



基于宽度混合森林回归的城市固废焚烧过程二英排放软测量

夏恒 汤健 崔璨麟 乔俊飞

Soft Sensing Method of Dioxin Emission in Municipal Solid Waste Incineration Process Based on Broad Hybrid Forest Regression

XIA Heng, TANG Jian, CUI Can-Lin, QIAO Jun-Fei

在线阅读 View online: <https://doi.org/10.16383/j.aas.c220012>

您可能感兴趣的其他文章

面向城市固废焚烧过程的二英排放浓度检测方法综述

Dioxin Emission Concentration Measurement Approaches for Municipal Solid Wastes Incineration Process: A Survey

自动化学报. 2020, 46(6): 1063–1089 <https://doi.org/10.16383/j.aas.c190005>

面向边缘计算应用的宽度孪生网络

Broad Siamese Network for Edge Computing Applications

自动化学报. 2020, 46(10): 2060–2071 <https://doi.org/10.16383/j.aas.c200555>

智能船舶综合能源系统及其分布式优化调度方法

Intelligent Ship Integrated Energy System and Its Distributed Optimal Scheduling Algorithm

自动化学报. 2020, 46(9): 1809–1817 <https://doi.org/10.16383/j.aas.c200176>

基于池的无监督线性回归主动学习

Unsupervised Pool-Based Active Learning for Linear Regression

自动化学报. 2021, 47(12): 2771–2783 <https://doi.org/10.16383/j.aas.c200071>

大数据智能决策

Big Data for Intelligent Decision Making

自动化学报. 2020, 46(5): 878–896 <https://doi.org/10.16383/j.aas.c180861>

基于改进随机森林算法的工业过程运行状态评价

Industrial Operation Performance Evaluation of Industrial Processes Based on Modified Random Forest

自动化学报. 2021, 47(9): 2214–2225 <https://doi.org/10.16383/j.aas.c190066>

基于宽度混合森林回归的城市固废焚烧过程二噁英排放软测量

夏恒^{1,2,3} 汤健^{1,2,3} 崔璨麟^{1,2,3} 乔俊飞^{1,2,3}

摘要 二噁英是城市固废焚烧过程排放的痕量有机污染物. 受限于相关技术的复杂度和高成本, 二噁英排放浓度检测的大时滞已成为制约城市固废焚烧过程优化控制的关键因素之一. 虽然具有低成本、快响应、高精度等特点的数据驱动软测量模型能够有效解决上述问题, 但二噁英建模方法必须要契合数据的小样本、高维度特性. 对此, 提出了由特征映射层、潜在特征提取层、特征增强层和增量学习层组成的宽度混合森林回归软测量方法. 首先, 构建由随机森林和完全随机森林构成的混合森林组进行高维特征映射; 其次, 依据贡献率对全联接混合矩阵进行潜在特征提取, 采用信息度量准则保证潜在有价值信息的最大化传递和最小化冗余, 降低模型的复杂度和计算消耗; 然后, 基于所提取潜在信息训练特征增强层以增强特征表征能力; 最后, 通过增量式学习策略构建增量学习层后采用 Moore-Penrose 伪逆获得权重矩阵. 在基准数据集和城市固废焚烧过程二噁英数据集上的实验结果表明了方法的有效性和优越性.

关键词 城市固废焚烧, 二噁英排放建模, 宽度学习, 宽度混合森林回归, 潜在特征, 增量学习

引用格式 夏恒, 汤健, 崔璨麟, 乔俊飞. 基于宽度混合森林回归的城市固废焚烧过程二噁英排放软测量. 自动化学报, 2023, 49(2): 343-365

DOI 10.16383/j.aas.c220012

Soft Sensing Method of Dioxin Emission in Municipal Solid Waste Incineration Process Based on Broad Hybrid Forest Regression

XIA Heng^{1,2,3} TANG Jian^{1,2,3} CUI Can-Lin^{1,2,3} QIAO Jun-Fei^{1,2,3}

Abstract Dioxin is a trace organic pollutant emitted from municipal solid waste incineration process. Limited by the complexity and high cost of relative technology, the big time delay of dioxin emission concentration detection has become one of the key factors restricting the optimize control of municipal solid waste incineration process. Although the data-driven soft sensing model with the characteristics of low cost, fast response and high precision can effectively solve the above problems, the dioxin modeling method must fit the small sample and high-dimensional characteristics of the modeling data. In this paper, a broad hybrid forest regression soft sensing method is proposed, which consists of feature mapping layer, latent feature extraction layer, feature enhancement layer and incremental learning layer. Firstly, a hybrid forest group composed of random forest and completely random forest is constructed for high-dimensional feature mapping. Secondly, the latent features extraction of the fully connected mixed matrix is carried out according to the contribution rate, and the information measurement criterion is used to ensure the maximum transmission and minimize redundancy of potential valuable information. Thus, the model complexity and computational consumption are reduced. Then, the feature enhancement layer is trained based on the extracted potential information to enhance the feature representation ability. Finally, the incremental learning layer is constructed by using incremental learning strategy, and the weight matrix is obtained by using the Moore-Penrose pseudo inverse. The experimental results on high-dimensional benchmark and dioxin datasets of municipal solid waste incineration process show the effectiveness and superiority of the proposed method.

Key words Municipal solid waste incineration, dioxin emission modeling, broad learning, broad hybrid forest regression, latent feature, incremental learning

Citation Xia Heng, Tang Jian, Cui Can-Lin, Qiao Jun-Fei. Soft sensing method of dioxin emission in municipal solid waste incineration process based on broad hybrid forest regression. *Acta Automatica Sinica*, 2023, 49(2): 343-365

收稿日期 2022-01-04 录用日期 2022-04-07

Manuscript received January 4, 2022; accepted April 7, 2022

国家自然科学基金 (62073006, 62173120, 62021003), 北京市自然科学基金资助项目 (4212032, 4192009), 科技创新 2030 —— “新一代人工智能”重大项目 (2021ZD0112301, 2021ZD0112302) 资助

Supported by National Natural Science Foundation of China (62073006, 62173120, 62021003), Beijing Natural Science Foundation (4212032, 4192009), National Key Research and Development Program of China (2021ZD0112301, 2021ZD0112302)

本文责任编辑 董峰

Recommended by Associate Editor Dong Feng

1. 北京工业大学信息学部 北京 100124 2. 北京工业大学智慧环保北京实验室 北京 100124 3. 北京工业大学智能感知与自主控制教育部工程研究中心 北京 100124

1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124 2. Beijing Laboratory of Smart Environmental Protection, Beijing University of Technology, Beijing 100124 3. Engineering Research Center of Intelligent Perception and Autonomous Control, Ministry of Education, Beijing University of Technology, Beijing 100124

本文缩写词的详细解释见表 1.

城市固废焚烧是目前世界范围内解决“垃圾围城”困境的主要方式之一,具有无害化、减量化和资源化等显著优势^[1-2].目前中国 MSWI 的处理能力占比已超过 50%, 污染排放监管力度也逐渐加强^[3].二噁英作为 MSWI 过程排放的有组织废气中具有持久性和剧毒性的有机污染物^[4],是造成焚烧建厂存在“邻避现象”的主要原因^[5],也是 MSWI 过程必须最小化控制的重要环保指标之一^[6-7].基于高分辨

气相色谱/高分辨质谱的离线化验分析方法是目前 DXN 排放浓度检测的主要手段^[8],存在技术难度大、时间滞后性大、人力与经济成本高等缺点,已成为阻碍 MSWI 过程实现实时优化控制的关键因素之一.因此,DXN 排放浓度的在线检测是 MSWI 过程的首要挑战问题^[9].

针对上述问题,利用可在线检测的 DXN 关联物构建映射模型进而获得 DXN 浓度的在线间接检测方法成为热点^[9-10];然而,其存在设备复杂、成本

表 1 符号说明

Table 1 Symbol description

缩写词	中文全称	英文全称
MSWI	城市固废焚烧	Municipal solid waste incineration
DXN	二噁英	Dioxin
HRGC/HRMS	高分辨色谱质谱联用	High-resolution chromatography combined with high-resolution mass spectrometry
NN	神经网络	Neural network
PCDDs/PCDFs	多氯二苯并二噁英/多氯二苯并呋喃	Polychlorinated dibenzo-p-dioxins/Polychlorinated dibenzofurans
BPNN	反向传播神经网络	Back-propagation neural network
SVR	支持向量回归	Support vector regression
SEN	选择性集成	Selective ensemble
RF	随机森林	Random forest
DFR	深度森林回归	Deep forest regression
BLS	宽度学习系统	Broad learning system
BLS-NN	神经网络宽度学习系统	Broad learning system neural network
BHFR	宽度混合森林回归	Broad hybrid forest regression
CRF	完全随机森林	Completely random forest
MSW	城市固废	Municipal solid waste
TEQ	毒性当量	Toxic equivalent quantity
G 1	烟气 1	Gas 1
NO _x	氮氧化物	Nitrogen oxides
HCL	氯化氢	Hydrogen chloride
HF	氟化氢	Hydrogen fluoride
SO ₂	二氧化硫	Sulfur dioxide
Pb	铅	Plumbum
Hg	汞	Mercury
Cd	镉	Cadmium
G 2	烟气 2	Gas 2
G 3	烟气 3	Gas 3
RSM	随机子空间法	Random subspace method
PCA	主成分分析	Principal components analysis
MI	互信息	Mutual information
CT	CT 切片在轴轴数据上的相对位置	The relative location of CT slices on the axial axis data
RB	住宅建筑数据	Residential building data
NIR	橙汁近红外光谱数据	The orange juice near infrared spectra data
RMSE	均方根误差	Root mean square error
MAE	平均绝对误差	Mean absolute error
R ²	决定系数	Coefficient of determination
DFR-clfc	基于跨层全连接的深度森林回归	Deep forest regression based on cross-layer full connection

高、干扰因素多、预测精度无法保证等问题^[11], 同时其在本质上也是一种结合数据建模的检测手段. 相较于离线分析和在线间接检测方法而言, 基于工业集散控制系统采集的易检测过程数据驱动的软测量技术是解决 DXN 无法在线检测问题的有效途径, 具有稳定、精准和快速响应等特点^[3]. 软测量技术已在石油、化工和炼钢等复杂工业过程的难测参数检测中广泛应用^[12-15].

目前, 面向 DXN 排放浓度的软测量研究可分为基于单学习器和基于集成多学习器两个方向. 针对前者: Chang 等^[16] 采用遗传规划结合神经网络对欧美等多个国家焚烧厂的 Polychlorinated dibenzo-p-dioxins/Polychlorinated dibenzofurans 排放进行建模, 其数据涉及多种不同类型的焚烧炉, 导致模型不具有好的应用性; Bunsan 等^[17] 通过多次重复实验确定反向传播神经网络结构后构建软测量模型, 但其不具有好的移植性且 BPNN 面对小样本时存在过拟合、稳定性差等问题; 针对上述基于 NN 的软测量方法存的问题, 肖晓东等^[18] 利用支持向量回归^[19] 构建基于国内 MSWI 过程实际数据的软测量模型, 但存在核函数、惩罚系数等超参数难以确定的问题; 进一步, 乔俊飞等^[20] 针对北京某 MSWI 电厂的高维过程数据, 设计了基于多层特征选择策略的软测量模型, 但约简特征模型的泛化性能有待提高.

从机理视角, DXN 排放浓度与 MSWI 过程的多个工艺阶段的众多过程变量均具有相关性, 并且在不同工况下也存在差异性; 此外, 获取 DXN 浓度检测真值存在难度大、成本高的缺点, 使得建模数据的小样本、高维度特性成为 DXN 软测量面临的主要问题. 上述因素导致基于单学习器的软测量模型难以获得较佳检测精度. 因此, 基于集成多学习器的软测量模型成为当前的研究热点, 其包括: 汤健等^[21] 基于选择性集成思想设计了一种自适应确定 SVR 超参数的软测量方法, 采用文献^[22] 的数据进行验证, 但测量精度有待提高; 在此基础上, 汤健等^[23] 采用变量投影重要性评价和设定特征约简比率的策略对北京某 MSWI 过程的 DXN 建模数据进行维数约简后, 构建能够自适应确定核参数的 SEN 软测量模型, 进一步提高了检测精度; 针对上述方法存在的放弃部分特征可能导致的信息丢失问题, Xia 等^[24] 通过随机森林和梯度提升树的混合集成策略进行 DXN 软测量模型构建, 但模型结构过于复杂且运行时间较长, 不适用于实际应用; 进一步, 借鉴 NN 模式和非 NN 模式深度学习在提取深层次表征特征方面的突出性能^[25-26], Tang 等^[27-28] 提出了面向小样本高维数据的深度森林回归算法并用于构建 DXN 软测量模型, 但检测精度仍有待进一步提高; 在此基础上, Xu 等^[29] 采用主成分分析对高

维过程数据进行特征提取后再基于 DFR 构建软测量模型, 虽然提升了模型的泛化性能但约简后的潜在特征已不具备物理意义. 显然, 上述集成学习策略均存在训练难度大、模型复杂度高以及收敛速度慢等问题.

近年来, 宽度学习通过先扩展模型网络“宽度”再采用 Moore-Penrose 逆矩阵^[30] 获取权重的方式构建模型, 具有收敛速度快、超参数少和精度高等优势^[31-32]. 现有的 BLS 研究主要基于 NN 模式, 已在图像识别^[33-35]、故障诊断^[36-38] 和工业过程控制^[39-41] 等多个领域广泛应用. 但是, 上述研究均是在低特征维数、大样本数据集的研究背景下进行的应用与探索^[39, 42]. 面向高特征维数、小样本数据集^[43] 的 BLS 研究还未见报道.

针对上述问题, 以 MSWI 过程 DXN 排放浓度检测为目标, 提出了基于宽度混合森林回归的建模算法. 主要创新工作为:

- 1) 基于 BLS 框架的优势, 提出了具有特征映射层、潜在特征提取层、特征增强层和增量学习层结构的 BHFR 建模算法.
- 2) 利用 RF 和完全随机森林组成的混合森林组替代 BLS-NN 神经元组, 实现对高维特征向量的映射.
- 3) 基于 PCA 的潜在特征提取和互信息度量准则, 以保证全联接混合矩阵中潜在有价值信息的最大化传递和最小化冗余.
- 4) 通过增量式学习策略构建以混合森林组作为最小单位的增量学习层, 采用 Moore-Penrose 伪逆获得权重矩阵.

在高维基准数据集和 MSWI 过程 DXN 数据集上的实验结果表明了本文方法的有效性和优越性.

1 面向 DXN 排放的 MSWI 过程描述

MSWI 过程包含固废储运、固废燃烧、余热交换、蒸汽发电、烟气净化和烟气排放等工艺阶段, 以日处理量 800 吨的北京某炉排式 MSWI 过程为例, 其工艺如图 1 所示.

结合 DXN 分解、生成、吸附和排放的全流程对各阶段的主要功能描述如下:

- 1) 固废储运阶段. 环卫车辆从城市各收集站点将城市固废运输至 MSWI 电厂, 经称重记录后从卸料平台倾倒至固废储存池中未发酵区, 然后由固废抓斗对其进行混合搅拌, 再抓取至发酵区, 经 3 ~ 7 天发酵和脱水以保证 MSW 焚烧的低位热值^[44]. 研究表明, 原生 MSW 中存在微量 DXN (约 0.8 ng TEQ/kg), 并含有 DXN 生成反应所需的多种含氯化合物^[5].

- 2) 固废燃烧阶段. 固废抓斗将发酵后的 MSW

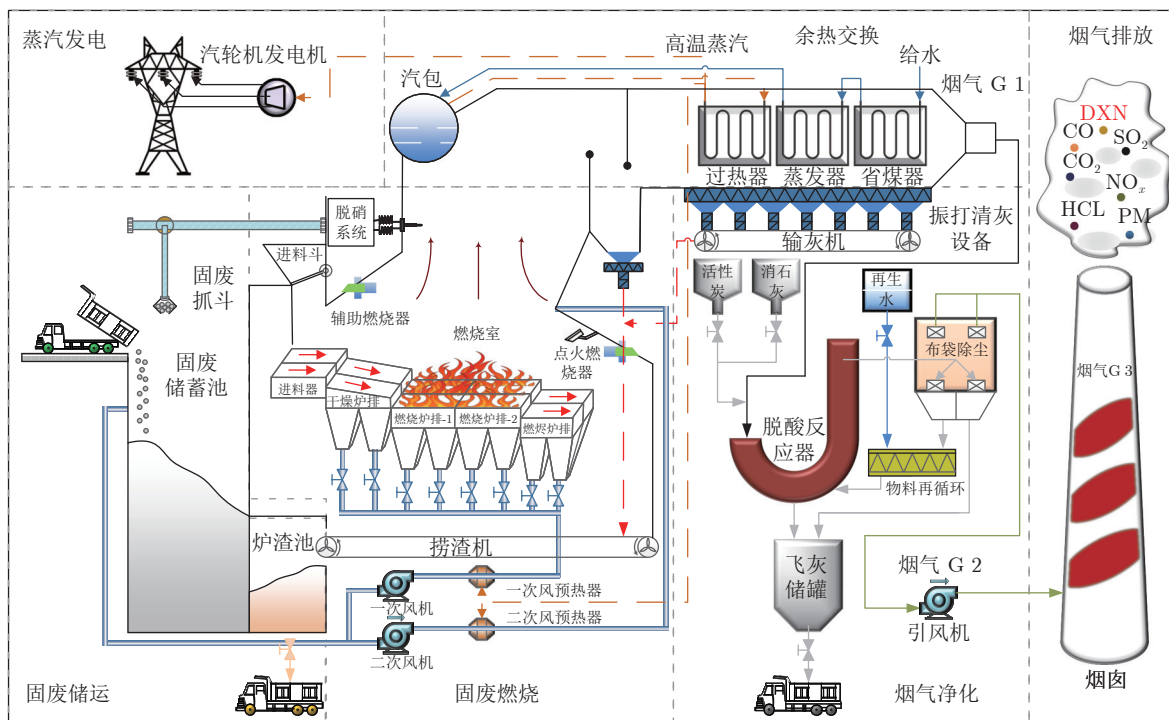


图1 城市固废焚烧工艺流程图

Fig.1 Process flow chart of municipal solid waste incineration process

投放至进料斗,经进料器将MSW推送到焚烧炉内,依次经过干燥、燃烧1、燃烧2和燃烬炉排后,MSW中的可燃成分随之完全燃烧;所需助燃空气由一次风机和二次风机从炉排下方和炉膛中部注入,最终燃烧产生的灰渣从燃烬炉排末端落至捞渣机,经水冷后送入炉渣池.研究表明,启炉和停炉期间因工况波动频繁导致所产生的DXN量级远远高于正常运行期间^[45].为保证原生MSW中含有的以及焚烧时产生的DXN在炉内能够被完全分解,炉膛燃烧过程需满足:严格控制烟气温度在850℃以上、高温烟气在炉内停留时间超过2秒、确保足够的烟气湍流度等工艺要求^[46].

3) 余热交换阶段.炉膛产生的高温烟气(高于850℃)经引风机抽吸进入余热锅炉系统,先后经过过热器、蒸发器和省煤器设备,与锅炉汽包液态水进行热交换后产生高温蒸汽,进而实现降温处理,使余热锅炉出口的烟气温度低于200℃(即烟气G1).从DXN生成机理的角度,高温烟气经余热锅炉降温时,导致DXN生成的化学过程包括高温气相合成(800℃~500℃)^[47]、前驱物合成(450℃~200℃)^[48]和从头合成(350℃~250℃)^[49]等反应,但目前还暂无统一的定论^[50].

4) 蒸汽发电阶段.利用余热锅炉产生的高温蒸汽推动汽轮发电机,机械能转变为电能,实现厂级用电的自给自足和剩余电量的上网供应,实现资源

化和获取经济效益.

5) 烟气净化阶段.MSWI过程的烟气净化主要包含脱硝、脱硫、脱重金属、吸附DXN和除尘(颗粒物)等一系列过程,进而实现焚烧烟气污染物排放达标的目的.采用活性炭喷射系统吸附烟气中的DXN是目前应用最为广泛的技术手段^[51],其吸附后的DXN富集于飞灰中.

6) 烟气排放阶段.经降温和净化处理后的含有微量DXN的焚烧烟气(即烟气G2)由引风机抽吸后经烟囱排放至大气中.MSWI过程的不间断、长时间的运行特性导致烟囱内壁颗粒物中附着大量DXN(即记忆效应^[52]),但在何种工况下存在释放的可能性还是目前业界未解决的研究难题.

目前,面向MSWI过程的DXN软测量检测研究主要针对排放阶段(即烟气G3)的DXN浓度检测,其对应的过程变量信息如表A1所示.本文研究重点是构建G3烟气处的软测量模型.

2 基于宽度混合森林回归的建模策略

本文BHFR建模策略包含特征映射层、潜在特征提取层、特征增强层和增量学习层4个主要部分,如图2所示.

图2中, $\{\mathbf{X}, \mathbf{y}\} \in \mathbf{R}^{N_{\text{Raw}} \times (M+1)}$ 表示原始数据集,由原始输入数据 $\mathbf{X} \in \mathbf{R}^{N_{\text{Raw}} \times M}$ 和输出真值 $\mathbf{y} \in \mathbf{R}^{N_{\text{Raw}} \times 1}$ 组成; $\{\text{DT}_1, \dots, \text{DT}_J\}$ 表示混合森林算法中的 J 个

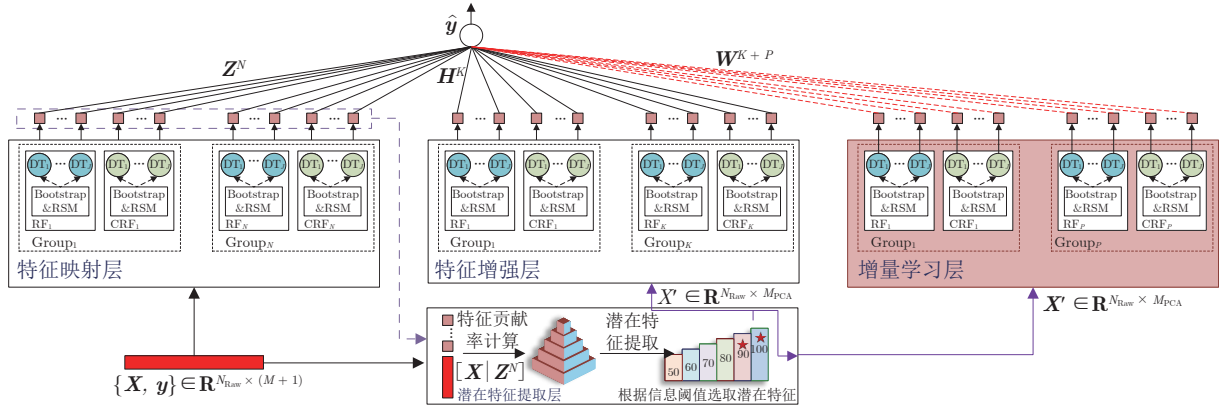


图2 宽度混合森林回归建模策略图

Fig.2 Modeling strategy of broad hybrid forest regression

决策树模型; Bootstrap^[53] 和随机子空间^[54] 表示对输入数据进行样本和特征采样; $\{\text{RF}_n, \text{CRF}_n\}$ 表示第 n 个混合森林组模型, RF_n 和 CRF_n 表示第 n 个 RF 和 CRF 模型; $\{\text{Group}_n\}_{n=1}^N$ 表示特征映射层中包含的 N 个混合森林组模型; \mathbf{Z}^N 表示特征映射层的输出; \mathbf{H}^K 表示特征增强层的输出; $[\mathbf{X}|\mathbf{Z}^N]$ 表示原始数据与 \mathbf{Z}^N 的全联接混合矩阵; $\mathbf{X}' \in \mathbf{R}^{N_{\text{Raw}} \times M_{\text{PCA}}}$ 表示经潜在特征提取后的新训练数据; $\{\text{Group}_k\}_{k=1}^K$ 表示特征增强层包含的 P 个混合森林组模型; $\{\text{Group}_p\}_{p=1}^P$ 表示增量学习层中包含的 K 个混合森林组模型; \mathbf{W}^{K+P} 表示最终的权重矩阵。

各部分的主要功能如下:

1) 特征映射层. 将原始输入数据 $\mathbf{X} \in \mathbf{R}^{N_{\text{Raw}} \times M}$ 通过特征映射层的 N 个混合森林组 $\{\text{RF}_n, \text{CRF}_n\}_{n=1}^N$ 进行特征映射, 得到映射输出矩阵 \mathbf{Z}^N ;

2) 潜在特征提取层. 利用 PCA 对由 \mathbf{X} 与特征映射层输出 \mathbf{Z}^N 组成的全联接混合矩阵 $[\mathbf{X}|\mathbf{Z}^N]$ 进行潜在特征提取, 去除特征空间的冗余信息, 进一步通过所提取的潜在特征与真值 \mathbf{y} 的互信息确定潜在特征维数并得到新的输入数据 $\mathbf{X}' \in \mathbf{R}^{N_{\text{Raw}} \times M_{\text{PM}}}$;

3) 特征增强层. 以 \mathbf{X}' 作为输入, 通过特征增强层的 K 个混合森林组 $\{\text{RF}_k, \text{CRF}_k\}_{k=1}^K$ 进行特征映射, 得到增强层输出矩阵 \mathbf{H}^K ;

4) 增量学习层. 以 \mathbf{X}' 作为输入, 以混合森林组为最小单位逐步增加并更新权重 \mathbf{W}^{K+P} , 直到训练误差收敛。

2.1 特征映射层

从本质上讲, BHFR 是以 RF 和 CRF 为基元构成的混合森林组作为基础映射单元取代原始 BLS 中的神经元。

以特征映射层的第 n 个混合森林组为例描述特征映射层的建模过程. 对 $\{\mathbf{X}, \mathbf{y}\}$ 进行 Bootstrap

和 RSM 采样, 获得混合森林组模型的 J 个训练子集, 表示如下:

$$\left\{ \mathbf{X}_{\text{Bootstrap}}^{n,j}, \mathbf{y}_{\text{Bootstrap}}^{n,j} \right\}_{j=1}^J = \varphi_n^{\text{FML}} \left(\varphi_n^{\text{FML}} (\{\mathbf{X}, \mathbf{y}\}, P_{\text{Bootstrap}}) \right) \quad (1)$$

式中, $\varphi_n^{\text{FML}}(\cdot)$ 和 $\varphi_n^{\text{FML}}(\cdot)$ 表示特征映射层中对第 n 个混合森林组的 Bootstrap 和 RSM 采样, $P_{\text{Bootstrap}}$ 表示 Bootstrap 采样概率。

基于 $\{\mathbf{X}_{\text{Bootstrap}}^{n,j}, \mathbf{y}_{\text{Bootstrap}}^{n,j}\}_{j=1}^J$ 训练包含 J 个决策树的混合森林算法, 其中特征映射层中的第 n 个混合森林组的第 j 个决策树表示如下:

$$f_{n,j}^{\text{DT}}(\cdot) = \sum_{l=1}^L c_l I(\mathbf{x}_{\text{Bootstrap}}^{n,j} \in \mathbf{R}_l), l = 1, 2, \dots, L \quad (2)$$

式中, L 表示决策树叶节点数量, $I(\cdot)$ 表示指示函数, 当预测样本属于当前叶节点的样本集时 ($\mathbf{x}_{\text{Bootstrap}}^{n,j} \in \mathbf{R}_l$), 即将 c_l 作为该样本的预测值. c_l 采用递归分裂方式计算, 其物理意义是: 第 l 个叶节点的预测值, 由第 l 个叶节点中所有样本的加权均值所得。

进一步, RF 和 CRF 的建模过程如图 3 所示。

RF 的基本思想为通过 Bootstrap 和 RSM 构建多个以平方误差为损失函数分裂生长的决策树模型后并行集成, 建模步骤如 RF 建模过程所示. RF 中决策树的分裂损失函数 $\Omega_i(\cdot)$ 可表示为:

$$\begin{aligned} \Omega_i(s, v) = \min & \left([\mathbf{y}_L - E[\mathbf{y}_L]]^2 + [\mathbf{y}_R - E[\mathbf{y}_R]]^2 \right) = \\ & \min \left(\sum_{\mathbf{x}_{\text{Bootstrap}}^{n,j} \in \mathbf{R}_L} (y_L^i - c_L)^2 + \right. \\ & \left. \sum_{\mathbf{x}_{\text{Bootstrap}}^{n,j} \in \mathbf{R}_R} (y_R^i - c_R)^2 \right) \quad (3) \end{aligned}$$

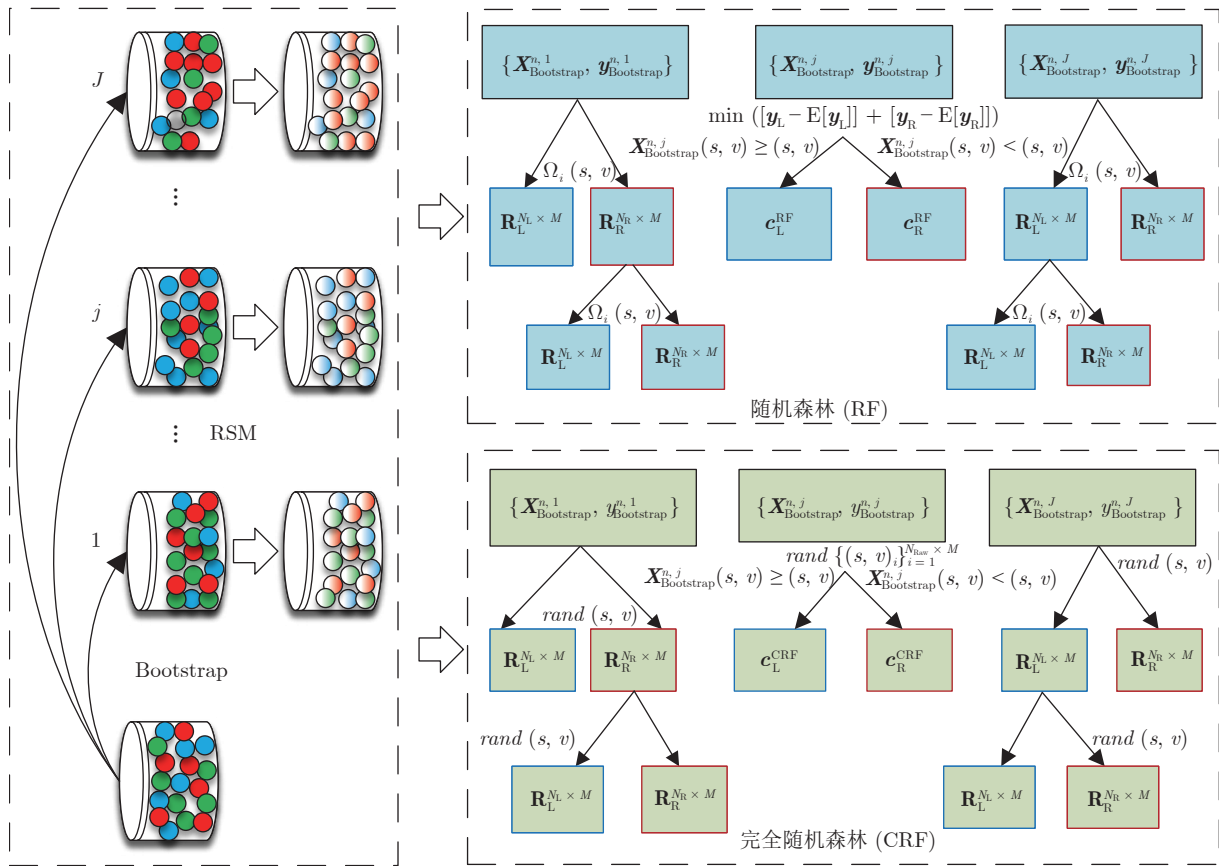


图 3 RF 和 CRF 的建模过程

Fig.3 Modeling process of RF and CRF

式中, $\Omega_i(s, v)$ 表示以第 s 个特征的值 v 作为切分准则的损失函数值, $\min(\cdot)$ 表示最小化函数, \mathbf{y}_L 和 $E[\mathbf{y}_L]$ 表示左节点数据集的真值向量和期望 (文中采用均值), \mathbf{y}_R 和 $E[\mathbf{y}_R]$ 表示右节点数据集的真值向量和期望 (文中采用均值), y_L^i 和 y_R^i 表示左和右节点数据集中第 i 个样本的真值, c_L 和 c_R 表示左和右节点数据集真值的期望, $\sum (y_L^i - c_L)^2$, $\mathbf{x}_{\text{Bootstrap}}^{n,j} \in \mathbf{R}_L$ 和 $\sum (y_R^i - c_R)^2$, $\mathbf{x}_{\text{Bootstrap}}^{n,j} \in \mathbf{R}_R$ 表示左和右节点数据集的平方误差。

通过最小化 $\Omega_i(s, v)$, 将训练集 $(\mathbf{X}_{\text{Bootstrap}}^{n,j}, \mathbf{y}_{\text{Bootstrap}}^{n,j})$ 切分为两个树节点, 如下:

$$\min \{\Omega_i(s, v)\}_{i=1}^{N_{\text{Raw}} \times M} \rightarrow \begin{cases} \mathbf{R}_L^{N_L \times M_{\text{RSM}}} \\ \mathbf{R}_R^{N_R \times M_{\text{RSM}}} \end{cases} \quad (4)$$

式中, $\mathbf{R}_L^{N_L \times M_{\text{RSM}}}$ 和 $\mathbf{R}_R^{N_R \times M_{\text{RSM}}}$ 表示切分后左右两个树节点所包含的样本集, N_L 和 N_R 分别表示 $\mathbf{R}_L^{N_L \times M_{\text{RSM}}}$ 和 $\mathbf{R}_R^{N_R \times M_{\text{RSM}}}$ 中的样本数量, M_{RSM} 表示 $\mathbf{R}_L^{N_L \times M_{\text{RSM}}}$ 和 $\mathbf{R}_R^{N_R \times M_{\text{RSM}}}$ 中的样本特征数量。

当前左右树节点的输出值为样本真值的期望如下:

$$\begin{cases} c_L^{\text{RF}} = E[\mathbf{y}_L], & \mathbf{y}_L \in \mathbf{R}_L^{N_L \times M_{\text{RSM}}} \\ c_R^{\text{RF}} = E[\mathbf{y}_R], & \mathbf{y}_R \in \mathbf{R}_R^{N_R \times M_{\text{RSM}}} \end{cases} \quad (5)$$

式中, \mathbf{y}_L 和 \mathbf{y}_R 表示 $\mathbf{R}_L^{N_L \times M_{\text{RSM}}}$ 和 $\mathbf{R}_R^{N_R \times M_{\text{RSM}}}$ 中的真值向量, $E[\mathbf{y}_L]$ 和 $E[\mathbf{y}_R]$ 表示 \mathbf{y}_L 和 \mathbf{y}_R 的数学期望。

CRF 的基本思想为通过 Bootstrap 和 RSM 构建多个完全随机分裂生长的决策树模型后并行集成, 建模步骤见 CRF 建模过程所示. CRF 中决策树的分裂采用完全随机选择方式, 可表示为:

$$\text{rand} \{(s, v)_i\}_{i=1}^{N_{\text{Raw}} \times M_{\text{RSM}}} \rightarrow \begin{cases} \mathbf{R}_L^{N_L \times M_{\text{RSM}}} \\ \mathbf{R}_R^{N_R \times M_{\text{RSM}}} \end{cases} \quad (6)$$

式中, $\text{rand} \{(s, v)_i\}_{i=1}^{N_{\text{Raw}} \times M_{\text{RSM}}}$ 表示完全随机选取第 s 个特征的值 v 作为切分点。

被随机分裂的左右树节点输出值为样本真值的期望, 如下:

$$\begin{cases} c_L^{\text{CRF}} = E[\mathbf{y}_L], & \mathbf{y}_L \in \mathbf{R}_L^{N_L \times M_{\text{RSM}}} \\ c_R^{\text{CRF}} = E[\mathbf{y}_R], & \mathbf{y}_R \in \mathbf{R}_R^{N_R \times M_{\text{RSM}}} \end{cases} \quad (7)$$

通过上述过程, 第 n 个混合森林组可表示为:

$$f_n^{\text{FML}}(\cdot) = \{f_{n, \text{RF}}^{\text{FML}}(\cdot), f_{n, \text{CRF}}^{\text{FML}}(\cdot)\} \quad (8)$$

进而, 第 n 个映射特征可表示为:

$$\mathbf{Z}_n = f_n^{\text{FML}}(\mathbf{X}) = \{f_{n,\text{RF}}^{\text{FML}}(\mathbf{X}), f_{n,\text{CRF}}^{\text{FML}}(\mathbf{X})\} = ((c_{1,l}^{n,\text{RF}}, c_{1,l}^{n,\text{RF}}), \dots, (c_{N_{\text{Raw}},l}^{n,\text{RF}}, c_{N_{\text{Raw}},l}^{n,\text{RF}})) \quad (9)$$

式中, $(c_{n_{\text{Raw}},l}^{n,\text{RF}}, c_{n_{\text{Raw}},l}^{n,\text{RF}})$ 表示第 n 组混合森林对原始输入数据第 n_{Raw} 个样本的映射特征.

最终, 特征映射层的输出 \mathbf{Z}^N 表示为:

$$\mathbf{Z}^N = (\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_N) \in \mathbf{R}^{N_{\text{Raw}} \times 2N} \quad (10)$$

式中, \mathbf{Z}^N 包含 N_{Raw} 个样本和 $2N$ 维特征.

算法 1. RF 建模步骤

输入. 原始数据集 $\{\mathbf{X}, \mathbf{y}\} \in \mathbf{R}^{N_{\text{Raw}} \times (M+1)}$, 森林中决策树数量 J , 叶节点最小样本数 N_{samples} , RSM 特征选择数量 M_{RSM} .

输出. RF 模型 $f_{n,\text{RF}}^{\text{FML}}(\cdot)$.

- 1) 基于式 (1) 获取 J 训练子集 $\{\mathbf{X}_{\text{Bootstrap}}^{n,j}, \mathbf{y}_{\text{Bootstrap}}^{n,j}\}_{j=1}^J$;
- 2) For $j = 1 : J$;
- 3) 基于式 (3) 和式 (4) 寻找树节点的最佳切分点 (s, v) ;
- 4) 递归运行步骤 3), 直到获得 L 个叶节点得到第 j 个决策树模型 $f_{n,j}^{\text{DT}}(\cdot) = \sum_{l=1}^L c_l I(\mathbf{x}_{\text{Bootstrap}}^{n,j} \in \mathbf{R}_l)$;
- 5) end;
- 6) 获得 J 个决策树模型, 即为 RF 模型 $f_{n,\text{RF}}^{\text{FML}}(\cdot)$.

算法 2. CRF 建模步骤

输入. 原始数据集 $\{\mathbf{X}, \mathbf{y}\} \in \mathbf{R}^{N_{\text{Raw}} \times (M+1)}$, 森林中决策树数量 J , 叶节点最小样本数 N_{samples} , RSM 特征选择数量 M_{RSM} .

输出. CRF 模型 $f_{n,\text{CRF}}^{\text{FML}}(\cdot)$.

- 1) 基于式 (1) 获取 J 训练子集 $\{\mathbf{X}_{\text{Bootstrap}}^{n,j}, \mathbf{y}_{\text{Bootstrap}}^{n,j}\}_{j=1}^J$;
- 2) For $j = 1 : J$;
- 3) 基于式 (3) 和 (6) 寻找树节点的最佳切分点 (s, v) ;
- 4) 递归运行步骤 3), 直到获得 L 个叶节点得到第 j 个决策树模型 $f_{n,j}^{\text{DT}}(\cdot) = \sum_{l=1}^L c_l I(\mathbf{x}_{\text{Bootstrap}}^{n,j} \in \mathbf{R}_l)$;
- 5) end;
- 6) 获得 J 个决策树模型, 即为 CRF 模型 $f_{n,\text{CRF}}^{\text{FML}}(\cdot)$.

2.2 潜在特征提取层

为了避免信息传递过程中的信息丢失导致的过拟合现象, 本文 BHFR 采用全联接策略实现特征映射层与特征增强层、增量学习层之间的信息传递. 同时, 为了保证模型训练过程中信息冗余的最小化, 此处采用 PCA 提取全联接混合矩阵的潜在特征, 再利用互信息进一步筛选与真值信息最大化相关的潜在特征, 进而实现对高维数据的降维处理.

首先, 原始输入数据 \mathbf{X} 与特征映射矩阵 \mathbf{Z}^N 组合得到全联接混合矩阵 \mathbf{A} , 可表示为:

$$\mathbf{A} = [\mathbf{X} | \mathbf{Z}^N] \in \mathbf{R}^{N_{\text{Raw}} \times (M+2N)} \quad (11)$$

式中, \mathbf{A} 含 N_{Raw} 个样本和 $(M+2N)$ 维特征.

接着, 考虑到 \mathbf{A} 的维数远高于原始数据, 此处利用 PCA 最小化 \mathbf{A} 中的冗余信息, 计算 \mathbf{A} 的相关矩阵 \mathbf{R} , 如下:

$$\mathbf{R} = \frac{1}{N_{\text{Raw}} - 1} \mathbf{A}^T \mathbf{A} \in \mathbf{R}^{(M+2N) \times (M+2N)} \quad (12)$$

进一步, 对 \mathbf{R} 进行奇异值分解, 得到 $(M+2N)$ 个特征值和相应特征向量, 如下:

$$\mathbf{R} = \mathbf{U}_{(M+2N)} \mathbf{\Sigma}_{(M+2N)} \mathbf{V}_{(M+2N)} \quad (13)$$

式中, $\mathbf{U}_{(M+2N)}$ 表示 $(M+2N)$ 阶正交矩阵, $\mathbf{\Sigma}_{(M+2N)}$ 表示 $(M+2N)$ 阶对角矩阵, $\mathbf{V}_{(M+2N)}$ 表示 $(M+2N)$ 阶正交矩阵.

$$\mathbf{\Sigma}_{(M+2N)} = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_{(M+2N)} & \\ & & & \ddots \end{bmatrix} \quad (14)$$

式中, $\sigma_1 > \sigma_2 > \dots > \sigma_{(M+2N)}$ 表示由大到小排列的特征值.

然后, 根据设定潜在特征贡献阈值 η , 确定最终的潜在特征数量 Q_{PCA} , 如下:

$$\eta = \frac{\sum_{q=1}^{Q_{\text{PCA}}} \sigma_q}{(M+2N) \sum_{q=1} \sigma_q} \quad (15)$$

式中, $Q_{\text{PCA}} \ll (M+2N)$.

基于上述确定的 Q_{PCA} 个潜在特征, 获得特征值集合 $\{\sigma_q\}_{q=1}^{Q_{\text{PCA}}}$ 对应的特征向量矩阵 $\mathbf{V}_{Q_{\text{PCA}}}$, 即 \mathbf{A} 的投影矩阵. 然后, 对 \mathbf{A} 进行特征投影以实现冗余信息的最小化处理, 将获得潜在特征记为 \mathbf{X}^{PCA} , 即:

$$\mathbf{X}^{\text{PCA}} = \mathbf{A} \mathbf{V}_{Q_{\text{PCA}}} \in \mathbf{R}^{N_{\text{Raw}} \times M_{\text{PCA}}} \quad (16)$$

式中, $\mathbf{V}_{Q_{\text{PCA}}} \in \mathbf{R}^{(M+2N) \times Q_{\text{PCA}}}$ 表示前 Q_{PCA} 个潜在特征对应的特征向量.

进一步, 计算所选潜在特征 $\mathbf{x}_q^{\text{PCA}}$ 与真值 $\mathbf{y} \in \mathbf{R}^{N_{\text{Raw}} \times 1}$ 间的互信息值 I_q^{MI} , 如下:

$$I_q^{\text{MI}}(\mathbf{x}_q^{\text{PCA}}, \mathbf{y}) = \sum p(\mathbf{x}_q^{\text{PCA}}, \mathbf{y}) \log_2 \frac{p(\mathbf{x}_q^{\text{PCA}}, \mathbf{y})}{p(\mathbf{x}_q^{\text{PCA}})p(\mathbf{y})} \quad (17)$$

式中, $p(\mathbf{x}_q^{\text{PCA}}, \mathbf{y})$ 表示第 q 个潜在特征 $\mathbf{x}_q^{\text{PCA}}$ 与真值 \mathbf{y} 的联合概率分布, $p(\mathbf{x}_q^{\text{PCA}})$ 表示第 q 个潜在特征 $\mathbf{x}_q^{\text{PCA}}$ 的边缘概率分布, $p(\mathbf{y})$ 表示真值 \mathbf{y} 的边缘概率分布.

接着, 通过信息最大化选择机制以保证所选择

潜在特征与真值间的相关性, 如下:

$$\{I_q^{\text{MI}}\}_{q=1}^{Q_{\text{PCA}}} \xrightarrow{I_q^{\text{MI}} \geq \zeta} \{I_q^{\text{MI}}\}_{q=1}^{Q_{\text{PCA}}^{\text{MI}}} \quad (18)$$

式中, $\{I_q^{\text{MI}}\}_{q=1}^{Q_{\text{PCA}}}$ 表示 Q_{PCA} 个潜在特征 $\mathbf{x}_q^{\text{PCA}}$ 与真值 \mathbf{y} 的互信息值, ζ 表示最大化信息的阈值, $\{I_q^{\text{MI}}\}_{q=1}^{Q_{\text{PCA}}^{\text{MI}}}$ 表示与真值 \mathbf{y} 信息相关度最大的 $Q_{\text{PCA}}^{\text{MI}}$ 个潜在特征.

最终, 获得包括 $Q_{\text{PCA}}^{\text{MI}}$ 个潜在特征的新数据集 $\{\mathbf{X}', \mathbf{y}\} \in \mathbf{R}^{N_{\text{Raw}} \times (Q_{\text{PCA}}^{\text{MI}} + 1)}$. 为下文统一描述, 此处设定提取后维数 $M_{\text{PCA}}^{\text{MI}} = Q_{\text{PCA}}^{\text{MI}}$.

2.3 特征增强层

首先对新数据集 $\{\mathbf{X}', \mathbf{y}\}$ 进行采样, 获取混合森林算法的训练子集, 如下:

$$\left\{ \mathbf{X}_{\text{Bootstrap}}^{k,j}, \mathbf{y}_{\text{Bootstrap}}^{k,j} \right\}_{j=1}^J = \varphi_k^{\text{FEL}}(\varphi_k^{\text{FEL}}(\{\mathbf{X}', \mathbf{y}\}, P_{\text{Bootstrap}})) \quad (19)$$

式中, $\varphi_k^{\text{FEL}}(\cdot)$ 和 $\varphi_k^{\text{FEL}}(\cdot)$ 表示对第 k 个混合森林组的 Bootstrap 和 RSM 采样.

接着, 以第 k 个混合森林组中第 j 个 RF 的构建为例, 如下:

$$\left\{ \mathbf{X}_{\text{Bootstrap}}^{k,j}, \mathbf{y}_{\text{Bootstrap}}^{k,j} \right\} \xrightarrow{\Omega_j(s,v)} \sum_{l=1}^L c_l I(\mathbf{x}_{\text{Bootstrap}}^{k,j} \in \mathbf{R}_l) = f_{k,j}^{\text{DT-RF}}(\cdot) \quad (20)$$

式中, $f_{k,j}^{\text{DT-RF}}(\cdot)$ 表示特征增强层中第 k 个混合森林组中 RF 的第 j 个决策树; L 表示决策树叶节点的数量; c_l 采用递归分裂方式计算, 具体过程见式 (3) ~ 式 (5).

进而, 可得到特征增强层中第 k 个混合森林组中的 RF 模型为:

$$f_{k,\text{RF}}^{\text{FEL}}(\cdot) = \left\{ f_{k,j}^{\text{DT-RF}}(\cdot) \right\}_{j=1}^J \quad (21)$$

然后, 类似地以第 k 个混合森林组中的第 j 个 CRF 的构建为例, 如下:

$$\left\{ \mathbf{X}_{\text{Bootstrap}}^{k,j}, \mathbf{y}_{\text{Bootstrap}}^{k,j} \right\} \xrightarrow{\text{rand}_j(s,v)} \sum_{l=1}^L c_l I(\mathbf{x}_{\text{Bootstrap}}^{k,j} \in \mathbf{R}_l) = f_{k,j}^{\text{DT-CRF}}(\cdot) \quad (22)$$

式中, $f_{k,j}^{\text{DT-CRF}}(\cdot)$ 表示特征增强层中第 k 个混合森林组中 CRF 的第 j 个决策树; c_l 采用递归分裂方式计算, 具体过程见式 (6)、式 (7).

进而, 可得到特征增强层中第 k 个混合森林组

的 CRF 模型, 其表示为:

$$f_{k,\text{CRF}}^{\text{FEL}}(\cdot) = \left\{ f_{k,j}^{\text{DT-CRF}}(\cdot) \right\}_{j=1}^J \quad (23)$$

通过上述过程, 得到第 k 个混合森林组 $f_k^{\text{FEL}}(\cdot)$, 第 k 个增强特征表示如下:

$$\mathbf{H}_k = f_k^{\text{FEL}}(\mathbf{X}') = [f_{k,\text{RF}}^{\text{FEL}}(\mathbf{X}'), f_{k,\text{CRF}}^{\text{FEL}}(\mathbf{X}')] = \left[(c_{1,l}^{k,\text{RF}}, c_{1,l}^{k,\text{CRF}}), \dots, (c_{N_{\text{Raw}},l}^{k,\text{RF}}, c_{N_{\text{Raw}},l}^{k,\text{CRF}}) \right] \quad (24)$$

式中, $(c_{n_{\text{Raw}},l}^{k,\text{RF}}, c_{n_{\text{Raw}},l}^{k,\text{CRF}})$ 表示第 k 个混合森林组对新数据中第 n_{Raw} 个样本的增强映射.

最后, 特征增强层的输出 \mathbf{H}^K 表示如下:

$$\mathbf{H}^K = [\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_K] \in \mathbf{R}^{N_{\text{Raw}} \times 2K} \quad (25)$$

当不考虑增量学习策略时, BHFR 模型的表示如下:

$$\mathbf{Y} = \mathbf{G}^K \mathbf{W}^K = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_N | \mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_K] \mathbf{W}^K \quad (26)$$

式中, \mathbf{G}^K 表示特征映射层与特征增强层输出的组合, 即 $\mathbf{G}^K = [\mathbf{Z}^N | \mathbf{H}^K]$, 其包含 N_{Raw} 个样本和 $(2N + 2K)$ 维特征; \mathbf{W}^K 表示特征映射层和特征增强层与输出层间的权重, 其计算如下:

$$\mathbf{W}^K = (\lambda \mathbf{I} + [\mathbf{G}^K]^T \mathbf{G}^K)^{-1} [\mathbf{G}^K]^T \mathbf{Y} \quad (27)$$

式中, \mathbf{I} 表示单位矩阵, λ 表示正则项系数. 相应地, \mathbf{G}^K 的伪逆计算可表示为:

$$[\mathbf{G}^K]^\dagger = (\lambda \mathbf{I} + [\mathbf{G}^K]^T \mathbf{G}^K)^{-1} [\mathbf{G}^K]^T = [\mathbf{Z}^N | \mathbf{H}^K]^\dagger \quad (28)$$

2.4 增量学习层

针对 BLS 模型达不到预期建模精度的问题, 文献 [31] 提出了增加特征映射节点和增加特征增强层节点两种策略. 本文的 BHFR 以混合森林组为基本单元, 依据训练误差的收敛程度实现增量学习, 具体如下.

首先, 对新数据集 $\{\mathbf{X}', \mathbf{y}\}$ 进行基于 Bootstrap 和 RSM 的采样, 获取混合森林算法训练子集, 过程如下:

$$\left\{ \mathbf{X}_{\text{Bootstrap}}^{p,j}, \mathbf{y}_{\text{Bootstrap}}^{p,j} \right\}_{j=1}^J = \varphi_p^{\text{ILL}}(\varphi_p^{\text{ILL}}(\{\mathbf{X}', \mathbf{y}\}, P_{\text{Bootstrap}})) \quad (29)$$

式中, $\varphi_p^{\text{ILL}}(\cdot)$ 和 $\varphi_p^{\text{ILL}}(\cdot)$ 表示增量学习层中第 p 个混合森林组的 Bootstrap 和 RSM 采样.

接着, 构建第 p 个混合森林组中的决策树 $f_{p,\text{RF}}^{\text{ILL}}(\cdot)$ 和 $f_{p,\text{CRF}}^{\text{ILL}}(\cdot)$, 其过程与特征映射层和特征增量层相

同, 此处不再赘述.

进一步, 当增加 1 个混合森林组后, 特征映射层、特征增量层和增量学习层的输出可表示如下:

$$\begin{aligned} \mathbf{G}^{K+1} &= [\mathbf{G}^K | f_1^{\text{FEL}}(\mathbf{X}')] = \\ &= [\mathbf{G}^K | \{f_{1,\text{RF}}^{\text{FEL}}(\mathbf{X}'), f_{1,\text{CRF}}^{\text{FEL}}(\mathbf{X}')\}] = \\ &= [\mathbf{G}^K | [(c_{1,l}^{1,\text{RF}}, c_{1,l}^{1,\text{CRF}}), \dots, (c_{N_{\text{Raw}},l}^{1,\text{RF}}, c_{N_{\text{Raw}},l}^{1,\text{CRF}})]] \end{aligned} \quad (30)$$

式中, $\mathbf{G}^K = [\mathbf{Z}^N | \mathbf{H}^K]$ 包含 N_{Raw} 个样本和 $(2N + 2K)$ 维特征, \mathbf{G}^{K+1} 包含 N_{Raw} 个样本和 $(2N + 2K + 2J)$ 维特征.

然后, 进行 \mathbf{G}^{K+1} 的 Moore-Penrose 逆矩阵的递推更新, 如下:

$$\mathbf{B}^T = \begin{cases} [\mathbf{C}]^\dagger, & \mathbf{C} \neq 0 \\ [1 + \mathbf{D}^T \mathbf{D}]^{-1} \mathbf{D}^T [\mathbf{G}^K]^\dagger, & \mathbf{C} = 0 \end{cases} \quad (31)$$

式中, 矩阵 \mathbf{C} 和矩阵 \mathbf{D} 的计算如下:

$$\mathbf{C} = \mathbf{H}_{K+1} - \mathbf{G}^K \mathbf{D} \quad (32)$$

$$\mathbf{D} = [\mathbf{G}^K]^\dagger f_1^{\text{ILL}}(\mathbf{X}') \quad (33)$$

进而, \mathbf{G}^{K+1} 的 Moore-Penrose 逆矩阵的递推公式如下:

$$[\mathbf{G}^{K+1}]^\dagger = \begin{bmatrix} [\mathbf{G}^K]^\dagger - \mathbf{D} \mathbf{B}^T \\ \mathbf{B}^T \end{bmatrix} \quad (34)$$

进一步, 特征映射层、特征增量层和增量学习层与输出层间权重的更新矩阵为:

$$\mathbf{W}^{K+1} = \begin{bmatrix} \mathbf{W}^K - \mathbf{D} \mathbf{B}^T \mathbf{Y} \\ \mathbf{B}^T \mathbf{Y} \end{bmatrix} \quad (35)$$

式中, $\mathbf{W}^K = (\lambda \mathbf{I} + [\mathbf{G}^K]^T \mathbf{G}^K)^{-1} [\mathbf{G}^K]^T \mathbf{Y}$.

由于采用上述伪逆更新策略只需要计算增量学习层混合森林组的伪逆矩阵, 因此能够实现快速的增量式学习.

进一步, 根据训练误差的收敛程度实现自适应增量学习. 此外, 定义误差的收敛阈值 θ_{Con} 用以确定增量学习中混合森林组的数量 p . 相应地, BHFR 模型的增量学习训练误差表示如下:

$$\begin{aligned} \ell &= \lim_{p \rightarrow \infty} \left| \frac{1}{N} \left(\sqrt{(\mathbf{G}^{K+p} \mathbf{W}^{K+p} - \mathbf{y})^2} - \right. \right. \\ &\quad \left. \left. \sqrt{(\mathbf{G}^{K+p+1} \mathbf{W}^{K+p+1} - \mathbf{y})^2} \right) \right| \leq \theta_{\text{Con}} \\ \text{s. t.} \\ \theta_{\text{Con}} &\geq 0 \end{aligned} \quad (36)$$

式中, ℓ 表示增量学习第 $p+1$ 个与第 p 个混合森林组训练差值, $((\mathbf{G}^{K+p} \mathbf{W}^{K+p} - \mathbf{y})^2)^{1/2}$ 和 $((\mathbf{G}^{K+p+1} \mathbf{W}^{K+p+1} - \mathbf{y})^2)^{1/2}$ 表示包含 p 个和 $p+1$ 个混合森林组的 BHFR 模型训练误差.

最终, 本文 BHFR 模型的预测输出为:

$$\hat{\mathbf{Y}} = \mathbf{G}^{K+P} \mathbf{W}^{K+P} \quad (37)$$

2.5 算法伪代码

BHFR 方法的伪代码如下:

算法 3. 算法伪代码

输入. 原始数据集 $\{\mathbf{X}, \mathbf{y}\} \in \mathbf{R}^{N_{\text{Raw}} \times (M+1)}$, 特征映射层混合森林组数量 N , 特征增强层混合森林组数量 K , 增量学习层混合森林组数量 P , 森林中决策树数量 J , 潜在特征贡献阈值 η , 训练误差收敛阈值 θ_{Con} .

输出. 权重矩阵 \mathbf{W}^{K+P} .

- 1) For $n = 1 : N$;
- 2) 基于式 (1) 获取训练子集 $\{\mathbf{X}_{\text{Bootstrap}}^{n,j}, \mathbf{y}_{\text{Bootstrap}}^{n,j}\}_{j=1}^J$;
- 3) 基于式 (3) ~ 式 (7) 构建混合森林组 $f_n^{\text{FML}}(\cdot) = \{f_{n,\text{RF}}^{\text{FML}}(\cdot), f_{n,\text{CRF}}^{\text{FML}}(\cdot)\}$;
- 4) 依据式 (9)、式 (10) 计算特征映射层的输出 \mathbf{Z}^N ;
- 5) end;
- 6) 联接原始输入数据 \mathbf{X} 与特征映射矩阵 \mathbf{Z}^N , 得到全联接混合矩阵 \mathbf{A} ;
- 7) 基于式 (12) 计算 \mathbf{A} 的相关矩阵 \mathbf{R} ;
- 8) 采用式 (13)、式 (14) 分解相关矩阵 \mathbf{R} , 得到特征值 $\Sigma_{(M+2N)}$ 和特征向量 $\mathbf{V}_{(M+2N)}$;
- 9) 基于式 (15) 确定最终的潜在特征矩阵 $\Sigma_{(Q_{\text{PCA}})}$ 和特征向量, 获得 Q_{PCA} 个潜在特征;
- 10) 基于式 (17)、式 (18) 计算并选择 Q_{PCA} 个潜在特征与真值 \mathbf{y} 之间的互信息最大的 $Q_{\text{PCA}}^{\text{MI}}$ 个潜在特征, 获得新训练数据 \mathbf{X}' ;
- 11) For $k = 1 : K$;
- 12) 基于式 (19) 获取训练子集 $\{\mathbf{X}_{\text{Bootstrap}}^{k,j}, \mathbf{y}_{\text{Bootstrap}}^{k,j}\}_{j=1}^J$;
- 13) 基于式 (20) ~ 式 (23) 构建混合森林组 $f_k^{\text{FEL}}(\cdot) = \{f_{k,\text{RF}}^{\text{FEL}}(\cdot), f_{k,\text{CRF}}^{\text{FEL}}(\cdot)\}$;
- 14) 依据式 (24) 计算特征映射层的输出 \mathbf{H}^K ;
- 15) end;
- 16) 根据式 (26) ~ 式 (28) 计算特征映射层和特征增强层的伪逆矩阵 $[\mathbf{G}^K]^\dagger$;
- 17) For $p = 1 : P$;
- 18) 基于式 (29) 获取训练子集 $\{\mathbf{X}_{\text{Bootstrap}}^{p,j}, \mathbf{y}_{\text{Bootstrap}}^{p,j}\}_{j=1}^J$, 并构建混合森林组 $f_p^{\text{ILL}}(\cdot) = \{f_{p,\text{RF}}^{\text{ILL}}(\cdot), f_{p,\text{CRF}}^{\text{ILL}}(\cdot)\}$;
- 19) 根据式 (30) 计算伪逆矩阵 $[\mathbf{G}^{K+P}]^\dagger$;

20) 根据式 (31) ~ 式 (34) 更新权重矩阵 \mathbf{w}^{K+P} ;

21) 依据式 (9) 判断增量学习过程中的训练误差是否达到收敛. 若是, 则结束并跳出循环; 反之, 继续循环;

22) end;

23) 获得增量学习结束后的权重矩阵 \mathbf{w}^{K+P} .

由上述伪代码可知, 特征映射层、特征增强层和增量学习层是在相同方向进行扩展, 算法程序在实现上按照与建模策略相对应的描述进行编写, 进而在软测量的工业应用中只需要通过更新增量学习层即可实现模型的在线更新.

3 实验研究与应用

3.1 实验数据和评价指标描述

本文先采用高维基准数据集验证, 再采用北京某 MSWI 电厂的实际 DXN 数据进行工业验证, 详细信息如表 2 所示.

表 2 实验数据统计结果

Table 2 Statistical results of experimental datasets

数据集	实际样本量	实际样本量			特征维数	
		总数	训练集	验证集		测试集
NIR	218	109	55	27	27	700
CT	583	117	59	29	29	291
RB	372	124	62	31	31	106
DXN	141	141	71	35	35	116

表 2 中, CT 切片数据集和住宅建筑数据集源自 UCI 平台数据库^[55]; 橙汁近红外光谱数据集源于 <http://www.ucl.ac.be>; DXN 数据源自于北京某 MSWI 电厂. 本文数据在文献 [23] 的基础上进行了样本扩充和特征删减, 共涵盖了 2009 ~ 2020 年的 DXN 排放浓度建模数据 141 组. 其中 DXN 真值为 2 小时采样化验后的折算浓度, 每组建模数据的过程变量均为同一时刻下的过程数据, 采样频率为小时, 对缺失数据和异常变量进行剔除后的输入变量为 116 维, 其相应取值为当前 DXN 真值采样时间段内的均值.

本文选取均方根误差、平均绝对误差和决定系数共三个评价指标比较不同方法的性能, 计算如下:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N - 1}} \quad (38)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (39)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (40)$$

3.2 基准验证

本文 BHFR 方法的参数设置为: 决策树叶节点最小样本数 N_{samples} 为 7、RSM 特征选择数量为 \sqrt{M} 、决策树的数量 N_{tree} 为 10、特征映射层和特征增强层中混合森林组的数量 N_{Forest} 均为 10、正则化参数 λ 为 1^{-10} .

首先, 根据特征映射层的输出 \mathbf{z}^N 和原始输入数据的全联接混合矩阵 \mathbf{A} 确定用于特征增强层和增量学习层潜在特征数量. 在 NIR、CT 和 RB 数据集中 \mathbf{A} 的特征维数分别为 900、491 和 303 维, 其前 60 个潜在特征的累计贡献率如图 4 所示.

根据图 4 所示, 当潜在特征贡献率阈值 η 为 0.9 时, NIR、CT 和 RB 数据集中选择的潜在特征数量分别为 3、19 和 12 个. 上述所选潜在特征与真值间的互信息值如图 5 所示.

如图 5 所示, 互信息阈值 ζ 分别为 0.70、0.85 和 0.80 时, NIR、CT 和 RB 数据集中被选的潜在特征数量分别为 2、11 和 1 个.

接着, 预设增量学习层的混合森林组单元数量为 100, BHFR 模型的训练误差与混合森林组数量间的关系如图 6 所示.

由图 6 所示的训练误差曲线可知, BHFR 在 NIR、CT 和 RB 数据集上的训练过程均可收敛至某一确定下限值.

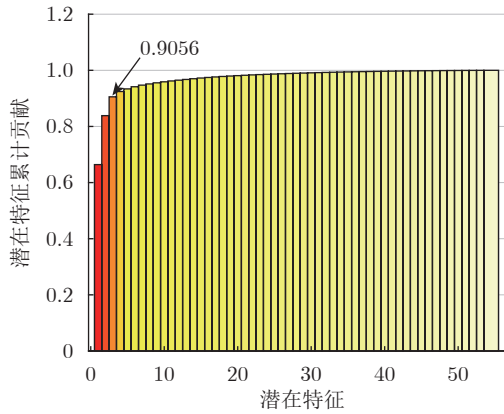
采用 RF、DFR、DFR-clfc、BLS-NN 与本文方法进行对比, 相应模型参数设置为:

1) RF, 决策树叶节点最小样本数 N_{samples} 为 3, RSM 特征选择数量为 \sqrt{M} , 决策树数量 N_{tree} 为 500; 2) DFR, 决策树叶节点最小样本数 N_{samples} 为 3, RSM 特征选择数量为 \sqrt{M} , 决策树数量 N_{tree} 为 500, 每层中 RF 和 CRF 模型的数量 N_{RF} 和 N_{CRF} 均为 2, 总层数设置为 50; 3) DFR-clfc, 决策树叶节点最小样本数 N_{samples} 为 3, RSM 特征选择数量为 \sqrt{M} , 决策树数量 N_{tree} 为 500, 每层中 RF 和 CRF 模型的数量 N_{RF} 和 N_{CRF} 均为 2, 总层数设置为 50; 4) BLS-NN, 特征节点数 N_m 为 5, 增强节点数 N_e 为 41, 神经元数量 N_n 为 9 和正则化参数 λ 为 2^{-30} .

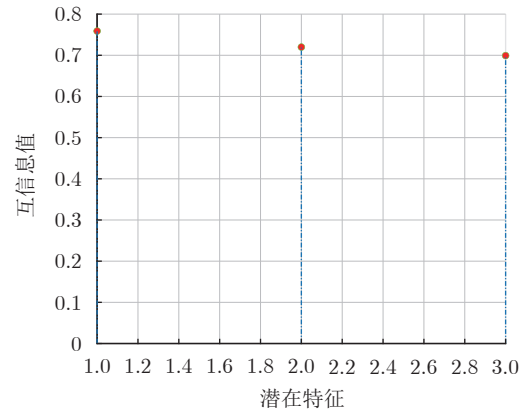
将上述方法在相同条件下重复 20 次实验, 其统计结果和预测曲线如表 3 ~ 表 5 和图 7 ~ 图 9 所示.

由表 3 ~ 表 5 和图 7 ~ 图 9 可知:

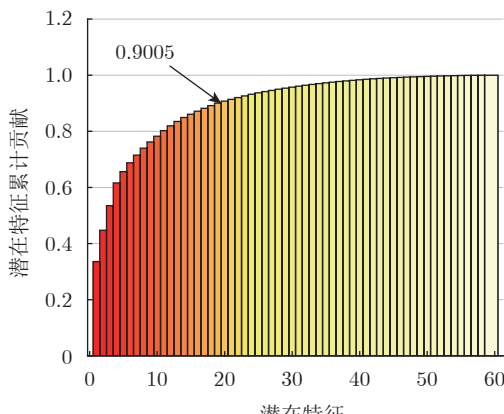
1) NIR 数据集. 在训练集中, RMSE、MAE 和



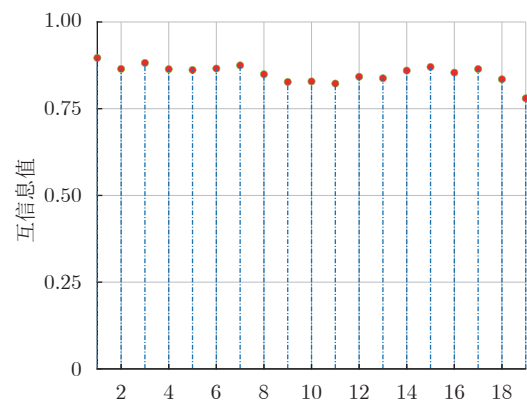
(a) NIR 数据集
(a) NIR dataset



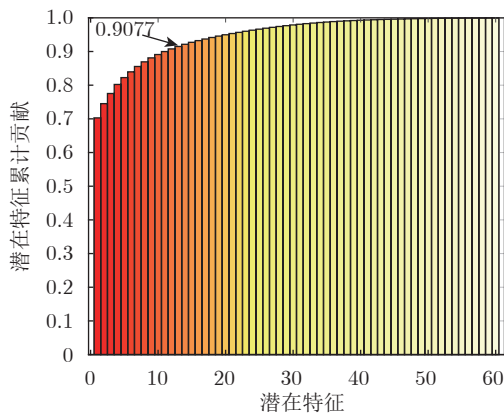
(a) NIR 数据集
(a) NIR dataset



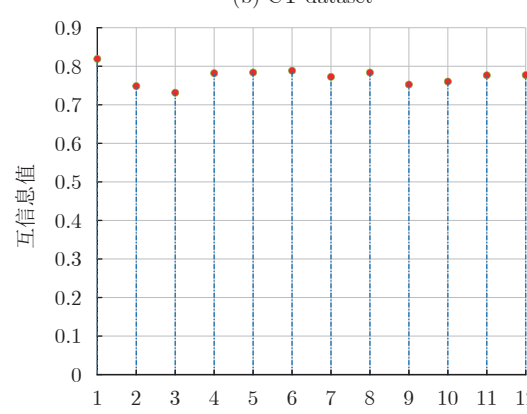
(b) CT 数据集
(b) CT dataset



(b) CT 数据集
(b) CT dataset



(c) RB 数据集
(c) RB dataset



(c) RB 数据集
(c) RB dataset

图 4 基准数据集中潜在特征的贡献率曲线

Fig.4 Contribution rate curves of latent features of benchmark datasets

图 5 基准数据集潜在特征与真值的互信息值

Fig.5 Mutual information values of latent features and true values of benchmark datasets

R^2 指标的均值统计结果均以 BLS-NN 最佳 (7.4226×10^{-1} 、 4.6530×10^{-1} 和 9.9476×10^{-1}), 但方差统计结果弱于 RF、DFR 和 DFR-clfc, 本文 BHFR 在各项指标的均值和方差统计结果上的表现与 BLS-NN 接近; 在验证集中, RF 的 RMSE 和 R^2 指标均

值效果最佳 (1.348×10^1 和 2.7181×10^{-1}), DFR-clfc 在 3 个指标的方差统计结果最优, 本文 BHFR 的性能整体不如 RF、DFR 和 DFR-clfc, BLS-NN 在各项指标中表现最差; 在测试集中, BHFR 在

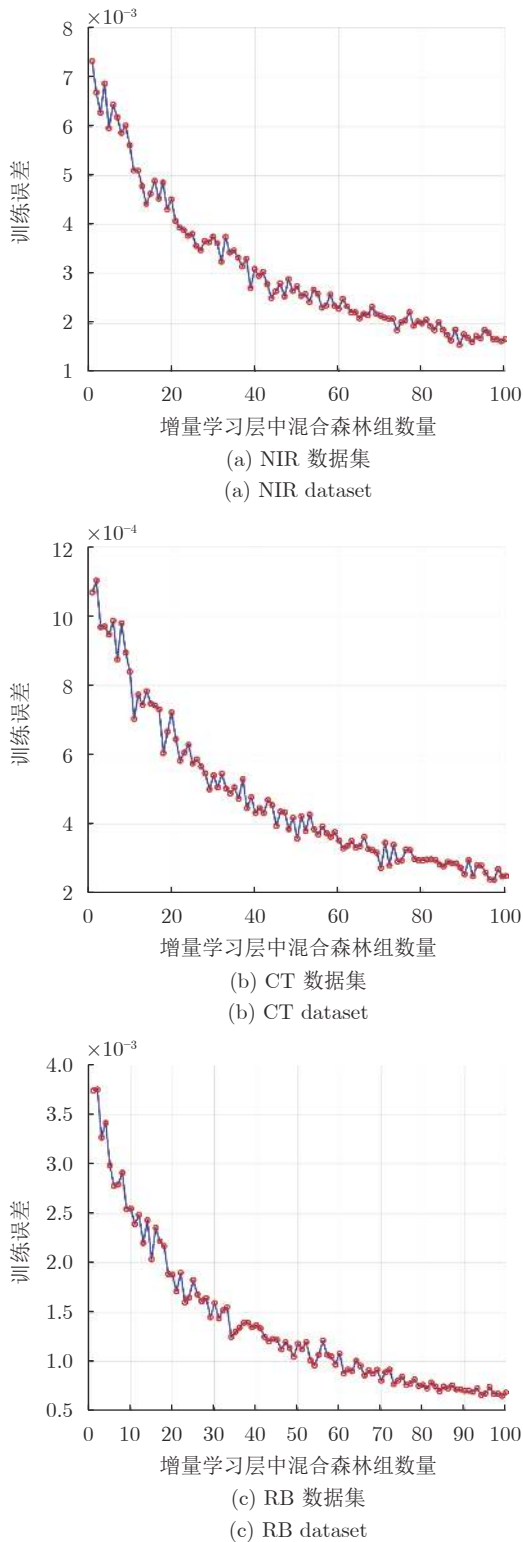


图 6 基准数据集训练误差收敛曲线

Fig.6 Training error convergence curve of benchmark datasets

RMSE、MAE 和 R^2 的均值统计结果优于其他方法,但其方差统计结果依旧低于 RF、DFR 和 DFR-

clfc; 上述结果表明, BHFR 针对高维谱数据的泛化性能最强,但稳定性有待进一步增强.

2) CT 数据集. 在训练集中, BLS-NN 的三个指标的均值和方差均表现最佳, BHFR 仅次于 BLS-NN, BHFR 的训练稳定性好于 RF 和 DFR-clfc; 在验证集中, 三个指标的均值统计结果均是本文 BHFR 最佳 (1.2459×10^0 、 8.5055×10^{-1} 和 9.8148×10^{-1}), 其方差仅弱于 DFR; 在测试集中, BHFR 的表现与验证集一样, 均值性能表现最佳, 稳定性优于 RF 和 DFR-clfc, DFR 性能最稳定. 上述结果表明: 相较于基于树的方法, BLS-NN 在训练时存在明显的过拟合现象; 本文 BHFR 的均值性能最佳, 但其稳定性有待加强.

3) RB 数据集. 训练集中, BLS-NN 的三个指标的均值和方差结果均为最佳; 本文 BHFR 的均值统计结果在验证和测试数据集中均优于其他方法, 但 BHFR 在模型稳定性方面依然弱于 DFR 方法. 上述结果表明, BLS-NN 存在过拟合, 本文 BHFR 稳定性有待增强.

综上可知, 上述小样本高维数据的实验结果表明了本文 BHFR 具有比 RF、DFR 及其改进版 DFR-clfc、BLS-NN 更佳的泛化性能.

3.3 工业验证

在 DXN 数据集中, 本文 BHFR 的参数设置为: 决策树叶节点最小样本数 N_{samples} 为 7, RSM 特征选择数量为 \sqrt{M} , 决策树的数量 N_{tree} 为 10, 特征映射层和特征增强层中混合森林组的数量 N_{Forest} 均为 10, 潜在特征贡献率阈值 η 为 0.9, 正则化参数 λ 为 2^{-10} .

类似基准数据集, 基于全联接混合矩阵 \mathbf{A} 确定用于特征增强层和增量学习层的潜在特征数量, 其维数为 316 维, 前 80 个潜在特征的累计贡献率如图 10 所示.

根据图 10 所示, 当潜在特征贡献率阈值 η 为 0.9 时, DXN 数据集中选择的潜在特征数量为 35 个. 接着, 计算 35 个潜在特征与 DXN 真值间的互信息值, 结果如图 11 所示.

根据图 11 所示, 将互信息阈值 ζ 设置为 0.75, DXN 数据集中被选的潜在特征数量为 6 个. 进一步, 预设增量学习层的混合森林组单元数量为 1000, 相应地 BHFR 模型的训练误差与混合森林组数量间的关系如图 12 所示.

由图 12 所示的训练误差曲线可知, BHFR 在 DXN 数据集上的训练过程可收敛至某一确定下限值.

然后, 采用 RF、DFR、DFR-clfc 和 BLS-NN

表 3 NIR 数据集实验结果
Table 3 Experimental results of NIR dataset

方法	数据集	RMSE		MAE		R^2	
		平均值	方差	平均值	方差	平均值	方差
RF	训练集	5.7142×10^0	2.1347×10^{-2}	3.8799×10^0	9.4469×10^{-3}	8.1131×10^{-1}	9.2887×10^{-5}
	验证集	1.3522×10^1	1.3660×10^{-2}	8.3125×10^0	7.2473×10^{-3}	2.6715×10^{-1}	1.6048×10^{-4}
	测试集	1.0925×10^1	3.2313×10^{-2}	8.1093×10^0	1.7493×10^{-2}	3.0986×10^{-1}	5.1625×10^{-4}
DFR	训练集	5.8836×10^0	1.4331×10^{-3}	4.1004×10^0	6.1766×10^{-4}	8.0007×10^{-1}	6.5971×10^{-6}
	验证集	1.3585×10^1	3.7257×10^{-3}	8.3057×10^0	1.5854×10^{-3}	2.6036×10^{-1}	4.4251×10^{-5}
	测试集	1.0842×10^1	1.4742×10^{-3}	8.0306×10^0	1.8258×10^{-3}	3.2046×10^{-1}	2.3176×10^{-5}
DFR-clfc	训练集	5.7742×10^0	1.7160×10^{-2}	4.0139×10^0	9.6271×10^{-3}	8.0735×10^{-1}	7.6485×10^{-5}
	验证集	1.3569×10^1	3.1520×10^{-3}	8.3265×10^0	1.3592×10^{-3}	2.6219×10^{-1}	3.7318×10^{-5}
	测试集	1.0793×10^1	2.5668×10^{-3}	7.9746×10^0	3.3925×10^{-3}	3.2664×10^{-1}	3.9879×10^{-5}
BLS-NN	训练集	7.4226×10^{-1}	3.7588×10^{-1}	4.6530×10^{-1}	1.5602×10^{-1}	9.9476×10^{-1}	3.2486×10^{-5}
	验证集	7.8288×10^3	4.9487×10^7	2.5450×10^3	4.6180×10^6	-4.3402×10^5	7.2220×10^{11}
	测试集	6.4945×10^3	1.4866×10^7	2.1689×10^3	1.5496×10^6	-3.2544×10^5	8.9669×10^{10}
BHFR	训练集	2.8931×10^0	5.5126×10^{-1}	2.1335×10^0	3.3004×10^{-1}	9.4864×10^{-1}	6.0585×10^{-4}
	验证集	1.3782×10^1	1.7263×10^0	9.2584×10^0	6.5031×10^{-1}	2.3224×10^{-1}	2.1525×10^{-2}
	测试集	9.9505×10^0	5.6804×10^{-1}	7.3675×10^0	2.4515×10^{-1}	4.2455×10^{-1}	7.3749×10^{-3}

表 4 CT 数据集实验结果
Table 4 Experimental results of CT dataset

方法	数据集	RMSE		MAE		R^2	
		平均值	方差	平均值	方差	平均值	方差
RF	训练集	1.5839×10^0	1.2389×10^{-2}	1.2676×10^0	6.1900×10^{-3}	9.7215×10^{-1}	1.4862×10^{-5}
	验证集	1.8907×10^0	2.2323×10^{-2}	1.4978×10^0	1.4923×10^{-2}	9.5761×10^{-1}	4.2488×10^{-5}
	测试集	2.2138×10^0	3.6173×10^{-2}	1.8254×10^0	2.1898×10^{-2}	9.5706×10^{-1}	5.3991×10^{-5}
DFR	训练集	4.1046×10^0	1.3520×10^{-3}	3.4579×10^0	9.5891×10^{-4}	8.1383×10^{-1}	1.1086×10^{-5}
	验证集	4.1864×10^0	1.2940×10^{-3}	3.5658×10^0	1.4003×10^{-3}	7.9338×10^{-1}	1.2646×10^{-5}
	测试集	4.8124×10^0	2.1197×10^{-3}	4.2219×10^0	1.4647×10^{-3}	7.9851×10^{-1}	1.4815×10^{-5}
DFR-clfc	训练集	3.7411×10^0	4.1351×10^{-2}	3.1392×10^0	3.2494×10^{-2}	8.4491×10^{-1}	2.8032×10^{-4}
	验证集	3.8251×10^0	3.9698×10^{-2}	3.2193×10^0	3.5889×10^{-2}	8.2708×10^{-1}	3.1969×10^{-4}
	测试集	4.3811×10^0	4.9342×10^{-2}	3.8122×10^0	4.5453×10^{-2}	8.3262×10^{-1}	2.8693×10^{-4}
BLS-NN	训练集	1.0447×10^{-7}	4.7677×10^{-15}	4.1145×10^{-8}	5.7001×10^{-16}	1.0000×10^0	7.4604×10^{-32}
	验证集	2.5527×10^0	6.9348×10^{-1}	1.7606×10^0	2.0416×10^{-1}	9.1542×10^{-1}	3.2479×10^{-3}
	测试集	2.9466×10^0	1.3019×10^0	1.8882×10^0	4.4686×10^{-1}	9.1371×10^{-1}	3.5944×10^{-3}
BHFR	训练集	4.4458×10^{-1}	8.4237×10^{-3}	3.7023×10^{-1}	4.5597×10^{-3}	9.9773×10^{-1}	1.6195×10^{-6}
	验证集	9.5728×10^{-1}	1.3354×10^{-2}	7.4450×10^{-1}	6.1644×10^{-3}	9.8905×10^{-1}	1.0451×10^{-5}
	测试集	9.3108×10^{-1}	1.2802×10^{-2}	6.7034×10^{-1}	1.2237×10^{-2}	9.9235×10^{-1}	5.1917×10^{-6}

与本文 BHFR 进行对比。参数设置为: 1) RF, 决策树叶节点最小样本数 N_{samples} 为 3, RSM 特征选择数量为 \sqrt{M} , 决策树的数量 N_{tree} 为 500; 2) DFR, 决策树叶节点最小样本数 N_{samples} 为 3, RSM 特征选择数量为 \sqrt{M} , 决策树的数量 N_{tree} 为 500, 每层中 RF 和 CRF 模型的数量 N_{RF} 和 N_{CRF} 均为 2, 总层数设置为 50; 3) DFR-clfc, 决策树叶节点最小样

本数 N_{samples} 为 3, RSM 特征选择数量为 \sqrt{M} , 决策树的数量 N_{tree} 为 500, 每层中 RF 和 CRF 模型的数量 N_{RF} 和 N_{CRF} 均为 2, 总层数设置为 50; 4) BLS-NN, 特征节点数 N_m 为 5, 增强节点数 N_e 为 41, 神经元数量 N_n 为 9 和正则化参数 λ 为 2^{-30} 。上述方法在相同条件下重复 20 次实验, 其统计结果和预测曲线如表 6 和图 13 所示。

表 5 RB 数据集实验结果
Table 5 Experimental results of RB dataset

方法	数据集	RMSE		MAE		R^2	
		平均值	方差	平均值	方差	平均值	方差
RF	训练集	5.3243×10^1	3.7463×10^0	4.2919×10^1	2.4224×10^0	8.4698×10^{-1}	1.2156×10^{-4}
	验证集	1.3211×10^2	6.6035×10^0	8.1855×10^1	2.9950×10^0	5.3824×10^{-1}	3.2140×10^{-4}
	测试集	7.5697×10^1	4.3622×10^0	6.0986×10^1	2.7466×10^0	7.0126×10^{-1}	2.7457×10^{-4}
DFR	训练集	4.8752×10^1	1.4944×10^{-1}	3.9200×10^1	8.4396×10^{-2}	8.7185×10^{-1}	4.1370×10^{-6}
	验证集	1.2600×10^2	1.6487×10^{-1}	7.8890×10^1	1.1939×10^{-1}	5.8012×10^{-1}	7.3315×10^{-6}
	测试集	7.4221×10^1	3.1170×10^{-1}	6.0387×10^1	2.4493×10^{-1}	7.1299×10^{-1}	1.8723×10^{-5}
DFR-clfc	训练集	3.0978×10^1	2.6657×10^1	2.4856×10^1	1.7056×10^1	9.4690×10^{-1}	3.4386×10^{-4}
	验证集	1.1830×10^2	4.9405×10^0	7.2901×10^1	2.7676×10^0	6.2977×10^{-1}	1.9820×10^{-4}
	测试集	6.9427×10^1	1.7460×10^0	5.5570×10^1	2.1741×10^0	7.4879×10^{-1}	9.2839×10^{-5}
BLS-NN	训练集	9.4877×10^{-4}	4.8003×10^{-6}	4.3563×10^{-4}	7.7161×10^{-7}	1.0000×10^0	1.1561×10^{-18}
	验证集	5.0098×10^2	1.3099×10^5	2.6163×10^2	2.6950×10^4	-8.9285×10^0	1.6631×10^2
	测试集	5.2093×10^2	1.5354×10^5	2.8934×10^2	3.2539×10^4	-2.0737×10^1	9.3226×10^2
BHFR	训练集	8.4893×10^0	6.4125×10^{-1}	6.4970×10^0	3.1874×10^{-1}	9.9608×10^{-1}	5.6404×10^{-7}
	验证集	9.9275×10^1	2.2922×10^0	5.4880×10^1	1.0407×10^0	7.3930×10^{-1}	6.3926×10^{-5}
	测试集	4.5645×10^1	1.7188×10^0	3.2284×10^1	9.0105×10^{-1}	8.9137×10^{-1}	3.8975×10^{-5}

由表 6 和图 13 可知, 1) RF 在训练、验证和测试中的 RMSE、MAE 和 R^2 指标均值统计结果均优于 DFR, 但在稳定性指标上弱于 DFR; 2) DFR 和 DFR-clfc 在建模精度上与 RF 接近, 同时建模稳定性要好于 RF, 其中 DFR-clfc 在训练、验证和测试集的精度略高于 DFR, 但 DFR 的稳定性更好; 3) BLS-NN 对训练数据出现了明显的过拟合, 其在验证和测试集中的泛化性能和稳定性上均表现最差, 表明 BLS-NN 难以适用于本文中的真实工业过程的小样本高维数据; 4) BHFR 在测试集中的 RMSE、MAE 和 R^2 指标的均值统计结果均为最佳, 稳定性仅弱于 DFR, 表明 BHFR 具有良好的泛化性能和稳定性。

综上可知, DXN 软测量建模实验表明本文 BHFR 具有比经典 RF、DFR 及其改进版 DFR-clfc 更好的学习能力, 同时在测试集上的建模精度和对数据的拟合程度也强于 RF、DFR、DFR-clfc 和 BLS-NN, 体现了其在构建 DXN 软测量模型中的明显优势。

3.4 算法时间复杂度分析

在 RF、DFR、DFR-clfc 和 BHFR 模型中采用的决策树算法均为二叉树, 该类型决策树算法的时间复杂度可表示为 $O_{\text{Tree}}(M \times N \times \log N)$ (一个决策树模型的时间复杂度记为 $O_{\text{Tree}}(\#Tree)$), 其中 M 表示数据的维度, N 表示数据的样本数量. 下面仅考虑模型的主要结构部分, 并给出了各模型的算法时间复杂度。

1) RF 的时间复杂度为:

$$T_{\text{RF}} = O_{\text{RF}}(N_{\text{tree}} \times \#Tree^*) \quad (41)$$

式中, $\#Tree^* = M_{\text{RSM}} \times N \times \log N$, N_{tree} 表示 RF 的决策树数量 (一个森林模型的时间复杂度记为 $O_{\text{RF}}(\#Forest)$).

2) DFR 的时间复杂度为:

$$T_{\text{DFR}} = O_{\text{DFR}}((N_{\text{RF}} + N_{\text{CRF}}) \times \text{Layers} \times \#Forest) \quad (42)$$

式中, N_{RF} 和 N_{CRF} 表示每层中 RF 模型和 CRF 模型的数量, Layers 表示 DFR 模型的深度。

3) DFR-clfc 的时间复杂度为:

$$T_{\text{DFR-clfc}} = O_{\text{DFR-clfc}}((N_{\text{RF}} + N_{\text{CRF}}) \times \text{Layers} \times \#Forest^*) \quad (43)$$

式中, $\#Forest^* = \#Forest \times \beta$, β 为森林模型复杂度的系数, 该系数由每层逐渐增加的特征数量决定。

4) BLS-NN 的时间复杂度可表示为:

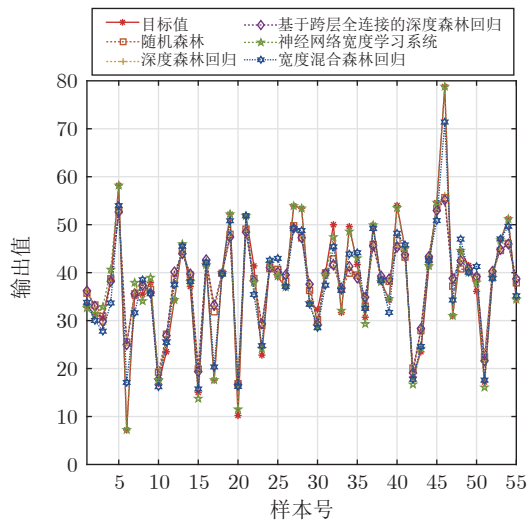
$$T_{\text{BLS-NN}} = O_{\text{BLS-NN}}(M \times N \times (N_m + N_e) \times N_n) \quad (44)$$

式中, N_m 表示特征节点数量, N_e 表示增强节点数量, N_n 表示每个节点中的神经元数量。

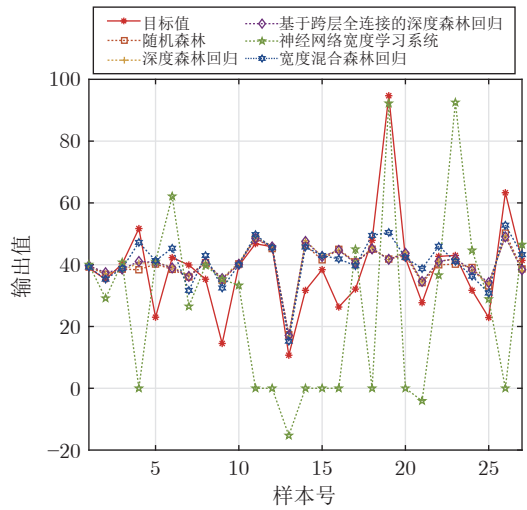
5) 本文 BHFR 的时间复杂度为:

$$T_{\text{BHFR}} = O_{\text{BHFR}}((\text{Group}_N + \text{Group}_K + \text{Group}_P) \times (N_{\text{RF}} + N_{\text{CRF}}) \times \#Forest^*) \quad (45)$$

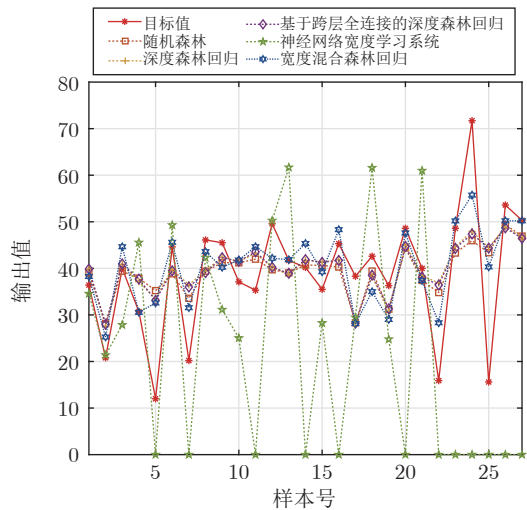
式中, $\#Forest^* = \#Forest \times \alpha$, α 表示森林模型复



(a) NIR 训练集

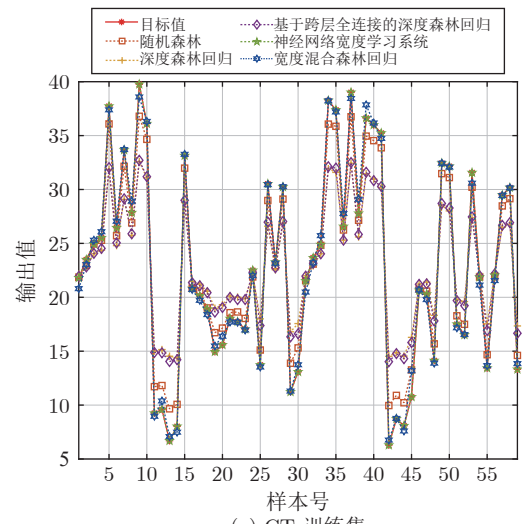


(b) NIR 验证集

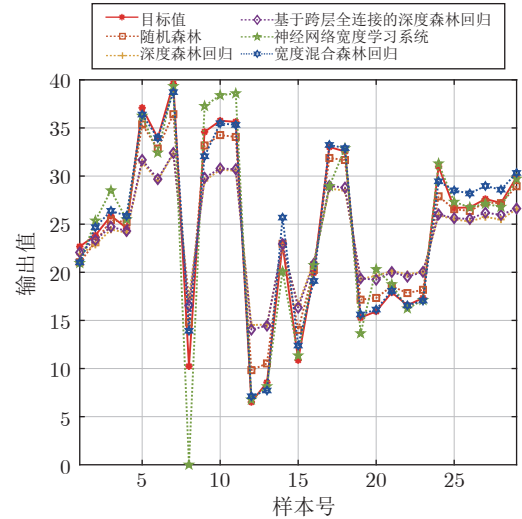


(c) NIR 测试集

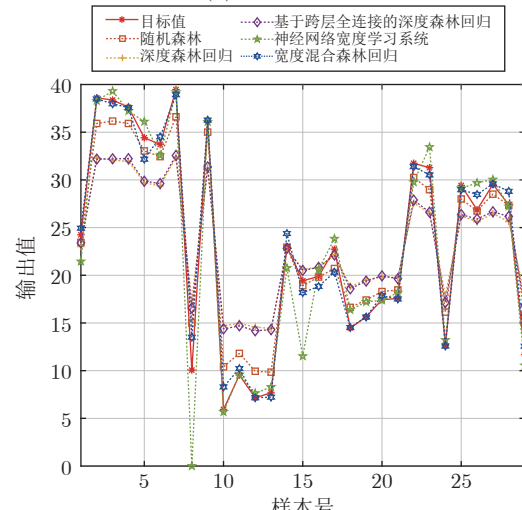
图 7 NIR 数据集的拟合曲线
Fig.7 Fitting curves of NIR dataset



(a) CT 训练集

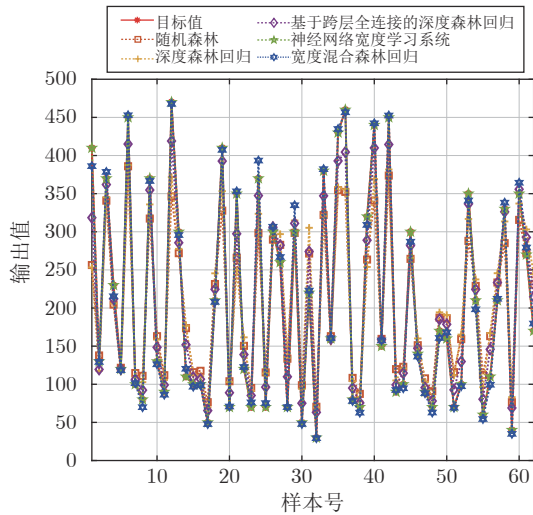


(b) CT 验证集

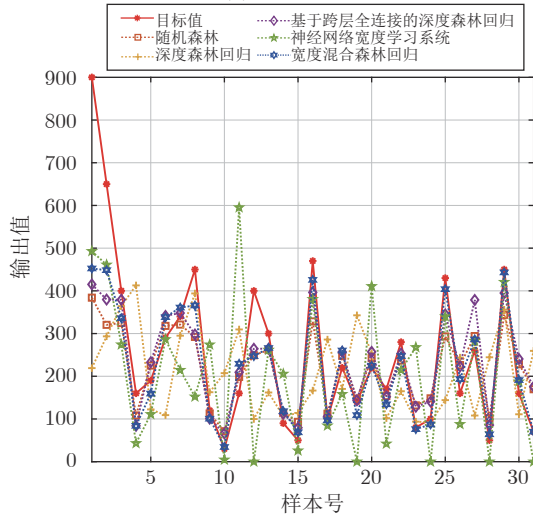


(c) CT 测试集

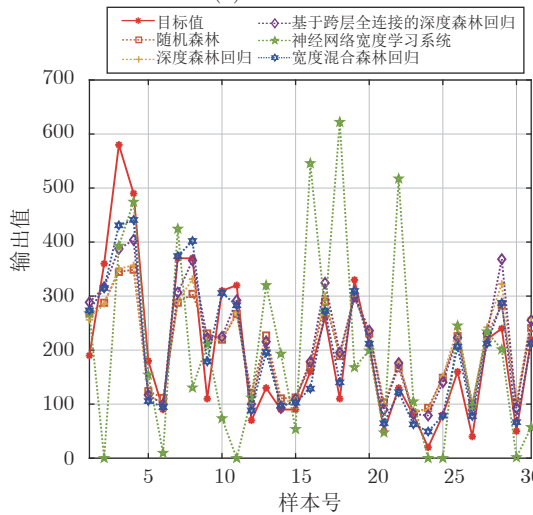
图 8 CT 数据集的拟合曲线
Fig.8 Fitting curves of CT dataset



(a) RB 训练集
(a) RB training set



(b) RB 验证集
(b) RB validation set



(c) RB 测试集
(c) RB test set

图 9 RB 数据集的拟合曲线
Fig.9 Fitting curves of RB dataset

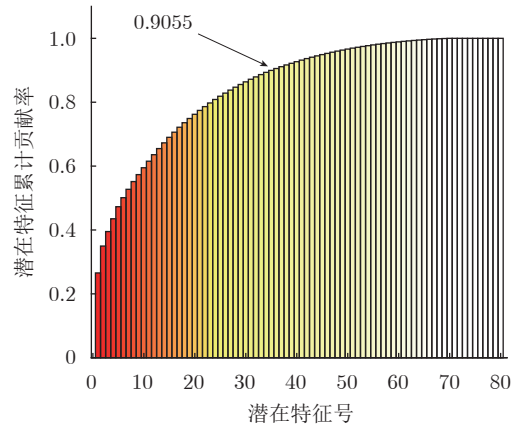


图 10 DXN 数据潜在特征的贡献率曲线

Fig.10 Contribution rate curve of latent feature of DXN dataset

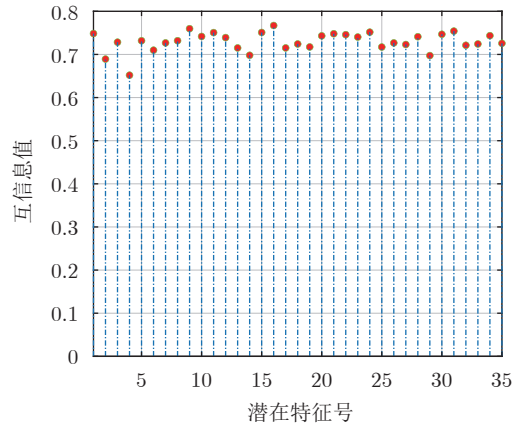


图 11 DXN 数据潜在特征与真值的互信息值

Fig.11 Mutual information value of latent feature and true value of DXN dataset

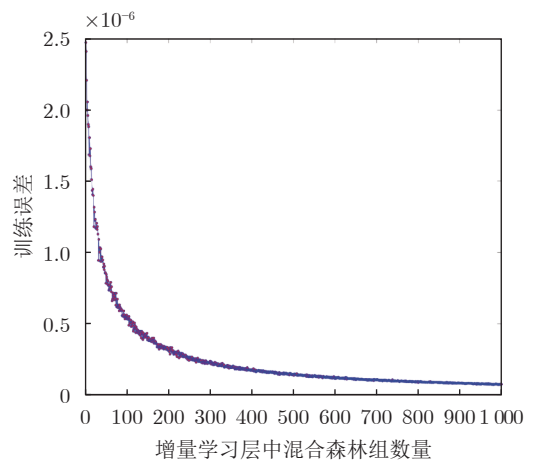


图 12 DXN 训练误差收敛曲线

Fig.12 Convergence curves of DXN training error

表 6 DXN 数据集实验结果
Table 6 Experimental results of DXN dataset

方法	数据集	RMSE		MAE		R^2	
		平均值	方差	平均值	方差	平均值	方差
RF	训练集	1.1159×10^{-2}	5.7497×10^{-8}	9.0221×10^{-3}	4.0684×10^{-8}	8.5346×10^{-1}	3.9360×10^{-5}
	验证集	2.0051×10^{-2}	1.8026×10^{-7}	1.4677×10^{-2}	8.2255×10^{-8}	5.0196×10^{-1}	4.3515×10^{-4}
	测试集	1.6922×10^{-2}	1.6150×10^{-7}	1.3548×10^{-2}	8.9520×10^{-8}	5.9001×10^{-1}	3.7817×10^{-4}
DFR	训练集	1.1493×10^{-2}	8.7413×10^{-9}	9.4568×10^{-3}	4.6626×10^{-9}	8.4463×10^{-1}	6.3663×10^{-6}
	验证集	2.0735×10^{-2}	9.7835×10^{-9}	1.5780×10^{-2}	1.1121×10^{-8}	4.6759×10^{-1}	2.5813×10^{-5}
	测试集	1.7791×10^{-2}	1.7308×10^{-8}	1.4608×10^{-2}	1.5235×10^{-8}	5.4701×10^{-1}	4.5066×10^{-5}
DFR-clfc	训练集	8.0852×10^{-3}	2.9078×10^{-6}	6.6040×10^{-3}	2.0819×10^{-6}	9.1986×10^{-1}	1.1887×10^{-3}
	验证集	2.0187×10^{-2}	1.4562×10^{-7}	1.5626×10^{-2}	2.3355×10^{-8}	4.9520×10^{-1}	3.6404×10^{-4}
	测试集	1.7025×10^{-2}	1.5755×10^{-7}	1.4068×10^{-2}	6.0233×10^{-8}	5.8501×10^{-1}	3.7843×10^{-4}
BLS-NN	训练集	1.2924×10^{-9}	1.5756×10^{-18}	9.5358×10^{-10}	7.2150×10^{-19}	1.0000×10^0	8.2358×10^{-29}
	验证集	6.8845×10^{-2}	7.0040×10^{-4}	5.3153×10^{-2}	3.3474×10^{-4}	-5.6928×10^0	3.7799×10^1
	测试集	7.8396×10^{-2}	6.7692×10^{-4}	6.0922×10^{-2}	4.1785×10^{-4}	-8.7153×10^0	4.7630×10^1
BHFR	训练集	6.0665×10^{-3}	1.6330×10^{-8}	3.9665×10^{-3}	8.4708×10^{-9}	9.5669×10^{-1}	3.3481×10^{-6}
	验证集	2.1551×10^{-2}	3.5181×10^{-8}	1.2384×10^{-2}	3.5083×10^{-8}	4.2484×10^{-1}	9.8731×10^{-5}
	测试集	1.6189×10^{-2}	2.2474×10^{-8}	1.1226×10^{-2}	1.0102×10^{-8}	6.2491×10^{-1}	4.8607×10^{-5}

杂度的系数且 $\alpha < 1$ (潜在特征提取层的降维作用使得特征增强层和增量学习层的输入特征维数显著减小).

由式 (41) ~ 式 (45) 可知, BHFR 的时间复杂度 T_{BHFR} 要远小于传统深度集成方法 DFR 和 DFR-clfc 的 T_{DFR} 和 $T_{\text{DFR-clfc}}$, 即相对于深度集成算法 BHFR 在时间成本上具有明显优势. 相较于 BLS-NN 方法的 $T_{\text{BLS-NN}}$, BHFR 的时间复杂度偏高, 其主要原因在于混合森林组的时间复杂度 $O(\#Forest)$ 要高于单一神经元. 因此, 进一步的研究是考虑在不损失模型精度的情况下降低 BHFR 的时间复杂度.

3.5 超参数敏感性分析

为进一步对本文 BHFR 进行性能评估, 本节对决策树叶节点最小样本数 N_{samples} 、RSM 特征选择数量 M_{RSM} 、决策树的数量 N_{tree} 、混合森林组的数量 N_{Forest} 、潜在特征贡献率阈值 η 、互信息阈值 ζ 和正则化参数 λ 等超参数进行敏感性分析, 其设置区间如表 7 所示.

实验中采用单因素分析策略, 即每次仅将单一参数作为可变项, 测试集的 R^2 实验结果如图 14 所示.

由图 14 可知:

1) 决策树叶节点最小样本数 N_{samples} . 如图 14(a) 所示, 其横坐标表示设置的叶节点最小样本数 N_{samples} (每个数据集的横坐标长度不同, 设置区间详见表 5), 纵坐标表示模型的性能评价指标 R^2 . 随着 N_{samples} 的逐渐增大, NIR、RB 和 DXN 数据集中

的 R^2 逐渐变小, 表明取相对小的 N_{samples} 时模型的泛化性能更佳, 反之则较差; 在 CT 数据集中, 模型性能不受 N_{samples} 变化的影响, 即当决策树仅包含一个根节点时 (即 N_{samples} 等于训练样本数量), 测试集具有非常好的泛化性能. 可知, 不同数据集对 BHFR 模型中 N_{samples} 的敏感程度具有差异性. 因此, 将 N_{samples} 设置在 [3, 10] 最为合适.

2) RSM 特征选择数量 M_{RSM} . 如图 14(b) 所示, 其横坐标表示设置的 RSM 特征选择数量 M_{RSM} (每个数据集的横坐标长度不同, 设置区间详见表 5), 纵坐标表示模型的性能评价指标 R^2 . 当 M_{RSM} 增加时, CT 和 RB 数据集中的 R^2 逐渐变大后保持平稳, NIR 和 DXN 数据集中的 R^2 曲线却具有明显的震荡趋势. 可知, NIR 和 DXN 数据集对 M_{RSM} 的数量更为敏感, CT 和 RB 数据集只需选取较少特征即可建立较高精度的模型, 即数据间存在差异性. 因此, 将 M_{RSM} 设置为原始特征数量的 1/10, 可兼顾模型的时间成本和泛化性能.

3) 决策树的数量 N_{tree} . 如图 14(c) 所示, 其横坐标表示设置的决策树的数量 N_{tree} (每个数据集的横坐标长度均为 5 ~ 100), 纵坐标表示模型的性能评价指标 R^2 . 随着 N_{tree} 的逐步增加, CT、RB 和 DXN 数据集中的 R^2 趋势总体平稳, NIR 数据集中的 R^2 的变化却较为明显. 可知, 针对 CT、RB 和 DXN 数据集, 决策树数量对模型的影响较小, 选取较少的决策树建模即可保证模型性能. 因此, 将 N_{tree} 设置为 40, 可使模型适用于不同数据集的建模需求.

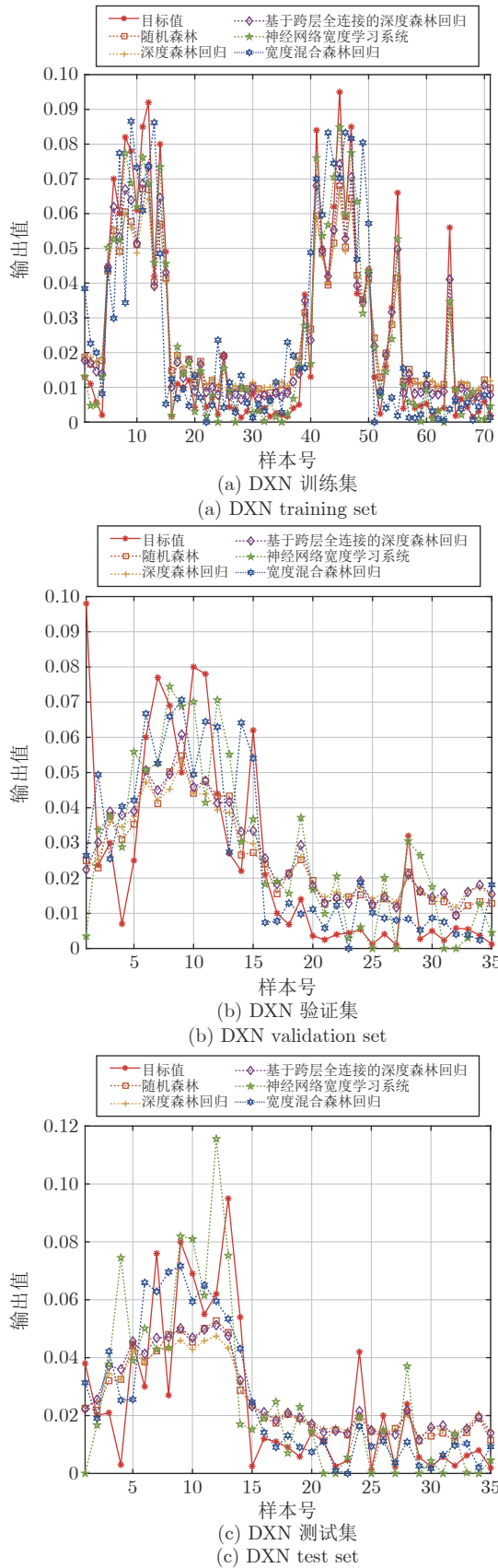


图 13 DDXN 数据集的拟合曲线
Fig. 13 Fit curves of DDXN dataset

表 7 BHFR 超参数及其区间设置

Table 7 Super parameters and interval setting of BHFR

模型超参数	符号	NIR	CT	RB	DXN
决策树叶 节点最小 样本数	N_{samples}	[1, 55]	[1, 59]	[1, 62]	[1, 71]
RSM 特征 选择数量	M_{RSM}	[1, 700]	[1, 291]	[1, 103]	[1, 116]
决策树数量	M_{tree}	[5, 100]	[5, 100]	[5, 100]	[5, 100]
混合森林组 数量	N_{Forest}	[5, 100]	[5, 100]	[5, 100]	[5, 100]
潜在特征 贡献率阈值	η	[0.7, 1]	[0.4, 1]	[0.71, 1]	[0.3, 1]
互信息阈值	ζ	[0.65, 0.75]	[0.75, 0.85]	[0.72, 0.8]	[0.68, 0.76]
正则化参数	λ	$[0.1^{-10}, 1]$	$[0.1^{-10}, 1]$	$[0.1^{-10}, 1]$	$[0.1^{-10}, 1]$

4) 混合森林组的数量 N_{Forest} . 如图 14(d) 所示, 其横坐标表示设置的混合森林组的数量 N_{Forest} (每个数据集的横坐标长度均为 5 ~ 100), 纵坐标表示模型的性能评价指标 R^2 . 随着 N_{Forest} 的增加, 各数据集的性能表现与超参数 N_{tree} 相一致. 可知, N_{Forest} 的设置不宜过大, 其值适当时模型的泛化性能最佳. 因此, 将 N_{Forest} 设置在 [25, 35] 区间, 模型建模精度较高.

5) 潜在特征贡献率阈值 η . 如图 14(e) 所示, 其横坐标表示设置的潜在特征贡献率阈值 η (每个数据集的横坐标长度不同, 设置区间详见表 5), 纵坐标表示模型的性能评价指标 R^2 . 其决定着从全连接混合矩阵特征空间 \mathbf{A} 中提取的潜在特征数量, η 越大表示潜在特征维数越高. 实验表明, 随着 η 增大, CT 数据集的 R^2 表现最稳定, NIR、RB 和 DXN 数据集的 R^2 在接近 $\eta = 1$ 时, 出现局部最小值 (即模型性能最差). 可知, 本文方法对超参数 η 的敏感程度相对较低. 因此, 将 η 设置为 0.9, 可确保选择合适数量的潜在特征, 进而保证建模精度.

6) 互信息阈值 ζ . 如图 14(f) 所示, 其横坐标表示设置的互信息阈值 ζ (每个数据集的横坐标长度不同, 设置区间详见表 5), 纵坐标表示模型的性能评价指标 R^2 . 理论上其决定特征增强层和增量学习层的输入特征维数, ζ 越大表示输入特征维数越低. 实验中, 随着 η 增大, CT 数据集的性能不变, RB 数据集的 R^2 存在极小值, DXN 数据集的 R^2 先变大后平稳, NIR 数据集的 R^2 逐渐变小后接近稳定. 可知, CT 和 DXN 数据集, 超参数 ζ 对模型性能的影响较小; 但 RB 和 NIR 数据集, 超参数 ζ 设置较小时模型性能更佳. 因此, 综合不同数据集的结果可知, 将 ζ 设置为 0.7 较为合适.

7) 正则化参数 λ . 如图 14(g) 所示, 其横坐标表示设置的正则化参数 λ (每个数据集的横坐标长度

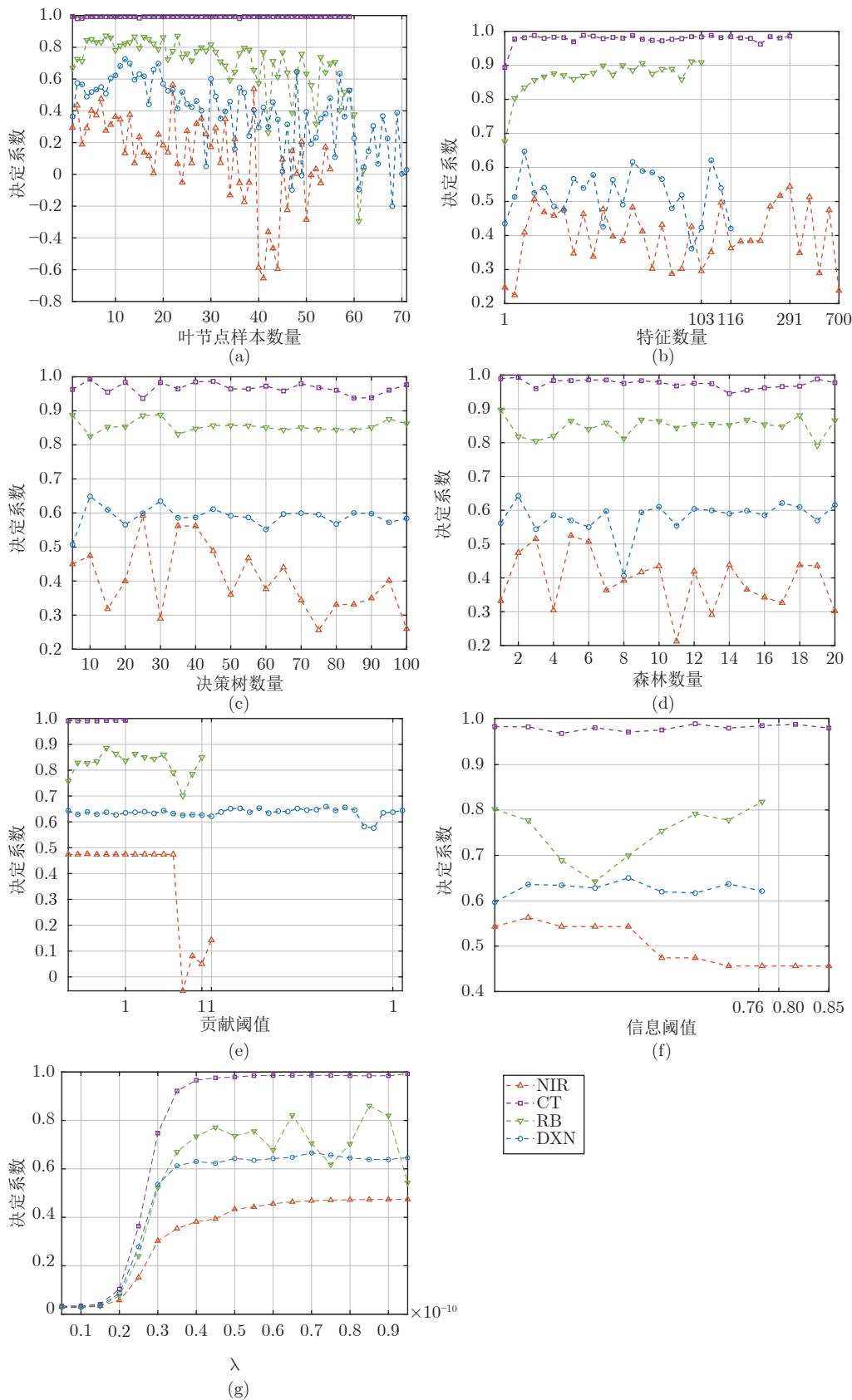


图 14 超参数敏感性分析曲线
Fig.14 Sensitivity analysis curves of super parameter

不同, 设置区间详见表 5), 纵坐标表示模型的性能评价指标 R^2 . 随着 λ 逐渐变小, NIR、CT、RB 和 DXN 数据集的总体表现为先逐渐增大再达到稳定, 其中 RB 数据集在模型达到平稳后表现出明显的震荡现象. 可知, 正则化参数 λ 对不同数据集的影响是一致的, 其中以 RB 数据集最为敏感. 因此, 将 λ 设置为 10^{-10} 可使模型效果达到最佳.

综上所述, 不同超参数针对不同数据集具有差异性, 有必要进行全局优化.

4 结论

针对 MSWI 过程关键指标参数 DXN 排放浓度难以实时准确检测的问题, 本文提出了一种基于 BHFR 的软测量方法, 其结合了宽度学习、集成学习和潜在特征提取等算法, 主要贡献: 1) 基于宽度学习系统框架, 采用非微分学习器构建了包含特征映射层、潜在特征提取层、特征增强层和增量学习层的软测量模型; 2) 利用信息全联接、潜在特征提取和互信息度量对 BHFR 模型内部信息进行处理, 有效保证了 BHFR 模型内部特征信息的传递最大化和冗余度最小化; 3) 采用混合森林组为映射单元实现建模过程的增量学习, 通过伪逆策略快速计算输出层权重矩阵, 再利用训练误差的收敛程度自适应调整增量学习, 实现了高精度的软测量建模. 基于北京某 MSWI 电厂的真实数据验证了 DXN 浓度软测量模型的有效性.

今后的研究工作是进行基于数值仿真的机理分析, 实现面向生成、吸附和排放等多阶段的 DXN 排放浓度软测量建模, 为 MSWI 过程污染排放运行优化控制提供有效支撑.

附录 A

表 A1 DXN 数据集的过程变量信息
Table A1 The process variable information for DXN datasets

序号	过程变量	单位
1	燃烧室左侧温度 1	℃
2	燃烧室左侧温度 2	℃
3	燃烧室右侧温度 3	℃
4	燃烧室左侧温度 4	℃
5	燃烧室右侧温度 5	℃
6	燃烧室右侧温度 6	℃
7	燃烬段炉排顶端气温左	℃
8	燃烬段炉排顶端气温右	℃
9	干燥炉排温度左入口	℃

表 A1 DXN 数据集的过程变量信息 (续表)

Table A1 The process variable information for DXN datasets (continued table)

序号	过程变量	单位
10	干燥炉排温度左出口	℃
11	干燥炉排温度右入口	℃
12	干燥炉排温度右出口	℃
13	干燥段与燃烧段炉排的炉墙左侧内温度	℃
14	干燥段与燃烧段炉排的炉墙左侧外温度	℃
15	干燥段与燃烧段炉排的炉墙右侧内温度	℃
16	干燥段与燃烧段炉排的炉墙右侧外温度	℃
17	燃烧炉排 1-1 左内侧温度	℃
18	燃烧炉排 1-1 左外侧温度	℃
19	燃烧炉排 1-1 右内侧温度	℃
20	燃烧炉排 1-1 右外侧温度	℃
21	燃烧炉排 1-2 左内侧温度	℃
22	燃烧炉排 1-2 左外侧温度	℃
23	燃烧炉排 1-2 右内侧温度	℃
24	燃烧炉排 1-2 右外侧温度	℃
25	燃烧炉排 2-1 左内侧温度	℃
26	燃烧炉排 2-1 左外侧温度	℃
27	燃烧炉排 2-1 右内侧温度	℃
28	燃烧炉排 2-1 右外侧温度	℃
29	燃烧炉排 2-2 左内侧温度	℃
30	燃烧炉排 2-2 左外侧温度	℃
31	燃烧炉排 2-2 右内侧温度	℃
32	燃烧炉排 2-2 右外侧温度	℃
33	二次燃烧室出口温度左	℃
34	二次燃烧室出口温度右	℃
35	一次空预器出口空气温度	℃
36	燃烧炉排入口空气温度	℃
37	干燥炉排入口空气温度	℃
38	二次空预器出口空气温度	℃
39	炉墙左侧冷却风出口温度	℃
40	炉墙右侧冷却风出口温度	℃
41	炉排左侧冷却风出口温度	℃
42	炉排右侧冷却风出口温度	℃
43	一级过热器出口蒸汽温度	℃
44	二级过热器出口蒸汽温度	℃
45	三级过热器出口蒸汽温度	℃
46	省煤器出口给水温度	℃
47	反应器入口温度	℃
48	布袋 A 入口温度	℃
49	布袋 B 入口温度	℃
50	流化风机出口温度	℃
51	FGD 出口烟气温度 A	℃
52	FGD 出口烟气温度 B	℃
53	烟道入口烟气温度	℃

表 A1 DXN 数据集的过程变量信息 (续表)

Table A1 The process variable information for DXN datasets (continued table)

序号	过程变量	单位
54	干燥炉排左 1 空气流量	km ³ N/h
55	干燥炉排右 1 空气流量	km ³ N/h
56	干燥炉排左 2 空气流量	km ³ N/h
57	干燥炉排右 2 空气流量	km ³ N/h
58	燃烧炉排左 1-1 空气流量	km ³ N/h
59	燃烧炉排右 1-1 空气流量	km ³ N/h
60	燃烧炉排左 1-2 空气流量	km ³ N/h
61	燃烧炉排右 1-2 空气流量	km ³ N/h
62	燃烧炉排左 2-1 空气流量	km ³ N/h
63	燃烧炉排右 2-1 空气流量	km ³ N/h
64	燃烧炉排左 2-2 空气流量	km ³ N/h
65	燃烧炉排右 2-2 空气流量	km ³ N/h
66	燃烬炉排左空气流量	km ³ N/h
67	燃烬炉排右空气流量	km ³ N/h
68	二次风流量	km ³ N/h
69	省煤器 No.2 给水流量	t/h
70	省煤器 No.1 给水流量	t/h
71	一级过热器冷却水流量	t/h
72	二级过热器冷却水流量	t/h
73	锅炉出口主蒸汽流量	t/h
74	烟道入口烟气流量	km ³ N/h
75	混合器水流量 A	kg/h
76	混合器水流量 B	kg/h
77	尿素溶剂供应流量	L/h
78	石灰储仓给料量	kg/h
79	活性炭储仓给料量	kg/h
80	一次风机出口空气压力	kPa
81	二次风机出口空气压力	kPa
82	省煤器出口压力	kPa
83	烟道入口烟气压力	Pa
84	汽包压力	MPa
85	三级过热器出口蒸汽压力	MPa
86	FGD 出口烟气压力	Pa
87	布袋压差 A	kPa
88	布袋压差 B	kPa
89	进料器左内侧速度	%
90	进料器左外侧速度	%
91	进料器右内侧速度	%
92	进料器右外侧速度	%
93	干燥炉排左内侧速度	%
94	干燥炉排左外侧速度	%
95	干燥炉排右内侧速度	%
96	干燥炉排右外侧速度	%
97	燃烧炉排 1 左内侧速度	%

表 A1 DXN 数据集的过程变量信息 (续表)

Table A1 The process variable information for DXN datasets (continued table)

序号	过程变量	单位
98	燃烧炉排 1 左外侧速度	%
99	燃烧炉排 1 右内侧速度	%
100	燃烧炉排 1 右外侧速度	%
101	燃烧炉排 2 左内侧速度	%
102	燃烧炉排 2 左外侧速度	%
103	燃烧炉排 2 右内侧速度	%
104	燃烧炉排 2 右外侧速度	%
105	燃烬炉排左侧速度	%
106	燃烬炉排右侧速度	%
107	汽包水位	mm
108	FGD 出口烟气压力-O ₂	%
109	出口烟气分析-O ₂	%
110	出口烟气分析-灰尘	mg/m ³ N
111	出口烟气分析-NO	mg/m ³ N
112	出口烟气分析-SO ₂	mg/m ³ N
113	出口烟气分析-HCL	mg/m ³ N
114	出口烟气分析-CO	mg/m ³ N
115	出口烟气分析-CO ₂	%
116	出口烟气分析-HF	mg/m ³ N

References

- 1 Kammen D M, Sunter D A. City-integrated renewable energy for urban sustainability. *Science*, 2016, **352**(6288): 922-928
- 2 Tobias W, Ludwig K L, Robert B, Esther E, Luca F, Bodo H, et al. Persistence of engineered nanoparticles in a municipal solid-waste incineration plant. *Nature Nanotechnology*, 2012, **7**(8): 520-524
- 3 Qiao Jun-Fei, Guo Zi-Hao, Tang Jian. Dioxin emission concentration measurement approaches for municipal solid wastes incineration process: A survey. *Acta Automatica Sinica*, 2020, **46**(6): 1063-1089
(乔俊飞, 郭子豪, 汤健. 面向城市固废焚烧过程的二噁英排放浓度检测方法综述. *自动化学报*, 2020, **46**(6): 1063-1089)
- 4 Wang B, Wang P L, Xie L H, Lin R B, Lv J, Li J R, et al. A stable zirconium based metal-organic framework for specific recognition of representative polychlorinated dibenzo-p-dioxin molecules. *Nature Communications*, 2019, **10**(1): 1-8
- 5 Wei J X, Li H, Liu J G. Fate of dioxins in a municipal solid waste incinerator with state-of-the-art air pollution control devices in China. *Environmental Pollution*, 2021, **289**: 1-10
- 6 Dai Wei, Lu Wen-Jie, Fu Jun, Ma Xiao-Ping. Multi-rate layered optimal operational control of industrial processes. *Acta Automatica Sinica*, 2019, **45**(10): 1946-1959
(代伟, 陆文捷, 付俊, 马小平. 工业过程多速率分层运行优化控制. *自动化学报*, 2019, **45**(10): 1946-1959)
- 7 Xue W Q, Fan J L, Victor G L, Li J N, Jiang Y, Chai T Y, et al. New methods for optimal operational control of industrial processes using reinforcement learning on two time scales. *IEEE Transactions on Industrial Informatics*, 2020, **16**(5): 3085-3099
- 8 National standard. HJ 77.2-2008. Ambient air and flue gas. Determination of polychlorinated dibenzo-p-dioxins (PCDDs) and polychlorinated dibenzofurans (PCDFs) Isotope dilution HRCC-HRMS. 2008

- (国家标准. HJ 77.2-2008. 环境空气和废气. 二噁英类的测定. 同位素稀释高分辨气相色谱-高分辨质谱法. 2008)
- 9 Li A-Dan, Hong Wei, Wang Jing. Online detection of dioxins and their related substances by laser desorption/laser ionization mass spectrometry. *Journal of Yanshan University*, 2015, **39**(6): 511-515
(李阿丹, 洪伟, 王晶. 激光解吸/激光电离-质谱法二噁英及其关联物的在线检测. 燕山大学学报, 2015, **39**(6): 511-515)
- 10 Nakui H, Koyama H, Takakura A, Watanabe N. Online measurements of low-volatile organic chlorine for dioxin monitoring at municipal waste incinerators. *Chemosphere*, 2011, **85**(2): 151-155
- 11 Zhang H J, Ni Y W, Chen J P, Zhang Q. Influence of variation in the operating conditions on PCDD/F distribution in a full-scale msw incinerator. *Chemosphere*, 2008, **70**(4): 721-730
- 12 Wang Tong, Duan Ze-Wen. Soft sensor modeling for dynamic liquid level of oil well based on fuzzy inference adaptive updating. *CIESC Journal*, 2019, **70**(12): 4760-4769
(王通, 段泽文. 基于模糊评估自适应更新的油井液面软测量建模. 化工学报, 2019, **70**(12): 4760-4769)
- 13 Qin Mei-Hua, Zhu Hong-Qiu, Li Yong-Gang, Chen Jun-Ming, Zhang Feng-Xue, Li Wen-Ting, et al. Soft-sensor method for ion concentration of electrochemical wastewater treatment based on STA-K-means clustering. *CIESC Journal*, 2019, **70**(9): 3458-3464
(秦美华, 朱红求, 李勇刚, 陈俊名, 张凤雪, 李文婷, 等. 基于 STA-K 均值聚类的电化学废水处理过程离子浓度软测量. 化工学报, 2019, **70**(9): 3458-3464)
- 14 Liu Zhuo, Tang Jian, Chai Tian-You, Yu Wen. Selective ensemble modeling approach for mill load parameter forecasting based on multi-modal feature sub-sets. *Acta Automatica Sinica*, 2021, **47**(8): 1921-1931
(刘卓, 汤健, 柴天佑, 余文. 基于多模态特征子集选择性集成建模的磨机负荷参数预测方法. 自动化学报, 2021, **47**(8): 1921-1931)
- 15 Qiao Jing-Hui, Chai Tian-You. Data and model-based soft measurement model of cement raw meal decomposition ratio. *Acta Automatica Sinica*, 2019, **45**(8): 1564-1578
(乔景慧, 柴天佑. 数据与模型驱动的水泥生料分解率软测量模型. 自动化学报, 2019, **45**(8): 1564-1578)
- 16 Chang N B, Chen W C. Prediction of pcdds/pcdfs emissions from municipal incinerators by genetic programming and neural network modeling. *Waste Management & Research*, 2000, **18**(4): 341-351
- 17 Bunsan S, Chen W Y, Chen H W, Chuang Y H, Grisdanurak N. Modeling the dioxin emission of a municipal solid waste incinerator using neural networks. *Chemosphere*, 2013, **92**(3): 258-264
- 18 Xiao Xiao-Dong, Lu Jia-Wei, Hai Jing, Liao Li. Prediction of dioxin emissions in flue gas from waste incineration based on support vector regression. *Renewable Energy Resources*, 2017, **35**(8): 1107-1114
(肖晓东, 卢加伟, 海景, 廖利. 垃圾焚烧烟气中二噁英类浓度的支持向量回归预测. 可再生能源, 2017, **35**(8): 1107-1114)
- 19 Wu J, Yang H. Linear regression-based efficient svm learning for large-scale classification. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26**(10): 2357-2369
- 20 Qiao Jun-Fei, Guo Zi-Hao, Tang Jian. Soft sensing of dioxin emission concentration in solid waste incineration process based on multi-layer feature selection. *Information and Control*, 2021, **50**(1): 75-87
(乔俊飞, 郭子豪, 汤健. 基于多层特征选择的固废焚烧过程二噁英排放浓度软测量. 信息与控制, 2021, **50**(1): 75-87)
- 21 Tang Jian, Qiao Jun-Fei. Dioxin emission concentration soft measuring approach of municipal solid waste incineration based on selective ensemble kernel learning algorithm. *CIESC Journal*, 2019, **70**(2): 696-706
(汤健, 乔俊飞. 基于选择性集成核学习算法的固废焚烧过程二噁英排放浓度软测量. 化工学报, 2019, **70**(2): 696-706)
- 22 Chang N B, Huang S H. Statistical modelling for the prediction and control of PCDDs and PCDFs emissions from municipal solid waste incinerators. *Waste Management & Research*, 1995, **13**(4): 379-400
- 23 Tang Jian, Qiao Jun-Fei, Xu Zhe, Guo Zi-Hao. Soft measuring approach of dioxin emission concentration in municipal solid waste incineration process based on feature reduction and selective ensemble algorithm. *Control Theory & Applications*, 2021, **38**(1): 110-120
(汤健, 乔俊飞, 徐喆, 郭子豪. 基于特征约简与选择性集成算法的城市固废焚烧过程二噁英排放浓度软测量. 控制理论与应用, 2021, **38**(1): 110-120)
- 24 Xia H, Tang J, Qiao J F, Yan A J, Guo Z H. Soft measuring method of dioxin emission concentration for mswi process based on rf and gbdt. In: Proceedings of the Chinese Control and Decision Conference. Hefei, China: IEEE, 2020. 2173-2178
- 25 Duan Yan-Jie, Lv Yi-Sheng, Zhang Jie, Zhao Xue-Liang, Wang Fei-Yue. Deep learning for control: The state of the art and prospects. *Acta Automatica Sinica*, 2016, **42**(5): 643-654
(段艳杰, 吕宜生, 张杰, 赵学亮, 王飞跃. 深度学习在控制领域的研究现状与展望. 自动化学报, 2016, **42**(5): 643-654)
- 26 Zhou Z H, Ji F. Deep forest. *National Science Review*, 2019, **6**: 74-86
- 27 Tang Jian, Xia Heng, Qiao Jun-Fei, Guo Zi-Hao. Modeling method of deep ensemble forest regression with its application. *Journal of Beijing University of Technology*, 2021, **47**(11): 1219-1229
(汤健, 夏恒, 乔俊飞, 郭子豪. 深度集成森林回归建模方法及应用. 北京工业大学学报, 2021, **47**(11): 1219-1229)
- 28 Tang J, Xia H, Zhang J, Qiao J F, Yu W. Deep forest regression based on cross-layer full connection. *Neural Computing and Applications*, 2021, **33**(15): 9307-9328
- 29 Xu W, Tang J, Xia H, Sun Z J. Prediction method of dioxin emission concentration based on PCA and deep forest regression. In: Proceedings of the 40th Chinese Control Conference. Shanghai, China: IEEE, 2021. 1212-1217
- 30 Graybill F A, Meyer C D, Painter R J. Note on the computation of the generalized inverse of a matrix. *Siam Review*, 1966, **8**(4): 522-524
- 31 Chen C L P, Liu Z. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(1): 10-24
- 32 Chen C L P, Liu Z, Feng S. Universal approximation capability of broad learning system and its structural variations. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, **30**(4): 1191-1204
- 33 Fan J C, Wang X, Wang X X, Zhao J H, Liu X. Incremental wishart broad learning system for fast polar image classification. *IEEE Geoscience and Remote Sensing Letters*, 2019, **16**(12): 1854-1858
- 34 Ye H L, Li H, Chen C L P. Adaptive deep cascade broad learning system and its application in image denoising. *IEEE Transactions on Cybernetics*, 2021, **51**(9): 4450-4463
- 35 Chu Y H, Lin H F, Liang Y, Zhang D Y, Diao Y F, Fan X C, et al. Hyperspectral image classification based on discriminative locality preserving broad learning system. *Knowledge-Based Systems*, 2020, **206**: 1-17
- 36 Cheng C, Wang W J, Chen H T, Zhang B C, Shao J J, Teng W X. Enhanced fault diagnosis using broad learning for traction systems in high-speed trains. *IEEE Transactions on Power Electronics*, 2020, **36**(7): 7461-7469
- 37 Pu X K, Li C G. Online semisupervised broad learning system for industrial fault diagnosis. *IEEE Transactions on Industrial*

Informatics, 2021, **17**(10): 6644–6654

- 38 Yu W K, Zhao C H. Broad convolutional neural network based industrial process fault diagnosis with incremental learning capability. *IEEE Transactions on Industrial Electronics*, 2020, **67**(6): 5081–5091
- 39 Sui S, Chen P L C, Tong S C, Feng S. Finite-time adaptive quantized control of stochastic nonlinear systems with input quantization: A broad learning system based identification method. *IEEE Transactions on Industrial Electronics*, 2020, **67**(10): 8555–8565
- 40 Feng J, Yao Y S, Lu S X, Liu Y. Domain knowledge-based deep-broad learning framework for fault diagnosis. *IEEE Transactions on Industrial Electronics*, 2021, **68**(4): 3454–3464
- 41 Han H G, Yang F F, Yang Y H, Wu X L. Type-2 fuzzy broad learning controller for wastewater treatment process. *Neurocomputing*, 2021, **459**(4): 188–200
- 42 Tang H, Dong P, Shi Y. A construction of robust representations for small data sets using broad learning system. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, **51**(10): 6074–6084
- 43 Qi G J, Luo J. Small data challenges in big data era: A survey of recent progress on unsupervised and semi-supervised methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, **44**(4): 2168–2187
- 44 Ying Yu-Xuan, Lin Xiao-Qing, Wu Ang-Jian, Li Xiao-Dong. Review and outlook on municipal solid waste smart incineration. *CIESC Journal*, 2021, **72**(2): 886–900
(应雨轩, 林晓青, 吴昂健, 李晓东. 生活垃圾智慧焚烧的研究现状及展望. *化工学报*, 2021, **72**(2): 886–900)
- 45 Xia H, Tang J, Aljerf L, Wang T Z, Qiao J F, Xu Q D, et al. Investigation on dioxins emission characteristic during complete maintenance operating period of municipal solid waste incineration. *Environmental Pollution*, 2023, **318**: Article No. 120949
- 46 Deng D Y, Qiao J Q, Liu M Q, Dorota K, Zhang M W, Dionysios D D F, et al. Detoxification of municipal solid waste incinerator fly ash by single-mode microwave irradiation: Addition of urea on the degradation of dioxin and mechanism. *Journal of Hazardous Materials*, 2019, **369**: 279–289
- 47 Stanmore B R. The formation of dioxins in combustion systems. *Combustion and Flame*, 2004, **136**(3): 398–427
- 48 Zhang S, Chen Z L, Lin X Q, Wang F, Yan J H. Kinetics and fusion characteristics of municipal solid waste incineration fly ash during thermal treatment. *Fuel*, 2020, **279**: 1–13
- 49 Zhang M M, Alfons G B. De novo synthesis of dioxins: A review. *International Journal of Environment and Pollution*, 2016, **60**: 63–110
- 50 Zhou H, Meng A H, Long Y Q, Li Q H, Zhang Y G. A review of dioxin-related substances during municipal solid waste incineration. *Waste Management*, 2015, **36**: 106–118
- 51 Peng Y Q, Lu S Y, Li X D, Yan J H, Cen K F. Formation, measurement, and control of dioxins from the incineration of municipal solid wastes: Recent advances and perspectives. *Energy & Fuels*, 2020, **34**: 13247–13267
- 52 Ma Y F, Xiao Q L, Chen Z L, Li X D, Lu S Y, Yan J H. Influence factors and mass balance of memory effect on PCDD/F emissions from the full-scale municipal solid waste incineration in China. *Chemosphere*, 2020, **239**: 1–18
- 53 Breiman L. Bagging predictors. *Journal of Machine Learning*, 1996, **24**: 123–140
- 54 Ho T K. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 1998, **20**(8): 832–844
- 55 Blake C L, Merz C J. UCI repository of machine learning databases [Online], available: http://www.archive.ics.uci.edu/ml/k_datasets.html, January 1, 2022



夏恒 北京工业大学信息学部博士研究生。主要研究方向为小样本数据建模, 城市固废焚烧过程二噁英排放预测。

E-mail: xiaheng@emails.bjut.edu.cn
(**XIA Heng** Ph.D. candidate at the Faculty of Information Technology, Beijing University of Technology. His research interest covers small sample data modeling and dioxin emission prediction of the municipal solid waste incineration process.)



汤健 北京工业大学信息学部教授。主要研究方向为小样本数据建模, 城市固废处理过程智能控制。本文通信作者。

E-mail: freeflytang@bjut.edu.cn
(**TANG Jian** Professor at the Faculty of Information Technology, Beijing University of Technology. His research interest covers small sample data modeling and intelligent control of municipal solid waste treatment process. Corresponding author of this paper.)



崔璨麟 北京工业大学信息学部硕士研究生。主要研究方向为城市固废焚烧过程风险预警。E-mail: cuicanlin@emails.bjut.edu.cn

(**CUI Can-Lin** Master student at the Faculty of Information Technology, Beijing University of Technology. His main research interest is risk warning of municipal solid waste incineration process.)



乔俊飞 北京工业大学信息学部教授。主要研究方向为污水处理过程智能控制, 神经网络结构设计与优化。

E-mail: junfeiq@bjut.edu.cn
(**QIAO Jun-Fei** Professor at the Faculty of Information Technology, Beijing University of Technology.

His research interest covers intelligent control of waste water treatment process and structure design & optimization of neural networks.)