

文章编号: 1671-251X(2025)07-0036-09

DOI: 10.13272/j.issn.1671-251x.2025060073

基于贝叶斯优化支持向量回归的 煤自燃温度预测模型

杨海燕, 胡新成, 蔡佳文, 余照阳

(贵州大学 矿业学院, 贵州 贵阳 550025)

摘要: 针对传统煤自燃温度预测模型未考虑指标气体与温度数据之间存在多重共线性、模型预测精度不足问题, 提出了一种基于贝叶斯优化(BO)算法改进支持向量回归(SVR)超参数(BO-SVR)的煤自燃温度预测模型。利用煤自燃程序升温实验, 对生成的指标气体数据进行收集与处理。利用 Spearman 相关性分析选择与煤温相关性较强的指标气体并分析指标气体生成量间的共线性; 对选择的指标气体进行主成分分析, 解决多重共线性问题的同时降低维数; 采用 5 折交叉验证方法划分训练集和测试集, 通过平均绝对误差(MAE)、均方根误差(RMSE)和判定系数(R^2)指标, 对 BO-SVR 模型的性能与 SVR、粒子群优化 SVR(PSO-SVR)和遗传算法优化 SVR(GA-SVR)模型进行定量评价。结果表明, BO-SVR 模型的 MAE 较其他 3 种模型分别降低了 74.2%, 36.7% 和 10.2%, RMSE 分别降低了 71.9%, 33.3% 和 11.4%, R^2 达 0.988 5, 高于其他模型。选取山西煤炭进出口集团河曲旧县露天煤业有限公司的烟煤煤样开展平行试验, BO-SVR 模型在新数据集上的 MAE 为 4.927 9 °C, RMSE 为 6.489 9 °C, R^2 达 0.985 3, 与原数据集预测结果保持高度一致性。表明 BO-SVR 模型具有较好的泛化性、预测精度和鲁棒性, 有助于提高预测煤自燃温度的准确性。

关键词: 煤自燃; 贝叶斯优化; 支持向量回归; 指标气体; 预测模型

中图分类号: TD752

文献标识码: A

Temperature prediction model for coal spontaneous combustion based on
Bayesian optimization support vector regression

YANG Haiyan, HU Xincheng, CAI Jiawen, YU Zhaoyang
(College of Mining, Guizhou University, Guiyang 550025, China)

Abstract: To address the issue that traditional coal spontaneous combustion temperature prediction models do not consider multicollinearity between indicator gases and temperature data and have insufficient prediction accuracy, a coal spontaneous combustion temperature prediction model using Support Vector Regression (SVR) with hyperparameters optimized by Bayesian Optimization (BO), abbreviated as BO-SVR, was proposed. A programmed heating experiment of coal spontaneous combustion was conducted to collect and process the generated indicator gas data. Spearman correlation analysis was used to select indicator gases with strong correlation to coal temperature and analyze the multicollinearity among the amounts of the generated indicator gases. Principal component analysis was performed on the selected indicator gases to resolve multicollinearity and reduce dimensionality simultaneously. Five-fold cross-validation was used to divide the training set and test set. The performance of the BO-SVR model was quantitatively evaluated in comparison with SVR, Particle Swarm Optimization SVR (PSO-SVR), and Genetic Algorithm-Optimized SVR (GA-SVR) models using Mean Absolute

收稿日期: 2025-06-27; 修回日期: 2025-07-18; 责任编辑: 王晖, 郑海霞。

基金项目: 国家自然科学基金地区基金项目(52364019, 52464017); 贵州省教育厅青年项目(黔教技〔2022〕106号)。

作者简介: 杨海燕(2001—), 女, 贵州松桃人, 硕士研究生, 主要研究方向为煤自燃灾害防治、机器学习等, E-mail: yanghy011122@163.com。通信作者: 胡新成(1988—), 男, 山东临沂人, 副教授, 主要研究方向为瓦斯与煤自燃复合灾害防治、机器学习、安全工程物联网等, E-mail: XinchengHu@hotmail.com。

引用格式: 杨海燕, 胡新成, 蔡佳文, 等. 基于贝叶斯优化支持向量回归的煤自燃温度预测模型[J]. 工矿自动化, 2025, 51(7): 36-43, 51.
YANG Haiyan, HU Xincheng, CAI Jiawen, et al. Temperature prediction model for coal spontaneous combustion based on Bayesian optimization support vector regression[J]. Journal of Mine Automation, 2025, 51(7): 36-43, 51.



扫码移动阅读

Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). Results showed that the MAE of the BO-SVR model decreased by 74.2%, 36.7%, and 10.2% compared with the other three models, respectively; the RMSE decreased by 71.9%, 33.3%, and 11.4%, respectively; and the R^2 reached 0.988 5, which was higher than other models. Parallel experiments were conducted using bituminous coal samples from Shanxi Coal Import and Export Group Hequ Jiuxian Open-pit Coal Industry Co., Ltd. The results showed that the BO-SVR model had an MAE of 4.927 9 °C, an RMSE of 6.489 9 °C, and an R^2 of 0.985 3 on the new dataset, which was highly consistent with the prediction results of the original dataset. This indicates that the BO-SVR model has good generalization ability, prediction accuracy, and robustness, contributing to improving the accuracy of coal spontaneous combustion temperature prediction.

Key words: coal spontaneous combustion; Bayesian optimization; support vector regression; indicator gas; prediction model

0 引言

煤自燃灾害作为煤炭安全生产的核心威胁之一^[1-2],不仅会造成煤炭资源的严重浪费,还可能引发火灾、爆炸等灾难性事故^[3]。目前,煤自燃灾害多通过指标气体进行预测预报。但是煤氧化产生的指标气体数据之间具有多重共线性,指标气体与煤温数据之间存在复杂的非线性关系,这导致预测预报温度的模型精度不够。因此,建立较为精准的预测模型以准确判定煤自燃温度,提前采取有效预防措施,对保障煤炭安全高效开采和降低经济损失具有重要意义。

在煤自燃预测预报领域,将机器学习算法与指标气体相结合进行煤自燃温度预测是重要研究方向^[4-6]。针对煤自燃的复杂非线性,咎军才等^[7]建立反向传播神经网络(Back Propagation Neural Network, BPNN)模型,探究煤自燃过程中气体对温度的反馈作用,从而对煤自燃温度进行预测。但是BPNN对样本需求量很大且训练过程易陷入过拟合^[8]。支持向量回归(Support Vector Regression, SVR)具有处理小样本、非线性问题的独特优势。SVR模型中,影响其预测性能的是正则化参数 C 和核函数尺度参数 γ ^[9-11]。为了实现更好的预测,学者们经常使用智能算法来优化SVR关键参数的选择。邓军等^[12-13]将SVR与粒子群优化(Particle Swarm Optimization, PSO)算法进行结合,通过选取最佳核函数和惩罚因子来提高预测精度。但是PSO算法易陷入局部最优且高度依赖参数设置。Guo Qing等^[14]使用遗传算法(Genetic Algorithm, GA)优化SVR的关键参数,优化后的模型在预测煤自燃温度方面具有明显的优势。但是GA局部搜索能力差且参数设置复杂。智能优化算法与SVR相结合的预测模型虽然在煤自燃预测中取得了一定的效果,但是仍存在易陷入局部最优、参数设置复杂问题,且预测精度有待进一步提升。

贝叶斯优化(Bayesian Optimization, BO)算法通

过构建目标函数的概率代理模型(如高斯过程),可在全局搜索空间中快速定位最优参数组合,且参数设置简单。BO算法与SVR模型结合,可以更高效定位全局最优,提升预测精度。因此,本文构建基于BO-SVR的煤自燃温度预测模型,以SVR为基础,通过径向基函数网络(Radial Basis Function Network, RBF)核函数挖掘煤自燃温度与指标气体特征间的非线性关系,结合BO对SVR关键参数进行迭代寻优。基于实验监测数据,利用Spearman相关性分析和主成分分析(Principal Component Analysis, PCA)方法对煤自燃氧化过程中的指标气体进行特征提取,经特征提取后的数据作为模型输入。引入平均绝对误差(Mean Absolute Error, MAE)作为参数优化目标函数。采用5折交叉验证法划分训练集和测试集,通过MAE、均方根误差(Root Mean Squared Error, RMSE)和判定系数(R^2)评价模型性能,并与SVR模型、PSO-SVR模型和GA-SVR模型进行对比分析。同时选取山西煤炭进出口集团河曲旧县露天煤业有限公司的烟煤煤样开展平行试验,验证BO-SVR模型性能。

1 理论基础

1.1 SVR

SVR是SVM在回归任务中的拓展应用。与传统回归算法不同,SVR通过引入“ ϵ -不敏感损失函数”,允许预测值在一定误差范围内与真实值存在偏差而不产生损失,能够有效处理小样本、非线性回归问题^[15-17]。

SVR的核心思想是在特征空间中寻找一个最优超平面,使得训练样本到该超平面的误差尽可能小。对于给定的训练数据集 $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$,其中 $\mathbf{x}_i$ 为第 i 个输入特征向量, n 为训练数据集的样本数量, $\mathbf{x}_i \in \mathbf{R}^d$, $\mathbf{R}^d$ 为 d 维实数空间, y_i 为第 i 个目标值, $y_i \in \mathbf{R}$,SVR试图构建一个回归函数 $f(\mathbf{x}) = \omega^T \Phi(\mathbf{x}) + c$,其中 $\mathbf{x}$ 为输入特征向量, ω 为权重向量, $\Phi(\mathbf{x})$ 为通过核函

数 $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$ 将原始输入空间映射到高维特征空间的映射函数, $\mathbf{x}_j$ 为第 j 个输入特征向量, c 为偏置项。

SVR 引入 ε -不敏感带概念, 定义损失函数:

$$L_\varepsilon(y_i, f(\mathbf{x}_i)) = \begin{cases} 0 & |y_i - f(\mathbf{x}_i)| \leq \varepsilon \\ |y_i - f(\mathbf{x}_i)| - \varepsilon & \text{其他} \end{cases} \quad (1)$$

SVR 的优化目标是最小化结构风险和经验风险, 构建带约束的优化问题:

$$\begin{cases} \min_{\omega, c, \xi_i, \xi_i^*} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{s.t. } y_i - f(\mathbf{x}_i) \leq \varepsilon + \xi_i \\ f(\mathbf{x}_i) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \forall i = 1, 2, \dots, n \end{cases} \quad (2)$$

式中 ξ_i 和 ξ_i^* 为松弛变量, 分别表示样本在 ε -不敏感带上和下方的误差。

通过拉格朗日乘子法将上述约束优化问题转换为对偶问题:

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i^* - \alpha_i) k(\mathbf{x}_i, \mathbf{x}_j) + c \quad (3)$$

式中 α_i 和 α_i^* 为拉格朗日乘子, 仅部分样本对应的乘子非零, 这些样本即为支持向量。

1.2 BO

BO 是一种基于概率模型的全局优化方法, 旨在高效地找到目标函数的最优解, 尤其适用于目标函数计算成本高昂、导数难以获取的场景。其核心思想是先构建目标函数的代理模型(通常为高斯过程), 再基于该模型的后验分布, 通过采集函数平衡对未知区域的探索与对已知最优解区域的利用, 从而逐步逼近最优解^[18-19]。高斯过程定义了一个关于函数的概率分布。对于输入空间中的任意有限个点 $a = a_1, a_2, \dots, a_n$, 其对应的函数值 $b_1, b_2, \dots, b_n$ 服从联合高斯分布:

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \sim N \left(\begin{bmatrix} \mu(a_1) \\ \mu(a_2) \\ \vdots \\ \mu(a_n) \end{bmatrix}, \begin{bmatrix} k'(a_1, a_1) & k'(a_1, a_2) & \cdots & k'(a_1, a_n) \\ k'(a_2, a_1) & k'(a_2, a_2) & \cdots & k'(a_2, a_n) \\ \vdots & \vdots & \ddots & \vdots \\ k'(a_n, a_1) & k'(a_n, a_2) & \cdots & k'(a_n, a_n) \end{bmatrix} \right) \quad (4)$$

式中: N 为正态分布; $\mu(a)$ 为均值函数, 通常设为零均值, 简化计算; $k'(a_i, a_j) = \sigma_f^2 \exp\left(-\frac{\|a_i - a_j\|^2}{2l^2}\right)$ 为核函数, 用于衡量第 i 个输入点 a_i 和第 j 个输入点 a_j 之间的相似性, σ_f^2 为信号方差, l 为高斯核的宽度。

在 BO 中, 通过已观测到的样本数据集 $D = \{(a_i, b_i)\}_{i=1}^n$ 来更新高斯过程的后验分布, 从而得到在新的输入点 a_* 处的预测分布:

$$p(b_* | D, a_*) = N(\mu_*(a_*), \sigma_*^2(a_*)) \quad (5)$$

式中: b_* 为 a_* 处对应的函数值; $\mu_*(a_*)$ 为 a_* 处对应的预测均值; $\sigma_*^2(a_*)$ 为 a_* 处对应的预测方差。

基于上述预测分布选择下一个最优采样点时, 需通过采集函数实现。常见的采集函数包括期望改善(Expected Improvement, EI)、上置信界(Upper Confidence Bound, UCB)和概率提升(Probability of Improvement, PI)等。与其他采集函数相比, EI 采集函数整合了改善幅度信息, 更能平衡探索与利用以高效收敛至全局最优。因此, 本文使用 EI 采集函数计算在点 a_i 处获得比当前最佳值更好结果的期望, 来衡量该点的价值, 值越大表明该点越值得探索。

$$EI(a_i) = \int_{-\infty}^{b_{\text{best}}} (b_{\text{best}} - z) p(z | D, a_i) dz \quad (6)$$

式中: b_{best} 为当前已观测到的最佳函数值; z 为点 a_i 处的潜在目标函数输出; $p(z | D, a_i)$ 为点 a_i 处的预测分布。

1.3 基于 BO-SVR 的煤自燃温度预测模型

为实现煤自燃温度的高精度预测, 构建基于 BO-SVR 的煤自燃温度预测模型, 通过优化模型超参数提升泛化性能。模型工作流程如图 1 所示。

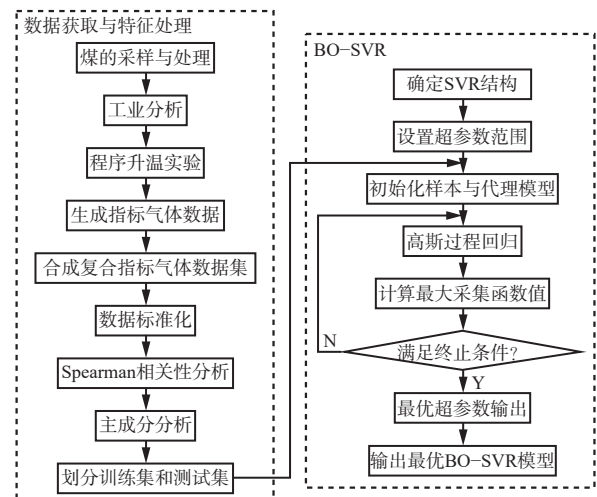


图 1 基于 BO-SVR 的煤自燃温度预测模型工作流程

Fig. 1 Workflow of BO-SVR-based coal spontaneous combustion temperature prediction model

1) 数据获取与特征处理。首先, 提取煤自燃程序升温实验收集的原始数据, 并根据气体数据比值构成复合数据集, 建立数据矩阵。考虑到不同变量间纲差异可能干扰模型训练, 采用 Z-score 标准化方法对数据进行归一化处理, 使各特征均值为 0、方

差为 1, 确保数据分布的一致性。然后, 运用 Spearman 相关系数分析方法, 通过绘制相关性热力图可视化变量间的依赖关系。该分析不仅能识别变量间的共线性问题, 还可量化各变量与煤自燃温度的关联程度, 剔除与煤温相关性较弱的冗余变量, 降低后续多变量分析的计算复杂度。最后, 针对剩余特征变量, 采用 PCA 进行降维处理。通过正交变换将原始高维特征映射到低维主成分空间, 在保留方差累积贡献率达 95% 的前提下, 有效消除多重共线性, 提取数据的核心特征, 提升模型训练效率与稳定性。并采用 5 折交叉验证法划分训练集和测试集。

2) BO-SVR 模型。选用 RBF 作为 SVR 模型的核函数, 以刻画煤自燃温度与特征变量间的非线性映射关系。RBF 核函数通过调整尺度参数 γ , 可灵活控制数据在特征空间的分布密度, 适应复杂的非线性关系。模型构建时, 将正则化参数 C 和 RBF 核函数尺度参数 γ 作为核心优化对象, 初始设定其搜索空间为 $C \in [10^{-3}, 10^3]$, $\gamma \in [10^{-3}, 10^3]$, 采用对数刻度保证参数搜索的均匀性。通过构建代理模型(高斯过程模型)拟合目标函数, 结合采集函数平衡探索与利用, 高效定位最优参数组合。在每次迭代中, 根据当前代理模型预测不同参数组合的 MAE, 并选取采集函数值最大的参数组合进行实际评估, 更新代理模型后继续迭代, 直至满足停止条件(迭代 100 次)。最终以最优参数 C^* 和 γ^* 构建基于 BO-SVR 的煤自燃温度预测模型。

2 实验与结果分析

2.1 数据采集与处理

实验选取在陕西大佛寺煤矿采集的煤样作为研究对象, 并进行工业分析和程序升温实验。该煤样的工业分析结果见表 1, 其中 M_{ad} 为水分, A_{ad} 为灰分, V_{ad} 为挥发分, FC_{ad} 为固定碳。程序升温实验装置包括供气装置、氧化反应装置、气体分析装置和数据收集装置 4 个部分, 如图 2 所示。采用 GC-4000 型气相色谱仪检测 O_2 , CO , CO_2 , CH_4 , C_2H_4 , C_2H_6 , C_3H_8 等气体。在实验中, 将原煤筛分为 60~80 目、80~100 目、100~120 目 3 种规格, 每种规格 3 g, 按 1:1:1 的比例混入煤样罐中。入口 O_2 体积分数设置为 21%, 空气流速为 100 mL/min。煤样以 0.8 $^{\circ}C/min$ 的速率从 30 $^{\circ}C$ 加热至 260 $^{\circ}C$, 煤温每升高 10 $^{\circ}C$, 用气相色谱仪分析气体成分与浓度。

在煤自燃程序升温实验期间收集的气体产物生成量与温度变化之间的关系如图 3 和图 4 所示。

从图 3 可看出, 出口 O_2 体积分数随着温度的升

表 1 实验用煤样的工业分析结果

Table 1 Summary of coal samples used in experiment				
$M_{ad}/\%$	$A_{ad}/\%$	$V_{ad}/\%$	$FC_{ad}/\%$	自燃倾向性
7.54	10.87	37.71	47.36	II

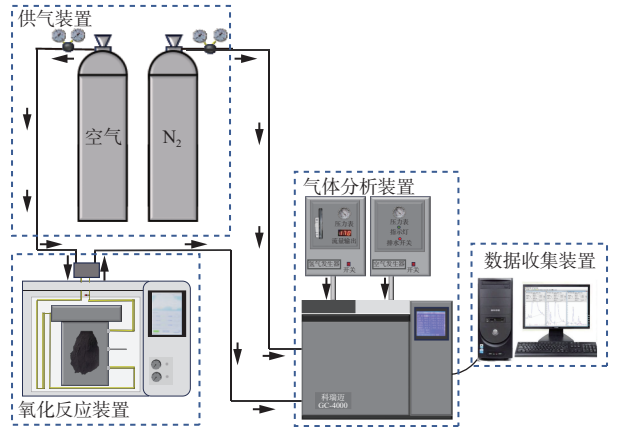


图 2 煤自燃程序升温实验流程

Fig. 2 Procedure of programmed heating experiment for coal spontaneous combustion

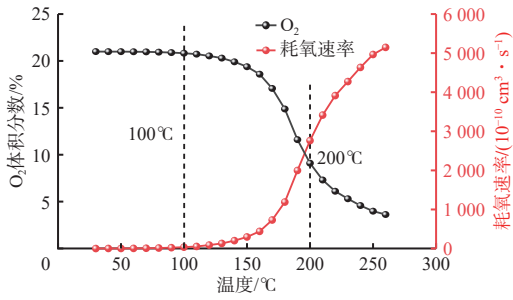


图 3 出口 O_2 体积分数和耗氧率随温度的变化规律

Fig. 3 Variation trends of outlet oxygen volume fraction and oxygen consumption rate with temperature

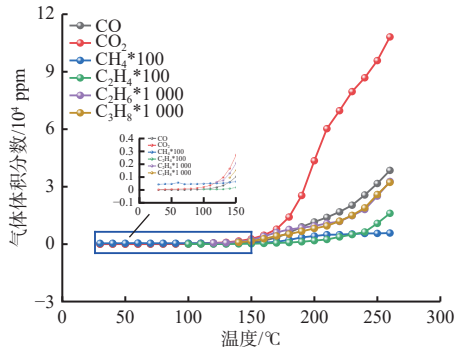


图 4 程序升温实验过程中的气体体积分数随温度的变化规律

Fig. 4 Variation trends of gas volume fractions with temperature during programmed heating experiment

高而逐渐降低。在 100 $^{\circ}C$ 之前, 出口 O_2 体积分数下降缓慢。100 $^{\circ}C$ 后, 出口 O_2 体积分数迅速下降。在 200 $^{\circ}C$ 后, 下降趋势逐渐趋于平缓。随着温度升高, 耗氧速率始终呈增加趋势。在 100 $^{\circ}C$ 之后呈近似指数增长趋势。在 200 $^{\circ}C$ 以后, 这一增长趋势开始减

缓。这说明煤自燃是一个多阶段、复杂的氧化过程。

从图 4 可看出,随着温度的升高,CO 和 CO₂ 体积分数逐渐增加,上升过程呈明显的阶段性变化。150 ℃ 之前,CO 和 CO₂ 体积分数均很小。150 ℃ 之后,二者体积分数上升速度显著加快,其中 CO₂ 体积分数呈近似指数上升趋势。随着温度升高,CH₄ 体积分数总体呈上升趋势。在 160 ℃ 之前,CH₄ 体积分数增加较为缓慢。在 160 ℃ 之后,CH₄ 体积分数增加速率变快。在 90 ℃ 时,C₂H₄ 出现,其体积分数先缓慢增加,200 ℃ 之后增加速率明显加快。在 110 ℃ 时,C₂H₆ 出现,体积分数随着温度的升高快速增加。240 ℃ 之后,C₂H₆ 和 C₂H₄ 体积分数再次显著增加,表明煤样开始进入燃烧阶段。C₃H₈ 在 130 ℃ 出现后其体积分数便迅速增加。

实验结果表明:O₂ 体积分数、耗氧速率、CO、CO₂、CH₄、C₂H₆、C₃H₈ 体积分数的变化与煤温存在明显的相关性,可以作为煤自燃预警指标。

煤自燃过程中释放的气体浓度易受外界环境(如通风条件、气压波动)影响,仅以单一指标气体判断煤自燃存在局限性。为克服这一局限性,引入复合指标气体(例如 CO/CO₂、C₂H₄/C₂H₆)。复合指标气体通过气体浓度的比率关系可以削弱环境因素的影响,显著提高预测的稳定性^[20]。复合指标气体比值随煤温变化关系曲线如图 5 和图 6 所示。

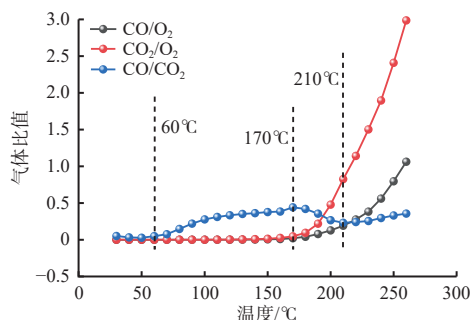


图 5 碳氧化物气体比值随温度的变化规律

Fig. 5 Variation trends of carbon oxide gases ratio with temperature

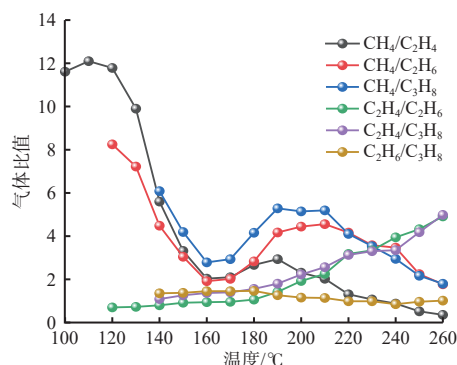


图 6 碳氢化物气体比值随温度的变化规律

Fig. 6 Variation trends of hydrocarbon gas ratio with temperature

从图 5 可看出,CO/CO₂ 比值随温度升高呈逐渐增加趋势。在低温氧化阶段(30~60 ℃),CO/CO₂ 比值较低。随着温度升高,在 60 ℃ 以后,CO/CO₂ 比值显著上升,在 170 ℃ 时,CO/CO₂ 比值达到最高,随后逐渐降低,在 210 ℃ 后再次缓慢增加。CO/O₂ 和 CO₂/O₂ 比值随温度升高而逐渐增加,尤其在 170 ℃ 以后,比值均显著上升。由此可看出,碳氧化物气体比值与煤温存在相关性。

从图 6 可看出,CH₄/C₂H₄、CH₄/C₂H₆、CH₄/C₃H₈ 比值刚开始均较高,随着温度的升高,CH₄/C₂H₄、CH₄/C₂H₆ 和 CH₄/C₃H₈ 的比值变化趋势为先降低后升高再降低。C₂H₄/C₂H₆ 和 C₂H₄/C₃H₈ 自出现后便呈上升趋势。在 180 ℃ 之前增加的较为缓慢,180 ℃ 后上升速率明显加快。C₂H₆/C₃H₈ 自出现后便呈下降趋势,总体变化不大。

实验结果表明:CO/CO₂、CO/O₂、CO₂/O₂、CH₄/C₂H₄、CH₄/C₂H₆、CH₄/C₃H₈、C₂H₄/C₂H₆、C₂H₄/C₃H₈、C₂H₆/C₃H₈ 的比值变化与煤温具有一定的相关性,可以作为煤自燃预警指标。

2.2 数据特征择优

在多元回归分析中,共线性问题是影响模型稳定性和解释性的重要挑战^[21]。煤自燃程序升温产生的气体及气体比值数据之间可能存在多重共线性,若直接将所有因素作为输入代入模型进行多元回归,一方面容易降低预测精度,另一方面由于其高维特性,会造成计算效率下降。因此,本研究先对变量进行相关系数分析,然后绘制相关系数热力图,分析并揭示单一指标气体和复合指标气体之间是否存在多重共线性关系及各指标气体与煤自燃温度之间的关系。

运用 Shapiro-Wilk 正态性检验对全部变量进行分布特性诊断。统计结果显示,除煤温外,其余变量的显著性 p 值均显著低于统计阈值 0.05,表明变量不满足正态分布。因此选择 Spearman 相关性分析方法,结果如图 7 所示,图中圆圈的大小及颜色的深浅表示了两变量之间的相关程度,其中红色表示正相关,绿色表示负相关。结果表明,程序升温实验生成的指标气体之间存在显著的共线性关系,并且均与煤温存在一定的相关性。其中,CH₄/C₂H₄ 与煤自燃温度的相关性较弱,不作为后续研究的变量,以减少计算复杂度。

2.3 数据特征降维

本研究通过煤自燃程序升温实验获取了 17 个与煤自燃温度相关的特征变量,经 Spearman 相关性分析后保留了 16 个变量(TZ1—TZ16),涵盖单一气体

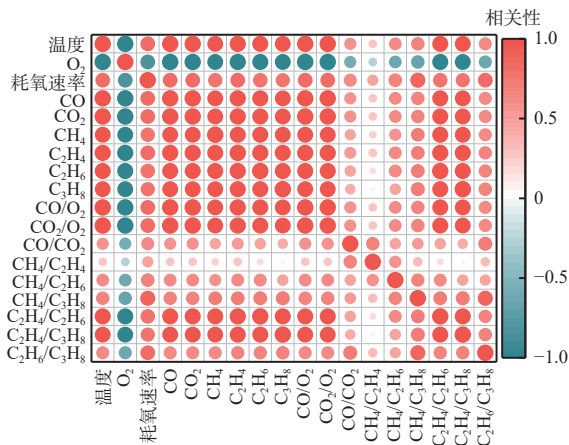


图 7 各指标气体与煤温的相关系数热力图

Fig. 7 Heat map of relationship between each indicator gas and coal temperature

指标与复合气体指标 2 类。由于这些变量间存在多重共线性问题,且高维数据易引发维数灾难,所以需通过 PCA 方法进行特征提取^[22-23]。采用 Matlab 软件实现 PCA 方法对变量的处理,处理后的部分主成分数据见表 2,其中 PC 表示主成分,各主成分方差贡献率及累计贡献率如图 8 所示。可看出前 4 个主成分的方差贡献率分别为 75.24%, 14.62%, 4.65%, 4.04%, 方差累计贡献率为 98.55%, 满足特征提取要求(>95% 阈值),故确定前 4 个主成分为影响煤自燃温度演化的关键控制因子。

表 2 PCA 降维后的部分主成分数据
Table 2 Partial principal component data after PCA dimensionality reduction

特征	PC1	PC2	PC3	PC4
TZ1	-0.278 4	-0.035 7	0.105 1	0.242 3
TZ2	0.282 1	-0.025 2	-0.060 6	-0.205 7
TZ3	0.283 9	-0.107 7	0.026 1	0.016 0
TZ4	0.282 3	-0.083 0	-0.020 9	-0.159 9
TZ5	0.275 1	0.016 3	-0.113 8	-0.294 8
TZ6	0.263 6	-0.196 8	0.130 3	0.243 9
TZ7	0.283 2	-0.067 0	0.025 5	0.143 0
TZ8	0.282 2	-0.108 9	0.033 4	0.122 1
TZ9	0.267 8	-0.200 0	0.121 1	0.179 4
TZ10	0.274 7	-0.180 4	0.084 9	0.055 0
TZ11	0.111 1	0.456 2	0.376 8	0.583 7
TZ12	0.082 5	0.420 6	0.681 0	-0.472 7
TZ13	0.142 9	0.498 4	-0.415 3	-0.150 0
TZ14	0.286 0	0.004 7	0.078 6	-0.045 4
TZ15	0.284 9	0.057 9	-0.109 2	0.037 8
TZ16	0.165 9	0.463 4	-0.366 1	0.256 3

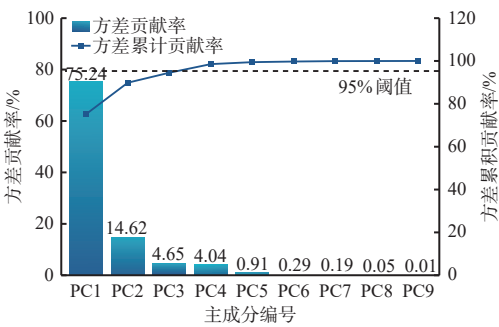


图 8 各主成分方差贡献率及累计贡献率

Fig. 8 Contribution rate and cumulative contribution rate graph of each principal component

3 模型性能对比与验证

3.1 模型性能评价指标

为客观评价模型性能,本文选择 3 个常见的模型性能评价指标。MAE 是预测值与真实值之间距离的平均值,能够衡量温度预测结果的精确性,其值越小说明预测越精准。RMSE 是预测值与真实值之间误差平方的平均值的平方根,可以很好地反映温度预测值中较大误差的影响程度,其值越小代表预测值与真实值的总体偏差越小,模型的预测性能越好。 R^2 是用于评估模型拟合优度和预测准确性的重要统计指标,可以衡量模型解释因变量变异的能力,取值范围为 $[0, 1]$ 。其值越接近 1,说明模型预测值与实际值的拟合程度越高,预测效果越好;其值越接近 0,表示模型的解释能力越弱,预测效果不佳。

3.2 模型性能对比

为验证所提出的基于 BO-SVR 的煤自燃温度预测模型的优势,与 SVR, PSO-SVR 和 GA-SVR 这 3 个常用的煤自燃温度预测模型进行对比。对比模型均选用 RBF 核函数映射机制挖掘煤自燃温度与指标气体数据间的非线性映射关系。其中,SVR 模型的正则化参数 C 设为 1,自动计算 RBF 核的 γ 参数。其他对比模型的正则化参数 C 和 RBF 核函数尺度参数 γ 初始设定其搜索空间为 $C \in [10^{-3}, 10^3]$, $\gamma \in [10^{-3}, 10^3]$,采用对数刻度保证参数搜索的均匀性,均迭代 100 次。此外,考虑到不同模型的准确性和稳定性及实验样本的多样性,减少单次划分训练集与测试集引入的偏差,因此采用 5 折交叉验证的 MAE, RMSE 和 R^2 的均值对模型性能进行量化评估,确保实验结果在不同样本划分下的稳健性。

不同模型对煤自燃温度的预测值与真实值对比如图 9 所示。从图形的空间分布特征分析,SVR 模型的样本预测值大多偏离真实值的位置,其预测误差波动很大,表明模型的预测精度较低,鲁棒性较

差。PSO-SVR 模型更多的预测值在真实值附近,但仍有一些预测值偏离真实值。GA-SVR 模型样本的预测值与真实值基本一致,但个别预测结果分布在远离真实值的位置,产生较大误差。BO-SVR 模型样本的预测值大多紧密分布在真实值附近,这充分表明 BO-SVR 模型更接近真实值的拟合,能够准确捕捉煤自燃温度的变化趋势,表明 BO-SVR 模型的预测精度、鲁棒性和泛化能力比其他模型更好。

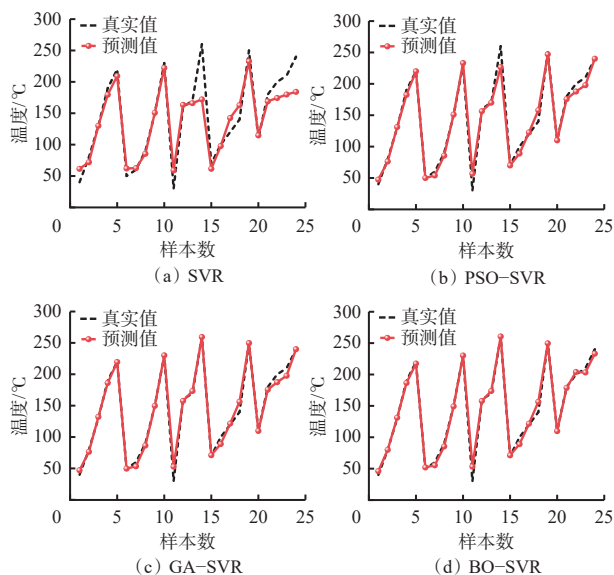


图 9 各个模型预测值与真实值的对比

Fig. 9 Comparison between predicted values and actual values of each model

不同模型的性能评价指标见表 3。可看出 BO-SVR 模型的 MAE 较 SVR 模型、PSO-SVR 模型和 GA-SVR 模型分别降低了 74.2%, 36.7%, 10.2%, RMSE 分别降低了 71.9%, 33.3%, 11.4%, R^2 值达 0.988 5, 高于其他模型, 说明 BO-SVR 模型的预测性能更优。

表 3 各个预测模型性能评价指标对比

Table 3 Comparison of performance evaluation indicators of each prediction model

模型	MAE/°C	RMSE/°C	R^2
SVR	17.371 7	22.651 9	0.810 5
PSO-SVR	7.081 8	9.528 7	0.973 0
GA-SVR	4.991 5	7.168 2	0.983 9
BO-SVR	4.480 8	6.354 3	0.988 5

3.3 模型性能验证

为了更好地说明 BO-SVR 模型预测煤自燃温度的准确性和实用性,选取山西煤炭进出口集团河曲旧县露天煤业有限公司的烟煤煤样进行平行试验。预测结果如图 10 所示,可看出预测值始终贴合真实值,表明 BO-SVR 模型具有优越的泛化性能。

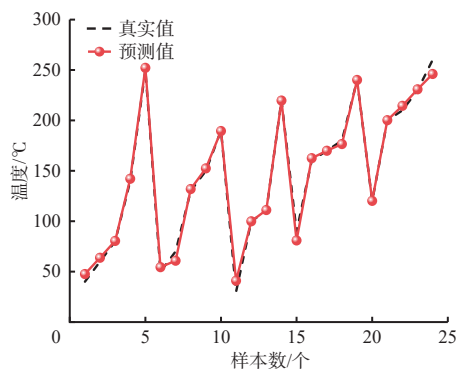


图 10 BO-SVR 模型预测煤温的预测值与真实值对比

Fig. 10 Comparison of predicted values and actual values of coal temperature predicted by BO-SVR model

BO-SVR 模型的 MAE 性能评价指标结果: 为 4.927 9 °C, RMSE 为 6.489 9 °C, R^2 为 0.985 3。

4 结论

1) 经 BO 优化后的 SVR 模型显著降低了预测模型的参数选取难度与结构复杂度,有效避免了局部最优问题。

2) 将 BO-SVR 模型与原始 SVR 模型、PSO-SVR 模型及 GA-SVR 模型进行对比。采用 5 折交叉验证法对各模型进行评价。结果显示,BO-SVR 模型的 MAE 为 4.480 8 °C,较其他 3 种模型分别降低 74.2%, 36.7% 和 10.2%; RMSE 为 6.354 3 °C,分别降低了 71.9%, 33.3% 和 11.4%; R^2 达 0.988 5。这表明 BO-SVR 模型能够更为精准地进行煤自燃温度预测,具有较强的泛化性和鲁棒性。

3) 选取国内其他煤矿的煤样开展平行试验。采用相同实验进行数据采集与处理,并应用 BO-SVR 模型进行温度预测。该模型在新数据集上的 MAE 为 4.927 9 °C, RMSE 为 6.489 9 °C, R^2 达 0.985 3,与原数据集测试结果保持高度一致性。结果表明,BO-SVR 模型具备良好的环境适应性和跨样本预测能力,能够为煤矿的自燃风险预警提供可靠技术支撑。

参考文献(References):

- [1] 张鹏宇,陈晓坤,赵亮,等.大面积采空区煤自燃环境下气体运移规律研究[J].煤矿安全,2022,53(9):122-128,136.
ZHANG Pengyu, CHEN Xiaokun, ZHAO Liang, et al. Study on gas migration law in large area goaf under coal spontaneous combustion environment[J]. Safety in Coal Mines, 2022, 53(9): 122-128, 136.
- [2] 易欣,张敏,邓军.煤自燃指标体系分析与优选实验研究[J].煤矿安全,2023,54(1):85-93.
YI Xin, ZHANG Min, DENG Jun. Analysis and

- optimization of coal spontaneous combustion index system[J]. *Safety in Coal Mines*, 2023, 54(1): 85-93.
- [3] 林柏泉, 李庆钊, 原德胜, 等. 彬长矿区低煤阶煤层气井的排采特征与井型优化[J]. *煤炭学报*, 2015, 40(1): 135-141.
- LIN Baiquan, LI Qingzhao, YUAN Desheng, et al. CBM production character and surface well selection in Binchang low rank coal field[J]. *Journal of China Coal Society*, 2015, 40(1): 135-141.
- [4] 仲晓星, 王建涛, 周昆. 矿井煤自燃监测预警技术研究现状及智能化发展趋势[J]. *工矿自动化*, 2021, 47(9): 7-17.
- ZHONG Xiaoxing, WANG Jiantao, ZHOU Kun. Monitoring and early warning technology of coal spontaneous combustion in coal mines: research status and intelligent development trends[J]. *Industry and Mine Automation*, 2021, 47(9): 7-17.
- [5] 邓军, 李鑫, 王凯, 等. 矿井火灾智能监测预警技术近 20 年研究进展及展望[J]. *煤炭科学技术*, 2024, 52(1): 154-177.
- DENG Jun, LI Xin, WANG Kai, et al. Research progress and prospect of mine fire intelligent monitoring and early warning technology in recent 20 years[J]. *Coal Science and Technology*, 2024, 52(1): 154-177.
- [6] 王玉怀, 胡硕鹏, 朱永兴, 等. 机器学习在煤自燃预测预报中的应用现状及展望[J]. *中国煤炭*, 2024, 50(10): 98-103.
- WANG Yuhuai, HU Shuopeng, ZHU Yongxing, et al. Application status and prospect of machine learning in coal spontaneous combustion prediction and early warning[J]. *China Coal*, 2024, 50(10): 98-103.
- [7] 咎军才, 魏成才, 蒋可娟, 等. 基于 BP 神经网络的煤自燃温度预测研究[J]. *煤炭工程*, 2019, 51(10): 113-117.
- ZAN Juncai, WEI Chengcai, JIANG Kejuan, et al. Prediction of coal spontaneous combustion temperature based on BP neural network[J]. *Coal Engineering*, 2019, 51(10): 113-117.
- [8] SAHINER B, CHAN H P, HADJIISKI L. Classifier performance estimation under the constraint of a finite sample size: resampling schemes applied to neural network classifiers[J]. *Neural Networks*, 2008, 21(2/3): 476-483.
- [9] PROTIĆ M, SHAMSHIRBAND S, PETKOVIĆ D, et al. Forecasting of consumers heat load in district heating systems using the support vector machine with a discrete wavelet transform algorithm[J]. *Energy*, 2015, 87: 343-351.
- [10] ZHANG Wei, ZHANG Linlin, YANG Jingxuan, et al. An experimental modeling of cyclone separator efficiency with PCA-PSO-SVR algorithm[J]. *Powder Technology*, 2019, 347: 114-124.
- [11] ZHUANG D Y, MA K, TANG C A, et al. Mechanical parameter inversion in tunnel engineering using support vector regression optimized by multi-strategy artificial fish swarm algorithm[J]. *Tunnelling and Underground Space Technology*, 2019, 83: 425-436.
- [12] 邓军, 雷昌奎, 曹凯, 等. 煤自燃预测的支持向量回归方法[J]. *西安科技大学学报*, 2018, 38(2): 175-180.
- DENG Jun, LEI Changkui, CAO Kai, et al. Support vector regression approach for predicting coal spontaneous combustion[J]. *Journal of Xi'an University of Science and Technology*, 2018, 38(2): 175-180.
- [13] DENG Jun, LEI Changkui, XIAO Yang, et al. Determination and prediction on "three zones" of coal spontaneous combustion in a gob of fully mechanized caving face[J]. *Fuel*, 2018, 211: 458-470.
- [14] GUO Qing, REN Wanxing, LU Wei. A method for predicting coal temperature using CO with GA-SVR model for early warning of the spontaneous combustion of coal[J]. *Combustion Science and Technology*, 2022, 194(3): 523-538.
- [15] CORTES C, VAPNIK V. Support-vector networks[J]. *Machine Learning*, 1995, 20(3): 273-297.
- [16] 黄贤振, 孙良仕, 高妮, 等. 基于 SFS-SVR 的高速铣削刀具剩余使用寿命预测[J]. *东北大学学报(自然科学版)*, 2023, 44(6): 824-831.
- HUANG Xianzhen, SUN Liangshi, GAO Wei, et al. Remaining useful life prediction of cutting tools based on SFS-SVR in high speed milling operations[J]. *Journal of Northeastern University (Natural Science)*, 2023, 44(6): 824-831.
- [17] 范勇, 胡名东, 杨广栋, 等. PCA-GWO-SVR 机器学习用于边坡爆破振动速度峰值预测研究[J]. *振动工程学报*, 2024, 37(8): 1431-1441.
- FAN Yong, HU Mingdong, YANG Guangdong, et al. PCA-GWO-SVR machine learning applied to prediction of peak vibration velocity of slope blasting[J]. *Journal of Vibration Engineering*, 2024, 37(8): 1431-1441.
- [18] 陈泉霖, 陈奕宇, 霍静, 等. 高维贝叶斯优化研究综述[J]. *软件学报*, 2025, 36(6): 2576-2603.
- CHEN Quanlin, CHEN Yiyu, HUO Jing, et al. Survey on high-dimensional bayesian optimization[J]. *Journal of Software*, 2025, 36(6): 2576-2603.
- [19] 吴永洪, 张智斌. 基于贝叶斯优化的 CNN-GRU 短期电力负荷预测[J]. *现代电子技术*, 2023, 46(20): 125-129.
- WU Yonghong, ZHANG Zhibin. CNN-GRU short-term power load forecasting based on Bayesian optimization[J]. *Modern Electronics Technique*, 2023, 46(20): 125-129.