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

汤 健^{1,2†}, 乔俊飞^{1,2}, 徐 喆¹, 郭子豪^{1,2}

(1. 北京工业大学 信息学部, 北京 100124;

2. 计算智能与智能系统北京市重点实验室, 北京 100124)

摘要:城市固废焚烧 (MSWI) 排放的污染物二噁英 (DXN) 对生态环境与人类健康具有极大危害,其浓度的实时 检测对实现 MSWI 过程运行优化和城市污染控制至关重要.具有痕量特性的 DXN 排放浓度不能实时检测,机理模 型难以构建,并且与过程变量间的映射关系复杂.针对上述问题,本文提出了一种基于特征约简和选择性集成算法 的 DXN 排放浓度软测量方法.首先,对在线采集的 MSWI 过程变量和离线化验的 DXN 排放浓度数据进行预处理, 获得具有小样本高维特性的建模样本;接着,基于变量投影重要性 (VIP) 值和特征约简比率值确定模型输入特征; 最后,基于操纵训练样本的集成构造策略构建自适应确定核参数的选择性集成模型.采用国外文献和国内工 业 MSWI 过程的 DXN 排放浓度数据仿真验证了所提方法的有效性.

关键词:城市固废焚烧 (MSWI); 二噁英 (DXN); 变量投影重要性 (VIP); 特征约简; 选择性集成算法 引用格式: 汤健, 乔俊飞, 徐喆, 等. 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, 2020, 37(x): xxx - xxx

DOI: 10.7641/CTA.2020.xxxxx

Soft measuring approach of dioxin emission concentration in municipal solid waste incineration process based on feature reduction and selective ensemble algorithm

TANG Jian ^{1,2†}, QIAO Jun-fei^{1,2}, XU Zhe¹, GUO Zi-hao^{1,2}.

Faculty of Information Technology, Beijing University of Technology, Beijing, 100124, China;
 Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China)

Abstract: Municipal solid waste incineration (MSWI) process produces a type of highly toxic and persistent pollutant, i.e., Dioxins (DXN), which has tremendous realistic and potential hazards to the ecological environment and human health. It is very important for optimizing operation of MSWI process and controlling urban pollution in terms of realization of continuous real-time measurement of DXN emission concentration. The generation mechanism of DXN is very complex. Thus, there is a complex non-linear mapping relationship between DXN and input/output variables of MSWI process. Aim at these problems, a soft measuring method of DXN emission concentration based on feature selection and selective ensemble strategy is proposed. Firstly, the process variables of MSWI process and easy-to-measure emission gas data are pretreated to obtain the modeling samples of DXN emission concentration. Then, the variable projection importance (VIP) value based on linear projection to latent structure algorithm and input feature selection ratio based on expert experience are used to select the input features. At last, by using ensemble construction strategy based on the selective ensemble kernel PLS soft measuring model with characteristic of adaptive select kernel parameter is constructed. This method is suitable to construct soft sensor model based on small sample data. The proposed method is simulated and validated by using the data of DXN emission concentration in reference and actual MSWI process.

Key words: Municipal solid waste incineration (MSWI); Dioxin (DXN); variable projection importance (VIP); feature reduction; selective ensemble

Citation: TANG Jian, QIAO Junfei, XU Zhe, et al. Soft measuring approach of dioxin emission concentration in municipal solid waste incineration process based on feature reduction and selective ensemble algorithm. *Control Theory & Applications*, 2020, 37(x): xxx – xxx

收稿日期: xxxx-xx-xx; 录用日期: xxxx-xx-xx.

[†] 通信作者. freeflytang@bjut.edu.cn; Tel.: +86 10 67392443.

本文责任编委:

国家重点研发计划 (2018YFC1900800-5), 国家自然科学基金 (61703089,61803191,61973226), 矿冶过程自动控制技术国家(北京市)重点实验室(BGRIMM-KZSKL-2020-02).

Supported by the National key R & D plan(2018YFC1900800-5), National Natural Science Foundation of China (61703089,61803191,61973226) and National & Beijing Key Laboratory of Process Automation in Mining & Metallurgy(BGRIMM-KZSKL-2020-02).

1 引言

城市固废焚烧 (municipal solid waste incineration, MWSI) 在生活垃圾的无害化、减量化和资源化处理 等方面优势显著. MSWI 过程排放被称为"世纪之毒" 的痕量特性污染物二噁英 (dioxin, DXN)^[1], 其在生物 体内具有显著地积累和放大效应,对生态环境与人类 健康具有巨大危害^[2].目前,DXN 排放浓度还无法在 线实时检测,其产生机理至今仍不清晰,难以建立数 学模型[3]. 通过利用高浓度的单氯苯等关联指示物进 行 DXN 排放浓度的在线测量是目前研究热点之 一[4-8];但该方法所具有的时间滞后性难以满 足 MSWI 过程运行优化与反馈控制的要求,并且关联 模型低精度、检测设备的高复杂性和低性价比等原 因导致该方式难以实际应用.目前工业界多采用具有 周期长、滞后大、成本高等缺点的在线采样与离线 化验相结合的方法[3],即:以月/季周期或不确定周期 由具有资质的检测机构在现场人工采样若干小时、 在实验室离线化验至少一周、单次采样化验费用 为 0.5~1 万元. 显然, 上述方式不能支撑以降低 DXN 排放浓度为直接设定目标的 MSWI 运行优化与反馈 控制.数据驱动软测量技术常用于在线预测依靠离线 化验或专家推断等方式才能获取测量值的难以检测 过程参数^[9,10]. MSWI 过程的炉内温度、炉排速度、 一/二次风量、烟气压力/温度、活性炭量等过程变量 以及 SO₂、HCL 等污染气体的浓度, 以秒为周期基 于分布式控制系统和在线测量系统进行采集;但排放 烟气中的 DXN 浓度却以月/季周期或不确定周期基 于实验室离线化验获得.显然,通过过程变量和 DXN 排放浓度的时序匹配仅能获得少量有标记建模样本. 此外, MSWI 全流程过程变量间存在较强共线性.因 此, DXN 排放浓度的软测量需要面对特征约简和小 样本数据非线性建模等问题.

面向 DXN 排放浓度预测问题, 基于国外机构采 集的关键过程变量, 文献[11] 针对不同类型的焚烧炉 构建线性回归模型; 文献[12] 构建了基于遗传编程的 非线性模型,结果表明预测性能强于多元线性回归和 误差逆向传播神经网络 (back propagation neural network, BPNN) 模型; 为进一步提高模型泛化性能, 文献[13] 构建了基于最小二乘 - 支持向量机 (least squares-support vector machine, LS-SVM) 的预测模 型. 采用国内焚烧过程数据, 文献[14] 提出了采用遗 传算法优化 BPNN 模型的软测量策略, 但该方法所固 有的随机特性导致难以针对小样本数据获得稳定的 预测性能;针对上述问题,文献[15]提出对小样本数 据进行重新抽样和噪声注入以增加样本数量,再构建 基于最大熵神经网络的 DXN 排放浓度预测模型的策 略.上述方法存在难以处理高维特征、面对小样本数 据时预测性能稳定性差等问题.

研究表明,潜结构映射算法 (projection to latent structure, PLS) 及其核版本能够提取线性/非线性潜在 变量构建模型,具有能够消除高维输入特征间的共线 性、降低对建模样本数量要求等优点[16];但存在过多 的输入特征会降低模型的泛化性能和可解释性、适 合建模样本特性的核参数难以有效选择等问题[17].针 对高维近红外谱数据, 文献[18] 提出组合特征选择与 核潜结构映射算法的建模策略,有效提高了模型的预 测性能;但其仅构建了泛化性能有待提升的单一模型, 并不适用于小样本数据. 研究表明, 基于选择性集 成 (selective ensemble, SEN) 算法的软测量模型具有 较佳的泛化性和鲁棒性[19]. 基于"训练样本重采样" 集成构造策略的 SEN 算法验证了集成部分可用候选 子模型可获得比集成全部候选子模型更好的泛化性 能^[20],但所采用的 BPNN 算法并不适合于小样本数 据建模. 面向小样本高维频谱数据, 文献[21] 提出了 综合考虑多源多尺度特征和多工况样本的双层 SEN 潜结构映射建模策略,但该方法所构建软测量模型存 在复杂度高、普适性弱等缺点,同时也未进行维数约 简.因此,将维数约简、核参数自适应选择、模型复 杂度可裁剪等功能为一体的 SEN 策略,应用于 DXN 排放浓度软测量的研究中未见报道.

综上,针对 DXN 排放浓度建模数据所固有的高 维、小样本、共线性和非线性等特性,本文提出了基 于特征约简与 SEN 算法的软测量方法.对预处理后 的建模数据采用变量投影重要性 (variable projection importance, VIP) 和输入特征约简比率进行维数约简, 基于预先给定训练子集数量、候选子模型的结构参 数及候选核参数,构建候选子模型并对其进行评价, 并基于集成子模型选择阈值和加权算法进行集成子 模型的选择与合并,对基于全部候选核参数的 SEN 核潜结构映射模型利用性能最佳准则获得最终 DXN 模型.采用文献和国内某 MSWI 厂多年的 DXN 排放 浓度数据验证了所提方法的有效性.

2 面向 DXN 排放的 MSWI 工艺描述

MWSI包括固废储运、炉内焚烧、蒸汽发电、烟 气处理等阶段,其中:焚烧炉将固废转变为灰烬、烟 气和热量,底部炉排促使固废在燃烧室内移动以便燃 烧更为有效和充分,余热锅炉产生用于发电的蒸汽, 烟气处理设备清除焚烧烟气中的部分污染物.

从产生机理的视角, DXN 排放浓度与炉内焚烧和烟气处理阶段的过程变量和排放至大气中的易检测 气体浓度等相关, 如图 1 所示.

图 1 表明, MSWI 包括 DXN 的产生、吸收和排放共 3 个阶段. 在焚烧炉和余热锅炉的"加热 - 燃烧 - 冷却"阶段, 为保证有机物的有效分解, 通常要求焚烧炉内的烟气达到至少 850°C 并保持 2 秒; 机理研究表

汤健等: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



图 1 面向 DXN 排放的城市固废焚烧 (MSWI) 过程描述

Fig. 1 DXN emissions-based municipal solid waste incineration (MSWI) process description

明,预热区域(20°C~500°C)、高温转换区 域 (800°C~500°C) 和低温转换区域 (500°C~250 °C) 均与 DXN 产生密切相关^[22], 不同温度区域的产 生机理也具有差异性;特别是在烟气冷却阶段,某些 被分解的 DXN 也可能会重新生成.显然,作为 DXN 产生阶段产物的烟气 G1 具有外部可测的最大 DXN 浓度,但目前 MSWI 企业均未对其进行监测. 在烟气 处理阶段,石灰和活性炭被喷射进入反应器以去除酸 性气体和吸附 DXN 及某些重金属, 进而使得 G1 烟 气中的 DXN 被分为两部分: 一部分被吸附后进入飞 灰储仓;另一部分经袋式过滤器过滤后留存于烟 气 G2 中, 通过引风机排入烟囱进而作为烟气 G3 排 放至大气中.因此,理论上 DXN 是与焚烧过程和烟气 处理过程的众多变量,以及 CO、HCL、SO₂、NOx 和HF等气体浓度相关.图1表明, MSWI的过程变 量和其排放的易检测气体浓度以秒为周期在线实时 采集,但 DXN 排放浓度需通过至少1周的时间才能 获得离线化验值.因此,DXN 排放浓度的建模数据是 特性难以描述的小样本高维数据, DXN 生成和吸收 阶段的机理复杂不清,不同阶段过程变量的特性也存 在差异性,难以进行基于机理知识的特征选择.因此, 基于约简特征构建基于"训练样本重采样" 集成构造 策略的 SEN 模型可有效用于 DXN 排放浓度的预测. 进一步,待构建的预测模型可表示为:

$$f_{\text{DXN}}^{\text{emission}}(\cdot) = \sum_{j_{\text{se}}=1}^{J_{\text{sel}}} w_{\text{DXN}}^{j_{\text{sel}}} \cdot f_{\text{DXN}}^{j_{\text{sel}}}(\cdot), \qquad (1)$$

其中, $f_{\text{DXN}}^{j_{\text{sel}}}(\cdot)$ 和 $w_{\text{DXN}}^{j_{\text{sel}}}$ 是第 j_{sel} th 个集成子模型及 其权重系数.

3 软测量策略及实现

本文提出的 DXN 排放浓度软测量策略包括数据 采集与预处理、基于 VIP 的特征约简和基于训练样 本集成构造的 SEN 软测量模型共 3 个模块, 如图 2 所示.

图 2 中, $\{x_n\}_{n=1}^N$ 和 $\{(x_{Sel})_n\}_{n=1}^N$ 表示原始和约 简特征后的模型输入数据; $y = \{y_n\}_{n=1}^N$ 表示模型输 出数据, 即 DXN 排放浓度; $f_{FeSel}(\cdot)$ 表示输入特征约 简模型; ρ_{FeSel} 表示基于经验确定的特征约简比率设 定值; $f_{SEN}(\cdot)$ 表示最终所构建的软测量模型; J 表示 训练子集数量, 也是候选子模型的数量; ρ_{KLV} 是候选 子 模 型 的 结 构 参 数, 即 核 潜 在 变 量 (kernel latent variable, KLV) 的数量; ρ_{SubSel} 是集成子模型选择阈 值设定值; $\{k_{er}^\ell\}_{\ell}^L$ 是核潜结构映射模型的候选核参数 的 预 设 定 集 合, 其 中 L 为 候 选 核 参 数 的 数 量; \hat{y} 为 DXN 模型的预测输出.

进一步,将方法中的学习参数表示为,

$$M_{\text{para}} = \{\rho_{\text{FeSel}}, J, \rho_{\text{KLV}}, \{k_{\text{er}}^{\ell}\}_{\ell}^{L}, \rho_{\text{SubSel}}\}.$$
 (2)

3.1 数据采集与预处理模块

将 DXN 产生阶段、吸收阶段和排放阶段在线采 集的与 DXN 排放浓度相关的过程变量或易检测气体 排放浓度确定为输入特征,结合 DXN 排放浓度的离 线化验值构建建模数据的输入输出样本对,并进行离 群点和缺失值的处理.全部建模样本可表示为:

$$\{\mathbf{X}, y\} = \{\{\boldsymbol{x}_n\}_{n=1}^N, \{y_n\}_{n=1}^N\} = \{(\boldsymbol{x}, y)_n\}_{n=1}^N, \quad (3)$$

其中, $X^{\in N \times M}$ 和 $y^{\in N \times 1}$ 表示用于构建 DXN 排放浓 度模型的输入和输出数据.

3.2 基于 VIP 的特征约简

首先构建基于全部输入特征的线性潜结构映射模型,再计算这些输入特征的 VIP 值并对其进行排序,



图 2 软测量策略

Fig. 2 Soft sensor strategy

最后基于依据经验设定的特征约简比率值选择输入特征. PLS 算法通过最大化输入输出数据间的协方差 在新的潜在变量空间构建具有线性结构的多层回归 模型,其层数即为潜在变量 (latent variable, LV) 的数 量. 算法描述如下所示.

算法 1. 线性 PLS 算法

输入.输出矩阵 X.

输出.输出矩阵 Y.

步骤 1. 标准化矩阵 X 和 Y 为零均值 1 方差;

步骤 2. 令 E0= X, F0= Y 和 h = 1;

步骤 3. 对于每一个潜在变量 (LV) h, 令 $u_h = y_{j_q}$, y_{j_q} 取 Γ_{h-1} 中的某一个值;

步骤 4. 计算矩阵 X 的权重向量 w_h^T =

- $u_h^T E_{h-1}/(u_h^T u_h)$;标准化 $w_h : w_h = w_h/||w_h||$; 步骤 5. 计算矩阵 X 的得分向量: $t_h = E_{h-1}w_h$; 步骤 6. 计算 Y 的载荷向量: $q_h^T =$
- $t_h^{T}F_{h-1}/(t_h^{T}t_h)$,标准化 $q_h : q_h = q_h/||q_h||$; 步骤 7. 计算矩阵 Y 的得分向量: $u_h = F_{h-1}q_h$; 步骤 8. 重复进行 Step3~ Step6 直至收敛. 比 较 step 5 中的 t_h 与上次循环中的值,如相等或误差 在某一范围之内,转至 step9, 否则转至步骤 step4.

步骤 9. 计算矩阵 X 的载荷向量: p_b^{T} =

 $oldsymbol{t}_h^{ ext{T}} ext{E}_{h-1}/(oldsymbol{t}_h^{ ext{T}}oldsymbol{t}_h)$,进行标准化: $oldsymbol{p}_h = oldsymbol{p}_h/\left\|oldsymbol{p}_h
ight\|$;

步骤 10. 计算回归系数: $b_h^T = u_h^T t_{h-1} / (t_h^T t_h)$; 步骤 11. 计算潜变量 h 的残差: $E_h =$

 $\mathbf{E}_{h-1} - \boldsymbol{t}_h \boldsymbol{p}_h^{\mathrm{T}}$, $\mathbf{F}_h = \mathbf{F}_{h-1} - b_h \boldsymbol{t}_h \boldsymbol{q}_h^{\mathrm{T}}$;

步骤 12. 令 *h* = *h* + 1, 返回 Step3 直到所有的 潜变量 LV 计算完毕. 由以上算法描述可知,线性 PLS 算法的外部模型 用于提取和输入/输出空间均相关的 LV 潜在变量,内 部模型通过这些 LV 构建回归模型,其可表示为:

$$\begin{cases} \hat{\mathbf{Y}} = \mathbf{X}\mathbf{B} + \mathbf{G} \\ \mathbf{B} = \mathbf{X}^{\mathrm{T}}\mathbf{U}(\mathbf{T}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{U})^{-1}\mathbf{T}^{\mathrm{T}} \end{cases}$$
(4)

其中,G是未建模动态.

进一步,将基于{X,y}构建的模型表示为{T,W,P,B,Q},如下所示:

$$\{\mathbf{X}, \boldsymbol{y}\} \rightarrow \{\mathbf{T}, \mathbf{W}, \mathbf{P}, \mathbf{B}, \mathbf{Q}\},$$
 (5)

其中, T = [t_1 , …, t_h , …, t_H], W = [w_1 , …, w_h , …, w_H], P = [p_1 , …, p_h , …, p_H], B = diag{ b_1 , …, b_h , …, b_H }和Q = [q_1 , …, q_h , …, q_H]表示与算法1中的向量所相应的矩阵; H为LV的数量.

VIP 指标能够表征每个输入特征对模型性能的影响,以第特征为例,其计算公式如下:

$$VIP_m = \sqrt{\frac{\sum_{h=1}^{H} w_{mh}^2 \cdot (\boldsymbol{b}_h^2 \boldsymbol{t}_h^{\mathrm{T}} \boldsymbol{t}_h) \cdot M}{\boldsymbol{b}_h^2 \mathrm{T}^{\mathrm{T}} \mathrm{T} \cdot H}}, \qquad (6)$$

其中, M 是特征数量, w_{mh} 表示第 hth LV 中第 mth 特征的权重, 即存在:

$$\boldsymbol{w}_h = [w_{1h}, \cdots, w_{mh}, \cdots, w_{Mh}]. \tag{7}$$

依据 VIP 值由大到小进行特征排序,并将标记 为 $[x_1, \dots, x_{p^*}, \dots, x_{P^*}]$,其中 P^* 表示排序后最后 特征的编号,其值为 M.

依据经验设定特征约简比率 ρ_{FeSel} 后,基于下式

确定输入特征数量,

$$P_{\rm sel} = f_{\rm int} (P^* \cdot \rho_{\rm FeSel}), \tag{8}$$

其中, $0 < \rho_{\text{FeSel}} \leq 1$, $f_{\text{int}}(\cdot) = \lfloor \cdot \rfloor$ 表示取整函数.

选择新排序特征 $[x_1, \dots, x_{p^*}, \dots, x_{P^*}]$ 的前 P_{sel} 个作为约简后的特征,其记为,

$$\boldsymbol{x}_{\text{Sel}} = [x_1, \cdots, x_{p^*}, \cdots, x_{P_{\text{sel}}}]. \tag{9}$$

特征约简模型可记为 $f_{\text{FeSel}}(\cdot)$, 即:

$$\begin{aligned} \mathbf{X}_{\mathrm{Sel}} &= \{(\boldsymbol{x}_{\mathrm{Sel}})_n\}_{n=1}^N \\ &= f_{\mathrm{FeSel}}(\{\boldsymbol{x}_n\}_{n=1}^N, VIP, \rho_{\mathrm{FeSel}}), \end{aligned} \tag{10}$$

其中, X^{€N×Psel} 为约简特征后的输入数据.

此 处, 将 特 征 约 简 后 的 模 型 输 入 样 本 记 为 $\{X_{Sel}, y\} = \{(x_{Sel}, y)_n\}_{n=1}^N$.

3.3 基于训练样本集成构造的 SEN 软测量模型

首先次采用操纵训练样本的集成构造策略产生训 练样本子集并基于候选核参数和结构参数值构建候 选子模型,接着对候选子模型进行评价并基于选择阈 值获得集成子模型,然后对基于相同核参数的集成子 模型进行加权以获得模型输出,最后基于预测性能获 得最终的 DXN 软测量模型.主要步骤描述如下.

(1) 训练子集构造

基于"训练样本采样"方式从训练样本 $\{(\boldsymbol{x}_{Sel}, y)_n\}_{n=1}^N$ 中产生数量为J的训练样本子集,即集成构造,可表示为:

$$\{ (\boldsymbol{x}_{\text{Sel}}, y)_n \}_{n=1}^N \\ J \\ \} \Rightarrow \begin{cases} \{ (\boldsymbol{x}_{\text{Sel}}^1, y^1)_n \}_{n=1}^N = \{ X_{\text{Sel}}^1, \boldsymbol{y}^1 \} \\ \dots \\ \{ (\boldsymbol{x}_{\text{Sel}}^j, y^j)_n \}_{n=1}^N = \{ X_{\text{Sel}}^J, \boldsymbol{y}^j \}, \\ \dots \\ \{ (\boldsymbol{x}_{\text{Sel}}^J, y^J)_n \}_{n=1}^N = \{ X_{\text{Sel}}^J, \boldsymbol{y}^J \} \end{cases}$$

其中, $\{X_{\text{Sel}}^{j}, \boldsymbol{y}^{j}\}$ 表示第 *j*th 训练样本子集.

(2) 候选子模型构建

采用核 PLS 算法构建基于上述训练样本子集的 DXN 排放浓度候选子模型. 依据先验选择的 L 个候选核参数,标记为 $\{k_{\rm er}^\ell\}_\ell^L$.将表征候选子模型复杂度的结构参数,即核潜在变量 (KLV),标记为 $H_{\rm KLV}$.

此处以第 jth 训练样本子集 $\{X_{Sel}^{j}, y^{j}\}$ 为例描述 候选子模型的构建过程.

首先将输入数据 X_{Sel}^{j} 基于第 ℓ th 核参数 k_{er}^{ℓ} 映射 至高维空间获得 $K^{\ell j}$, 按下式进行处理,

$$\tilde{\mathbf{K}}^{\ell j} = (\mathbf{I} - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^{\mathrm{T}}) \mathbf{K}^{\ell j} (\mathbf{I} - \frac{1}{k} \mathbf{1}_N \mathbf{1}_N^{\mathrm{T}}), \quad (12)$$

其中, I 是 N 维的单位矩阵, 1_N 是值为 1 长度为 N 的向量.

接着,通过如下所示的算法构建候选子模型.

算法 2. 采用核 PLS 算法构建基于 k_{er}^{ℓ} 、 H_{KLV} 和 $\{X_{\text{Sel}}^{j}, \boldsymbol{y}^{j}\}$ 的候选子模型

步骤 0. 记 H_{KLV} 为可获得的全部 KLVs 的数量. 重复 $h_{\text{KLV}} = 1$ 到 H_{KLV} . 步骤 1. 令 $h_{\text{KLV}} = 1$, $\tilde{K}_{h_{\text{KLV}}}^{\ell j} = \tilde{K}^{\ell j}$, $y_{h_{\text{KLV}}}^{j} = y^{j}$; 步骤 2. 随机初始化 $[u'_{h_{\text{KLV}}} 等于 y_{h_{\text{KLV}}}^{j}$ 中的任 何一列;

步骤 3. 计算得分向量: *t*′_{*h*_{KLV}} =

 $ilde{\mathbf{K}}_{h_{\mathrm{KLV}}}^{\ell j} \boldsymbol{u'}_{h_{\mathrm{KLV}}}, \boldsymbol{t'}_{h_{\mathrm{KLV}}} \leftarrow \boldsymbol{t'}_{h_{\mathrm{KLV}}} / ||\boldsymbol{t'}_{h_{\mathrm{KLV}}}||;$ 步骤 4. 计算载荷向量: $\boldsymbol{c'}_{h_{\mathrm{KLV}}} = (\boldsymbol{y}_{h_{\mathrm{KLV}}}^{j})^{\mathrm{T}} \boldsymbol{t'}_{h_{\mathrm{KLV}}}$:

步骤 5. 计算得分向量: **u'**_{hĸĿv} =

 $oldsymbol{y}_{h_{ ext{KLV}}}^{j}oldsymbol{c}'_{h_{ ext{KLV}}},oldsymbol{c}'_{h_{ ext{KLV}}} \leftarrow oldsymbol{c}'_{h_{ ext{KLV}}}/||oldsymbol{c}'_{h_{ ext{KLV}}}||\,;$

步骤 6. 如果 t'_{hKLV} 收敛,转到 Step (7); 否则转 到 Step (3);

步骤 7. 按下式计算残差: $\tilde{K}_{h_{\text{KLV}}}^{\ell j} \leftarrow \left(I - \boldsymbol{t'}_{h_{\text{KLV}}}(\boldsymbol{t'}_{h_{\text{KLV}}})^{\text{T}}\right) \tilde{K}_{h_{\text{KLV}}}^{\ell j} \left(I - \boldsymbol{t'}_{h_{\text{KLV}}}(\boldsymbol{t'}_{h_{\text{KLV}}})^{\text{T}}\right),$ $\boldsymbol{y}_{h_{\text{KLV}}}^{j} \leftarrow \boldsymbol{y}_{h_{\text{KLV}}}^{j} - \boldsymbol{t'}_{h_{\text{KLV}}}(\boldsymbol{t'}_{h_{\text{KLV}}})^{\text{T}} \boldsymbol{y}_{h_{\text{KLV}}}^{j};$ 步骤 8. 令 $h_{\text{KLV}} = h_{\text{KLV}} + 1$, 如果 $h_{\text{KLV}} \geq$

少寐 \mathfrak{o} 、 $\forall n_{\text{KLV}} = n_{\text{KLV}} + 1$,如未 $n_{\text{KLV}} \ge H_{\text{KLV}}$ 终止;否则转到 Step (2).

采用上述算法, 基于 k_{er}^{ℓ} 构建的候选子模型集合 可表示为:

$$\begin{cases} \{\mathbf{X}_{\text{Sel}}^{j}, \boldsymbol{y}^{j}\}_{j=1}^{J} \\ k_{\text{er}}^{\ell} \\ H_{\text{KLV}} \end{cases} \Rightarrow \{f_{\text{can}}^{\ell j}(\cdot)\}_{j=1}^{J} = S_{\text{can}}^{\ell}, \quad (13)$$

其中, $f_{can}^{\ell j}(\cdot)$ 表示基于 k_{er}^{ℓ} 的第 *j*th的候选子模型集合.

最后, 基于
$$f_{can}^{\ell j}(\cdot)$$
的预测输出为,

$$\hat{\boldsymbol{y}}^{j} = \mathrm{K}^{\ell}\mathrm{U}'\left(\mathrm{T}'\mathrm{K}^{\ell}\mathrm{U}'\right)^{-1}\mathrm{T}'^{1}\boldsymbol{y}^{j},\qquad(14)$$

$$\tilde{\mathbf{K}}^{\ell} = (\mathbf{K}^{\ell}\mathbf{I} - \frac{1}{N}\mathbf{1}_{N}\mathbf{1}_{N}^{\mathrm{T}}\tilde{\mathbf{K}}^{\ell j})(\mathbf{I} - \frac{1}{N}\mathbf{1}_{N}\mathbf{1}_{N}^{\mathrm{T}}), \quad (15)$$

其 中, U' = [$\boldsymbol{u}'_1, \cdots, \boldsymbol{u}'_{h_{\text{KLV}}}, \cdots, \boldsymbol{u}'_{H_{\text{KLV}}}$], T' = [$\boldsymbol{t}'_1, \cdots, \boldsymbol{t}'_{h_{\text{KLV}}}, \cdots, \boldsymbol{t}'_{H_{\text{KLV}}}$],为原始训练样本的核矩阵.

(3) 候选子模型评价

对基于 k_{er}^{ℓ} 的 J 个候选子模型的输出进行合并以构成全集成模型, 其第 nth 样本的输出为:

$$\bar{f}_{EnAll}^{\ell}(\boldsymbol{x}_{sel}^{n}) = \sum_{j=1}^{J} \alpha_{j}^{\ell} f_{can}^{\ell j}(\boldsymbol{x}_{sel}^{n}), \qquad (16)$$

其中, α_i^{ℓ} 是第 *j*th 理想权重系数.

候选子模型和全集成模型的学习误差分别采用以 下两式计算:

$$E_j^{\ell} = \frac{1}{N} \sum_{n=1}^{N} \left(f_{\text{can}}^{\ell j}(x_{\text{sel}}^n) - y^n \right)^2, \qquad (17)$$

$$E_{\rm En}^{\ell} = \frac{1}{N} \sum_{n=1}^{N} \left(f_{\rm EnALL}(x_{\rm sel}^n) - y^n \right)^2, \qquad (18)$$

此处定义候选子模型输出与真值间的偏差函数 为 $m_{\text{can}}^{\ell_j}(x_{\text{sel}}^n) \equiv y^n - f_{\text{can}}^{\ell_j}(x_{\text{sel}}^n)$,其均方误差可记
 6
 控制理论与应用

 为 $MSE[f_{can}^{\ell j}] = E[(m_{can}^{\ell j})^2]$. 进一步,式(16)可改 系数,
 写为:

$$\bar{f}_{\text{EnAll}}^{\ell}(\boldsymbol{x}_{\text{sel}}^{n}) = \sum_{j=1}^{J} \alpha_{j}^{\ell} f_{\text{can}}^{\ell j}(\boldsymbol{x}_{\text{sel}}^{n}) \\
= y^{n} + \sum_{j=1}^{J} \alpha_{j}^{\ell} m_{\text{can}}^{\ell j}(\boldsymbol{x}_{\text{sel}}^{n})$$
(19)

定义候选子模型误差函数间的相关系数矩阵 为 $C_{is}^{\ell} = E[m_{can}^{\ell j} m_{can}^{\ell s}]$,其中s = 1, ..., J.理论上, 通过最小化 $\bar{f}_{EnAll}^{\ell}(\boldsymbol{x}_{sel}^n)$, 可求得 α_i^{ℓ} ,

$$\boldsymbol{\alpha}_{\rm opt}^{\ell} = \arg\min\left(\mathrm{MSE}[\bar{f}_{\rm EnAll}^{\ell}(\boldsymbol{x}_{\rm sel}^{n})]\right) \\ = \arg\min\left(\sum_{j,s} \alpha_{j}^{\ell} \alpha_{s}^{\ell} C_{js}^{\ell}\right) \qquad (20)$$

基于简化约束 $\sum_{j=1}^{J} \alpha_j^{\ell} = 1$,采用拉格朗日乘子法 求解上式,可获得理想权重向量的解 析 解 $\boldsymbol{\alpha}_{\text{opt}}^{\ell} = \{\boldsymbol{\alpha}_{\text{opt},j*}^{\ell}\}_{j*=1}^{J}$,其 中 $\boldsymbol{\alpha}_{\text{opt},j*}^{\ell}$ 表 示 第 j*th值. 上述方法的缺点是需要求解误差函数相关 系数矩阵的逆.

由于不同候选子模型间的存在相关性,为避免求 逆过程的不稳定,此处采用文献 [20] 所提方法:首先 针对每个候选子模型赋予随机权重向量,然后采用遗 传算法对这些权重进行优化,目标是使得全集成模型 的适应度(即预测性能)最佳,进而获得每个候选子模 型的优化权重向量 $\{\boldsymbol{\alpha}_{\text{opt},i*}^{\ell}\}_{i*=1}^{J}$.

(4) 集成子模型选择

为简化模型结构和提升预测性能,针对基于 k^ℓ_{ar} 的候选子模型,选择优化权重值大于 ρ_{SubSel} 的作为 集成子模型,将其数量记为 J^ℓ_{sel},即集成模型尺寸 为 J_{sel}^{ℓ} . 相应的, 将第 j_{sel}^{ℓ} th 集成子模型的输出记为:

$$\hat{\boldsymbol{y}}_{\text{sel}}^{j_{\text{sel}}^{\ell}} = f_{\text{sel}}^{j_{\text{sel}}^{\ell}}(\mathbf{X}^{\text{Sel}}), \qquad (21)$$

其中, $j_{\text{sel}}^{\ell} = 1, 2, \cdots, J_{\text{sel}}^{\ell}$.

进一步,将集成子模型的输出表示为:

$$\hat{\mathbf{Y}}^{\ell} = [\hat{\boldsymbol{y}}^{1_{\text{sel}}^{\ell}}, \cdots, \hat{\boldsymbol{y}}^{j_{\text{sel}}^{\ell}}, \cdots, \hat{\boldsymbol{y}}^{J_{\text{sel}}^{\ell}}], \qquad (22)$$

进而,集成子模型的集合可表示为:

$$S_{\rm sel}^{J_{\rm sel}^{\ell}} = \{ f_{\rm sel}^{j_{\rm sel}^{\ell}}(\cdot) \}_{j_{\rm sel}^{\ell}=1}^{J_{\rm sel}^{\ell}},$$
(23)

由上可知,集成子模型和候选子模型间的关系 为 $S_{\text{sel}}^{J_{\text{sel}}^{\ell}} \in S_{\text{can}}^{\ell}$,即前者为后者的子集.

(5) 集成子模型合并

基于 k_{ar}^{ℓ} 的 SEN 模型 $f_{SEN}^{J_{sel}^{\ell}}(\cdot)$ 的输出为:

$$\hat{\boldsymbol{y}}^{\ell} = f_{\text{SEN}}^{J_{\text{sel}}^{\ell}}(\cdot) = \sum_{j_{\text{sel}}^{\ell}=1}^{J_{\text{sel}}} w^{j_{\text{sel}}^{\ell}} \hat{\boldsymbol{y}}^{j_{\text{sel}}^{\ell}}, \qquad (24)$$

其中, w^{j^ℓ_{sel}} 为采用自适应加权融合算法计算得到的权

$$w^{j_{\text{sel}}^{\ell}} = \frac{1}{\left(\sigma^{j_{\text{sel}}^{\ell}}\right)^2 \sum_{j^{\ell} = -1}^{J_{\text{sel}}^{\ell}} \frac{1}{\left(\sigma^{j_{\text{sel}}^{\ell}}\right)^2}},$$
(25)

其中, $\sigma^{j_{sel}^\ell}$ 是基于 k_{er}^ℓ 的集成子模型的预测输 出 $\hat{\boldsymbol{y}}^{j_{sel}^{\ell}}$ 的标准差.

(6) 软测量模型选择

重复上述过程,构建基于 $\{k_{\rm er}^\ell\}_{\ell}^L$ 的全部 SEN 模 型,并记为 $\{f_{SEN}^{J_{sel}^{\ell}}(\cdot)\}_{\ell=1}^{L}$. 采用如下公式自适应选择 具有最优泛化性能的 SEN 模型作为最终 DXN 预测 模型 $f_{\text{SEN}}(\cdot)$,即

$$\hat{y} = f_{\text{SEN}}(\hat{y}^1, \cdots, \hat{y}^\ell, \cdots, \hat{y}^L),$$
 (26)

$$s.t. \begin{cases} \min\{\{RMSE(\hat{y}^{\ell})\}_{\ell=1}^{L}\} \\ RMSE(\hat{y}^{\ell}) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{y}^{\ell})^{n} - y^{n})^{2}} \\ \hat{y}^{\ell} = \{(\hat{y}^{\ell})^{n}\}_{n=1}^{N} = f_{\text{SEN}}^{J_{\text{sel}}^{\ell}}(X_{\text{Sel}}) \end{cases}$$
(27)

其中, $RMSE(\hat{y}^{\ell})$ 表示基于 k_{er}^{ℓ} 的 $f_{SEN}^{J_{sel}^{\ell}}(\cdot)$ 预测输出 的均方根误差, $(\hat{y}^{\ell})^n$ 表示 $f_{SEN}^{J_{sel}^{\ell}}(\cdot)$ 对第 nth 样本的 预测输出.

上述 DXN 预测模型的构建如图 3 所示.





Fig. 3 SEN soft sensor model construction process based on ensemble construction by using training samples

4 实验验证

4.1 国外文献数据验证

(1)数据描述

利用水冷壁焚化炉数据^[12-13]构建 DXN 排放浓 度模型,其输入特征按编号从1到8依次为:蒸汽负

第 x 期

竺 1 公

荷 (tone/h)、 烟 气 中 H₂O 含 量 (%)、 烟 道 温 度 (°C)、 烟 气 流 量 (Nm³/min)、CO 浓 度 (ppmv)、HCl 浓 度 (ppml)、 颗 类 物 (PM) 浓 度 (mg/Nm³) 和燃烧室上方温度 (°C) (注: 文献中对 该数据集无更为详细的描述). 此处将 28 个样本 的 70% 和 30% 分别作为建模数据和测试数据.

(2)实验结果

首先,构建 PLS 模型,其全部 8 个 LV 的方差贡 献率如表 1 所示.

表1 文献 DXN 数据基于 PLS 模型的方差贡献率统 计表

 Table 1
 Statistical table of variance contribution rate

 based on PLS model for ref DXN data

LV编号	输	入数据	输出数据			
	贡献率(%)	累积贡献率(%)	贡献率(%)	累积贡献率(%)		
1	64.42	64.42	30.26	30.26		
2	14.04	78.46	31.02	61.28		
3	10.87	89.34	12.89	74.17		
4	4.99	94.33	1.40	75.57		
5	2.20	96.53	0.57	76.14		
6	2.98	99.50	0.29	76.43		
7	0.43	99.93	0.57	77.00		
8	0.07	100.00	0.09	77.08		

如表1所示,全部LV提取的输入和输出数据累积方差贡献率为100%和77.08%,表明该文献选择的输入特征是合理的.



接着,计算输入特征的 VIP 值, 如图 4 所示.

Fig. 4 VIP value of input feature for ref DXN data

图 4 表明, 全部 8 个特征的 VIP 值均大于 1, 其中 第 5 个特征(CO 浓度)具有最大值, 其他特征按 VIP 值依次排序为:第3个(烟道温度)、第1个(蒸汽负荷)、第7个(PM浓度)、第2个(烟气中H₂O含量)、第4个(烟气流量)、第8个(燃烧室上方温度)、第6个(燃烧室上方温度).

接着,设定J=20, $\rho_{\text{SubSel}}=0.05$, $m_{\text{SubCom}}=1$, $\{k_{\text{er}}^{\ell}\}_{\ell}^{L}$ =[0.1,0.3,0.5,0.7,0.9,1,3,5,7,9,10,30,50,70,100, 300, 500,700,900,1000]. 采用遗传算法工具箱确定不同候选子模型的理想权重. 同时,考虑到遗传算法存在的随机性,基于每个候选参数的候选子模型均运行20次进行均值统计.

最后,考虑 ρ_{FeSel} 和 ρ_{KLV} 是影响模型输入特征和 集成子模型结构的关键参数,采用网格搜索法进行选 择.结合图 4 可知,在取值为 1、0.8、0.6、0.4 和 0.2 时 所 选 择 的 输 入 特 征 编 号 的 子 集 分 别 为 5、3、1、7、2、4、8、6、5、3、1、7、2、4、 5、3、1、5、3、5. 设定 ρ_{KLV} 的范围为 1-10,其最 大值高于原始输入特征数量原因在于核技术扩展了 变量维数.上述两个参数与模型预测性能间的关系如 图 5 所示.

由图 5 可知: 针对训练数据, $\rho_{\text{FeSel}} = 1$ 时的平均预 测误差稍弱于 $\rho_{\text{FeSel}} = 0.8$; 但是针对测试数据, $\rho_{\text{FeSel}} = 1$ 且 KLV=7 具有最佳平均和最小预测误差; 针对训 练和测试数据, 选择较少的输入特征建模时预测性能 都具有较大的跳跃性, 这表明该算法针对小样本低维 数 据 存 在 较 大 的 随 机 性. 可 见, 适 当 选 择 ρ_{FeSel} 和 ρ_{KLV} 的值是非常必要的.

(3)方法比较

本文方法与文献,以及线性 PLS 基准方法等进行 比较,结果如表 2 所示.

由表 2 可知:本文方法在 ρ_{FeSel} =1 和 KLV=7 时, 具有最佳预测性能,其训练数据和测试数据的平 均 RMSE 分别为 23.79 和 78.30,较文献 [12] 的 GP 模型在精度上至少提高1倍多,较文献 [13] 的 LS-SVM 模型和 PLS 基准方法也有较大提高;此 外,测试数据的预测范围波动比较大,其最小平均预 测误差值仅为 70.98.上述结果表明,选择适当的输入 特征和模型参数是非常关键的.

上述研究表明,本文方法是有效的.基于我 国 MSWI 过程的 DXN 排放数据构建软测量模型将 更具有实际意义.



Fig. 5 Relationship between ρ_{FeSel} and ρ_{KLV} and predicted performance for ref DXN data modeling

Tuest 2 Companies of president of the first of different fiberious for fer Drift duta										
方法		RMSE(训练数据)			RMS	备注				
		最大值	均值	最小值	最大值	均值	最小值			
文献 [12]		_	110.6	_	_	182.6	_	GP		
文献 [13]		—	101.2	_	_	79.8	—	LS-SVM		
线性 PLS	$ ho_{\mathrm{FeSel}}=1$	—	90.91	_	_	155.7	_	LV=3		
	$ ho_{\mathrm{FeSel}}=0.8$	—	94.85	_	_	161.4	_	LV=3		
	$ ho_{\mathrm{FeSel}}=0.6$	—	103.2	_	_	128.7	_	LV=3		
本文	$ ho_{\mathrm{FeSel}}=1$	30.49	23.79	14.57	88.19	78.3	70.98	KLV=7		
	$ ho_{\mathrm{FeSel}}=0.8$	69.02	61.06	51.62	128.3	110.2	94.96	KLV=4		
	$\rho_{\rm FeSel}{=}0.6$	80.62	67.44	55.85	277	153.4	92.5	KLV=8		

	表 2 文献 DXN 数据不同方法的预测误差比较结果
Table 2	Comparison of prediction errors of different methods for ref DXN data

4.2 国内工业数据验证

(1)数据描述

本文数据源于北京某 MSWI 焚烧企业,涵盖 了 20122018 年测试的 DXN 排放浓度检测样本 34 组,其中: DXN 取值为 3 次采样化验的均值,过程 变量取值为采样当天 24 小时的均值(包含有效过程 变量的数量为 287);训练和测试样本等分,各为 17 组.

(2)实验结果

首先,构建 PLS 模型,其前 10 个 LV 的方差贡 献率如表 3 所示. 表 3 国内 DXN 数据基于 PLS 模型的方差贡献率 统计表

 Table 3 Statistical table of variance contribution rate

 based on PLS model for China DXN data

LV	输	入数据	输出数据			
编号	贡献率 (%)	累积贡献率 (%)	贡献率 (%)	累积贡献率 (%)		
1	31.73	31.73	34.33	34.33		
2	22.59	54.32	29.85	64.18		
3	7.33	61.64	22.23	86.41		
4	8.99	70.63	5.49	91.9		
5	3.26	73.89	6.32	98.22		
6	3.19	77.08	1.58	99.8		
7	4.48	81.56	0.14	99.94		
8	3.52	85.08	0.04	99.98		
9	2.72	87.8	0.02	100		
10	2.46	90.25	0	100		

如表 3 所示,前 10 个 LV 所提取的输入和输出 数据的方差累积贡献率为分别为 90.25% 和 100%, 表明较少 LV 能够蕴含建模数据中的多数变化.

接着,计算全部输入特征的 VIP 值,如图 6 所示.





图 6 表 明, 输入特征 VIP 值 的 最小值为 0. 9230e-4、均值为 0.1999、最大值为 0.7051, 可见 输入特征间具有较大差异性, 其中排在前 10 的过 程变量是:燃烬炉排左侧速度、燃烬炉排右侧速 度、汽包炉水电导率、燃烧炉排右空气流量、炉墙 左侧外温度、发电机前轴承振动、给水泵出口流 量、烟囱排放 CO 浓度、给水泵出口给水压力、二 次燃烧室左侧温度,取值范围为 0.7051-0.5034;其 中,除"汽包炉水电导率"和"汽包炉水电导率"外, 均与 DXN 的生成、吸收和排放阶段直接相关,这 表明了国内 DXN 数据的可用性,以及采用 VIP 选 择输入特征的合理性.

接着,采用与文献DXN数据相同的J 和 $\{k_{er}^{\ell}\}_{\ell}^{L}$ 参数运行20次进行统计;取 ρ_{KLV} =112; 结合图6设定 ρ_{FeSel} 的取值为0.01,0.03,0.05,0.08, 0.1,其所对应的输入特征数量分别为2,8,14,22,28, 采用网格寻优法获取 ρ_{FeSel} 和 ρ_{KLV} 值并分析其与 模型预测性能间的关系.上述参数与模型预测性能 间的关系如图7所示.

由图7可知:针对训练数据,在 ρ_{FeSel} =0.1 和KLV=9时具有最佳的平均预测性能;针对测试 数据, ρ_{FeSel} =0.1和KLV=1时具有最佳预测性能, 其RMSE的均值为0.01792.核参数与模型预测性 能间的关系如图8所示.



图 8 国内 DXN 数据建模时核参数与预测性能间的关系 Fig. 8 VIP value of input feature for China DXN data

由图 8 可知,核参数的取值为 300.因此,选择 适当的输入特征和学习参数很必要.

(3)方法比较

此处与基准线性 PLS/核 PLS 方法进行比较,结果如表 4 所示.



图 7 国内 DXN 数据建模时 ho_{FeSel} 和 ho_{KLV} 与预测性能间的关系

Fig. '	7	Relationship	between ρ_{FeSel}	and p	$ ho_{\mathrm{KLV}}$ and	d predicted	l performance	for Chi	na DXN	data modeling
--------	---	--------------	-------------------------------	-------	--------------------------	-------------	---------------	---------	--------	---------------

r · · · · · · · · · · · · · · · · · · ·										
方法		RMSE(训练数据)			RM	ISE(测试数	备注			
		最大值	均值	最小值	最大值	均值	最小值			
线性 PLS	1	_	0.0233	_	_	0.02536	_	LV=1		
	0.1		0.01619			0.01846		LV=1		
核 PLS	0.1	_	0.01543	_		0.01862	-	Ker=3, KLV=1		
本文	0.1	0.01537	0.01384	0.01243	0.01994	0.01797	0.01643	KLV=1		
	0.08	0.01713	0.01524	0.01354	0.02153	0.01912	0.01783	KLV=1		
	0.05	0.01943	0,01823	0.01752	0.02182	0.02124	0.02023	KLV=1		
	0.03	0.02142	0.01952	0.01826	0.02261	0.02073	0.01991	KLV=1		

表 4 国内 DXN 数据不同方法的预测误差比较结果 Table 4 Comparison of prediction errors of different methods for China DXN data

由表 4 可知: (1) 在选择 28 个输入特征 (ρ_{FeSel} =0.1) 和 KLV=1 时,本文所提方法具有最佳预测性能,其最小 RMSE 为 0.01643,其维数与原输入特征相比降低了近 10 倍,表明特征约简和 SEN 建模策略的有效性; (2) 采用线性 PLS 和核 PLS 建模时,在 LV 和 KLV 数量为 1 时具有最佳预测性能,其第 1 个 KLV 所提取的输入和输出数据的方差贡献率分别为 56.76% 和 71.76%,比 PLS 方法提高了 20%-30%;但针对预测性能,却只是提升训练样本而未提升测试样本,这与特征约简的 VIP 值为线性和建模样本的数量稀少相关; (3) 所提方法的最小预测误差较核 PLS 方法在预测性能上具有较大提升,表明 SEN 策略是有效的.

综上可知,本文所提方法在如何进行非线性特 征选择、如何同时优化特征约简与模型参数选择机 制、如何降低模型的随机性等方面的研究还有待于 深入进行.

5 结论

本文提出了一种新的基于特征约简和选择性集成算法的二噁英排放浓度软测量方法,其主要创新 点是:基于线性潜结构映射模型获取的变量投影重 要性以及依经验设定的特征约简比率因子确定输入 特征;基于"训练样本重采样"的集成构造策略构建 核参数自适应选择和模型复杂度依需求设定的选择 性集成核潜结构映射模型.采用文献和国内 的 DXN 排放浓度数据验证了所提方法的有效性. 第 x 期

进一步的研究包括非线性特征选择、特征约简 与模型参数优化机制、以及如何进行分阶段多相态 二噁英浓度预测,以更好的为固废焚烧过程的运行 控制提供支撑.

参考文献:

- HOYOS A, COBO M, ARISTIZA B. Total suspended particulate (TSP), polychlorinated dibenzodioxin (PCDD) and polychlorinated dibenzofuran (PCDF) emissions from medical waste incinerators in Antioquia, Colombia. *Chemosphere*, 2008, 73(1): 137-42.
- [2] LUO Aqun, LIU Shaoguang, LIN ensong, et al. Progress of formation mechanisms and emission reduction methods of PCDD/Fs. *Chemical Industry and Engineering Progress*, 2016, 35(3): 910-916.
 (罗阿群, 刘少光, 林文松, 等. 二噁英生成机理及减排方法研究进展. 化工进展, 2016, 35(3): 910-916.)
- [3] QIAO Junfei, GUO Zihao, 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] YIN Xuefeng, LI Xiaodong, LU Shengyong. On-line real-time monitoring of trace organic pollutant formation in the simulated flue gas. *Proceedings of the CSEE*, 2007, 27(17): 29-33.
 (尹雪峰,李晓东,陆胜勇. 模拟烟气中痕量有机污染物生成的在线 实时监测. 中国电机工程学报, 2007, 27(17): 29-33.)
- [5] GULLETT B K, OUDEJANS L, TABOR D. Near-real-time combustion monitoring for PCDD/PCDF indicators by GC-REMPI-TOFMS. *Environmental Science & Technology*, 2012, 46(2):923-928.
- [6] GUO Ying, CHEN Tong, YANG Jie. Study on on-line detection of dioxins based on correlation model. *Chinese Journal of Environmental Engineering*, 2014, 8(8): 3524-3529.
 (郭颖,陈彤,杨杰. 基于关联模型的二恶英在线检测研究. 环境工程 学报, 2014, 8(8): 3524-3529.)
- [7] LI Adan, HONG Wei, WANG Jing. Online detection of dioxin and dioxin-related substances using laser desoption / laser ionizationmass 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 - 质谱法二恶英及其关联

物的在线检测. 燕山大学学报, 2015, 39(6):511-515.)

- [8] CAO Xuan, SHANG Fanjie, PAN Denggao. Gas chromatographymass spectrometry transmission line system for on-line detection of dioxins. China, CN206378474U, 2017-08-04. (曹轩, 尚凡杰, 潘登皋. 用于二恶英在线检测的气相色谱质谱间传 输线系统. CN206378474U, 公开/公告日: 2017-08-04.)
- WANG W, CHAI T Y, YU W. Modeling component concentrations of sodium aluminate solution via hammerstein recurrent neural networks. *IEEE Transactions on Control Systems Technology*, 2012, 20(4): 971-982.
- [10] TANG J, CHAI T Y, YU W, et al. Modeling load parameters of ball mill in grinding process based on selective ensemble multisensor information. *IEEE Transactions on Automation Science & Engineering*, 2013, 10(3): 726-740.
- [11] 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.
- [12] 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): 41-351.

- [13] TANG Jian, QIAO Junfei. Dioxin emission concentration soft measuring approach of municipal solid waste incineration based on selective ensemble kernel learning algorithm, *Journal of Chemical Industry and Engineering(China)*, 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.)
- [14] WANG Hairui, ZHANG Yong, WANG Hua. A study of GA BP based prediction model of Dioxin emis s ion from MSW incinerator. *Microcomputer Information*, 2008, 24(21):222-224. (王海瑞,张勇,王华. 基于 GA 和 BP 神经网络的二噁英软测量模 型研究. 微计算机信息, 2008, 24(21): 222-224.)
- [15] HU Wenjin, SU Yingying, TANG Yi, et al. Soft sensing modeling of dioxins for waste incineration based on small data sets. In: Proceedings of the 21th Chinese Process Control Conference, Xiamen, China, 2012.
 (胡文金, 苏盈盈, 汤毅, 等. 基于小样本数据的垃圾焚烧二恶英软测 量建模, 中国过程控制会议. 中国厦门, 2012.)
- [16] LIU Qiang, QIN Sizhao. Perspectives on Big Data Modeling of Process Industries, Acta Automatica Sinica, 2016, 42(2): 161-171. (刘强, 秦泗钊. 过程工业大数据建模研究展望. 自动化学报, 2016, 42(2): 161-171.)
- [17] TANG J, LIU Z, ZHANG J, et al. Kernel latent feature adaptive extraction and selection method for multi-component nonstationary signal of industrial mechanical device, *Neurocomputing*, 2016, 216(C): 296-309.
- [18] LEE J, CHANG K, JUN C H. Kernel-based calibration methods combined with multivariate feature selection to improve accuracy of nearinfrared spectroscopic analysis. *Chemometrics & Intelligent Laboratory Systems*, 2015, 147:139-146.
- [19] TANG Jian, TIAN Fuqing, JIA Meiying. Soft Measurement of Rotating Machinery Equipment Load Based on Spectrum Data Drive.
 Beijing: National Defense Industry Press, 2015.
 (汤健,田福庆,贾美英.基于频谱数据驱动的旋转机械设备负荷软 测量. 北京:国防工业出版社, 2015.)
- [20] ZHOU Z H, WU J, TANG W. Ensembling neural networks: many could be better than all. *Artificial Intelligence*, 137(1-2): 239-263, 2002.
- [21] TANG J, QIAO J F, WU Z W, et al. Vibration and acoustic frequency spectra for industrial process modeling using selective fusion multi-condition samples and multi-source features. *Mechanical Systems and Signal Processing*, 2018, 99:142-168.
- [22] GUO Z H, TANG J, QIAO J F. Mathematic Simulation Model of Dioxin Emission Concentration of Municipal Solid Waste Incineration Based on Aspen-plus Software, *the 37th Chinese Control Conference (CCC2018)*, 25-27 July 2018, Wuhan, China.

作者简介:

汤 健 教授,博士生导师,目前研究方向为小样本数据建模,固

废处理过程智能控制的研究; E-mail: freeflytang@bjut.edu.cn

乔俊飞 教授,博士生导师,目前研究方向为智能控制与智能信息

处理、复杂过程建模与优化控制的研究; E-mail: junfeq@bjut.edu.cn

徐 喆 副教授,硕士生导师,目前研究方向为事检测装置与感知 技术研究; E-mail: xu2002@bjut.edu.cn

郭子豪硕士研究生,目前研究方向为事虚拟样本生成、固废焚烧过程智能建模的研究; E-mail: miller94@163.com