

基于动态集成算法的航空发动机气路参数预测

张一震¹, 钟诗胜^{1,2}, 付旭云², 林琳¹

(1. 哈尔滨工业大学 机电工程学院, 哈尔滨 150001;

2. 哈尔滨工业大学(威海) 船舶与海洋工程学院, 山东 威海 264209)

摘 要: 针对单一学习机对航空发动机气路参数预测困难的问题, 提出了基于动态加权核密度估计(DWKDE)组合方法的集成预测算法, 该组合方法选择测试样本的近邻样本, 通过评估学习机在近邻样本的局部性能动态确定各学习机的权值, 并基于该权值利用加权核密度估计实现数据序列的集成预测。该组合方法不易受离群值和样本不对称分布的影响, 将该组合方法用于 AdaBoost.RT 和 AdaBoost.R2 算法, 获得了改进后的集成学习算法。实验证明: 相比于神经网络和原始集成学习算法, 改进后的集成学习算法较好地提高了航空发动机气路参数序列的预测精度, 方均根误差(RMSE)指标至少可降低 27%。

关 键 词: 气路参数预测; 局部性能评估; 动态集成算法; 加权核密度估计; 时间序列

中图分类号: V267; TP206⁺.3

文献标志码: A

Aeroengine gas path parameter prediction based on dynamic ensemble algorithm

ZHANG Yizhen¹, ZHONG Shisheng^{1,2}, FU Xuyun², LIN Lin¹

(1. School of Mechatronics Engineering,

Harbin Institute of Technology, Harbin 150001, China;

2. School of Naval Architecture and Ocean Engineering,

Harbin Institute of Technology (Weihai), Weihai Shandong 264209, China)

Abstract: For the problem that single global modeling methods cannot get satisfactory aeroengine gas path parameter prediction results, ensemble algorithms with a combination method called dynamic weighted kernel density estimation (DWKDE) were proposed. Neighboring samples of the test samples were chosen. The weights of the base learners were dynamically calculated by evaluating the base learners' local performance in the neighboring samples. Based on the calculated weights, the integrated prediction of the time series was realized by the weighted kernel density estimation. The proposed combination method was insensitive to outliers and deviations from normality. Applying this combination method to AdaBoost.RT and AdaBoost.R2, experiments were conducted on gas path parameters of aeroengine, verifying that the proposed algorithms can achieve higher prediction accuracy than the single neural network and the traditional ensemble learning algorithms, for example, the

收稿日期: 2017-04-12

基金项目: 国家自然科学基金重点项目(U1533202); 山东省自主创新及成果转化专项(2014CGZH1101);

民航科技创新引导资金项目(MHRD20150104)

作者简介: 张一震(1993—), 男, 硕士生, 主要从事时间序列预测, 集成学习方面的研究。

通讯作者: 钟诗胜(1964—), 男, 教授、博士生导师, 博士, 主要从事先进制造, 航空发动机健康管理和维修决策技术研究。E-mail: zhongss@hit.edu.cn

引用格式: 张一震, 钟诗胜, 付旭云, 等. 基于动态集成算法的航空发动机气路参数预测[J]. 航空动力学报, 2018, 33(9): 2285-2295.
ZHANG Yizhen, ZHONG Shisheng, FU Xuyun, et al. Aeroengine gas path parameter prediction based on dynamic ensemble algorithm[J]. Journal of Aerospace Power, 2018, 33(9): 2285-2295.

root mean square error (RMSE) can be reduced by at least 27%.

Key words: gas path parameters prediction; local performance evaluation; dynamic ensemble algorithm; weighted kernel density estimation; time series

航空发动机性能参数预测是航空发动机实施视情维修的基础,现有航空发动机性能参数预测主要有支持向量机预测^[1]、离散过程神经网络预测^[2]、组合优化相关向量机(CORVM)的概率预测方法^[3]、融合预测^[4]、集成学习算法预测^[5-6]等。基于单一学习模型的气路参数预测,属于一种全局建模方法,这会导致模型复杂且易陷入局部最优。相比于单个学习机,将多个学习机集成起来的集成学习算法往往可获得更高的预测精度。许多学者从理论上探讨集成学习的有效性,如 VC (Vapnik-Chervonenkis) 维^[7]、Margin^[8]等。近年来,集成学习越来越多地应用在时间序列预测^[9-10]和航空发动机性能参数预测问题中^[5, 11-12]。相比于单个学习机,集成学习机往往可以获得更高的预测精度。

现有的集成学习算法所采用的组合方法大多属于平均、取中值或它们的加权形式。Kourentzes 等^[13]指出,平均及其加权形式的组合方法易受到离群值的影响,分布的不对称性对取平均和取中值及其加权形式都有影响,而核密度估计(kernel density estimation, KDE)组合方法对上述两种情况均不敏感,但是该文献^[13]未将基学习机的局部性能考虑进来。本文通过量化评估各学习机的局部性能,提出了动态加权核密度估计组合方法,采用该组合方法的集成学习算法称为动态加权核密度估计(dynamic weighted kernel density estimation, DWKDE)集成学习算法。将该组合方法应用于 AdaBoost. RT 和 AdaBoost. R2 算法分别得到 DWRT (DWKDE-based AdaBoost. RT) 算法和 DWR2 (DWKDE-based AdaBoost. R2) 算法,将该算法运用于 Mackey-Glass 时间序列以验证该算法的有效性,然后将之运用于航空发动机气路参数序列预测任务中,实验结果表明本文提出的 DWRT 和 DWR2 算法能有效提高集成学习算法的预测精度。

1 相空间重构

相空间重构理论(phase space reconstruction theory, PSRT)是时间序列分析的基础。在预测时间序列时需要对序列重构以发现序列内部规律。

预测模型的建立与使用都是在相空间内完成的。

对于时间序列,假设 z_{m+h} ($h \in \mathbf{Z}^+$) 可采用之前的 p ($p \in \mathbf{Z}^+$) 个历史数据 $z_{m-p+1}, \dots, z_{m-1}, z_m$ 预测,如式(1)。

$$z_{m+h} = F(z_{m-p+1}, \dots, z_{m-1}, z_m) \tag{1}$$

对时间序列进行预测也就等价于采用学习算法得到式(1)。当 $h=1$ 时,称为对时间序列进行一步预测,即短期预测。而长期预测($h>1$)可以看作是由多次一步预测迭代而成。本文采用一步预测。

选择合适的嵌入维数 p , 对时间序列 $\{z_i\}_{i=1}^N$ 进行相空间重构,生成用于一步预测的样本集(包括输入和输出),如式(2)和式(3)

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-p} \end{bmatrix} = \begin{bmatrix} z_1 & z_2 & \cdots & z_p \\ z_2 & z_3 & \cdots & z_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ z_{N-p} & z_{N-p+1} & \cdots & z_{N-1} \end{bmatrix} \tag{2}$$

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N-p} \end{bmatrix} = \begin{bmatrix} z_{p+1} \\ z_{p+2} \\ \vdots \\ z_N \end{bmatrix} \tag{3}$$

时间序列 $\{z_i\}_{i=1}^N$ 重构之后的序列为 $\{x_i, y_i\}_{i=1}^{N-p}$, 可将之用作学习机的训练集和测试集。训练集的输入、输出分别记为 $\mathbf{X}_{\text{train}}, \mathbf{Y}_{\text{train}}$, 测试集的输入输出分别记为 $\mathbf{X}_{\text{test}}, \mathbf{Y}_{\text{test}}$ 。

2 AdaBoost. RT 算法分析

AdaBoost. RT^[14]算法是一个将 AdaBoost. M1 应用于回归问题的算法。引入一个相对误差绝对值的阈值,将训练样本分为两类:预测正确的和预测错误的。实验证明,通常情况下 AdaBoost. RT 的预测效果要优于另外一种常用的 AdaBoost. R2^[14],尤其是在数据中含有离群样本时。

AdaBoost. RT 算法简述如下:

1) 输入

用于训练的集合为 $G = \{x_i, y_i\}_{i=1}^M, y \in R$;

选择基学习算法 $\{f_t\}_{t=1}^T$;

指定总迭代次数 T (也表示了所生成基学习机个数);

指定相对误差绝对值 (absolute relative error, ARE) 的阈值 ϕ 。根据 ϕ 将训练样本分为预测正确的和预测错误的样本。

2) 初始化

令初始迭代次数 $t=1$;

设有 M 个训练样本, 令训练样本权值分布

$$D_t(i)=\frac{1}{M}, i=1, \cdots, M;$$

令误差率 $\epsilon_t=0$ 。

3) 迭代过程

第 1 步 在带有权值 D_t 的训练样本集 G 上训练 f_t ;

第 2 步 记第 t 个学习机 f_t 对第 j 个样本 x_j 的预测结果为 $f_t(x_j)$, 而实际真实值为 y_j 。计算 f_t 的误差率 $\epsilon_t = \sum_j D_t(j)$, j 满足 $\left| \frac{f_t(x_j) - y_j}{y_j} \right| > \phi$;

第 3 步 设置 $\beta_t = \epsilon_t^n$, n 可以为 1, 2 或 3, 本文取 1;

第 4 步 更新样本权值 D_t :

$$D_{t+1}(j) = \frac{D_t(j)}{Z_t} \times \begin{cases} \beta_t & \left| \frac{f_t(x_j) - y_j}{y_j} \right| \leq \phi \\ 1 & \text{其他} \end{cases}$$

其中 Z_t 是标准化因子, 保证 $D_{t+1}(j)$ 是一个分布, 即 $\sum_j D_{t+1}(j) = 1$;

第 5 步 若 $i=T$, 则转 4); 否则, $i=i+1$, 转 3)。

4) 输出

对于新样本 x_{new} , 将每个基学习机 $\{f_t\}_{t=1}^T$ 的预测结果集成起来, 得到

$$f_{\text{fin}}(x_{\text{new}}) = \frac{\sum_t \log(1/\beta_t) f_t(x_{\text{new}})}{\sum_t \log(1/\beta_t)}$$

算法流程图如图 1。

第 1 步中, 要求基学习机能对具有特定分布的数据进行学习, 可以使用“重赋权法”(re-weighting); 对于不能处理带有权重的样本的基学习机, 应该采用“重采样法”(re-sampling), 即每次都要根据样本权值进行重采样, 对重采样得到的样本进行训练, 然后根据训练结果更新训练样本权值(第 4 步)^[15]。本文选择基学习机为神经网络, 神经网络本身不能处理带有权重的样本, 故选择“重采样法”。

样本权值更新步骤(第 4 步)中, 在第 t 步被

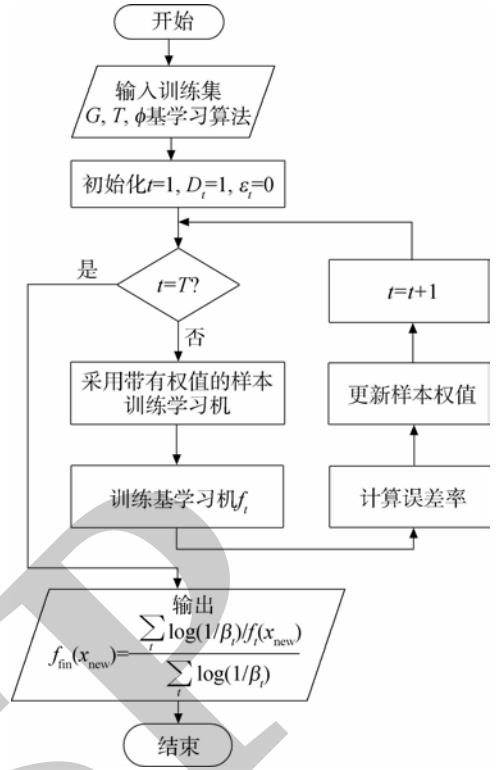


图 1 Adaboost. RT 算法流程图

Fig. 1 Flowchart of Adaboost. RT algorithm

基学习机 $f_t(x)$ 预测正确的样本的权值会变小, 而被预测错误的样本的权值将会相对变大。则在第 $t+1$ 步进行重采样时, 第 t 步预测正确的样本被重采样选中的概率变小, 预测错误的样本被重采样选中的概率变大。由此可见, 该算法将集中“精力”处理那些被预测错误的“难样本”。此处“难样本”指的是单个学习机预测效果较差的样本。

从上述的 AdaBoost. RT 算法中, 可以看出, 对任何一个测试样本 x_{new} 的预测结果 $f_t(x_{\text{new}})$ 来说, 第 t 个基学习机输出结果所占的权重始终为 $\frac{\log(1/\beta_t)}{\sum_t \log(1/\beta_t)}$, 它不随测试样本的变化而发生变化。这种在训练结束后, 基学习机的权值就确定的集成算法称为基于静态权值的集成学习算法。

算法每次迭代过程中, 根据当次迭代时的样本权值进行 m (令 $m=M$) 次有放回的抽样, 每次抽到的训练样本未包含全部 M 个样本^[5], 则每个基学习机的训练样本都是原始样本集的一个子集, 且训练各学习机所使用的子集各不相同。因此, 通过 AdaBoost. RT 算法得到的基学习机只是由部分训练样本训练得到的局域模型。训练样本的输入有多个维度, 可张成一个高维空间(设对于待预测样本 x_{new} , 在高维空间表现为一个点, 记

为 Q), 则各学习机的训练样本分布在不同空间, 因此各基学习机在不同局部空间的学习能力是不一致的^[16]。在不同局部空间内各学习机应该有不同的权值, 以保证在该局部空间内预测性能较好的基学习机的权值较大, 反之, 基学习机的权值较小。所以应将基学习机的局部性能因素考虑进来, 以更好地预测新样本。因此, 每个学习机的权值不再固定, 而是随预测样本所在局部空间而变化, 这种集成算法称为动态集成算法, 一般可以获得比静态集成算法更好的效果^[5]。

另外, AdaBoost, RT 的组合方法是一种加权平均的方式。平均及其加权形式的组合方法易受到离群值的影响, 分布的不对称性对取平均和取中值及其加权形式的组合方法都有影响, 而 KDE 组合方法对上述两种情况均不敏感。

本文将学习机的局部性能评估引入核密度估计, 得到加权核密度估计, 且其权值可根据不同测试样本而动态改变。下文介绍将加权核密度估计和基学习机的局部性能评估结合起来的动态加权核密度估计组合方法。

3 动态加权核密度估计集成学习算法

3.1 加权核密度估计

核密度估计是一种根据样本数据来估计密度函数表达式的方法。设 T 个基学习机 ($\{f_i\}_{i=1}^T$) 对待预测样本 x_{new} (点 Q) 的预测结果集合为 $\{s_i\}_{i=1}^T$, 则其核密度函数为 $p(s)$ 。则 $p(s)$ 的加权核密度估计的表示形式为式(4)。

$$\hat{p}(s) = \frac{1}{h} \cdot \sum_{i=1}^T w(s_i) \cdot K\left(\frac{s-s_i}{h}\right) \quad (4)$$

式中 T 为基学习机个数, 为了保证核密度估计效果, 一般要求 $T \geq 30$; $K(\cdot)$ 为核函数; h 为窗宽; $w(s_i)$ 为权值。

从定义中可以看出, 核函数是一种权函数, 在估计点 s 的核密度时, 该函数利用数据点 s_i 到 s 的距离 $(s-s_i)$ 和 s_i 对应的 $w(s_i)$ 来决定 s_i 的重要性大小。决定加权核密度估计表达式的因素有核函数、窗宽和 $w(s_i)$ 。

通常, 核函数的选择并不是核密度估计中最关键的因素, 而 $\hat{p}(s)$ 的光滑程度主要是由窗宽 h 决定的, 直接决定了估计效果^[17]。常用的一维核函数有高斯核函数、均匀核函数、指数核函数等。本文采用常用的高斯核函数, 其定义如式(5)。

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (5)$$

窗宽 h 的选择对估计的 $\hat{p}(s)$ 至关重要, 若 h 太大, 则 $\hat{p}(s)$ 会显得过于平滑, 忽略一些细节; 若 h 太小, 则 $\hat{p}(s)$ 曲线 (尤其是尾部) 会出现很大的波动。其主要选择方法有: 拇指法、最大似然估计法、最优理论窗宽和最小二乘交叉验证法等。用 Silverman^[18] 提出的拇指法计算出来的窗宽对于高斯核函数来说已经足够^[13]。窗宽计算公式如式(6)。

$$h = \left(\frac{4\hat{\sigma}^5}{3T}\right)^{\frac{1}{5}} \quad (6)$$

式中 $\hat{\sigma}$ 是 $\{s_i\}_{i=1}^T$ 的标准差。

使得核密度函数估计 $\hat{p}(s)$ 取得最大值的 s 记作 s_{mode} , 即各学习机集成之后的值, 其计算公式如式(7)。同时文献[13]指出, 核密度估计需要大量的预测值 (即大量的基学习机), 一般情况下在基学习机数目大于 30 的情况下能取得比较理想的结果。

$$s_{\text{mode}} = \arg \max_s \frac{1}{h} \cdot \sum_{i=1}^T w(s_i) \cdot K\left(\frac{s-s_i}{h}\right) \quad (7)$$

则 $\hat{p}(s)$ 的核心计算过程是确定 $w(s_i)$, 即需要评估基学习机的局部性能。

3.2 学习机局部性能评估

学习机对新样本 Q 的预测能力与学习机在其近邻样本上的学习能力有关^[7], 则学习机对其预测的好坏可以通过对其近邻样本预测的好坏来评估, 进而由在近邻样本上的预测效果确定基学习机对应的权值。

因此, 确定学习机对应的权值主要有以下 3 个步骤: ①根据一定的准则在训练集中寻找 x_{new} (Q) 的近邻样本; ②评估各学习机在其近邻样本上的预测性能; ③确定基学习机权值。

1) 在训练集中寻找 x_{new} 的近邻样本

在训练集中寻找新样本 x_{new} 的近邻样本的过程就是衡量不同序列之间相似性的过程。时间序列距离度量方法主要有闵可夫斯基 (Minkowski) 距离、最长公共子串 (longest common subsequence, LCS) 距离、余弦相似度、动态时间弯曲 (dynamic time warping, DTW) 距离^[19] 等。最常用的是闵可夫斯基距离, 假设两个样本的输入向量为 $\langle v_1, v_2, \dots, v_p \rangle$ 和 $\langle v'_1, v'_2, \dots, v'_p \rangle$, 则闵可夫斯基距离可表达为式(8)。

$$L = \left[\sum_{i=1}^p |v_i - v'_i|^n \right]^{\frac{1}{n}} \quad (8)$$

当 $n=1$, 式(8)成为曼哈顿距离; 当 $n=2$, 成

为欧几里德距离;当 $p=\infty$ 称为 Chebyshev 距离,即 $L=\max_{i=1,\dots,p}\{|v_i-v'_i|\}$ 。本文采用欧几里德距离来度量序列之间的距离。

计算所有训练样本与 Q 的距离,从小到大的顺序进行排列,按照 K 近邻法(K -nearest neighbor, k -NN)选择 K 个近邻样本。

2) 评估各学习机在其近邻样本上的性能

计算这 K 个近邻样本的到 Q 的欧几里得距离为 $d_i, i=1, \dots, K$ 。

归一化后的距离为式(9)。

$$d'_i = \frac{d_i}{\sum_{j=1}^K d_j} \quad i=1, \dots, K \quad (9)$$

设 i 个近邻样本的真实输出为 $y_i, i=1, \dots, K$, 设第 t 个基学习机对第 i 个近邻样本的预测值为 $f_{ti}, t=1, \dots, T$ 。计算第 t 个基学习机对样本 Q 的第 i 个近邻样本的预测误差 $e_{ti}=f_{ti}-y_i$ 。计算加权平均绝对值误差(weighted mean absolute error, WMAE), 在本文的定义为(10)。

$$e_{iQ}^{wmae} = \sum_{i=1}^K d'_i \cdot |e_{ti}| \quad (10)$$

可见,式(10)同时将学习机在 Q 的近邻样本上的绝对值误差和近邻样本到 Q 的距离考虑进来。距离 Q 越近的样本,则其对学习机性能评估影响越大。在 Q 的邻域内,预测性能好的基学习机对应的 e_{iQ}^{wmae} 较小;反之,较大。则可以根据 WMAE 确定基学习机的权值。

3) 确定权值

对于待预测样本 Q ,基于 WMAE 计算第 t 个基学习机对应的权值,如式(11)。

$$w_{tQ} = U \cdot \frac{1}{e_{iQ}^{wmae}} \quad (11)$$

式中 U 为标准化因子,使 $\sum_{t=1}^T w_{tQ} = 1$ 。

计算得到的 w_{tQ} 也即是文中第 3.1 节中的 $w(s_t)$ 。

3.3 动态加权核密度估计组合方法及集成学习算法

将核密度估计组合方法和学习机局部性能评估结合起来,得到动态加权核密度估计(DWKDE)组合方法。该组合方法可动态调整权值,在 Q 所在局部空间内预测精度高的基学习机的贡献权值较大;反之,贡献权值较小,而该权值正是加权核密度估计所需权值。可以总结出 DWKDE 组合方法的框架,如图 2。图 2 以三维数据(x_1 ,

x_2, x_3)为例,并取 5 个近邻样本,且与测试样本 Q 的欧式距离分别为 d_1, d_2, d_3, d_4, d_5 。

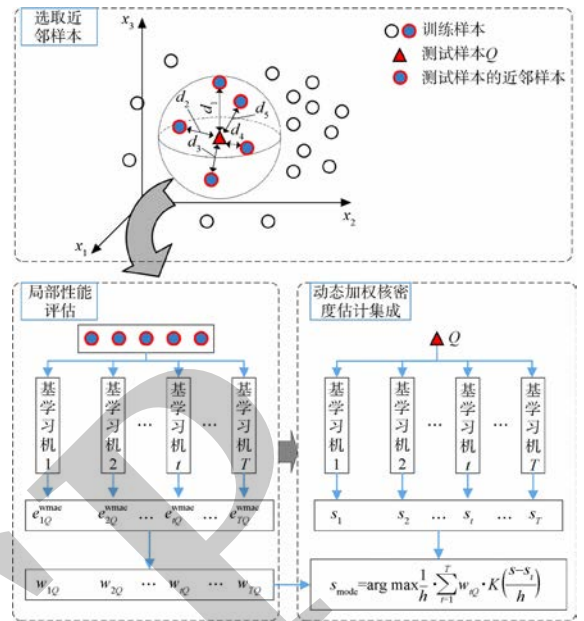


图 2 DWKDE 组合方法

Fig. 2 Combination method based on DWKDE

将 DWKDE 组合方法应用于 AdaBoost. RT 算法,得到 DWRT 算法。DWRT 算法具体步骤如下:

1) 输入

用于训练的集合为 $G=\{x_i, y_i\}_{i=1}^M, y \in R$;

选择基学习算法 $\{f_t\}_{t=1}^T$;

指定总迭代次数 T (也表示最终生成基学习机个数);

指定相对误差绝对值(absolute relative error, ARE)的阈值 ϕ 。在训练时,根据 ϕ 将训练样本分为预测正确的样本和预测错误的样本。

2) 初始化

令初始迭代次数 $t=1$;

设有 M 个训练样本,令第一次训练时,训练样本权值分布 $D_t(i)=\frac{1}{M}, i=1, \dots, M$;

令误差率 $\epsilon_t=0$;

令绝对值误差矩阵 E 为 $M \times T$ 的零矩阵,矩阵的每一列存储了每个学习机对所有训练样本预测的绝对值误差。

3) 迭代过程

第 1 步 在权值分布为 D_t 的训练样本集 G 上训练第 t 个基学习机;

第 2 步 计算 E 的第 t 列——第 t 个学习机

对训练样本预测的绝对值误差 $E(j,t) = |f_t(x_j) - y_j|, j=1, \cdots, M;$

第 3 步 记第 t 个学习机 f_t 对第 j 个样本 x_j 的预测结果为 $f_t(x_j)$, 而实际真实值为 y_j . 计算 f_t 的误差率

$$\epsilon_t = \sum_j D_t(j), j: \left| \frac{f_t(x_j) - y_j}{y_j} \right| > \phi$$

第 4 步 设置 $\beta_t = \epsilon_t^n, n$ 可以为 1、2 或 3, 本文取 1;

第 5 步 更新样本权值 D_t

$$D_{t+1}(j) = \frac{D_t(j)}{Z_t} \times \begin{cases} \beta_t & \text{若 } \left| \frac{f_t(x_j) - y_j}{y_j} \right| \leq \phi \\ 1 & \text{其他} \end{cases}$$

其中 Z_t 是标准化因子, 保证 $D_{t+1}(j)$ 是一个分布, 即 $\sum_j D_{t+1}(j) = 1;$

第 6 步 若 $i = T$ 则转 4); 否则, $i = i + 1$ 转 3)。

4) 选取近邻样本

对于待预测样本 x_{new} , 根据 K -NN 算法在训练集中寻找出 K 个近邻样本。

5) 基学习机局部性能评估及权值确定

从绝对值误差矩阵 E 中找到每个基学习机对 K 个近邻样本预测的绝对值误差, 并根据式 (9) 和式 (10) 计算加权平均绝对值误差 e_{iQ}^{wmac} ;

根据式 (11) 计算第 t 个基学习机对 Q 预测结果的权值 w_{tQ} 。

6) 输出

将每个基学习机 $\{f_t\}_{t=1}^T$ 的预测结果 $\{s_t\}_{t=1}^T$ 集成起来

$$s_{\text{mode}} = \arg \max_s \frac{1}{h} \cdot \sum_{t=1}^T w_{tQ} \cdot K\left(\frac{s - s_t}{h}\right)$$

将 DWKDE 组合方法运用于 Adaboost. R2 集成学习算法, 可得到 DWR2 算法。算法步骤与 DWRT 类似, 本文不再赘述。

4 应用实例

4.1 有效性验证

本文采用的算法有神经网络、原始 AdaBoost. RT 方法、采用 KDE 组合方法的 AdaBoost. RT 算法(式 (4) 里的每个基学习机预测结果对应的权值 w_{iQ} 均为 $1/T$, 没有局部性能评估步骤, 其余的计算步骤与本文 DWRT 算法一致)、采用 DWKDE 组合方法的 AdaBoost. RT 算法、原始 AdaBoost. R2 方法、采用 KDE 组合方法的 AdaBoost. R2 算法和采用 DWKDE 组合方法

的 AdaBoost. R2 算法, 为了表述方便, 这些算法分别记为 NN(Neural Network)、ORT、KDERT(kernel density estimation AdaBoost. RT)、DWRT、OR2、KDER2(kernel density estimation AdaBoost. R2) 和 DWR2。

本文采用常用的 Mackey-Glass 时间序列验证所提出算法的有效性, 该序列可由式 (12) 产生。采用以下 4 个指标来评估预测效果: 平均绝对误差(mean absolute error, MAE, 量符号记为 E_{ma})、方均根误差(root mean square error, RMSE, 量符号记为 E_{rms})、平均绝对百分比误差(mean absolute percentage error, MAPE, 量符号记为 E_{map}) 和平均绝对比例误差(mean absolute scaled error, MASE, 量符号记为 E_{mas})^[20]。

$$\frac{dz}{dt} = \frac{\alpha z(t - \tau)}{1 + z(t - \tau)^{10}} - \beta z(t) \tag{12}$$

按照文献[21]设置 $\alpha = 0.2, \beta = 0.1, \tau = 17,$ $z(0) = 1.2$ 且当 $t < 0$ 时 $z(t) = 0$ 。根据式 (12) 生成序列, 选取 $t = 118 - 1117$ 共 1000 个数据点, 记作 $\{z_i\}_{i=1}^{1000}$, 如图 3 所示。选用前面 500 个点来预测后面 500 个点。选定神经网络作为基学习机, 采用一步预测法预测。

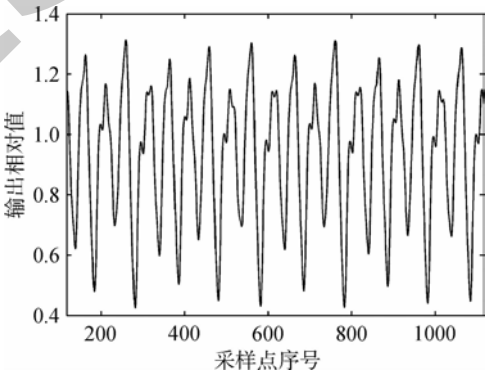


图 3 Mackey-Glass 时间序列
Fig. 3 Mackey-Glass time series

选择神经网络的结构为 8-30-1(表示第 1 层、第 2 层、第 3 层神经元数量分别为 8、30、1, 后文同), 则相空间重构时 $p = 8$ 。其中具体设置为: 训练最大迭代次数为 200, 训练精度为均方误差(mean squared error, MSE) 达到 0.000 01, 学习率为 0.1, 训练算法选择为 Levenberg-Marquardt (LM) 算法, 隐藏层激活函数为 Log-Sigmoid 型函数, 输出层激活函数为线性函数。实验设置总结如表 1。

所有实验均重复 10 次取平均值。对测试集

的预测结果如表 2。DWRT 相比于单个 NN、ORT 和 KDERT 预测精度提升比例如表 3 所示。

表 1 Mackey-Glass 时间序列实验参数设置
Table 1 Parameters setting of experiment for Mackey-Glass time series

模型	隐藏层神经 元个数	学习机 个数	阈值	近邻样 本个数
单个 NN	30	1		
ORT	30	40	0.0018	2
KDERT	30	40	0.0018	2
DWRT	30	40	0.0018	2

表 2 Mackey-Glass 时间序列预测效果(T=40)
Table 2 Prediction result of Mackey-Glass time series (T=40)

模型	平均绝对 误差	方均根 误差	平均绝对 百分比误差	平均绝对 比例误差
单个 NN	0.00179	0.00226	0.19354	0.06508
ORT	0.00152	0.00181	0.16112	0.05525
KDERT	0.00152	0.00186	0.15985	0.05507
DWRT	0.00062*	0.00091*	0.06410*	0.02239*

注:加“*”项表示各指标最优的数值。

表 3 DWRT 预测精度提升比例(T=40)
Table 3 Ratio improvement of prediction accuracy of DWRT (T=40)

模型	平均绝对 误差	方均根 误差	平均绝对 百分比 误差	平均绝对 比例 误差
单个 NN	0.6536	0.5967	0.6688	0.6560
ORT	0.5921	0.4972	0.6021	0.5948
KDERT	0.5921	0.5108	0.5990	0.5934

从表 2 和表 3 可知,相比于单个 NN,ORT 和 KDERT 均能有效提升预测精度,但是提升幅度不大,且 KDERT 和 ORT 的预测精度基本一致。而本文提出的 DWRT 能大幅提高预测精度。

在预测测试集的第 1 个数据点时,将 40 个基学习机的预测结果画成直方图,也将各集成学习机的计算结果表示出来,如图 4 所示。其中 DWRT 和 KDERT 核密度估计曲线的纵坐标为核密度,各基学习机预测结果的统计结果纵坐标为频数。从中可以看出采用 KDE 组合方法的 KDERT 算法和采用 DWKDE 组合方法的

DWRT 算法的核密度估计曲线的差异。DWRT 预测结果最接近真实值,KDERT 次之,ORT 的预测结果距离真实值最远。在 KDERT 算法估计核密度曲线时,式(4)里的权值 w_{iQ} 均为 $1/T$,而在 DWRT 算法估计核密度曲线时,在式(4)里的权值 w_{iQ} 的计算公式为式(11)。从图 4 中可以看出,虽然 KDERT 的核密度曲线比 DWRT 的核密度曲线更接近真实的预测结果的分布,但是由于每个学习机对该样本所在局部空间内的预测精度不一致性,可得到新的 w_{iQ} ,进而将 KDERT 的核密度曲线调整为 DWRT 的核密度曲线。可见虽然 DWRT 的核密度曲线不能较好地反映预测结果的分布,但是由于 DWRT 算法考虑了基学习机的局部性能差异性,赋予了每个基学习机预测结果不同的权值,使得在测试样本局部空间内预测效果好的学习机有更大的权值,由此得到的核密度估计曲线能有效反映计及学习机局部性能的预测结果分布,因此 DWRT 算法能获得较好的预测结果。

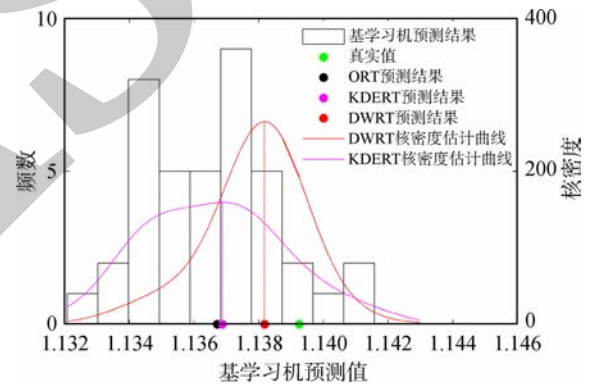


图 4 各算法对测试集的第 1 个数据点预测结果
Fig. 4 Prediction results of the first data point of test set by the algorithms

为研究基学习机数目 T 对于集成学习机的影响,初始设置 $T=5$,然后设置 T 从 10 增加至 150,每一次的增量为 10,其余实验设置和表 1 一致,每次实验均重复 10 次,将 10 个预测结果的均值作为最终预测结果。MAE、RMSE、MAPE 和 MASE 随基学习机数目 T 的变化趋势如图 5 所示。

从图 5 中可以看出,在每个评价指标上,每种集成学习机均能有效提高预测精度,表明集成学习机相对于单个学习机可有效提高预测精度,并且随着基学习机数目的增大,各集成算法均能有效提高预测精度。但是当基学习机数目过大时,预测精度提升幅度变小甚至会发生震荡。基学习

机数目 T 从 5 增加至 40 时,各集成学习机均能大幅提高预测精度,ORT、KDERT 和 DWRT 在 RMSE 指标上分别降低了 7.49%、13.97% 和 36.49%。基学习机个数 T 从 40 增加至 70 时,各集成学习机能小幅提高预测精度,ORT、KDERT 和 DWRT 在 RMSE 指标上分别降低了 4.94%、5.31% 和 10.08%。当 T 由 70 增至 150 时,ORT、KDERT 和 DWRT 在 RMSE 指标上仅分别降低了 2.09%、3.47% 和 7.28%,可见此时集成学习机的预测精度提升幅度很有限,过程中甚至会发生小幅震荡。

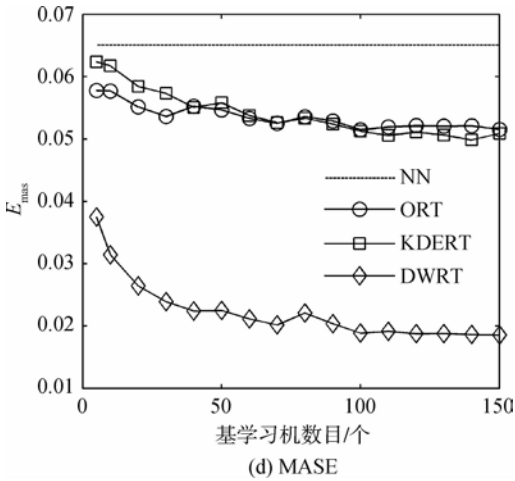
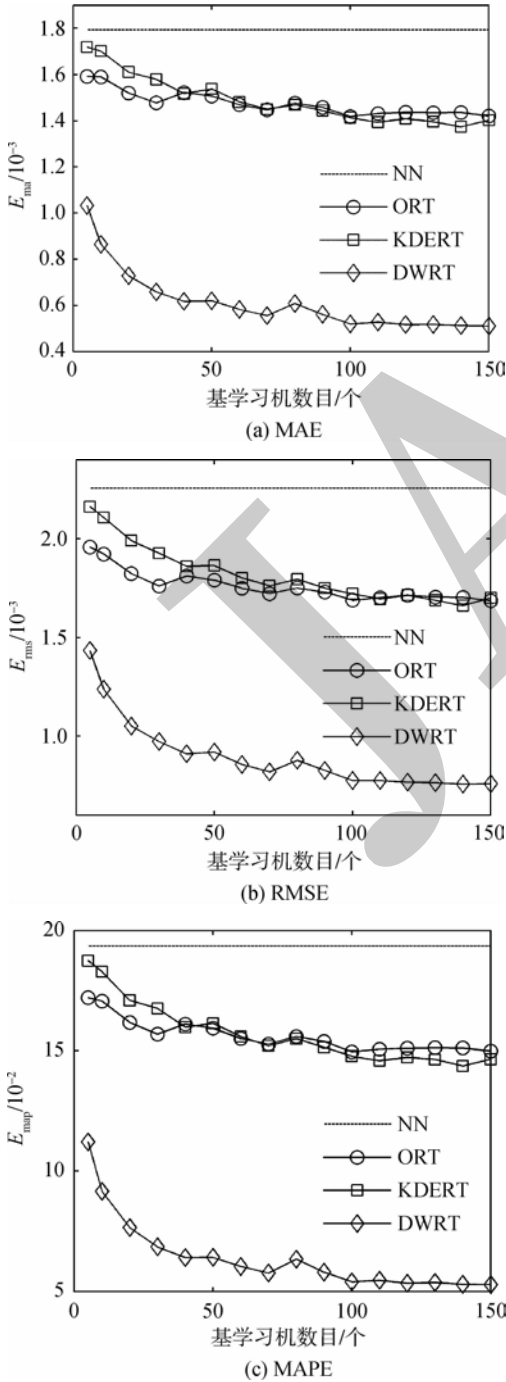


图 5 各学习机预测精度随基学习机数目变化趋势
Fig. 5 Prediction accuracy varying with the number of base learners obtained by the algorithms

在基学习机数量 $T < 40$ 时, KDERT 算法的预测精度要低于 ORT 算法。这主要是由于 KDERT 采用 KDE 组合方法,核密度估计在数据点较少时会得到不可靠的结果。当 $40 < T < 100$ 时, KDERT 的预测精度和 ORT 的预测精度基本一致,这是由于当数据点较多时,核密度估计会得到比较可靠的结果,这使得 KDE 组合方法可以成为一种有效的组合方法。当 $T > 100$ 时,根据该 4 种指标, KDERT 的预测精度要略高于 ORT,这是由于继续增多数据点可以进一步提高核密度估计的可靠性,可取得较好的预测效果。

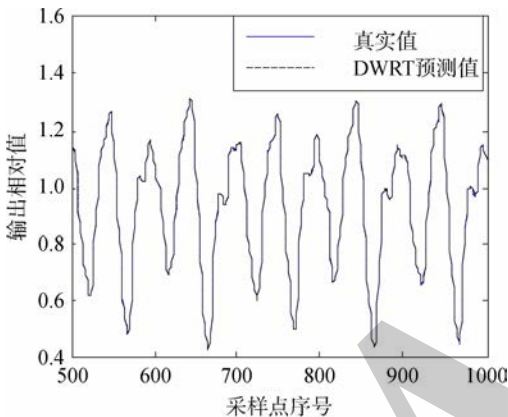
从图 5 中可以看出,总体上无论 T 取任何值,相比于 NN、ORT 和 KDERT,采用本文提出的 DWKDE 组合方法的 DWRT 算法均能取得最好的结果;相比于 ORT 和 KDERT,增大 T 能使得 DWRT 算法的预测精度得到更大幅度的提升。此外,从图 5(b)中可以看出,在 RMSE 指标上, ORT 在 $T = 150$ 时取得最小值为 0.001 687, KDERT 在 $T = 140$ 时取得最小值为 0.001 661, 而 DWRT 在 $T = 5$ 时对应的 RMSE 为 0.001 434。由此可知,相比于 $T = 150$ 时的 ORT 算法和 $T = 140$ 时的 KDERT 算法, DWRT 仅需 5 个基学习机就能获得更好的预测结果,极大地降低了算法的复杂度。

本文采用了与 LI 等^[21] 相同的时间序列,并有相同的训练集和测试集划分,基于模糊神经网络(neuro-fuzzy system, NFS)和自回归积分滑动平均模型(auto-regressive integrated moving average model, ARIMA)提出了 NFS-ARIMA(4, 0, 0)和 NFS-ARIMA(4, 1, 0)模型在测试集上的

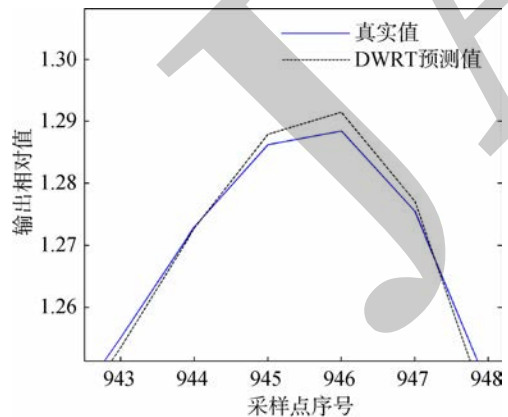
RMSE 分别为 0.001 3 和 0.000 86。本文提出的 DWRT 在 $T=40, 80, 100$ 时计算得到的 RMSE 分别为 0.000 91、0.000 88 和 0.000 78。在图 5 (b) 中, DWRT 所能取到最小的 RMSE 为 0.000 76, 此时 $T=140$ 。

综上, 本文提出的采用 DWKDE 组合方法的 DWRT 算法是一种较优的集成学习机。相比于 ORT 和 KDERT 算法, DWRT 算法能获得最好的预测效果。本文建议采用 DWRT 算法时, 学习机数目可以取在 40~70 之间, 此时在使用较少的 NN 的同时仍能获得较高的预测精度。

$T=100$ 时, DWRT 预测结果如图 6 所示。



(a) DWRT对Mackey-Glass测试集预测结果



(b) 绝对误差最大处放大图

图 6 DWRT 对 Mackey-Glass 测试集预测结果
($T=100$)

Fig. 6 DWRT prediction result of Mackey-Glass test set ($T=100$)

另外, 采用 OR2、KDER2 和 DWR2 做对比实验。本文设置 AdaBoost. R2 采用线性形式计算损失函数。令 $T=40$, 其余实验设置均和 AdaBoost. RT 的实验设置一致, 重复实验 10 次取均值。对测试集的预测结果如表 4。

表 4 Mackey-Glass 时间序列预测效果
($T=40$, 基于 AdaBoost. R2)

Table 4 Prediction result of Mackey-Glass time series ($T=40$, based on AdaBoost. R2)

模型	平均绝对 误差	方均根 误差	平均绝对 百分比 误差	平均绝对 比例 误差
单个 NN	0.001 79	0.002 26	0.193 54	0.065 08
OR2	0.006 07	0.007 19	0.730 74	0.220 20
KDER2	0.001 42	0.001 72	0.154 57	0.051 38
DWR2	0.000 59*	0.000 86*	0.063 75*	0.021 59*

注: 加“*”项表示各指标最优的数值。

从表 4 中可看出, 相对于 OR2, KDER2 算法的 RMSE 下降了 76.08%, 可见 KDE 组合方法的有效性。同时, 采用基于 DWKDE 组合方法的 DWR2 算法时, 其 RMSE 相对于 OR2 下降了 88.04%, 比 KDER2 下降了 50.00%, 比单个 NN 下降了 61.95%。可见, OR2 的效果较差, 甚至差于单个 NN 的效果, 这主要是因为 OR2 的样本权值更新策略会导致该方法对异常样本非常敏感, 异常样本得到足够的训练而正常样本却没有得到很好的训练。

对比表 2 和表 4 可知, 在该 4 个评价指标上, KDER2 预测精度要高于 KDERT, DWR2 预测精度要高于 DWRT。OR2 的组合方法属于加权中值的形式, 该组合方法易受样本不对称性分布的影响, 且 OR2 算法本身的权值更新方式使得其对异常样本敏感, 而 KDE 组合方法和 DWKDE 组合方法均能弥补 OR2 算法的此缺陷, 且 DWR2 可取得比 KDER2 更好的结果。这表明预测精度往往差于 AdaBoost. RT 的 AdaBoost. R2 借助于 KDE 和 DWKDE 组合方法可实现预测精度的“超越”。

由此可知, DWKDE 组合方法同样适用于 AdaBoost. R2, 可提高其预测精度。

4.2 航空发动机气路参数预测

航空发动机故障类型复杂, 超过 90% 的发动机故障是气路部件故障, 维修气路部件费用占发动机总维修费用的 60%^[22]。因此, 气路部件故障诊断在 EHM(engine health management)系统中占有重要的地位。发动机重要的气路参数为排气温度(EGT, 量符号记为 T_{eq})、低压转子转速(N_1 , 量符号记为 N_1)、核心机转速(N_2 , 量符号记为 N_2)和燃油流量(FF, 量符号记为 F_f), 其中 N_1 是发动机的

推力参数。本文中,航空发动机气路参数预测对象为排气温度裕度(EGTM,量符号记为 M_{egt})、核心机转速偏差值(ΔN_2)、燃油流量偏差值(ΔF_f)^[23]。

选取 1300 循环的 M_{egt} 序列、 ΔN_2 序列和 ΔF_f 序列进行实验,选定神经网络作为基学习机,采用一步预测法预测。对于该 3 种参数序列采用的神经网络的拓扑结构分别为 6-10-1、8-15-1、6-15-1。采用的实验的具体设置为:训练最大迭代次数为 200,训练精度为均方误差分别达到 0.000 1、0.000 01和 0.000 01,学习率为 0.05,训练算法选择为 Levenberg-Marquardt(LM)算法,设置输出层激活函数为线性函数,隐藏层激活函数为 Log-Sigmoid 型函数。实验的其他相关配置如表 5。

表 5 气路参数时间序列预测实验配置
Table 5 Experiment setting of gas path parameters time series prediction

气路参数	学习机个数	阈值	近邻样本个数
M_{egt}	40	0.000 05	7
ΔN_2	40	0.002 20	7
ΔF_f	40	0.000 46	7

采用单个神经网络、ORT、KDERT、DWRT、OR2、KDER2 和 DWR2 算法对该 3 种气路参数进行实验,每个实验均重复 10 次,取平均值作为最终的预测结果。在测试集上的实验结果如表 6~表 8 所示。

从表 6~表 8 的实验结果中可知,采用本文所提出的 DWKDE 组合方法的集成学习算法(DWRT 和 DWR2)均能取得最好结果,而采用 KDE 组合方法的集成学习算法的预测精度和原始集成学习算法一致或预测精度提升幅度不大。

表 6 M_{egt} 时间序列预测效果
Table 6 Prediction result of M_{egt} time series

模型	平均绝对误差	方均根误差	平均绝对百分比误差	平均绝对比例误差
单个 NN	0.009 46	0.011 87	0.020 09	0.061 19
ORT	0.006 14	0.007 74	0.013 00	0.039 74
KDERT	0.007 07	0.009 38	0.014 96	0.045 75
DWRT	0.004 04	0.005 60	0.008 58	0.026 13
OR2	0.023 39	0.027 30	0.049 73	0.151 26
KDER2	0.006 28	0.008 31	0.013 28	0.040 62
DWR2	0.003 83*	0.005 49*	0.008 14*	0.024 77*

注:加“*”项表示各指标最优的数值。

表 7 ΔN_2 时间序列预测效果
Table 7 Prediction result of ΔN_2 time series

模型	平均绝对误差	方均根误差	平均绝对百分比误差	平均绝对比例误差
单个 NN	0.004 75	0.007 29	1.942 26	0.726 06
ORT	0.003 32	0.005 44	1.378 38	0.508 18
KDERT	0.003 38	0.005 27	1.371 96	0.517 17
DWRT	0.002 17	0.003 76	0.903 58	0.331 33
OR2	0.014 98	0.019 92	5.748 40	2.291 46
KDER2	0.003 33	0.005 48	1.379 33	0.509 59
DWR2	0.002 00*	0.003 42*	0.835 93*	0.306 25*

注:加“*”项表示各指标最优的数值。

表 8 ΔF_f 序列预测效果
Table 8 Prediction result of ΔF_f time series

模型	平均绝对误差	方均根误差	平均绝对百分比误差	平均绝对比例误差
单个 NN	0.001 42	0.001 46	0.065 91	0.399 15
ORT	0.001 30	0.001 32	0.060 06	0.365 78
KDERT	0.001 70	0.001 78	0.078 76	0.477 78
DWRT	0.000 41*	0.000 49*	0.018 97*	0.115 11*
OR2	0.006 91	0.007 00	0.319 30	1.941 59
KDER2	0.002 17	0.002 25	0.100 47	0.610 60
DWR2	0.000 50	0.000 57	0.022 94	0.139 90

注:加“*”项表示各指标最优的数值。

由此可见,采用本文提出的 DWKDE 组合方法的 DWRT 和 DWR2 算法可适用于航空发动机气路参数预测任务中,并能有效提高预测精度——相比于其他算法,DWRT 和 DWR2 的方均根误差至少可降低 27%。

5 结 论

1) 针对基于平均、取中值的组合方法易受离群值和不对称性分布影响的问题,将基学习机局部性能评估和加权核密度估计结合,提出了一种组合方法——动态加权核密度估计(DWKDE)组合方法。基于该组合方法提出了 DWRT 和 DWR2 集成学习算法。

2) 确定了根据测试样本的近邻样本来计算基学习机局部性能的步骤,并与加权核密度估计结合以构建 DWKDE 组合方法。

3) 在 Mackey-Glass 序列上的实验结果表明 DWRT 和 DWR2 算法可有效提高预测精度,并在使用较少的基学习机的情况下仍能取得较高的预

测精度,减小了计算复杂度。借助于该组合方法,Adaboost.R2 可实现对 Adaboost.RT 的“超越”。

4) 航空发动机气路参数序列(M_{egt} 、 ΔN_2 和 ΔF_t)的预测实验结果表明 DWRT 和 DWR2 算法可适用于航空发动机气路参数预测任务,可有效提高预测精度。

参考文献:

[1] 吴瑞. 航空发动机状态预测与健康管理中的气路数据挖掘方法研究[D]. 四川 广汉:中国民用航空飞行学院,2015.
WU Rui. Research on gas path data mining method in aeroengine[D]. Guanghan Sichuan: Civil Aviation Flight University of China. (in Chinese)

[2] 钟诗胜,雷达,丁刚. 卷积和离散过程神经网络及其在航空发动机排气温度预测中的应用[J]. 航空学报,2012,33(3):438-445.
ZHONG Shisheng, LEI Da, DING Gang. Convolution sum discrete process neural network and its application in aeroengine exhausted gas temperature prediction[J]. Acta Aeronautica et Astronautica Sinica, 2012, 33(3): 438-445. (in Chinese)

[3] 范庚,马登武. 基于组合优化相关向量机的航空发动机性能参数概率预测方法[J]. 航空学报,2013,34(9):2110-2121.
FAN Geng, MA Dengwu. Probabilistic prediction method for aerogengine performance parameters based on combined optimum relevance vector machine[J]. Acta Aeronautica et Astronautica Sinica, 2013, 34(9): 2110-2121. (in Chinese)

[4] LIU K, GEBRAEEL N Z, SHI J. A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis[J]. Transactions on Automation Science and Engineering, 2013, 10(3): 652-664.

[5] 雷达. 基于智能学习模型的民航发动机健康状态预测研究[D]. 哈尔滨:哈尔滨工业大学,2013.
LEI Da. Civil aircraft engine health condition prediction based on intelligent learning models[D]. Harbin: Harbin Institute of Technology, 2013. (in Chinese)

[6] 张一震. 基于集成学习机的航空发动机气路参数预测方法研究[D]. 哈尔滨:哈尔滨工业大学,2017.
ZHANG Yizhen. Study on aeroengine gas parameters prediction based on ensemble learning[D]. Harbin: Harbin Institute of Technology, 2017. (in Chinese)

[7] 张春霞. 集成学习中有关算法的研究[D]. 西安:西安交通大学,2010.
ZHANG Chunxia. Research on some algorithms in ensemble learning[D]. Xi'an: Xi'an Jiaotong University, 2010. (in Chinese)

[8] SCHAPIRE R E, FREUND Y, BARTLETT P, et al. Boosting the margin: a new explanation for the effectiveness of voting methods[J]. The Annals of Statistics, 1998, 26(5): 1651-1686.

[9] WU X, KUMAR V, QUINLAN J R, et al. Top 10 algorithms in data mining[J]. Knowledge and Information Sys-

tems, 2008, 14(1): 1-37.

[10] LIM C P, GOH W Y. The application of an ensemble of boosted Elman networks to time series prediction: a benchmark study[J]. Computational Intelligence, 2005, 3(2): 119-126.

[11] SUN J, ZUO H, YANG H, et al. Study of ensemble learning-based fusion prognostics[R]. Macao: IEEE Prognostics and Health Management Conference, 2010.

[12] HU C, YOUN B D, WANG P, et al. Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life[J]. Reliability Engineering and System Safety, 2012, 103(3): 120-135.

[13] KOURENTZES N, BARROW D K, CRONE S F. Neural network ensemble operators for time series forecasting[J]. Expert Systems with Applications, 2014, 41(9): 4235-4244.

[14] SOLOMATINE D P, SHRESTHA D L. AdaBoost. RT: a boosting algorithm for regression problems[R]. Budapest: IEEE International Joint Conference on Neural Networks, 2004.

[15] 周志华. 机器学习[M]. 北京:清华大学出版社,2016.

[16] XUE F, SUBBU R, BONISSONE P. Locally weighted fusion of multiple predictive models[R]. Vancouver: IEEE International Joint Conference on Neural Networks, 2006.

[17] 王姚瑶. 基于非参数核密度估计的沪深 300 股指期货收益率研究[D]. 南京:南京大学,2015.
WANG Yaoyao. Research on yield rate of HS300 index futures based on non-parametric kernel density estimation [D]. Nanjing: Nanjing University, 2015. (in Chinese)

[18] SILVERMAN B W. Density estimation for statistics and data analysis [M]. London: Chapman and Hall/CRC (Chemical Rubber Company) Press, 1986.

[19] 孙建乐. 基于时间序列相似性的股价趋势预测研究[D]. 重庆:重庆交通大学,2014.
SUN Jianle. Stock price trend prediction research based on time series similarity[D]. Chongqing: Chongqing Jiaotong University, 2014. (in Chinese)

[20] HYNDMAN R J, KOEHLER A B. Another look at measures of forecast accuracy[J]. International Journal of Forecasting, 2006, 22(4): 679-688.

[21] LI C, HU J W. A new ARIMA-based neuro-fuzzy approach and swarm intelligence for time series forecasting [J]. Engineering Applications of Artificial Intelligence, 2012, 25(2): 295-308.

[22] 李业波,李秋红,黄向华,等. 航空发动机气路部件故障融合诊断方法研究[J]. 航空学报,2014,35(6):1612-1622.
LI Yebo, LI Qiu hong, HUANG Xianghua, et al. Research on gas fault fusion diagnosis of aero-engine component[J]. Acta Aeronautica et Astronautica Sinica, 2014, 35(6): 1612-1622. (in Chinese)

[23] 徐亮. 航空发动机气路性能参数趋势预测研究[D]. 四川 广汉:中国民用航空飞行学院,2016.
XU Liang. Research ontrend prediction of gas path parameters of aeroengine[D]. Guanghan Sichuan: Civil Aviation Flight University of China, 2016. (in Chinese)