mathbf{R}_{k-1}^{-1})^{-1} \mathbf{R}_{k-1}^{-1} \end{aligned} \quad (8)$$

$$\mathbf{R}_k^{-1} = (\mathbf{R}_k^*)^{-1} - \frac{\varphi(e_k) (\mathbf{R}_k^*)^{-1} \mathbf{h}_k^T \mathbf{h}_k (\mathbf{R}_k^*)^{-1}}{1 + \varphi(e_k) \mathbf{h}_k (\mathbf{R}_k^*)^{-1} \mathbf{h}_k^T} \quad (9)$$

令 $\mathbf{P}_k = \mathbf{R}_k^{-1}$, $\mathbf{P}_k^* = (\mathbf{R}_k^*)^{-1}$, 则式(8)和式(9)可分别重写为

$$\begin{aligned} \mathbf{P}_k^* &= \lambda^{-1} \mathbf{P}_{k-1} - \delta(1-\lambda)/\lambda^2 \\ &\quad \cdot \mathbf{P}_{k-1} (\mathbf{I} + \delta(1-\lambda)/\lambda \mathbf{P}_{k-1})^{-1} \mathbf{P}_{k-1} \end{aligned} \quad (10)$$

$$\mathbf{P}_k = \mathbf{P}_k^* - \frac{\varphi(e_k) \mathbf{P}_k^* \mathbf{h}_k^T \mathbf{h}_k \mathbf{P}_k^*}{1 + \varphi(e_k) \mathbf{h}_k \mathbf{P}_k^* \mathbf{h}_k^T} \quad (11)$$

然后综合式(3), 式(4)和式(5), 可推导出

$$\beta_k = \beta_{k-1} - \delta(1-\lambda) \mathbf{P}_k \beta_{k-1} + \varphi(e_k) \mathbf{P}_k \mathbf{h}_k^T \varepsilon_k \quad (12)$$

式中, $\varepsilon_k = y_k - \mathbf{h}_k \beta_{k-1}$ 称为先验误差, $e_k = y_k - \mathbf{h}_k \beta_k$ 称为后验误差。

式(10), 式(11), 式(12)构成了 VFF-M-OSELM 中输出权值 β_k 的递推计算公式。这里需要注意的是, 在式(12)中, 输出权值 β_k 的更新计算需要以后验误差 e_k 作为已知条件, 而 e_k 在 β_k 未知的情况下是不可得的, 显然, β_k 与 e_k 之间存在相互依赖关系。事实上, 在 β_k 未知的情形下先验误差 ε_k 可作为后验误差 e_k 的一个良好近似估计量, 即在式(11), 式(12)中, 我们可使用 ε_k 替代暂时未知的 e_k 来进行计算。

2.3 离群值在线检测

如 2.1 节所述, VFF-M-OSELM 模型采用了具有鲁棒特性的 M-estimator 方法来替代传统的 LS 方法, 以增强对离群值的在线处理能力。根据式(2)中 M-estimator 函数的定义, 这里主要通过比较当前预测误差与阈值参数 α 的大小关系来检测出潜在的离群值, 为此, 选择一个恰当的参数 α 就显得尤为重要。本文采用一种在线的参数估计方法以连续地估计阈值 α 。

根据统计学分析和中心极限定理, 可假设在非离群值环境下学习模型的估计误差服从均值为 0, 方差为 $\hat{\sigma}_k^2$ 的正态分布^[16], 即 $e_k \sim N(0, \hat{\sigma}_k^2)$ 。基于该

假设, 接下来有 $e'_k = e_k/\hat{\sigma}_k$ 服从标准正态分布, 即 $e'_k \sim N(0,1)$, 同时 $|e_k|$ 大于给定阈值 α 的概率 $\theta_\alpha = P_r\{|e_k| > \alpha\} = P_r\{|e'_k| > \alpha/\hat{\sigma}_k\}$ 。根据标准正态分布的性质, 有 $P_r\{e'_k < \alpha/\hat{\sigma}_k\} = \int_{-\infty}^{\alpha/\hat{\sigma}_k} e^{-x^2/2} dx / \sqrt{2\pi} = 1 - \theta_\alpha/2$, 即 $\alpha/\hat{\sigma}_k = \Theta^{-1}(1 - \theta_\alpha/2)$, 这里 $\Theta^{-1}(\cdot)$ 表示标准正态分布表中关于某个概率的分位数。不同的 θ_α 将产生不同的置信度, 如果 θ_α 设置为 0.01, 则有 99% 的置信度可检测并拒绝潜在的离群值, 此时对应的阈值参数 α 为

$$\alpha = \Theta^{-1}(0.995) \hat{\sigma}_k \approx 2.576 \hat{\sigma}_k \quad (13)$$

这里 $\hat{\sigma}_k$ 为非离群值环境下估计误差的标准差。关于 $\hat{\sigma}_k$ 的一个常用鲁棒估计量为 $1.483 \text{MAD}(e_k)$, 其中 $\text{MAD}(\cdot)$ 为中位数绝对偏差^[17], 但该计算方法涉及大量的取中位数操作, 计算复杂度很高。基于鲁棒回归理论^[17]和滑动窗口策略, 本文采用一种新的具有同样鲁棒特性但计算效率更高的估值计算方法来连续估计 $\hat{\sigma}_k$:

$$\begin{aligned} \hat{\sigma}_k &= \lambda \hat{\sigma}_{k-1} + (1-\lambda) 1.483(1 + 5/(L-1)) \\ &\quad \cdot \sqrt{\text{med}(W_k(e))} \end{aligned} \quad (14)$$

式中, λ 为遗忘因子, $\text{med}(\cdot)$ 表示取中位数操作, $W_k(e) = \{e_k^2, e_{k-1}^2, \dots, e_{k-L+1}^2\}$ 为当前 k 时刻关于估计误差的滑动窗口, L 为滑动窗口长度, $1.483(1 + 5/(L-1))$ 为有限样本条件下的修正因子。

2.4 可变遗忘因子方法

为了更好地跟踪时变系统的动态变化, 我们设计了一种新的基于相对误差的可变遗忘因子方法并整合到前述的 VFF-M-OSELM 模型中。首先定义一个关于相对误差的中间变量:

$$\phi_k = \mu_1 \phi_{k-1} + \mu_2 \left| \frac{y_k - \hat{y}_k}{y_k} \right| \quad (15)$$

式中, y_k 和 \hat{y}_k 分别为 k 时刻的真实值和预测值, $\left| \frac{y_k - \hat{y}_k}{y_k} \right|$ 为相对误差的绝对值。更新变量 ϕ_k 表示最近一段时间内相对误差的缓冲平均值, 这里 $0 < \mu_1, \mu_2 < 1$, 其中 μ_1 为误差平衡系数, 主要用于控制当前相对误差 $\left| \frac{y_k - \hat{y}_k}{y_k} \right|$ 和 ϕ_{k-1} 在 ϕ_k 中所占的比例, μ_2 为误差敏感系数, 主要用于控制 ϕ_k 趋近于 0 的速率, 对于收敛误差较大的系统, μ_2 的取值应相对较小, 反之亦然。一般地, μ_1 通常取接近于 1 的小数, μ_2 通常取接近于 0 的小数, 同时为了保证 ϕ_k

在收敛后期的单调下降特性，应满足 $\mu_1 + \mu_2 < 1$ 。事实上，更新变量 ϕ_k 通过描述相对误差的缓冲变化，能够很好地表征时变系统的实时状态，以 ϕ_k 为基础，遗忘因子可按式(16)进行更新计算：

$$\lambda_k = \left[\frac{1}{1 + \phi_k} \right]_{\lambda_-}^{\lambda_+} \quad (16)$$

式中， λ_+ 和 λ_- 分别表示遗忘因子的上限和下限，其中 λ_+ 通常取 1 或略小于 1 的小数，而 λ_- 的取值则与具体问题相关，并且 λ_- 的取值一般不宜过小以免产生数值不稳定性。

根据式(15)，式(16)容易分析得到，当系统状态发生显著变化时，当前预测误差将急剧增大，则 ϕ_k 在短时间内迅速增大，相应地 λ_k 迅速减小，使得旧的失效样本被加速遗忘，并利用新近样本建立新的学习模型，从而快速跟踪系统的最新状态。而随着新样本的不断学习，新的学习模型将逐渐趋于收敛，在此过程中，随着预测误差的降低， ϕ_k 逐渐减小并趋近于 0，相应地 λ_k 逐渐增大并趋近于 1，使得有效数据窗口长度增加，从而保证系统具有较高的稳态精度。因此从理论上分析，该 VFF 方法同时兼具突变环境下的快速跟踪能力和稳态环境下的持续学习能力，对于跟踪时变系统的实时状态是积极有效的，而且其实现过程也非常简洁高效。

3 实验

为了验证 VFF-M-OSELM 算法的有效性，本节通过两个典型实例对其进行了实验验证，并将实验结果与 R-OSELM^[8]，FF-OSELM^[11]，DFF-OSELM^[14]进行了比较与分析。

3.1 测试实例与实验设置

第 1 个测试实例是一个人工数据集，来源于文献[18]中的 Artificial data set 1，该数据集共有 2500 个数据样本，其中每 500 个样本的数据分布是显著不同的。第 2 个测试实例是一个典型的非线性时变系统，来源于文献[19]中的 Simulation 1，该实例包含时变系统在 3000 个时刻点的状态信息，其中系统的状态每隔 1000 个时刻点发生一次突变。

对于上述两个测试实例，分别使用 R-OSELM, FF-OSELM, DFF-OSELM 和本文提出的 VFF-M-OSELM 这 4 种在线学习算法对其进行在线建模与预测，两个实例的初始训练样本数分别为 200 和 500。对于所有 4 种在线学习算法，统一使用相同的 Sigmoid 激活函数，即 $G(\mathbf{a}, b, \mathbf{x}) = 1/(1 + \exp(-(\mathbf{a} \cdot \mathbf{x} + b)))$ ，其中输入权值向量 \mathbf{a} 和偏移 b 为范围在 $[-1, 1]$ 之间的随机数；隐层节点个数 $n = 100$ ，正则化参数 $\delta = 10^{-8}$ 。此外，各种算法的其他参数设置如下：对于 FF-OSELM，取遗忘因子 $\lambda = 0.98$ 和 $\lambda = 0.95$ 这两种典型取值分别进行实验；对于 DFF-OSELM，其参数设置参照文献[14]；对于 VFF-M-OSELM，设置误差窗口长度 $L = 10$ ， $\mu_1 = 0.9, \mu_2 = 0.08, \lambda_+ = 1, \lambda_- = 0.8$ 。

3.2 实验结果比较与分析

(1)跟踪能力验证：图 1 和图 2 分别给出了 4 种在线学习算法对实例 1 和实例 2 进行在线学习与预测所得到的预测误差曲线图，图中横坐标为时间，纵坐标为绝对误差(实际值与预测值之差的绝对值)。由实验结果可以看出，当系统状态发生突变时，R-OSELM 的预测误差急剧增大，并且受历史数据

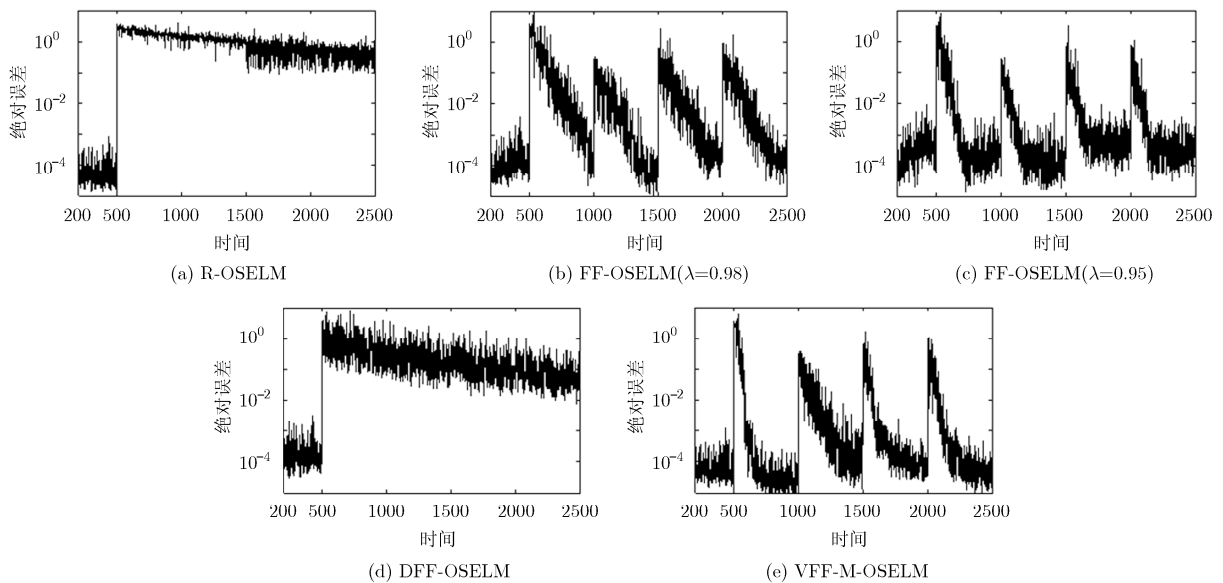


图 1 实例 1 上的预测误差

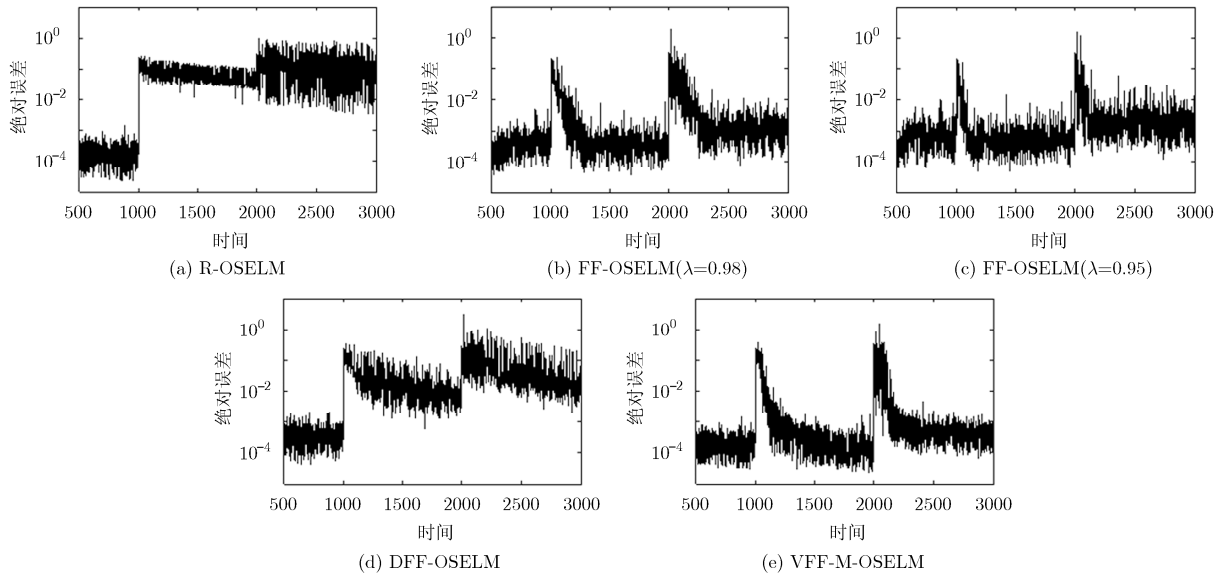


图 2 实例 2 上的预测误差

的影响,当前学习模型很难快速跟踪新的系统状态,预测误差长时间维持在较高水平,下降趋势非常缓慢。相对而言,具有遗忘机制的 FF-OSELM 和 VFF-M-OSELM 能够较快地摆脱旧的失效样本的影响,并利用新的学习样本建立新的学习模型,从而能快速跟踪时变系统的最新状态,故其预测误差从高位快速下降,并逐渐收敛到一个稳定水平。此外,从图中可以看出,基于方向遗忘因子的 DFF-OSELM 算法尽管对跟踪突变系统的新状态具有一定的作用,但其效果不够理想。

另外,仔细比较图 1(及图 2)中的(b), (c)可以发现,对于 FF-OSELM,遗忘因子越小,模型对突变系统的快速跟踪能力越强,预测误差的下降速度越快;但在系统趋于平稳时,学习模型所能利用的有效学习样本也越少,收敛误差越大,反之亦然。也就是说,采用固定遗忘因子的 FF-OSELM 算法在快速跟踪能力和稳态误差之间存在矛盾。相比而言,基于可变遗忘因子策略的 VFF-M-OSELM 算法在在线学习过程中能够根据系统的实时变化而进行遗忘因子的自适应调整(见图 3),故其既具有良好的快

速跟踪能力,同时也能够取得更高的稳态预测精度。

(2)鲁棒性验证:为进一步测试 VFF-M-OSELM 对离群值的鲁棒性,我们在实例 1 的 300, 800, 1300, 1800, 2300 时刻处,实例 2 的 700, 1300, 1800, 2300, 2800 时刻处添加了离群值,并重新进行了实验。由于 R-OSELM, FF-OSELM, DFF-OSELM 这 3 种在线算法对离群值的处理能力相似,作为代表,这里仅选择 FF-OSELM ($\lambda = 0.95$)作为对比算法。图 4 和图 5 分别给出了 FF-OSELM, VFF-M-OSELM 在实例 1 和实例 2 上的在线预测结果,容易看出,FF-OSELM 对于离群值非常敏感,一旦在线样本中出现离群值,其预测误差在随后的一段时间内持续偏大;相比而言,本文的 VFF-M-OSELM 通过采用 M-estimator 技术有效地抑制了离群值对在线学习模型造成干扰,因此可以看到离群值的存在并未对模型的预测性能造成明显影响。

图 6 给出了 VFF-M-OSELM 在突变离群值环境下遗忘因子的变化曲线图,与图 3 (非离群值环境)相比,两种情形下对应的实验结果基本一致,表明离群值也未对突变拐点的判断造成干扰。

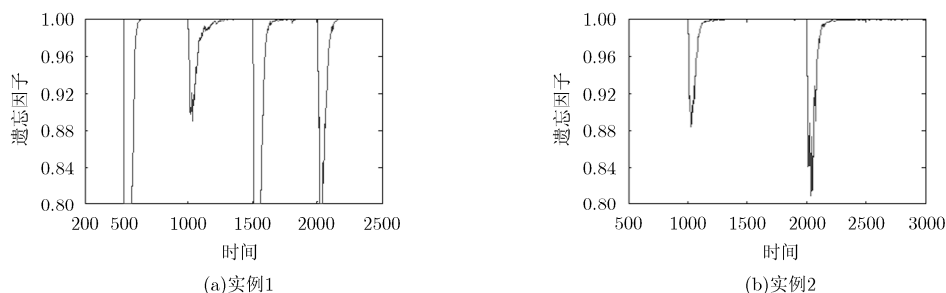
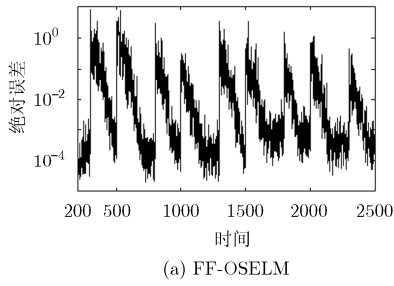
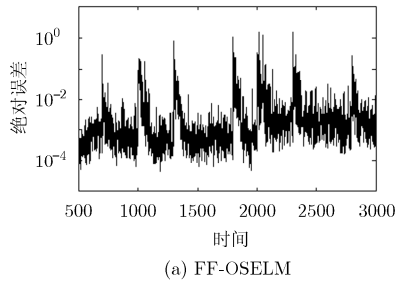


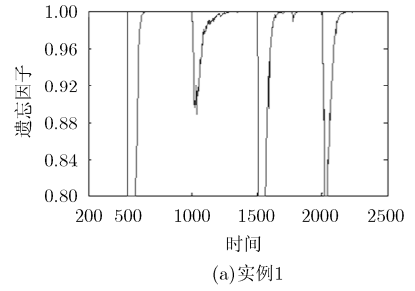
图 3 VFF-M-OSELM 中遗忘因子的变化曲线图



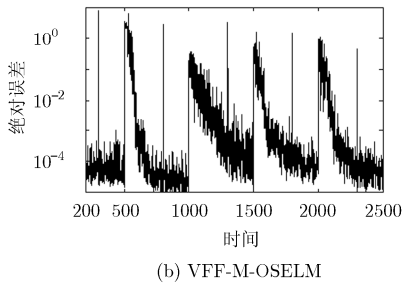
(a) FF-OSELM



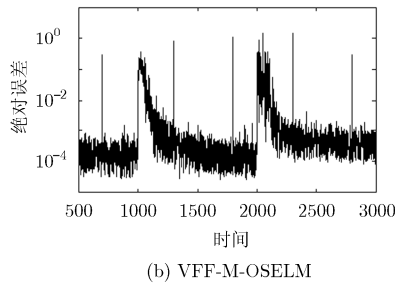
(a) FF-OSELM



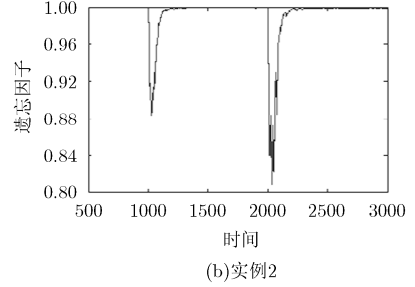
(a)实例1



(b) VFF-M-OSELM



(b) VFF-M-OSELM



(b)实例2

图4 实例1(包含离群值)上的预测误差

图5 实例2(包含离群值)上的预测误差

图6 VFF-M-OSELM 中遗忘因子的变化曲线图(包含离群值)

(3)学习效率比较: 表1给出了4种在线算法在两个实例数据集上进行在线学习和预测所花费的CPU时间,单位为s。实验结果显示, DFF-OSELM的学习效率最高, R-OSELM次之, 相对而言, FF-OSELM和VFF-M-OSELM这两种基于广义正则化技术^[11]的在线学习算法在提升算法稳定性的同时也增加了计算开销, 故其学习效率有所下降。与FF-OSELM相比, 尽管本文的VFF-M-OSELM算法由于在在线学习过程中需要进行额外的离群值在线检测和遗忘因子在线自适应调整, 从而使得算法的在线预测时间有所增加, 但其增幅不大。

表1 4种算法的在线预测时间(s)

| 在线算法 | 实例1 | 实例2 |
|-------------|------|------|
| R-OSELM | 1.75 | 1.93 |
| FF-OSELM | 4.51 | 4.93 |
| DFF-OSELM | 1.02 | 1.12 |
| VFF-M-OSELM | 5.29 | 5.94 |

3.3 关键参数对 VFF-M-OSELM 性能的影响

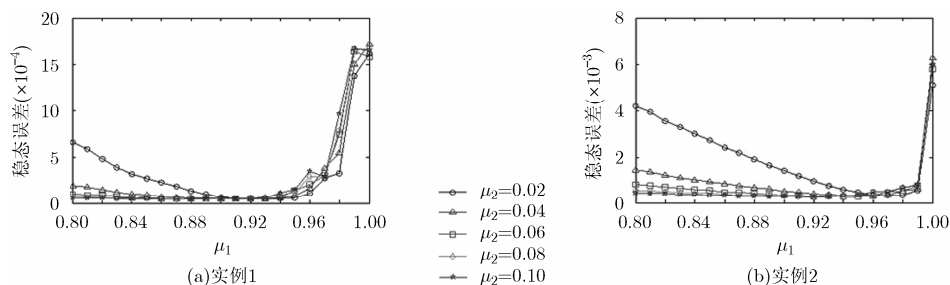
在VFF-M-OSELM中, 遗忘因子的迭代调整主要由误差平衡系数 μ_1 和误差敏感系数 μ_2 来控制(式(15))。为了更详尽和直观地说明这两个关键参数对VFF-M-OSELM性能的影响, 图7给出了VFF-M-OSELM在不同参数条件下的稳态误差, 这里, μ_1 的典型取值范围为 $\{0.80, 0.81, 0.82, \dots, 0.99, 1.00\}$, μ_2 的典型取值范围为 $\{0.02, 0.04, 0.06, 0.08, 0.10\}$ 。由图7可得: 首先, 对于每个不同的 μ_2 , 稳态误差

的总体变化趋势是一致的, 即随着 μ_1 的增加, 对应的稳态误差刚开始逐渐下降, 随后又不断上升, 并且当 μ_2 较大时, VFF-M-OSELM的性能对 μ_1 的敏感度较低, 即 μ_1 在较大的变化范围内VFF-M-OSELM都能取得较为平稳的稳态精度; 其次, 最优参数组合 (μ_1, μ_2) 并不唯一, 对于每个不同的 μ_2 , 我们总能找到一个最优的 μ_1 以取得最小的稳态误差, 并且当 μ_2 较大时, μ_1 的最优值相对较小; 第三, 对于不同的最优参数组合 (μ_1, μ_2) , 其对应的稳态收敛误差基本相当。基于以上发现本文认为, 对于一个给定的时变系统, 首先可从稳态收敛精度的角度出发预选出多组最优的参数组合 (μ_1, μ_2) , 然后结合系统的突变特性作进一步选择: 对于变化幅度较大的情形, 可选择 μ_1 较小, μ_2 较大的参数组合, 以增强模型对于突变的快速跟踪能力, 反之亦然。

另外, VFF-M-OSELM中的误差窗口长度 L 是进行离群值在线检测的关键参数(式(14))。为了评估 L 对VFF-M-OSELM性能的影响, 表2给出了VFF-M-OSELM在不同 L 下的预测误差和预测时间。从表2容易看出, VFF-M-OSELM在不同 L 下的预测误差和预测时间基本相当, 表明 L 对VFF-M-OSELM的预测性能和学习效率几乎没有影响, 换言之, L 能以较为宽松的方式进行选择, 根据我们的实验经验, L 在 $[10, 50]$ 区间内均能取得较为理想的实验结果。

4 结论

本文针对时变离群值环境下的在线学习问题, 提出了一种基于M-estimator和可变遗忘因子的在

图7 VFF-M-OSELM在不同 μ_1, μ_2 下的稳态误差表2 VFF-M-OSELM在不同 L 下的预测误差和预测时间

| 误差窗口 长度(L) | 实例 1 | | 实例 2 | |
|-------------------|------|-------------|-------|-------------|
| | 预测误差 | 预测时间 (s) | 预测误差 | 预测时间 (s) |
| 10 | 0.40 | 5.17 | 0.067 | 5.75 |
| 20 | 0.42 | 5.19 | 0.067 | 5.79 |
| 30 | 0.43 | 5.19 | 0.064 | 5.80 |
| 40 | 0.44 | 5.20 | 0.067 | 5.81 |
| 50 | 0.43 | 5.21 | 0.068 | 5.81 |

线贯序超限学习机算法 VFF-M-OSELM。VFF-M-OSELM 在保留 OSELM 算法良好在线学习能力的同时, 通过引入 M-estimator 技术和可变遗忘因子方法, 进一步增强了其对于离群值的抵抗能力和对于时变系统的动态跟踪能力。仿真实验结果表明, VFF-M-OSELM 具有较对比算法更好的自适应跟踪能力和鲁棒性, 对于时变离群值环境下的在线建模与预测问题是积极有效的。

参考文献

- [1] ZHANG Q, ZHANG P, LONG G, *et al.* Online learning from trapezoidal data streams[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(10): 2709–2723. doi: 10.1109/TKDE.2016.2563424.
- [2] LUGHOFER E. On-line active learning: A new paradigm to improve practical useability of data stream modeling methods[J]. *Information Sciences*, 2017, 415(1): 356–376. doi: 10.1016/j.ins.2017.06.038.
- [3] 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. doi: 10.1109/TNN.2006.880583.
- [4] HUANG G B, ZHU Q Y, and SIEW C K. Extreme learning machine: Theory and applications[J]. *Neurocomputing*, 2006, 70(1): 489–501. doi: 10.1016/j.neucom.2005.12.126.
- [5] LU X, ZHOU C, HUANG M, *et al.* Regularized online sequential extreme learning machine with adaptive regulation factor for time-varying nonlinear system[J]. *Neurocomputing*, 2016, 174(1): 617–626. doi: 10.1016/j.neucom.2015.09.068.
- [6] WANG X and HAN M. Online sequential extreme learning machine with kernels for nonstationary time series prediction [J]. *Neurocomputing*, 2014, 145(12): 90–97. doi: 10.1016/j.neucom.2014.05.068.
- [7] WANG X and HAN M. Improved extreme learning machine for multivariate time series online sequential prediction[J]. *Engineering Applications of Artificial Intelligence*, 2015, 40(4): 28–36. doi: 10.1016/j.engappai.2014.12.013.
- [8] HUYNH H T and WON Y. Regularized online sequential learning algorithm for single-hidden layer feedforward neural networks[J]. *Pattern Recognition Letters*, 2011, 32(14): 1930–1935. doi: 10.1016/j.patrec.2011.07.016.
- [9] SUN L, CHEN B, TOH K A, *et al.* Sequential extreme learning machine incorporating survival error potential[J]. *Neurocomputing*, 2015, 155(5): 194–204. doi: 10.1016/j.neucom.2014.12.029.
- [10] SUN J, FUJITA H, CHEN P, *et al.* Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble[J]. *Knowledge-Based Systems*, 2017, 120(C): 4–14. doi: 10.1016/j.knosys.2016.12.019.
- [11] 郭威, 徐涛, 汤克明, 等. 具有广义正则化与遗忘机制的在线贯序超限学习机[J]. *控制与决策*, 2017, 32(2): 247–254. doi: 10.13195/j.kzyjc.2015.1385.
- [12] GUO Wei, XU Tao, TANG Keming, *et al.* Online sequential extreme learning machine with generalized regularization and forgetting mechanism[J]. *Control and Decision*, 2017, 32(2): 247–254. doi: 10.13195/j.kzyjc.2015.1385.
- [13] CELAYA E and AGOSTINI A. Online EM with weight-based forgetting[J]. *Neural Computation*, 2015, 27(5): 1142–1157. doi: 10.1162/NECO_a_00723.
- [14] LIM J, LEE S, and PANG H. Low complexity adaptive forgetting factor for online sequential extreme learning machine (OS-ELM) for application to nonstationary system estimations[J]. *Neural Computing and Applications*, 2013, 22(3-4): 569–576. doi: 10.1007/s00521-012-0873-x.

- on-line Extreme Learning Machines with variable forgetting factor for dynamic system prediction[J]. *Neurocomputing*, 2016, 171(C): 693–707. doi: 10.1016/j.neucom.2015.07.035.
- [15] GOLUB G H and VAN LOAN C F. Matrix Computations[M]. Baltimore: JHU Press, 2012: 65.
- [16] BARRETO G A and BARROS A. A robust extreme learning machine for pattern classification with outliers[J]. *Neurocomputing*, 2016, 176(C): 3–13. doi: 10.1016/j.neucom.2014.10.095.
- [17] ROUSSEEUW P J and LEROY A M. Robust Regression and Outlier Detection[M]. New York: John Wiley & Sons, 2005: 43–44.
- [18] PEREZ-SANCHEZ B, FEONTENLA-ROMERO O, GUIJARRO-BERDINAS B, *et al.* An online learning algorithm for adaptable topologies of neural networks[J]. *Expert Systems with Applications*, 2013, 40(18): 7294–7304. doi: 10.1016/j.eswa.2013.06.066.
- [19] ZHOU X, LIU Z, and ZHU C. Online regularized and kernelized extreme learning machines with forgetting mechanism[J]. *Mathematical Problems in Engineering*, 2014, 2014(1): 1–11. doi: 10.1155/2014/938548.
- 郭威: 男, 1983年生, 讲师, 博士, 研究方向为机器学习、数据挖掘.
- 徐涛: 男, 1962年生, 教授, 博士生导师, 研究方向为数据挖掘、智能信息处理.
- 于建江: 男, 1975年生, 教授, 博士, 研究方向为神经网络、网络化控制系统.
- 汤克明: 男, 1965年生, 教授, 博士, 研究方向为数据挖掘、智能计算.