

具有广义正则化与遗忘机制的在线贯序超限学习机

郭威^{1,3}, 徐涛^{1,2†}, 汤克明³, 于建江³

(1. 南京航空航天大学 计算机科学与技术学院, 南京 210016; 2. 中国民航大学 计算机科学与技术学院, 天津 300300; 3. 盐城师范学院 信息工程学院, 江苏 盐城 224002)

摘要: 针对非平稳时间序列预测问题, 提出一种具有广义正则化与遗忘机制的在线贯序超限学习机算法. 该算法以增量学习新样本的方式实现在线学习, 以遗忘旧的失效样本的方式增强对非平稳系统的动态跟踪能力, 并通过引入一种广义的 l_2 正则化使其具有持续的正则化功能, 从而保证算法的持续稳定性. 仿真实例表明, 所提出算法具有较同类算法更好的稳定性和更小的预测误差, 适用于具有动态变化特性的非平稳时间序列在线建模与预测.

关键词: 在线贯序超限学习机; 广义正则化; 遗忘因子; 时间序列预测

中图分类号: TP183

文献标志码: A

Online sequential extreme learning machine with generalized regularization and forgetting mechanism

GUO Wei^{1,3}, XU Tao^{1,2†}, TANG Ke-ming³, YU Jian-jiang³

(1. School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; 2. School of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China; 3. School of Information Engineering, Yancheng Teachers University, Yancheng 224002, China)

Abstract: To solve the prediction problem of nonstationary time series, this paper proposes an online sequential extreme learning machine with forgetting and generalized regularization(OSELM-FGR). The proposed OSELM-FGR is able to learn the newly arrived samples incrementally by a recursive fashion, and has the improved ability to track the dynamic behavior of time-varying systems by forgetting the outdated samples in the learning process. Moreover, a generalized l_2 regularization is introduced into the OSELM-FGR to make the proposed algorithm have a persistent stability. Detailed performance comparisons of the OSELM-FGR with its counterparts are carried out. The experimental results show that, the proposed OSELM-FGR has better performance in the sense of stability and prediction accuracy, which can be applied to the online modeling and prediction of nonstationary time series with dynamic changes.

Keywords: online sequential extreme learning machine; generalized regularization; forgetting factor; time series prediction

0 引言

时间序列预测由于具有广泛的应用价值而受到学者的普遍关注. 目前, 已有的时间序列预测研究大多基于离线处理方式^[1-4]. 然而, 在股价预测、天气预测、交通流量预测等现实应用中, 时间序列数据往往以数据流的形式贯序到达, 且数据的潜在分布和变化趋势随时间不断发生变化. 在此情形下, 具有增量更新能力的在线学习算法更适合处理这种非平稳时间序列预测问题^[5]. 文献[6]综述了目前时间序列在线

预测的主要方法, 包括重新建模方法、动态神经网络方法、在线支持向量回归方法和核自适应滤波器方法等, 但这些方法仍存在执行效率较低、非平稳环境下预测精度不高等不足.

在线贯序超限学习机(OSELM)是近年来提出的一种新兴在线学习算法^[7], 能够以逐一或逐块的方式增量学习数据样本, 并采用递归最小二乘方法实现模型的在线更新. 与其他在线学习算法相比, OSELM具有学习速度快、泛化能力强、实现简单等

收稿日期: 2015-11-10; 修回日期: 2016-02-24.

基金项目: 国家自然科学基金项目(61603326, 61379064, 61273106); 国家科技支撑计划课题(2014BAJ04B02); 中央高校基本科研业务费专项资金项目(3122014D032); 中国民航信息技术科研基地开放课题(CAAC-ITRB-201401).

作者简介: 郭威(1983—), 男, 博士生, 从事数据挖掘、机器学习的研究; 徐涛(1962—), 男, 教授, 博士生导师, 从事数据挖掘、智能信息处理等研究.

†通讯作者. E-mail: txu@cauc.edu.cn

优点,已在时间序列预测等问题中获得了成功应用^[8-9],但同时也存在稳定性较差的缺点.由于OSELM在每次迭代学习过程中都涉及矩阵求逆计算,一旦隐层输出矩阵的自相关矩阵为奇异或病态,OSELM的泛化能力将严重下降.正则化技术是解决病态矩阵求逆问题的常用方法,基于此,文献[10]将Tikhonov正则化与OSELM相结合,提出一种正则化OSELM(R-OSELM)以提高算法的稳定性.在此基础上,为了更好地跟踪混沌时间序列的动态变化特性,文献[9]引入遗忘因子(FF)的概念,提出一种具有遗忘机制的R-OSELM(FR-OSELM).与OSELM和R-OSELM对旧样本采取等权处理方式不同,FR-OSELM在在线学习过程中更加重视新近样本的贡献,以遗忘旧样本的方式间接突出新近样本的作用,使更新后的预测模型更加接近时变系统的当前状态,从而进一步提高了模型的预测精度.然而,由于FR-OSELM采用了一种特殊的指数遗忘正则化,在FF的作用下,其正则化功能将随着时间的推移而逐渐被削弱直至完全失效,最终导致FR-OSELM出现同原始OSELM同样的由于病态矩阵求逆而引起的算法发散问题,严重影响了算法的稳定性和实用性.最近,文献[11]提出了一种新的方向遗忘因子OSELM(DF-OSELM),但该算法并未考虑其中潜在的病态矩阵求逆问题,且算法中FF更新策略较为繁杂.

面向非平稳时间序列预测问题,本文提出一种具有广义正则化与遗忘机制的在线贯序超限学习机(OSELM-FGR).OSELM-FGR在保留FF动态跟踪能力的同时,通过引入一种广义的 l_2 正则化使得算法具有持续的正则化功能,从而保证了算法的持续稳定性;同时本文推导了广义正则化条件下OSELM-FGR的递推计算公式,使其适用于在线学习与预测.两个典型非平稳时间序列在线预测实例验证了OSELM-FGR的有效性和实用性.

1 OSELM算法

OSELM是超限学习机^[12](ELM)的在线学习版本.对于 N 个任意的相异样本 $(x_j, t_j) \in R^d \times R^m$,具有 n 个隐层节点的单隐层前馈神经网络(SLFN)的数学模型为

$$\sum_{i=1}^n \beta_i g_i(x_j) = \sum_{i=1}^n \beta_i G(a_i, b_i, x_j), \quad j = 1, 2, \dots, N. \quad (1)$$

其中: a_i 和 b_i 分别是第 i 个隐层节点的输入权值和偏移, β_i 是连接第 i 个隐层节点与输出层的输出权值, $g_i(x_j) = G(a_i, b_i, x_j)$ 表示第 i 个隐层节点关于输入

向量 x_j 的输出.

该SLFN能以零误差逼近这 N 个样本,这意味着存在 (a_i, b_i) 和 β_i 使得下式成立:

$$\sum_{i=1}^n \beta_i G(a_i, b_i, x_j) = t_j, \quad j = 1, 2, \dots, N. \quad (2)$$

该式可写成如下矩阵形式:

$$H\beta = T. \quad (3)$$

其中

$$H = [h_1^T \quad h_2^T \quad \dots \quad h_N^T]^T = \begin{bmatrix} G(a_1, b_1, x_1) & \dots & G(a_n, b_n, x_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & \dots & G(a_n, b_n, x_N) \end{bmatrix}_{N \times n}, \quad (4)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_N^T \end{bmatrix}_{n \times m}, \quad T = \begin{bmatrix} t_1^T \\ t_2^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}. \quad (5)$$

根据文献[12]中的插值理论,给定一个SLFN和 N 个任意不同样本,只要隐层激活函数无限可微,隐层节点参数 (a_i, b_i) 可随机赋值并保持不变,此时 H 为一常数矩阵,则SLFN的训练问题便转化为求解线性系统 $H\beta = T$ 的最小二乘解 $\hat{\beta}$,即

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

ELM采用最小模小二乘解作为输出权值,即

$$\hat{\beta} = H^\dagger T, \quad (7)$$

其中 H^\dagger 为 H 的Moore-Penrose广义逆.如果 $H^T H$ 非奇异,则式(7)可写为

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

为了适应实时在线学习的需要,Liang等^[7]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OSELM.OSELM的学习过程包括一个初始的ELM批量学习过程和一个连续的贯序学习过程,具体如下.

在初始化阶段,给定初始训练子集 $\Omega_{k-1} = \{(x_j, t_j) | j = 1, 2, \dots, k-1\}$,由式(8)计算初始输出权值为

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

其中

$$P_{k-1} = (H_{k-1}^T H_{k-1})^{-1}, \quad T_{k-1} = [t_1 \quad t_2 \quad \dots \quad t_{k-1}]^T, \\ H_{k-1} = [h_1^T \quad h_2^T \quad \dots \quad h_{k-1}^T]^T.$$

在贯序学习阶段,每当获取到新的训练样本

(x_k, t_k) ,按照下式递推计算输出权值:

$$P_k = P_{k-1} - \frac{P_{k-1}h_k^T h_k P_{k-1}}{1 + h_k P_{k-1} h_k^T},$$

$$\beta_k = \beta_{k-1} + P_k h_k^T (t_k - h_k \beta_{k-1}), \quad (10)$$

其中 $h_k = [G(a_1, b_1, x_k) \cdots G(a_n, b_n, x_k)]$.

2 具有广义正则化与遗忘机制的OSELM

2.1 FR-OSELM存在的问题

尽管 OSELM 是一种优秀的在线学习算法,但其自身的不稳定性以及对所有学习样本采取等权处理方式使其很难直接应用于非平稳时间序列预测问题. FR-OSELM 以 OSELM 为基础,合并使用正则化技术和遗忘机制,在保持 OSELM 优越的在线学习能力的同时,进一步增强了算法的稳定性和动态跟踪能力.然而,由于 FF 的作用,FR-OSELM 的正则化功能也将随着时间的推移而被逐渐“遗忘”,即算法仅在有限时间内稳定有效.进一步的理论分析如下.

文献[9]从新旧学习样本对预测模型“贡献度”不同的直观分析出发,采用遗忘的方式对旧样本进行加权处理,并经过推演计算得到 FR-OSELM 算法.理论上,FR-OSELM 算法等价于最小化如下带有 FF 和正则化项的最小二乘代价函数:

$$J_{\text{FR}}(\beta_k) = \sum_{i=1}^k \lambda^{k-i} |e_i|^2 + \delta \lambda^k \|\beta_k\|^2 = \sum_{i=1}^k \lambda^{k-i} |t_i - h_i \beta_k|^2 + \delta \lambda^k \|\beta_k\|^2. \quad (11)$$

其中: λ 和 δ 分别为 FF 和正则化参数, λ 用来对旧样本进行遗忘加权, δ 用来提高解的稳定性.运用递归最小二乘方法^[13]对式(11)进行求解,可得到与文献[9]完全一致的关于 β_k 的递推计算公式,即

$$P_k = \frac{P_{k-1}}{\lambda} - \frac{P_{k-1}h_k^T h_k P_{k-1}}{\lambda(\lambda + h_k P_{k-1} h_k^T)},$$

$$\beta_k = \beta_{k-1} + P_k h_k^T (t_k - h_k \beta_{k-1}). \quad (12)$$

由式(11)可以发现,FR-OSELM 等价使用了一种特殊的指数遗忘正则化项 $\delta \lambda^k \|\beta_k\|^2$,这种特殊的选择使得 FR-OSELM 的递推计算公式(式(12))具有与原始 OSELM(式(10))同样的简洁性,且二者的计算复杂度基本相同,均为 $O(n^2)$, n 为隐层节点个数.然而,由于当 $0 < \lambda < 1$ 时,正则化项 $\delta \lambda^k \|\beta_k\|^2$ 呈指数递减,随着时间 k 的不断增大, λ^k 单调递减并趋近于 0,即其正则化功能将逐渐弱减直至完全失效,从而最终导致 FR-OSELM 出现同原始 OSELM 同样的由于病态矩阵求逆而引起的算法发散问题.

2.2 具有广义正则化与遗忘机制的OSELM

为有效解决 FR-OSELM 中存在的问题,进一步提高算法的稳定性和实用性,本文提出一种具有广义正则化与遗忘机制的在线贯序超限学习机(OSELM-FGR). OSELM-FGR 在其代价函数中引入一种新的常系数正则化项 $\delta \|\beta_k\|^2$ 来替代式(11)中的指数遗忘正则化项 $\delta \lambda^k \|\beta_k\|^2$,使其正则化功能不会受遗忘因子 λ 的影响而逐渐被削弱,从而保证算法具有持续的稳定性. OSELM-FGR 的代价函数表示如下:

$$J_{\text{FGR}}(\beta_k) = \sum_{i=1}^k \lambda^{k-i} |t_i - h_i \beta_k|^2 + \delta \|\beta_k\|^2. \quad (13)$$

下面通过推导广义正则化条件下 β_k 的递推计算表达式来满足在线学习的要求.将代价函数 $J_{\text{FGR}}(\beta_k)$ 对 β_k 求微分,并令 $\partial J_{\text{FGR}}(\beta_k) / \partial \beta_k = 0$,可得

$$R_k \beta_k = S_k. \quad (14)$$

其中

$$R_k = \sum_{i=1}^k \lambda^{k-i} h_i^T h_i + \delta I = \lambda R_{k-1} + \delta(1 - \lambda)I + h_k^T h_k, \quad (15)$$

$$S_k = \sum_{i=1}^k \lambda^{k-i} h_i^T t_i = \lambda S_{k-1} + h_k^T t_k. \quad (16)$$

定义

$$R_k^* = \lambda R_{k-1} + \delta(1 - \lambda)I, \quad (17)$$

则式(15)变为

$$R_k = R_k^* + h_k^T h_k. \quad (18)$$

将 Sherman-Morrison-Woodbury 公式和 Sherman-Morrison 公式^[14]分别应用于式(17)和(18),可得

$$(R_k^*)^{-1} = (\lambda R_{k-1} + \delta(1 - \lambda)I)^{-1} = \frac{1}{\lambda} R_{k-1}^{-1} - \frac{\delta(1 - \lambda)}{\lambda^2} R_{k-1}^{-1} \left(I + \frac{\delta(1 - \lambda)}{\lambda} R_{k-1}^{-1} \right)^{-1} R_{k-1}^{-1}, \quad (19)$$

$$R_k^{-1} = (R_k^* + h_k^T h_k)^{-1} = (R_k^*)^{-1} - \frac{(R_k^*)^{-1} h_k^T h_k (R_k^*)^{-1}}{1 + h_k (R_k^*)^{-1} h_k^T}. \quad (20)$$

令 $P_k = R_k^{-1}$, $P_k^* = (R_k^*)^{-1}$,则式(19)和(20)可分别重写为

$$P_k^* = \frac{1}{\lambda} P_{k-1} - \frac{\delta(1 - \lambda)}{\lambda^2} P_{k-1} \left(I + \frac{\delta(1 - \lambda)}{\lambda} P_{k-1} \right)^{-1} P_{k-1}, \quad (21)$$

$$P_k = P_k^* - \frac{P_k^* h_k^T h_k P_k^*}{1 + h_k P_k^* h_k^T}. \quad (22)$$

由式(14)和(16)可得

$$\beta_k = R_k^{-1} S_k = P_k (\lambda S_{k-1} + h_k^T t_k) = P_k (\lambda R_{k-1} \beta_{k-1} + h_k^T t_k). \quad (23)$$

将式(15)代入(23),可得

$$\begin{aligned} \beta_k &= P_k ((R_k - \delta(1-\lambda)I - h_k^T h_k) \beta_{k-1} + h_k^T t_k) = \\ &\beta_{k-1} - \delta(1-\lambda) P_k \beta_{k-1} - P_k h_k^T h_k \beta_{k-1} + P_k h_k^T t_k = \\ &\beta_{k-1} + P_k h_k^T (t_k - h_k \beta_{k-1}) - \delta(1-\lambda) P_k \beta_{k-1}. \end{aligned} \quad (24)$$

式(21),式(22)和(24)构成OSELM-FGR算法中输出权值 β_k 的递推计算公式.

与FR-OSELM的输出权值递推计算公式(式(12))相比,由于OSELM-FGR采用了广义的 l_2 正则化,其得到的递推计算公式较FR-OSELM稍显复杂,且式(21)还涉及对维度为 n 的矩阵进行求逆计算,使得算法在每个迭代步的计算复杂度达到 $O(n^3)$.幸运的是,在实际应用中,由于正则化参数 δ 一般取数量级在 10^{-4} 甚至更小的正实数^[9](取值过大相当于引入额外的噪音而影响建模的准确性),遗忘因子 λ 一般取略小于1的正常数^[13],则 $\delta(1-\lambda)/\lambda$ 为一很小的实数,因此可近似认为 $(\delta(1-\lambda)/\lambda)^2 \approx 0$,由此可得

$$\left(I + \frac{\delta(1-\lambda)}{\lambda} P_{k-1}\right) \left(I - \frac{\delta(1-\lambda)}{\lambda} P_{k-1}\right) \approx I, \quad (25)$$

进一步可得

$$\left(I + \frac{\delta(1-\lambda)}{\lambda} P_{k-1}\right)^{-1} \approx \left(I - \frac{\delta(1-\lambda)}{\lambda} P_{k-1}\right). \quad (26)$$

因此式(21)可近似表示为

$$P_k^* \approx \frac{1}{\lambda} P_{k-1} - \frac{\delta(1-\lambda)}{\lambda^2} P_{k-1} \left(I - \frac{\delta(1-\lambda)}{\lambda} P_{k-1}\right) P_{k-1}. \quad (27)$$

这时,OSELM-FGR算法在每个迭代步的计算复杂度降至 $O(n^2)$,这与原始OSELM算法是相同的.

综上,OSELM-FGR算法的完整描述如下:

假定数据样本 (x_j, t_j) 以数据流的形式逐一到达,激活函数为 $G(a, b, x)$,隐层节点个数为 n ,正则化参数为 δ ,遗忘因子为 λ .

Step 1: 初始化. 对于给定的初始训练样本子集 $\Omega_{k-1} = \{(x_j, t_j) | j = 1, \dots, k-1\}$,进行如下操作:

- 1) 随机生成隐层节点参数 (a_i, b_i) , $i = 1, 2, \dots, n$;
- 2) 按照式(4)计算初始的隐层输出矩阵 H_{k-1} ;
- 3) 计算初始的输出权值

$$\beta_{k-1} = P_{k-1} H_{k-1}^T T_{k-1}. \quad (28)$$

其中

$$P_{k-1} = (H_{k-1}^T H_{k-1} + \lambda I)^{-1},$$

$$T_{k-1} = [t_1 \ t_2 \ \dots \ t_{k-1}]^T.$$

Step 2: 在线学习与预测. 对于每个新样本 (x_k, t_k) ,进行如下操作:

- 1) 计算该样本输入 x_k 对应的隐层输出向量为

$$h_k = [G(a_1, b_1, x_k), G(a_2, b_2, x_k), \dots, G(a_n, b_n, x_k)]; \quad (29)$$

- 2) 预测 x_k 对应的输出值,即 t_k 的预测值为

$$\hat{t}_k = h_k \beta_{k-1}; \quad (30)$$

- 3) 在得到真实输出值 t_k 后,更新计算输出权值,即

$$\begin{aligned} P_k^* &= \frac{1}{\lambda} P_{k-1} - \frac{\delta(1-\lambda)}{\lambda^2} P_{k-1} \left(I - \frac{\delta(1-\lambda)}{\lambda} P_{k-1} \right) P_{k-1}, \\ P_k &= P_k^* - \frac{P_k^* h_k^T h_k P_k^*}{1 + h_k P_k^* h_k^T}, \\ \beta_k &= \beta_{k-1} + P_k h_k^T (t_k - h_k \beta_{k-1}) - \delta(1-\lambda) P_k \beta_{k-1}; \end{aligned} \quad (31)$$

- 4) 置 $k = k + 1$,返回Step 2.

3 仿真实验

为了验证OSELM-FGR算法的有效性和实用性,下面将其应用于两个非平稳时间序列的在线预测,并将实验结果与OSELM, R-OSELM, FR-OSELM和DFF-OSELM进行比较.以上5种在线学习算法均使用同样的Sigmoid激活函数,即 $G(a, b, x) = 1/(1 + \exp(-(ax + b)))$,其中输入权值向量 a 与偏移 b 为范围在 $[-1, 1]$ 之间的随机数.由于隐层节点个数 n 是影响神经网络学习精度和泛化能力的重要因素,参照文献[9],对每种学习算法分别选择 $n = 20$, $n = 50$ 和 $n = 100$ 三种情形进行实验,3种情形对应的初始训练样本数分别为50,100和200.此外,各种算法的其他参数设置如下:对于R-OSELM算法, $\delta = 10^{-8}$;对于FR-OSELM算法和OSELM-FGR算法, $\delta = 10^{-8}$, $\lambda = 0.98$;对于DFF-OSELM算法,其参数设置参照文献[11],并额外增加了一个相同的正则化参数($\delta = 10^{-8}$)来增强算法的稳定性.对于每个测试实例,给出的预测精度和学习时间均为30次独立实验的平均值,算法性能评价标准采用均方根误差,即

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{j=1}^N (t_j - \hat{t}_j)^2}. \quad (32)$$

其中: t_j, \hat{t}_j 分别为真实值和预测值; N 为测试样本数.

3.1 动态时变系统

在动态时变环境中,数据的分布随时间持续变化.这里采用一个非平稳时间序列数据集来模拟时变系统的这种动态特性.该数据集包含4000个样本,每个样本的输入由服从均值为0、标准差为0.1的正态分布的4个随机变量构成,样本输出为输入向量与系数向量的内积.为了构建非平稳的时间序列,设系数向量的初始值 $a(0) = [0.5 \ 0.2 \ 0.7 \ 0.8]$,其值随后按下式变化:

$$a_j(i) = a_j(i-1) + 2 \times 10^{-6}(\exp(a_j(i-1)))^{\frac{1}{2}},$$

$$i = 1, 2, \dots, 4000, j = 1, 2, 3, 4. \quad (33)$$

其中: i 代表样本序号, j 代表系数向量的分量下标.

表1给出了5种在线学习算法在执行200,500,1000,2000和3000步在线预测时所得到的预测RMSE的平均值.表1中“Inf”表示无穷大.由表1可见,OSELM由于在迭代计算过程中受病态矩阵求逆计算的影响,预测结果不够稳定,特别是当隐层节点个数为50和100时,得到的预测结果极不可靠.R-OSELM通过引入正则化技术避免了病态矩阵求逆问题,得到的预测RMSE持续稳定有效.尽管FR-OSELM也引入了正则化,但其正则化功能仅在短时间内有效,随着预测步长的增加,算法完全失效,得到同OSELM类似的无意义的预测结果.相比而言,基于广义正则化的OSELM-FGR具有持续的稳定性,得到的预测结果同R-OSELM一样稳定可靠.此外,OSELM-FGR和DFF-OSELM通过摒弃旧的失效样本的影响,能更准确地描述出动态时变系统的当前状态,从而取得较R-OSELM更小的预测RMSE,且大多数情况下OSELM-FGR的预测误差最小.

表1 5种算法对动态时变系统预测RMSE的平均值

隐层节点	在线算法	预测步长				
		200	500	1000	2000	3000
20	OSELM	2.98e-04	2.60e-04	2.94e-04	3.82e-04	5.19e-04
	R-OSELM	2.83e-04	2.82e-04	2.81e-04	3.92e-04	5.29e-04
	FR-OSELM	3.06e-04	2.81e-04	2.52e-04	2.40e-04	2.63e-04
	DFF-OSELM	2.85e-04	2.58e-04	2.60e-04	2.39e-04	2.62e-04
	OSELM-FGR	2.70e-04	2.73e-04	2.16e-04	2.67e-04	2.42e-04
50	OSELM	1.30e+18	2.32e+13	8.84e+19	1.49e+08	2.82e+19
	R-OSELM	7.51e-05	1.14e-04	1.84e-04	3.35e-04	4.95e-04
	FR-OSELM	6.12e-05	Inf	Inf	Inf	Inf
	DFF-OSELM	7.60e-05	9.25e-05	1.24e-04	1.63e-04	1.93e-04
	OSELM-FGR	9.73e-05	9.55e-05	1.03e-04	1.02e-04	1.07e-04
100	OSELM	4.77e-04	5.35e+18	6.48e+01	1.89e-01	2.48e+28
	R-OSELM	8.53e-05	1.36e-04	2.07e-04	3.56e-04	5.25e-04
	FR-OSELM	6.65e-05	Inf	Inf	Inf	Inf
	DFF-OSELM	8.30e-05	1.05e-04	1.45e-04	1.91e-04	2.32e-04
	OSELM-FGR	6.87e-05	6.67e-05	6.63e-05	6.42e-05	6.59e-05

为了更直观地观察比较这5种算法的预测精度和稳定性,本文给出了隐层节点个数为100时各种算法执行1000步在线预测的绝对误差(真实值与预测值之差),如图1所示.

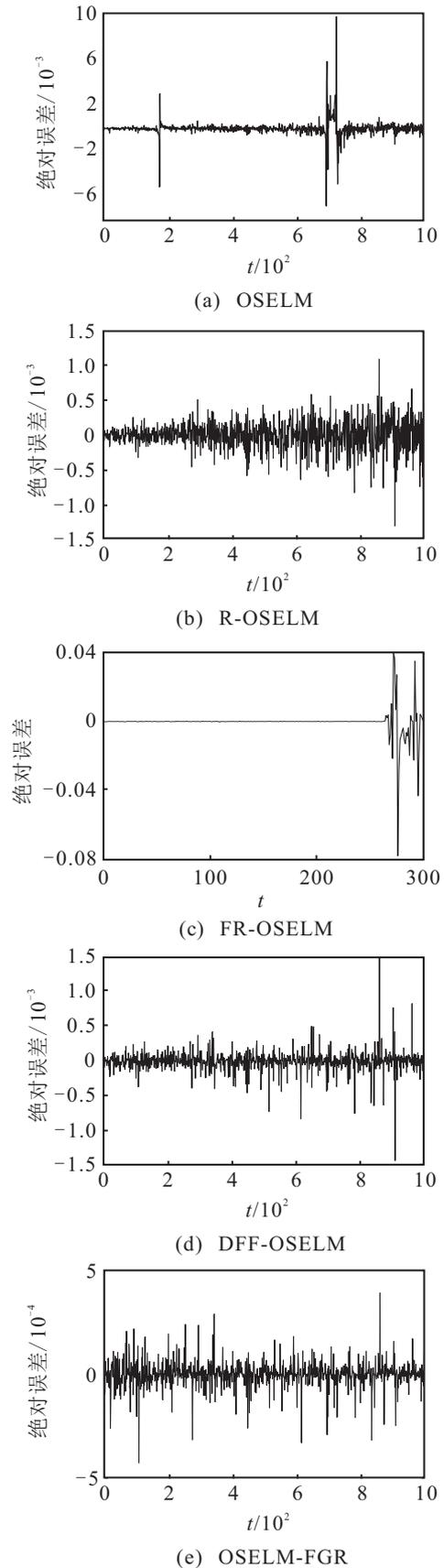


图1 绝对误差曲线

由于FR-OSELM在预测步长较大时得到的预测误差为无穷大,图1(c)仅给出前300步的部分结果.由图1可知,OSELM的预测误差存在较多较大的尖峰值,性能不够稳定;FR-OSELM仅短时有效,当预测步长超过270时,预测误差急剧增加,算法迅速失效;通过比较图1(b),图1(d)和图1(e)容易看出,OSELM-FGR取得了最优的预测性能,显示出了更好的动态跟踪能力.

表2给出了5种算法执行不同步长的在线预测的学习时间,单位为s.对于单个算法,隐层节点个数越多,预测步长越长,需要的学习时间越长.从这5种算法的纵向比较来看,OSELM-FGR的模型更新计算较其他算法稍复杂,其在线学习过程所花费的时间也稍长.

表2 5种算法对动态时变系统的学习时间

隐层节点	在线算法	预测步长				
		200	500	1000	2000	3000
20	OSELM	0.015 1	0.051 0	0.094 8	0.180 2	0.201 2
	R-OSELM	0.023 4	0.052 6	0.094 3	0.180 7	0.203 1
	FR-OSELM	0.020 8	0.050 5	0.100 5	0.190 1	0.243 8
	DFF-OSELM	0.025 5	0.065 1	0.114 6	0.223 4	0.267 7
	OSELM-FGR	0.029 2	0.066 7	0.128 1	0.258 3	0.317 2
50	OSELM	0.039 1	0.107 3	0.214 1	0.417 7	0.626 6
	R-OSELM	0.041 7	0.102 6	0.205 2	0.412 5	0.633 9
	FR-OSELM	0.044 8	0.115 6	0.231 8	0.471 9	0.718 2
	DFF-OSELM	0.041 7	0.110 4	0.222 9	0.448 4	0.690 1
	OSELM-FGR	0.077 1	0.203 6	0.418 2	0.805 7	1.252 6
100	OSELM	0.205 7	0.499 0	0.983 9	2.166 1	3.288 5
	R-OSELM	0.186 5	0.480 2	1.040 6	2.158 9	3.274 0
	FR-OSELM	0.217 2	0.550 5	1.096 9	2.345 3	3.618 2
	DFF-OSELM	0.124 5	0.315 1	0.617 7	1.238 0	1.841 5
	OSELM-FGR	0.423 4	1.025 5	2.125 5	4.322 9	6.469 8

3.2 变参数Logistic混沌时间序列

在工程实践中,系统参数可跟随时间不断变化,这类系统随时间延长展现出复杂丰富的动力学特性,且由于参数的变化,系统的动力学特性也在不断发生改变.考虑如下参数 μ 发生渐变的Logistic混沌系统:

$$x(t+1) = \mu(t)x(t)(1-x(t)),$$

$$\mu(t+1) = \mu(t) + 10^{-4}(1-0.2\sin(t)). \quad (34)$$

取初始值 $x(1) = 0.512, \mu(1) = 3.4, t = 1, 2, \dots$,生成5000个时间序列.为了消除瞬态的影响,去掉前1000个时间序列,并选择嵌入维度为4,时间延迟为1,对剩

下的4000个时间序列进行相空间重构以得到最终的实验数据集.

表3和表4分别给出了采用5种在线学习算法执行30次独立的在线预测实验所得到的预测RMSE的平均值和标准差.

表3 5种算法对Logistic混沌系统预测RMSE的平均值

隐层节点	在线算法	预测步长				
		200	500	1000	2000	3000
20	OSELM	1.05e-04	3.43e-03	2.23e-02	9.51e-02	1.42e-01
	R-OSELM	1.02e-05	1.39e-03	1.48e-03	2.47e-03	6.02e-03
	FR-OSELM	7.88e-06	Inf	Inf	Inf	Inf
	DFF-OSELM	1.12e-05	2.22e-03	2.02e-03	2.24e-03	4.81e-03
	OSELM-FGR	1.10e-05	1.00e-03	1.12e-03	1.39e-03	2.54e-03
50	OSELM	2.63e-06	5.63e-03	2.04e-02	5.48e-02	9.91e-02
	R-OSELM	5.90e-06	7.33e-04	6.38e-04	8.92e-04	1.29e-03
	FR-OSELM	4.31e-06	Inf	Inf	Inf	Inf
	DFF-OSELM	5.74e-06	1.21e-03	8.40e-04	7.04e-04	1.02e-03
	OSELM-FGR	6.04e-06	6.65e-04	5.76e-04	6.61e-04	9.41e-04
100	OSELM	4.38e-06	7.85e-03	2.13e-02	5.74e-02	1.08e-01
	R-OSELM	5.71e-06	5.09e-04	4.93e-04	5.63e-04	7.03e-04
	FR-OSELM	3.76e-06	Inf	Inf	Inf	Inf
	DFF-OSELM	5.18e-06	7.46e-04	5.27e-04	4.52e-04	6.18e-04
	OSELM-FGR	3.96e-06	4.81e-04	4.58e-04	5.17e-04	6.90e-04

表4 5种算法对Logistic混沌系统预测RMSE的标准差

隐层节点	在线算法	预测步长				
		200	500	1000	2000	3000
20	OSELM	3.29e-04	1.80e-03	9.68e-03	2.23e-01	2.00e-01
	R-OSELM	2.93e-06	2.94e-04	4.60e-04	6.21e-04	1.55e-03
	FR-OSELM	2.29e-06	Inf	Inf	Inf	Inf
	DFF-OSELM	3.28e-06	4.06e-04	4.67e-04	1.10e-03	2.67e-03
	OSELM-FGR	3.43e-06	2.17e-04	3.55e-04	3.32e-04	1.66e-03
50	OSELM	8.47e-08	1.80e-03	6.55e-03	2.18e-02	1.98e-02
	R-OSELM	1.43e-06	9.83e-05	7.87e-05	1.02e-04	2.34e-04
	FR-OSELM	6.78e-07	Inf	Inf	Inf	Inf
	DFF-OSELM	1.47e-06	2.32e-04	1.41e-04	9.21e-05	1.51e-04
	OSELM-FGR	1.74e-06	1.13e-04	6.92e-05	8.49e-05	1.59e-04
100	OSELM	4.20e-08	2.25e-03	5.27e-03	1.16e-02	1.55e-02
	R-OSELM	2.32e-06	6.76e-05	4.58e-05	4.65e-05	6.49e-05
	FR-OSELM	2.11e-06	Inf	Inf	Inf	Inf
	DFF-OSELM	1.87e-06	1.16e-04	5.64e-05	5.68e-05	7.44e-05
	OSELM-FGR	2.26e-06	3.16e-05	3.38e-05	4.04e-05	4.73e-05

表3的实验结果与表1类似, OSELM由于其自身的不稳定性, 得到的预测RMSE明显大于其他算法, 特别是当预测步长较长时差距较大; FR-OSELM由于只具有短时的正则化功能, 仅能在较短的预测步长内给出有意义的预测结果; 相比而言, R-OSELM, DFF-OSELM和OSELM-FGR得到的预测结果持续稳定有效, 且大多数情况下OSELM-FGR的预测误差最小. 由表4可进一步看出, 具有正则化功能的R-OSELM, DFF-OSELM和OSELM-FGR得到的预测RMSE的标准差相对其平均值较小, 即30次实验结果的波动较小, 可显示出良好的算法稳定性.

为验证正则化参数 δ 对R-OSELM, DFF-OSELM和OSELM-FGR算法性能的影响, 本文给出了典型正则化参数条件下3种算法执行3000步在线预测的预测RMSE折线图, 如图2所示.

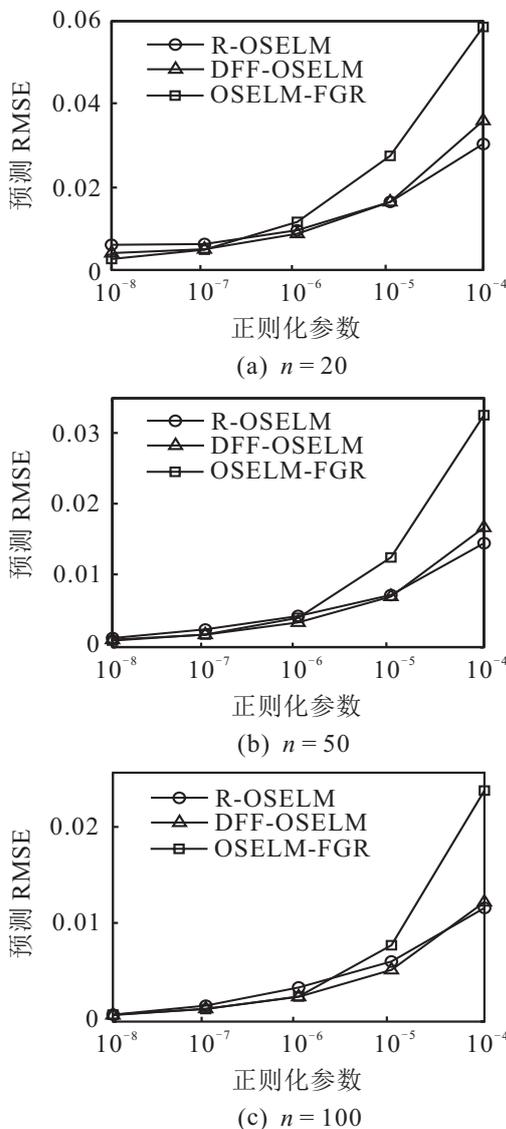


图2 正则化参数对算法性能的影响

图2中 n 为隐节点个数. 通过实验发现, 在3种不同隐层节点情形下, 3种算法均在相同的正则化参数

下($\delta = 10^{-8}$)取得最小的预测RMSE, 当 $\delta < 10^{-8}$ 时会起不到正则化的作用, 而得到无效的预测结果, 随着 δ 的进一步增加, 3种算法的预测误差也随之变大, 因此, 选择一个合适的正则化参数是很重要的. 在实际应用中, 正则化参数的选取通常可采用黄金分割法或L-曲线法^[10].

类似地, 本文分析和验证了遗忘因子 λ 对FR-OSELM, DFF-OSELM和OSELM-FGR算法性能的影响. 实验结果表明, 对于FR-OSELM, λ 越小, 算法失效越快, 反之亦然. 在DFF-OSELM中, λ 可在预测过程中保持动态更新, 这有助于提高算法的预测精度, 但 λ 的更新过程需要较多的辅助参数, 增加了算法应用的难度, 并且预设的辅助参数并不能保证该算法总能取得理想的预测效果. 如表3所示, 当预测步长较短时, DFF-OSELM的预测结果甚至不如R-OSELM. 相对而言, OSELM-FGR是一种更为简单实用的在线学习与预测方法, 其中 λ 的取值范围通常为 $[0.95 \ 1]$, 且对于变化较快的时变系统, λ 可取相对较小的值, 反之亦然. 图3给出了OSELM-FGR中 λ 与预测RMSE的对应关系, 在本实例中, λ 取0.98或0.99时预测效果较好.

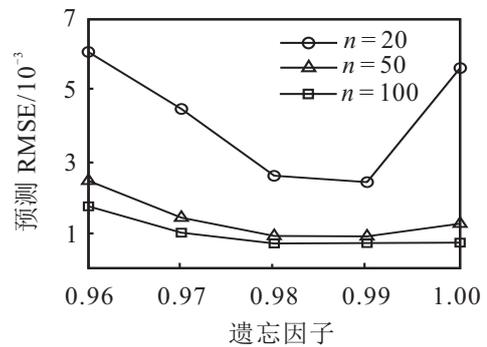


图3 遗忘因子对OSELM-FGR性能的影响

4 结 论

现实应用中存在大量非线性非平稳时间序列在线预测问题, 要求预测模型不仅具有持续稳定的在线增量学习能力, 而且能及时跟踪时序系统的动态变化. 为此, 本文提出一种具有广义正则化与遗忘机制的在线贯序超限学习机算法. 该算法采用递推模式计算输出权值, 使其具备了对新样本的在线学习能力以保持模型的动态更新; 此外, 该算法在通过遗忘旧的失效样本进一步增强模型动态跟踪能力的同时, 引入一种广义的 l_2 正则化技术, 使得算法具有持续的正则化功能, 从而保证了算法的持续稳定性, 并解决了FR-OSELM中因正则化功能被逐渐遗忘而导致的算法失效的问题. 两个典型非平稳时间序列在线预测实例表明, 与对比算法相比, 本文算法良好的动态跟

踪能力和持续的稳定性使其更适用于具有动态变化特性的非平稳时间序列在线建模与预测.

参考文献(References)

[1] Wang L, Zeng Y, Chen T. Back propagation neural network with adaptive differential evolution algorithm for time series forecasting[J]. Expert Systems with Applications, 2015, 42(2): 855-863.

[2] 龙文, 梁昔明, 龙祖强, 等. 基于混合进化算法的RBF神经网络时间序列预测[J]. 控制与决策, 2012, 27(8): 1265-1268.
(Long W, Liang X M, Long Z Q, et al. RBF neural network time series forecasting based on hybrid evolutionary algorithm[J]. Control and Decision, 2012, 27(8): 1265-1268.)

[3] Chen T T, Lee S J. A weighted LS-SVM based learning system for time series forecasting[J]. Information Sciences, 2015, 299: 99-116.

[4] Chen M Y, Chen B T. A hybrid fuzzy time series model based on granular computing for stock price forecasting[J]. Information Sciences, 2015, 294: 227-241.

[5] Pérez-Sánchez B, Fontenla-Romero O, Guijarro-Berdiñas B, et al. An online learning algorithm for adaptable topologies of neural networks[J]. Expert Systems with Applications, 2013, 40(18): 7294-7304.

[6] 义日贵. 基于KRLS的在线时间序列预测方法及其应用研究[D]. 哈尔滨: 哈尔滨工业大学电气工程及自动化学院, 2014.
(Yi R G. Research on online time series prediction and application based on KRLS[D]. Harbin: School of

Electrical Engineering and Automation, Harbin Institute of Technology, 2014.)

[7] Liang N Y, Huang G B, Saratchandran P, et al. A fast and accurate online sequential learning algorithm for feedforward networks[J]. IEEE Trans on Neural Networks, 2006, 17(6): 1411-1423.

[8] Wang X, Han M. Online sequential extreme learning machine with kernels for nonstationary time series prediction[J]. Neurocomputing, 2014, 145: 90-97.

[9] 张弦, 王宏力. 具有选择与遗忘机制的极端学习机在时间序列预测中的应用[J]. 物理学报, 2011, 60(8): 68-74.
(Zhang X, Wang H L. Selective forgetting extreme learning machine and its application to time series prediction [J]. Acta Physica Sinica, 2011, 60(8): 68-74.)

[10] Huynh H T, Won Y. Regularized online sequential learning algorithm for single-hidden layer feedforward neural networks[J]. Pattern Recognition Letters, 2011, 32(14): 1930-1935.

[11] Soares S G, Araújo R. An adaptive ensemble of on-line extreme learning machines with variable forgetting factor for dynamic system prediction[J]. Neurocomputing, 2016, 171: 693-707.

[12] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2006, 70(1): 489-501.

[13] Simon H. Adaptive filter theory[M]. Upper Saddle River: Prentice Hall, 2002: 442-444.

[14] Golub G H, Van Loan C F. Matrix computations[M]. Baltimore: JHU Press, 2013: 65.

(责任编辑: 滕 蓉)

下 期 要 目

复杂系统DEMATEL算法研究进展评述 孙永河, 等

有序势博弈及其在智能体无线网络中的应用 于永渊, 等

基于分解和差分进化的多目标粒子群优化算法 李 飞, 等

DoS攻击下电力网络控制系统脆弱性分析及防御 王轶楠, 等

动态矩阵控制中基于Kalman滤波的开环预测方法 谢亚军, 等

基于全概率更新的改进RANSAC算法 王 可, 等

多策略改进的多目标粒子群优化算法 杨景明, 等

一种彩色图像的量子描述方法及应用 李盼池, 等

基于数值符号和形态特征的时间序列相似性度量方法 李海林, 等

基于云相似度的语言偏好信息多属性大群体决策方法 徐选华, 等

带扰动补偿的无抖振离散重复控制器设计 周文委, 等