

电网碳钢材料腐蚀预测模型研究

何成¹, 游溢¹, 王宗江¹, 黄路遥^{2,3*}, 张强^{2,3}, 陈云^{2,3},
卢壹梁^{2,3}, 杨丙坤^{2,3}, 王晓芳^{2,3}

(1. 国网新疆电力有限公司电力科学研究院, 乌鲁木齐 830011; 2. 中国电力科学研究院有限公司, 北京 100192; 3. 国网智能电网研究院有限公司, 北京 102209)

摘要: **目的** 研究新疆地区典型环境因素对典型电网用碳钢材料腐蚀的影响。**方法** 根据采集到的新疆地区155组腐蚀及环境数据, 通过支持向量回归(SVR)、梯度提升算法(GBoost)、皮尔逊相关系数(PCC)、逐点互信息(PMI)和随机森林(RF)等5种方式对输入变量进行特征选择, 分析11种典型环境因素对典型电网用碳钢材料腐蚀速率的影响, 选择重要性排序前5输入变量并集, 进行Sobol敏感性分析。采用经主成分分析进行降维处理构建的SVR模型和利用梯度提升算法优化损失函数构建的GBoost模型预测腐蚀速率, 研究2种模型的性能和预测能力。**结果** 对新疆地区典型电网用碳钢材料碳钢大气腐蚀影响较为重要的前5个特征依次是年降水量、年均湿度、年均温差、PM₁₀和O₃, GBoost模型相比于SVR模型在预测影响碳钢腐蚀的环境因素方面表现出较高的准确性和可靠性。**结论** GBoost模型具有更好地预测泛化能力和模型解释力, 能够有效捕捉碳钢腐蚀速率与环境因素之间的复杂关系。

关键词: 电网; 碳钢; 大气腐蚀; 机器学习; 寿命预测; 预测精度

中图分类号: TG172 **文献标志码:** A **文章编号:** 1672-9242(2025)02-0151-09

DOI: 10.7643/issn.1672-9242.2025.02.017

Corrosion Prediction Model of Carbon Steel for Power Grid

HE Cheng¹, YOU Yi¹, WANG Zongjiang¹, HUANG Luyao^{2,3*}, ZHANG Qiang^{2,3}, CHEN Yun^{2,3},
LU Yiliang^{2,3}, YANG Bingkun^{2,3}, WANG Xiaofang^{2,3}

(1. Electric Power Science and Research Institute of State Grid Xinjiang Electric Power Co., Ltd., Urumqi 830011, China;
2. China Electric Power Research Institute, Beijing 100192, China; 3. State Grid Smart Grid Research
Institute Co., Ltd., Beijing 102209, China)

ABSTRACT: The work aims to investigate the influence of typical environmental factors on the corrosion rate of carbon steel used in Xinjiang power grid. Based on 155 sets of corrosion and environmental data collected from Xinjiang, feature selection of input variables was accomplished through the application of five distinct methods, namely Support Vector Regression (SVR), Gradient Boosting algorithm (GBoost), Pearson Correlation Coefficient (PCC), Pointwise Mutual Information (PMI), and Random Forest (RF). The influence of 11 typical environmental factors on the corrosion rate of typical carbon steel used in power

收稿日期: 2024-07-18; 修订日期: 2024-08-13

Received: 2024-07-18; Revised: 2024-08-13

基金项目: 国网新疆电力有限公司科技项目(5230DK230016)

Fund: Science and Technology Project of State Grid Xinjiang Electric Power Co., Ltd. (5230DK230016)

引文格式: 何成, 游溢, 王宗江, 等. 电网碳钢材料腐蚀预测模型研究[J]. 装备环境工程, 2025, 22(2): 151-159.

HE Cheng, YOU Yi, WANG Zongjiang, et al. Corrosion Prediction Model of Carbon Steel for Power Grid[J]. Equipment Environmental Engineering, 2025, 22(2): 151-159.

*通信作者(Corresponding author)

grids was analyzed. The top five input variables ranked by their significance, were chosen and amalgamated for the Sobol sensitivity examination. Principal component analysis (PCA) was conducted to construct a SVR model and GBoost was used to optimize loss function to construct a GBoost model to predict the corrosion rate. The performance and prediction ability of the two models were studied. The result showed that the top five environmental factors are, in turn, annual precipitation, annual humidity, annual temperature difference, PM₁₀ and O₃. Compared with the SVR model, the GBoost model shows higher accuracy and reliability in predicting the correction rate of carbon steel in Xinjiang. In conclusion, the GBoost model has better predictive generalization ability and model explanatory ability, and can effectively capture the complex relationship between the corrosion rate of carbon steel and environmental factors.

KEY WORDS: power grid; carbon steel; atmospheric corrosion; machine learning; service life prediction; prediction accuracy

随着特高压工程在不同地区电力输送的发展,我国输电线路长度、输电规模及电力需求日渐扩大,电网的安全问题日益突出。各类电网设备金属材料在不同的服役环境,导致输电网中金属部件存在不同的腐蚀失效问题^[1-2]。近年来,关于电网设备金属部件腐蚀失效的报道不断增多,严重影响电网的安全稳定运行^[3]。这一现象在新疆地区更为明显,新疆电网涵盖范围广,许多变电站、输电线路经过多风强风、沙漠地带,恶劣的自然环境对输变电设备产生不同程度的腐蚀。目前电网设备金属材料主要采用碳钢作为基础材料^[4],电网设备暴露于大气环境中,受大气中腐蚀介质的影响,其中的金属材料随着时间的推移不断受到腐蚀,其腐蚀类型和速度受到大气条件的显著影响^[5]。

在大气腐蚀发生的过程中,暴露在大气中的钢材会吸附空气中的水分子及颗粒物,形成薄液膜,发生化学、电化学或物理作用。这种过程会导致材料失厚或失质量,从而使其性能下降,甚至失效,广泛发生于制造业、建筑业、交通运输业等领域。金属发生腐蚀后,其表面会附着一层锈蚀层,随着腐蚀程度的加深,腐蚀产物也在不断演化。因此,正确认识并预测钢材的腐蚀行为和腐蚀速率,对于预防腐蚀失效、降低腐蚀成本具有重要意义。过去的研究通常使用室内挂片试验或室外暴露试验来进行钢材腐蚀研究,主要集中在分析钢材大气腐蚀的环境因素、材料因素和时间因素^[6]。然而,传统的长期挂片试验存在数据时间和空间上的碎片化特点,数据积累周期长、多种环境因素耦合不稳定使挂片试验所收集的数据往往具有一定的滞后性^[7]。

近年来,随着人工智能技术的迅速发展,机器学习、深度学习等技术为处理复杂的高维数据、探索特征之间的关系以及精准预测腐蚀速率提供了创新的方法和理念^[8-10]。支元杰等^[11]通过对低合金钢在6种不同环境中长期暴露后腐蚀数据的分析,建立了随机森林(Random Forest, RF)模型。研究结果显示,相较于人工神经网络、支持向量回归和逻辑回归等方法,随机森林能取得最佳的泛化能力。而后他进一步将影响碳钢的腐蚀速率和长期暴露试验中的12个环

境因素,利用随机森林和斯皮尔曼相关分析确定的关键环境变量作为输入参数,构建了SVR(Support Vector Regression)预测模型。与最大信息系数(Maximal Information Coefficient, MIC)和主成分分析(Principal Component Analysis, PCA)降维后的模型相比,采用混合方法的支持向量机模型表现出更高的预测精度^[12]。Pei等^[13]基于随机森林模型发现,相对湿度、温度和降雨量对初始大气腐蚀的影响程度高于空气中的颗粒物、二氧化硫、二氧化氮、一氧化碳和臭氧。在预测瞬时大气腐蚀方面,RF模型展现出比人工神经网络和支持向量回归模型更高的准确性。Diao等^[14]采用机器学习算法建立了腐蚀速率预测模型,并结合梯度提升决策树、Kendall相关分析以及主成分分析等特征选择方法,识别出影响腐蚀速率的关键因素,从而显著提升了模型的泛化能力。

随着各个地区、各个省份电力输送的发展,输变电电压材料面临复杂的服役环境。其中,新疆各地区气候的跨度极大,研究其环境条件下电网设备的腐蚀状态具有重大意义。本研究采用了支持向量回归、梯度提升算法、皮尔逊相关系数、逐点互信息和随机森林等5种方法对输入变量进行特征选择,旨在分析11个环境因素对碳钢腐蚀速率的影响,并进行重要性排序。首先,从这些方法中选取重要性排名前5的输入变量,并取其并集,随后使用Sobol敏感性分析方法确定关键的5个输入变量。随后,采用主成分分析进行降维处理,用构建SVR的模型和利用梯度提升算法优化损失函数构建的GBoost模型预测腐蚀速率,比较2个模型的优劣。最后,研究比较了这2种模型的性能和预测能力,以评估它们在腐蚀速率预测中的优劣表现。

1 试验

机器学习的核心在于数据集,数据集的质量、规模和代表性直接决定了模型的训练效果和应用能力。一个优秀的数据集应该涵盖目标领域内各种情况和变化,从而能够帮助模型学习并推广到新的数据。有效的数据预处理和特征工程可以显著改善模型的性

能, 训练出精度较高、泛化能力较强的模型。本节描述了腐蚀数据集和环境数据的采集和处理过程, 进一步基于环境特征与腐蚀速率的相关性进行特征降维, 为模型训练提供数据集。

1.1 数据采集

本研究采集新疆多地变电站的实验数据样本共

155 条, 样本数据包含 11 组特征, 分别为 SO_2 浓度 c_{SO_2} 、年平均温度 (t_{avg})、年均湿度 (h_{avg})、年最高温度 (t_{max})、年最低温度 (t_{min})、年降水量 (d)、年均温差 (t_{diff})、空气质量 AQI、PM10、 c_{NO_2} 和 c_{O_3} , 样本的目标变量为碳钢腐蚀速率 (v)。表 1 截取了部分数据 (10 条) 进行展示, 构建出多变量控制下的腐蚀速率预测模型。

表 1 部分腐蚀与环境数据
Tab.1 Part of corrosion and environmental data

$c_{\text{SO}_2}/(\mu\text{g}\cdot\text{m}^{-3})$	$t_{\text{avg}}/^\circ\text{C}$	$h_{\text{avg}}/\%$	$t_{\text{max}}/^\circ\text{C}$	$t_{\text{min}}/^\circ\text{C}$	d/mm	$t_{\text{diff}}/^\circ\text{C}$	AQI	PM10	c_{NO_2}	c_{O_3}	$v/(\text{mm}\cdot\text{a}^{-1})$
14	14	39.9	14.6	13.2	95.9	1.3	119	160	43.9	59	15.129 9
14	16	34.9	16.8	15.2	15.9	1.48	111	132	45	55.9	4.817 2
14	12.2	43	13	11.6	128	1.29	117	156	43.9	58	11.751 1
14	8.2	53	8.8	7.6	192	1.31	109	126	44	54.9	62.880 8
14	10.6	48	11.2	10	160	1.29	115	146	43.9	57	9.023 7
14	8.3	51	8.9	7.7	320	1.13	113	138	44	56.9	74.930 4
14	7.7	52	8.3	7.1	352	1.1	114	140	44	56.9	63.206 4
12	9	49	9.6	8.4	272	1.26	99	120	40	59.9	59.189 9
14	11.2	45	11.8	10.6	192	1.27	116	150	43.9	57	76.219 5
14	8.8	50	9.4	8.2	288	1.23	114	140	44	56.9	63.762 8

1.2 数据预处理

在数据集中常常还存在空白值(缺失值)的问题, 通常包括使用均值填充或删除包含空白值的行。在本试验中, 选择的是删除含有空白值的行, 而对于异常值的处理是基于 IQR (四分位数间距) 方法删除数据集中异常值。不同的特征可能具有不同的量纲和范围, 需要将所有变量在 0 到 1 之间归一化, 使用公式 (1)。

$$x_{\text{norm}} = \frac{x - x_{\text{norm}}}{x_{\text{max}} - x_{\text{norm}}} \quad (1)$$

归一化通过线性变换将数据映射到统一的尺度, 通常是 [0, 1] 或者 [-1, 1]。确保它们具有相同的权重, 避免某些特征由于其数值较大而主导学习过程, 加快算法的收敛速度。

2 模型和方法

在本研究过程中, 重点是研究环境因素对碳钢腐蚀速率的影响。本文涉及的特征选择方法有支持向量回归、梯度提升算法、皮尔逊相关系数、逐点互信息和随机森林。数据挖掘要求不仅分析单个因素, 还需探索各因素间的相关性以及它们对腐蚀速率的影响程度。因此, 对筛选出来的特征进行敏感性分析, 进一步过滤噪声, 提取影响腐蚀速率的关键输入向量。

2.1 特征选择

1) SVM 回归模型适用于解决非线性问题, 通过寻找最佳的超平面来使得离超平面最近的样本点到

超平面的距离最大化, 利用少量样本数据就可以有效地捕捉样本的统计特性, 提高回归分析的准确性。Wen 等^[15]用支持向量回归和粒子群优化 (Particle Swarm Optimization, PSO) 相结合的方法对 3C 钢在 5 种不同海水环境因子下的腐蚀速率进行了预测, 证明它比反向传播神经网络 (Back Propagation Neural Network, BPNN) 模型更准确, 具有优异的泛化性。

2) 梯度提升算法 (Gradient Boosting, Gboost)^[16] 是集成学习算法, 通过组合多个弱预测模型来生成更强大的预测模型, 基于梯度提升决策树的基础上进行了改进和扩展。它通过使用多棵决策树的集成来进行预测和建模, 从而显著提升预测性能。GBoost 被广泛应用于多个应用场景, 包括腐蚀预测和特征重要性计算等。Li 等^[17]利用 XGBoost 模型, 以材料成分、环境因素以及暴露时间为输入变量, 预测了不同钢材在海洋环境中水下、潮汐和飞溅区的腐蚀速率。Seghier 等^[18]开发了一种实用的集成学习模型, 用于准确预测石油和天然气管道的内腐蚀油气管道腐蚀速率, 对管道系统的安全运行和控制能力有着重要的影响。Wang 等^[19]基于统计分析、散粒噪声理论和小波分析方法, 从数据中提取了 11 个特征参数。然后, 利用梯度提升决策树 (GBDT) 模型对提取的参数组成的数据样本进行了分析。结果表明, 所提出的 GBDT 模型可以有效和准确地地区分腐蚀类型的数据样本, 包含 X65 钢和 304SS, GBDT 对腐蚀类型的判别结果与腐蚀形貌分析结果一致。可见, GBoost 在腐蚀预测方面具有较强的应用前景。

3) 皮尔逊相关系数 (Pearson Correlation Coef-

ficient, PCC) 是一种流行的线性方法, 用于检验 2 个变量之间的相关性, 适合于量化数据和连续数据的相关性分析, 由著名统计学家卡尔皮尔逊提出^[20], 见式(2)。

$$\gamma(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X]\text{Var}[Y]}} \quad (2)$$

在机器学习中, 可以使用它来计算特征与类别之间的相似性, 从而判断所提取的特征与类别之间的关系是正相关、负相关还是无相关。当皮尔逊相关系数接近 1 时, 表示 2 个数据呈正相关关系; 当接近-1 时, 表示 2 个数据呈负相关关系; 当接近 0 时, 表示 2 个数据之间没有明显的线性关系。

4) 逐点互信息 (Pointwise Mutual Information, PMI)^[21] 是一种信息论的度量, 用于衡量 2 个事件共同发生的概率与这 2 个事件独立发生的概率的乘积之比。逐点互信息 (Point-wise, MI)^[22-23] 的公式为:

$$\text{MI}(x, y) = \log \frac{P(x, y)}{P(x)P(y)} \quad (3)$$

PMI 不依赖于数据的分布或关系类型, 而是直接基于事件的共现频率, 适用于评估分类数据或文本数据中词汇和事件之间的关联程度。

5) 随机森林算法是一种结合了决策树的 bagging 集成学习算法, 它由多个树模型组成, 每个树模型都能对给定的输入特征进行预测^[24]。由于其深层次的模型结构, RF 模型对高变异性数据具有出色的处理能力。Pei 等^[25]改进后的随机森林模型在预测碳钢大气腐蚀方面表现出色, 其预测精度明显优于神经网络和支持向量回归模型。

重要性排序是指在数据分析和建模过程中, 对输入变量按其对于输出变量的影响程度进行排名, 其意义在于帮助确定哪些变量对于模型预测或解释目标变量最为重要, 从而优化模型的性能或增进对数据的理解。本研究中采用了支持向量回归、梯度提升算法、皮尔逊相关系数、逐点互信息和随机森林等方法进行特征选择。采用这些方法分别评估了每个环境因素对碳钢腐蚀速率的影响, 并对它们进行了重要性排序。通过这种方式, 研究可以确定哪些因素对腐蚀速率具有显著影响, 从而更有效地构建预测模型或进行进一步的数据分析。

2.2 变量敏感性分析

对综合了 5 种特征重要性分析得到的前 5 个输入变量取并集后, 进行 Sobol 敏感性分析, 评价各个变量对预测腐蚀速率的影响。主要目的是确定不同变量如何影响腐蚀速率的 Sobol 敏感性。

Sobol 指数法的核心思想是利用方差分解, 将函数模型分解为单个参数及其组合, 以分析参数的敏感性。该方法通过计算单个输入参数或参数集对总输出方差的贡献来评估它们对模型输出的影响程度^[26]。当

模型具有多个参数且这些参数之间可能存在相互作用时, 进行参数敏感性分析尤为重要^[27]。

2.3 腐蚀速率预测模型

支持向量机模型在处理小样本、非线性和高维模式识别问题时具有独特的优势^[28], 它会选择合适的核函数将数据集转换到更高维核空间, 之后通过找到最优的支持向量来包围转换后的数据集, 从而进行预测。基于群智能优化算法优化的 GBoost 适用于处理各种具有复杂特征关系、高维度的数据集。GBoost 在处理高维度数据时效果显著, 因为它能够有效地处理特征的稀疏性和相关性。这种组合能够提高模型的预测能力和泛化能力, 同时能够有效地应对现实世界中各种复杂的数据挑战。因此, 选择支持向量和 GBoost 模型预测本研究中用于腐蚀预测的数据集处理的工具。

本文经过 5 种特征选择分别对影响碳钢腐蚀速率的 11 个输入变量进行重要性排序, 取前 5 个输入变量, 最终合并得到 9 个变量。进一步进行 Sobol 敏感性分析, 最终得到前 5 个输入参量对碳钢腐蚀速率的敏感性, 然后分别将数据放入支持向量机和 GBoost 模型中进行预测。

将数据放入支持向量机模型训练之前, 首先进行了主成分分析降低数据集维度, PCA 通常是用于通过一种方法将多个可能相关的变量转换为较少数量的不相关变量, 降低数据维度^[29]。在主成分分析中, 通过几个主成分揭示多个变量之间潜在的内部结构, 这些主成分尽可能多地保留原始变量的信息^[30]。PCA 分析的一般步骤如下, 对于数据矩阵 \mathbf{X} , 计算其协方差:

$$\mathbf{C} = \frac{1}{N} \mathbf{X}\mathbf{X}^T \quad (4)$$

式中: N 为样本的数量。为了执行特征值分解, 以降序获得特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$, 其中 d 是样本的维数, 对应的特征向量是 w_1, w_2, \dots, w_d 。

累积贡献因子的参数定义:

$$P_{d'} = \sum_{i=1}^{d'} P_i \quad (5)$$

$$P_i = \frac{\lambda_i}{\sum_j \lambda_j} \quad (6)$$

当 $P_{d'}$ 大于或等于 90% 时, 主成分很好地代表了原始样品, d' 是此时主成分的数量。

为了比较筛选输入变量之后对训练效果提升的作用, 本文将对未筛选时的 11 个变量进行敏感性分析, 得到前 5 的变量分别放入支持向量机预测腐蚀模型和集成学习 GBoost 模型进行训练, 支持向量机预测腐蚀模型基本建模步骤如下:

首先, 数据预处理后对于训练集与测试集的划分。先打乱数据集, 将数据分割为训练集和测试集, 其中训练集用于训练算法模型, 测试集用于评估模型性能, 划分比例是 80% 训练集和 20% 测试集。

然后, 主成分分析选择保留前 3 的主成分, 将输入向量维度降低至三维, 随后寻找最优解参数。第 1 轮寻优, 惩罚系数 C 的值, 范围为 $10^{-2} \sim 10^2$, 共 11 个值; γ 参数的值, 范围为 $2^{-5} \sim 2^5$, 共 11 个值, 其他剩余的参数在 Python scikit-learning 机器学习库中设置为默认值。经过网格搜索 (Grid Search) 10 折交叉验证来寻找最佳的 SVR 模型参数组合, 得到最优惩罚系数 C 和 γ , 进入第 2 轮寻优。同样的过程, 缩小惩罚系数 C 和 γ 的范围, 进一步获得最优解并绘制热度图。

最后, 测试模型。使用训练集训练机器学习算法模型, 并生成训练集的预测值。将数据反归一化后计算预测值的误差评价指标, 即决定系数 R^2 、均方跟误差 R_{RMSE} , 见式 (7)、(8), 并绘制出结果的训练图和预测图。

$$R^2 = 1 - \frac{\sum_{n=1}^N (\hat{y}_n - y_n)^2}{\sum_{n=1}^N (\bar{y}_n - y_n)^2} \quad (7)$$

$$R_{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{N}} \quad (8)$$

R^2 值越接近 1, 说明模型能够更好地解释目标变量的变化, 拟合度越好; 接近 0 则表示模型的预测能力较弱。RMSE 值是实际观测值与预测值之间差异的平方和的平均值取平方根, RMSE 值越小越好, 因为它衡量了模型预测值与实际观测值之间的离散程度。

GBoost 模型的建模基本步骤与 SVR 模型类似如下: 先对数据进行预处理, 划分数据后, 初始化基础模型, 建立一个简单的回归树, 通常使用梯度下降 (Gradient Descent) 来优化损失函数, 确定最小化损失函数, 通过迭代增强调整权重优化模型, 同样使用 10 折交叉验证来调整 GBoost 的最优超参数组合, 最后测试模型。

3 结果与讨论

3.1 误差分析

在确定关键环境因子时, 需要考虑不同因子之间的相关性。密切相关的因子可能会包含重复的信息, 这可能会降低模型的准确性。通过计算每对环境因子之间的 PMI, 获得的相关性热图如图 1 所示。二氧化氮与空气质量指数之间的相关性约为 87%。这表明, NO_2 浓度的变化往往会显著影响到空气质量的整体评估。另外, 平均温度与最高温度以及最低温度之间的相关性超过 68%。这意味着当平均温度发生变化时, 最高和最低温度通常也会有相似的趋势变化。

通过计算每对环境因子之间的 PCC, 获得的相关性热图如图 2 所示。与图 1 中的 PMI 相关热度图结果一致, 空气质量指数与二氧化氮之间的密切相关性为 86%。此外, 平均温度与最高温度以及最低温度之

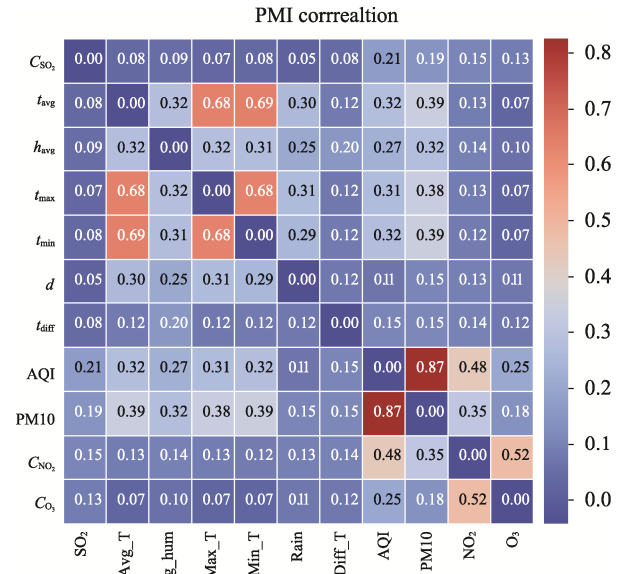


图 1 PMI 相关热度图
Fig.1 PMI related heat map

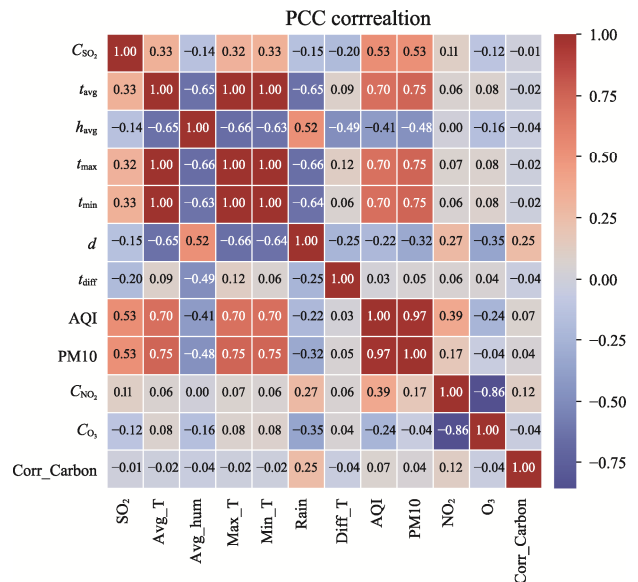


图 2 PCC 相关热度图
Fig.2 PCC related heat map

间的相关性达到 100%。然而, 不同于皮尔逊相关系数特征选择后的热度图, 逐点互信息绘制的热度图增加了一个额外相关性项目, 即各个变量与腐蚀速率的相关性。如图 2 所示, 降雨量与腐蚀速率的相关性最高, 达到 25%, 其次是二氧化氮的浓度。相比之下, 平均温度、最高温度和最低温度与腐蚀速率的相关性仅为 2%。因此, 根据表 2 的数据, 决定将平均温度、最高温度和最低温度从分析中剔除。

采用支持向量回归 (SVR)、梯度提升算法 (GBoost)、皮尔逊相关系数 (PCC)、逐点互信息 (PMI) 和随机森林筛选环境变量, 通过计算 11 个环境变量对腐蚀速率的重要性, 筛选出影响碳钢腐蚀的前 5 环境变量, 即: d 、 h_{avg} 、 t_{diff} 、PM10、 c_{SO_2} 、 c_{O_3} 、AQI、 c_{NO_2} 及 t_{avg} 。

表 2 重要性排序
Tab.2 Order of importance

SVR		Gboost		PCC		PMI		RF	
环境变量	重要占比/%	环境变量	重要占比/%	环境变量	重要占比/%	环境变量	重要占比/%	环境变量	重要占比/%
<i>d</i>	1.29	<i>d</i>	111.54	<i>d</i>	25.42	AQI	72.40	<i>d</i>	27.91
<i>t_{diff}</i>	0.16	<i>t_{avg}</i>	55.05	<i>c_{NO₂}</i>	12.20	PM10	67.70	<i>c_{O₃}</i>	16.62
AQI	0.13	<i>c_{O₃}</i>	17.24	AQI	6.79	<i>t_{diff}</i>	64.30	<i>t_{diff}</i>	9.09
PM10	0.05	<i>c_{SO₂}</i>	13.96	<i>c_{O₃}</i>	4.30	<i>h_{avg}</i>	58.80	AQI	8.99
<i>h_{avg}</i>	0.04	<i>t_{diff}</i>	12.73	<i>t_{diff}</i>	4.12	<i>c_{NO₂}</i>	58.50	PM10	8.67
<i>c_{SO₂}</i>	0.00	AQI	7.73	PM10	4.00	<i>c_{O₃}</i>	55.90	<i>h_{avg}</i>	5.83
<i>c_{O₃}</i>	-0.02	<i>c_{NO₂}</i>	6.14	<i>h_{avg}</i>	3.95	<i>t_{min}</i>	51.50	<i>t_{min}</i>	5.62
<i>t_{max}</i>	-0.06	PM10	5.93	<i>t_{max}</i>	2.09	<i>t_{avg}</i>	51.40	<i>t_{max}</i>	5.45
<i>t_{avg}</i>	-0.07	<i>h_{avg}</i>	4.34	<i>t_{avg}</i>	2.09	<i>t_{max}</i>	50.60	<i>t_{avg}</i>	5.16
<i>t_{min}</i>	-0.07	<i>t_{min}</i>	1.54	<i>t_{min}</i>	1.66	<i>d</i>	46.40	<i>c_{NO₂}</i>	4.83
<i>c_{NO₂}</i>	-0.26	<i>t_{max}</i>	1.16	<i>c_{SO₂}</i>	1.23	<i>c_{SO₂}</i>	38.90	<i>c_{SO₂}</i>	1.82

3.2 敏感性分析

综合上述前 5 种特征重要性分析的结果,可得到共计 9 个变量的并集。使用 SALib 中的 Sobol 方法进行全局灵敏度分析后,得到了特征敏感性指标,见表 3。第 1 列列出了各个变量,第 2 列展示了每个环境参数对于最终输出的第一阶灵敏度指数,即考虑单个输入变量时该变量对输出的影响程度。前 5 个敏感性较高的特征依次是:*d*、*h_{avg}*、*t_{diff}*、PM10、*c_{SO₂}*。第 3 列显示了每个环境参数对于最终输出(可能是某种影响或结果)的总体灵敏度指数。具体来说,它们表示每个参数在模型输出中的贡献程度,敏感性排序前 5 项的特征依次为:*d*、*h_{avg}*、*t_{diff}*、PM10、*c_{O₃}*。结合图 2 所示信息,最终选出敏感性分析前 5 的特征为总体灵敏度指数前 5 特征:*d*、*h_{avg}*、*t_{diff}*、PM10、*c_{O₃}*。

通常情况下,钢材在大气中的腐蚀过程属于电化学腐蚀。这是一种在薄液膜下经历干湿交替的电化学反应。在这种环境中,溶解的气体如 CO₂、NO₂、SO₂,都会加速钢材的腐蚀进程。研究发现,影响碳钢大气腐蚀的 5 个重要特征依次为年降水量、年均湿度、年均温差、PM10 和 O₃,与华南地区筛选出的关键特征——温度、相对湿度、二氧化硫浓度、二氧化氮浓度和臭氧浓度有所不同。这种差异可能源于新疆地区的气候特点:温差大、降水少和湿度较低,表明钢材在大气中的腐蚀确实受到多种环境因素的影响。与支元杰等^[11]初期对中国 10 个城市的 12 个环境参数进行了长达 8 a 的关键因子的筛选结果类似,在 8 a 的暴露测试期间,RH、pH 和 TEMP 都表现出与腐蚀速率的强相关性,SO₂ 仅在第 1 年被选中。

华南地区与新疆地区在气候、地理和污染源上存在显著差异。华南地区的气候较为湿润,降水量较大,而新疆则属于干旱或半干旱地区,年降水量相对较少。这种气候差异会直接影响到腐蚀环境的形成和发展。华南地区的高湿度和降水频率可能促使酸性气体

(如 SO₂ 和 NO₂) 在空气中与水分结合,形成酸性液膜,加速腐蚀。新疆的年均温差较大,这种温差可能导致温度变化引起的湿度变化,以及水膜的干湿循环,可能导致钢材表面的微小裂缝,从而加速腐蚀的发生。湿度直接影响到表面水膜的形成和厚度。在新疆,尤其是内陆地区,湿度相对较低,这可能会限制腐蚀介质的存在。而在湿度较高的瞬间或季节,腐蚀仍可能加速。因此,年均湿度在腐蚀过程中的重要性体现在它对水膜形成的影响上。SO₂、NO₂ 和 CO₂ 等酸性气体在新疆地区的浓度相对较低,这可能是由于该地区相对较少的工业排放和污染源。这意味着这些气体对腐蚀的贡献相对较小。因此,尽管它们在腐蚀机理中扮演着重要角色,但在新疆的实际情况下,其影响可能被年降水量、湿度和温差等因素所掩盖。

表 3 特征敏感性指标
Tab.3 Sensitivity index of features %

Variables	1st order Sobol index	Total Sobol index
<i>d</i>	23.69	56.13
<i>h_{avg}</i>	13.35	25.86
<i>t_{diff}</i>	5.39	16.59
PM10	4.44	13.98
<i>c_{O₃}</i>	4.22	10.71
<i>t_{avg}</i>	2.63	6.12
AQI	2.15	8.73
<i>c_{NO₂}</i>	1.91	7.30
<i>c_{SO₂}</i>	1.44	12.79

3.3 模型预测分析

经过第 2 轮参数优化后的支持向量机(SVR)模型的热度图如图 3 所示,展示了不同的参数 C 和 gamma 组合对模型性能的影响。从图 3 中可观察到,当使用 11 个变量时,最优的惩罚系数 C=0.215 4, gamma 值为 0.793 7;而使用 5 个变量时,最优的惩

罚系数 $C=0.1$, γ 值为 0.793 7。这些结果为进一步调整模型参数和优化预测性能提供了重要参考。

经优化后, SVR 和 GBoost 腐蚀预测模型训练和预测的拟合结果如图 4 所示, 每个图的横坐标是实际

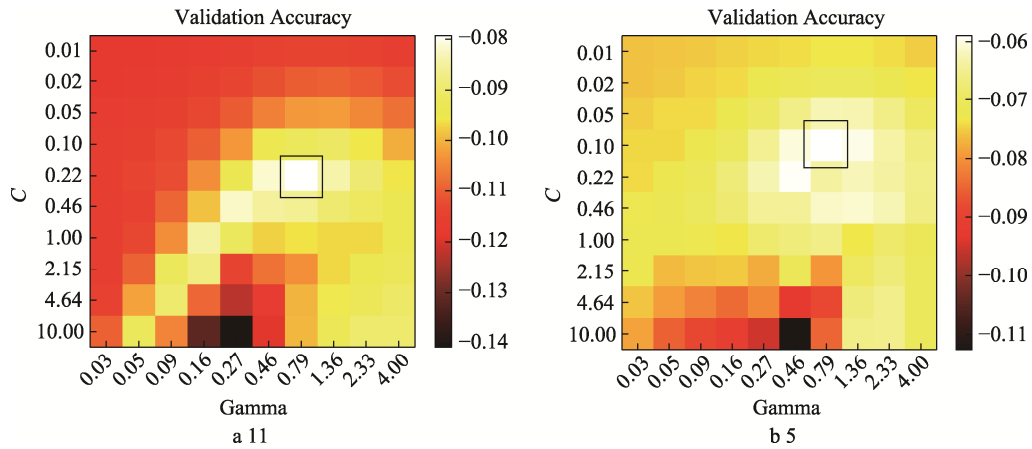


图 3 不同个数变量的寻优 C 和 γ 图
Fig.3 Optimization of C and γ plots for different variables

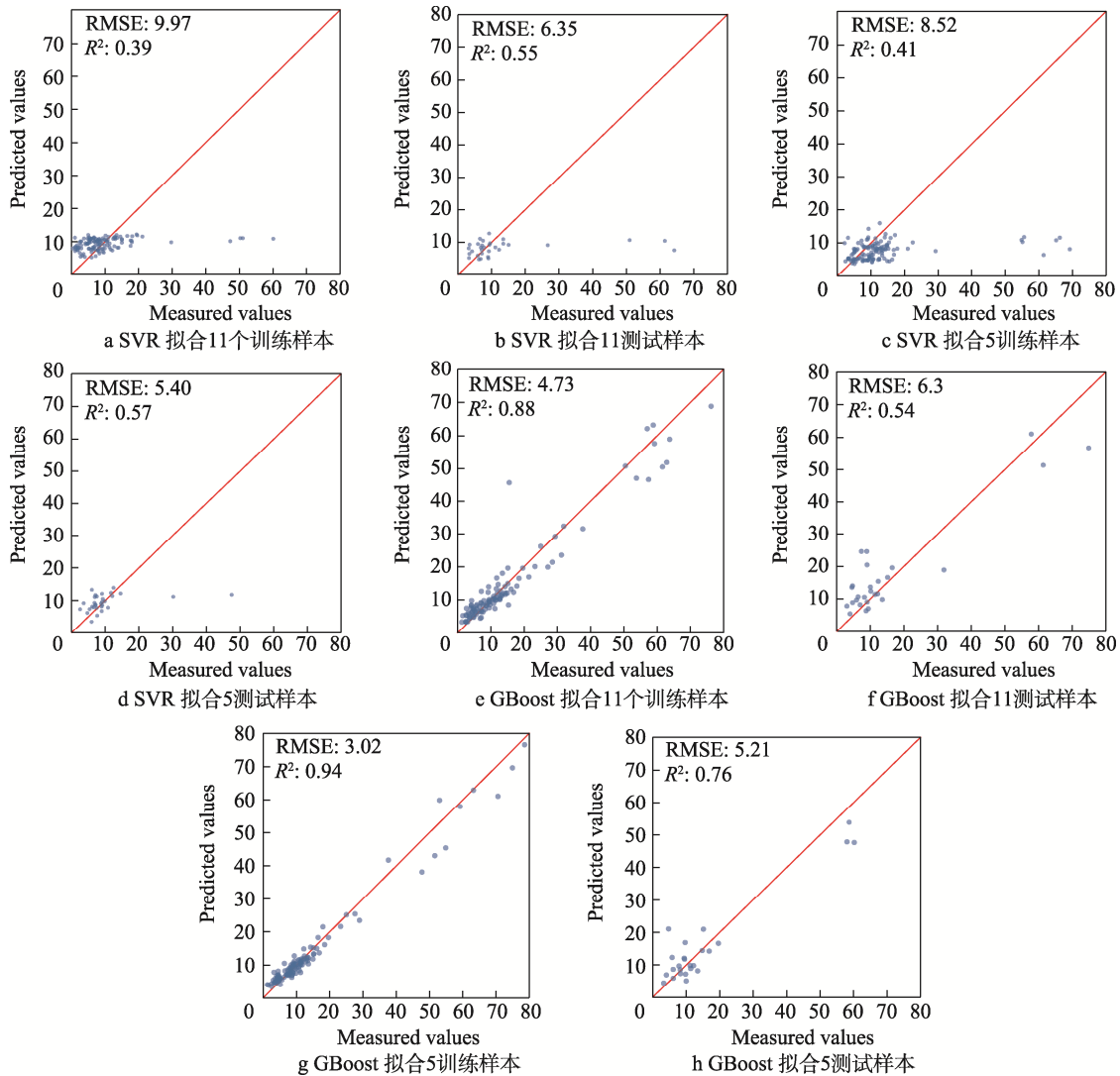


图 4 通过 SVR 和 GBoost 模型对 11 和 5 个训练样本、测试样本进行拟合

Fig.4 Fitting of 11 and 5 training samples and test samples by SVR and GBoost models: a) fitting of 11 training samples by SVR; b) fitting of 11 test samples by SVR; c) fitting of 5 training samples by SVR; d) fitting of 5 test samples by SVR; e) fitting of 11 training samples by GBoost; f) fitting of 11 test samples by GBoost; g) fitting of 5 training samples by GBoost; h) fitting of 11 test samples by GBoost

值,而纵坐标表示基于模型预测值。对角线表示预测值等于相应真实值的预测线,靠近红色对角线的点代表较小的预测误差。根据数据分析,当从 11 个变量减少到 5 个变量时,SVR 和 GBoost 腐蚀预测模型的表现都有所改善。具体来说, RMSE 和 R^2 这 2 个评估指标显示了明显的改善趋势,表明模型在训练数据上对输出变量(碳钢腐蚀速率)的解释能力得到了显著增强。

对于包含 5 个变量的预测模型,SVR 模型的最小均方根误差为 5.40,最小 R^2 值为 0.57;而 GBoost 模型的最小均方根误差为 5.21,最小 R^2 值为 0.76。相较于 SVR 模型,GBoost 模型在本次学习中表现更佳,具有更小的均方根误差和更高的决定系数。表 4 展示了部分实际碳钢腐蚀速度数据与模型计算结果的对比。通过对输入变量和目标变量的分析,可以更直观地观察到 GBoost 模型在预测碳钢腐蚀速率方面的高准确性,具有更强的预测泛化能力。

表 4 部分实际碳钢腐蚀速度与模型计算结果对比

Tab.4 Comparison between the actual corrosion rate of carbon steel with the calculated results of the model

SVR		GBoost	
Input variables	Target variable	Input variables	Target variable
2.4	11.27	4.49	7.62
4.95	8.99	5.89	5.81
5.21	9.69	7.58	6.70
7.12	10.22	8.36	7.80
9.97	12.09	9.34	8.64
10.96	13.84	11.69	10.89
12.71	10.60	12.71	10.68
14.42	13.78	14.42	12.38
32.95	12.39	59.32	45.15
63.22	12.70	62.88	58.77

总体而言,通过减少变量数目,并应用机器学习模型进行预测,可以提高模型的简洁性和解释力,同时增强预测的准确性和泛化能力,这对于理解环境因素对碳钢腐蚀速率影响的研究具有重要意义。因此,可以得出结论,对碳钢腐蚀影响较大的前 5 个特征是 d 、 h_{avg} 、 t_{diff} 、PM10、 CO_3 。

4 结论

1) 经过 5 种特征选择和敏感性分析后确定的结果显示,对碳钢大气腐蚀影响较为重要的前 5 个特征依次是年降水量、年均湿度、年均温差、PM10 和 O_3 。这些特征被认定为在影响模型预测碳钢腐蚀过程中起着关键作用。

2) GBoost 模型相比于支持向量模具有更好的预测泛化能力和模型解释力,能够更有效地捕捉碳钢腐

蚀速率与环境因素之间的复杂关系。该模型通过对输入数据的有效处理和参数调整,使其能够在验证集上达到较高的决定系数 (R^2),表明模型具有较好的泛化能力。

参考文献:

- [1] 夏晓健,金焱,乔汉文,等. 输变电设备腐蚀状况调查与分析[J]. 腐蚀科学与防护技术, 2019, 31(2): 121-127. XIA X J, JIN Y, QIAO H W, et al. Survey on Corrosion of Power Transmission and Transformation Equipment[J]. Corrosion Science and Protection Technology, 2019, 31(2): 121-127.
- [2] 郝文魁,陈新,徐铃铃,等. 电网碳钢、镀锌钢大气腐蚀等级图绘制研究[J]. 中国腐蚀与防护学报, 2023, 43(4): 795-802. HAO W K, CHEN X, XU L L, et al. Drawing of Atmospheric Corrosion Map of Carbon Steel and Galvanized Steel for Power Grid[J]. Journal of Chinese Society for Corrosion and Protection, 2023, 43(4): 795-802.
- [3] 陈云翔,倪清钊,林德源,等. 大气环境下电网设备金属材料的腐蚀及服役寿命预测研究进展[J]. 材料导报, 2016, 30(21): 89-96. CHEN Y X, NI Q Z, LIN D Y, et al. Research Progress in Corrosion and Service Life Prediction of Metal Materials in Grid Equipment under Atmospheric Environment[J]. Materials Review, 2016, 30(21): 89-96.
- [4] 刘栓,周开河,方云辉,等. 石墨烯重防腐涂层在国网输电铁塔防护的应用研究[J]. 中国材料进展, 2017, 36(6): 442-447. LIU S, ZHOU K H, FANG Y H, et al. Research on the Graphene Based Heavy Coating in Protection of Transmission Power Tower[J]. Materials China, 2017, 36(6): 442-447.
- [5] 李文翰,尹学涛,周学杰,等. 电网输变电设备钢结构和镀锌构件的大气腐蚀与防护措施[J]. 材料保护, 2018, 51(10): 114-118. LI W H, YIN X T, ZHOU X J, et al. Summary on Atmospheric Corrosion and Protection Measure of Steel Components and Galvanized Components for Transmission and Distribution Projects[J]. Materials Protection, 2018, 51(10): 114-118.
- [6] 王莎莎,马帅杰,车琨,等. 机器学习在自然环境腐蚀评估与预测领域的应用现状[J]. 中国腐蚀与防护学报, 2023, 43(3): 441-451. WANG S S, MA S J, CHE K, et al. Application Status of Machine Learning in Field of Natural Environment Corrosion Assessment and Prediction[J]. Journal of Chinese Society for Corrosion and Protection, 2023, 43(3): 441-451.
- [7] 裴梓博. 碳钢大气环境腐蚀大数据研究及主要影响因素作用规律[D]. 北京: 北京科技大学, 2021. PEI Z B. Research on Big Data of Atmospheric Corrosion of Carbon Steels and Action Law of Main Influencing

- Factors[D]. Beijing: University of Science and Technology Beijing, 2021.
- [8] COELHO L B, ZHANG D W, VAN INGELGEM Y, et al. Reviewing Machine Learning of Corrosion Prediction in a Data-Oriented Perspective[J]. *NPJ Materials Degradation*, 2022, 6: 8.
- [9] 谢建新, 宿彦京, 薛德祯, 等. 机器学习在材料研发中的应用[J]. *金属学报*, 2021, 57(11): 1343-1361.
XIE J X, SU Y J, XUE D Z, et al. Machine Learning for Materials Research and Development[J]. *Acta Metallurgica Sinica*, 2021, 57(11): 1343-1361.
- [10] FU H D, ZHANG H T, WANG C S, et al. Recent Progress in the Machine Learning-Assisted Rational Design of Alloys[J]. *International Journal of Minerals, Metallurgy and Materials*, 2022, 29(4): 635-644.
- [11] ZHI Y J, FU D M, ZHANG D W, et al. Prediction and Knowledge Mining of Outdoor Atmospheric Corrosion Rates of Low Alloy Steels Based on the Random Forests Approach[J]. *Metals*, 2019, 9(3): 383.
- [12] ZHI Y J, JIN Z H, LU L, et al. Improving Atmospheric Corrosion Prediction through Key Environmental Factor Identification by Random Forest-Based Model[J]. *Corrosion Science*, 2021, 178: 109084.
- [13] PEI Z B, ZHANG D W, ZHI Y J, et al. Towards Understanding and Prediction of Atmospheric Corrosion of an Fe/Cu Corrosion Sensor *via* Machine Learning[J]. *Corrosion Science*, 2020, 170: 108697.
- [14] DIAO Y P, YAN L C, GAO K W. Improvement of the Machine Learning-Based Corrosion Rate Prediction Model through the Optimization of Input Features[J]. *Materials & Design*, 2021, 198: 109326.
- [15] WEN C, ZHANG Y, WANG C X, et al. Machine Learning Assisted Design of High Entropy Alloys with Desired Property[J]. *Acta Materialia*, 2019, 170: 109-117.
- [16] HUANG L Y, CHEN W T, HAO W K, et al. Self-Healing Anti-Corrosion Coatings: A Mechanism Study Using Computational Materials Science[J]. *Electrical Materials and Applications*, 2024, 1(1): e12006.
- [17] LI Z, LONG Z L, LEI S, et al. Evaluating the Corrosion Resistance of Marine Steels under Different Exposure Environments *via* Machine Learning[J]. *Physica Scripta*, 2023, 98(1): 015402.
- [18] EL AMINE BEN SEGHIER M, HÖCHE D, ZHELUDKEVICH M. Prediction of the Internal Corrosion Rate for Oil and Gas Pipeline: Implementation of Ensemble Learning Techniques[J]. *Journal of Natural Gas Science and Engineering*, 2022, 99: 104425.
- [19] LI Q S, WANG J H, WANG K, et al. Determination of Corrosion Types from Electrochemical Noise by Gradient Boosting Decision Tree Method[J]. *International Journal of Electrochemical Science*, 2019, 14(2): 1516-1528.
- [20] 孙曙光, 纪学玲, 杜太行, 等. 机械振动下交流接触器电寿命预测失效特征量提取[J]. *仪器仪表学报*, 2019, 40(3): 114-125.
- SUN S G, JI X L, DU T H, et al. Failure Feature Extraction for Electrical Life Prediction of AC Contactor under Mechanical Vibration[J]. *Chinese Journal of Scientific Instrument*, 2019, 40(3): 114-125.
- [21] MILANI A, RAJDEEP N, MANGAL N, et al. Sentiment Extraction and Classification for the Analysis of Users Interest in Tweets[J]. *International Journal of Web Information Systems*, 2018.
- [22] CHEN Y N, YE W B, LI D. Spectral Clustering Community Detection Algorithm Based on Point-Wise Mutual Information Graph Kernel[J]. *Entropy*, 2023, 25(12): 1617.
- [23] GONG L J, GAO D G, ZHANG X X, et al. REL-NPMI: Exploring Genotype and Phenotype Relationship of Pancreatitis Based on Improved Normalized Point-by-Point Mutual Information[J]. *Computers in Biology and Medicine*, 2023, 158: 106868.
- [24] 张明, 付冬梅, 张达威, 等. 基于综合智能模型的碳钢大气腐蚀重要变量提取和依赖关系挖掘[J]. *工程科学学报*, 2023, 45(3): 407-418.
ZHANG M, FU D M, ZHANG D W, et al. Extraction of Important Variables and Mining of Dependencies of Atmospheric Corrosion of Carbon Steel Based on a Comprehensive Intelligent Model[J]. *Chinese Journal of Engineering*, 2023, 45(3): 407-418.
- [25] PEI Z B, CHENG X Q, YANG X J, et al. Understanding Environmental Impacts on Initial Atmospheric Corrosion Based on Corrosion Monitoring Sensors[J]. *Journal of Materials Science & Technology*, 2021, 64: 214-221.
- [26] 任启伟, 陈洋波, 周浩澜, 等. 基于 Sobol 法的 TOPMODEL 模型全局敏感性分析[J]. *人民长江*, 2010, 41(19): 91-94.
REN Q W, CHEN Y B, ZHOU H L, et al. Global Sensitivity Analysis of TOPMODEL Parameters Based on Sobol Method[J]. *Yangtze River*, 2010, 41(19): 91-94.
- [27] MOUSIVAND A, MENENTI M, GORTE B, et al. Global Sensitivity Analysis of the Spectral Radiance of a Soil-Vegetation System[J]. *Remote Sensing of Environment*, 2014, 145: 131-144.
- [28] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述[J]. *电子科技大学学报*, 2011, 40(1): 1-10.
DING S F, QI B J, TAN H Y. An Overview on Theory and Algorithm of Support Vector Machines[J]. *Journal of University of Electronic Science and Technology of China*, 2011, 40(1): 1-10.
- [29] SONG D D, GAO J, LI X G, et al. Evaluation of Aging Behavior of Polypropylene in Natural Environment by Principal Component Analysis[J]. *Polymer Testing*, 2014, 33: 131-137.
- [30] LU L, LI X G, GAO J. Evaluation of Aging Behavior of Medium Density Polyethylene in Natural Environment by Principal Component Analysis[J]. *Journal of Applied Polymer Science*, 2012, 125(1): 19-23.