# 钢材大气腐蚀预测算法模型比选

沈 坚<sup>1,2</sup>,吴柯娴<sup>1,2,3\*</sup>,何晓宇<sup>1,2</sup>,方兴龙<sup>4</sup>

3. 浙江大学结构工程研究所, 杭州 310058; 4. 浙江海港内河港口发展有限公司, 杭州 310005)

[摘 要]针对我国各地区的各类钢材,提出了一种基于机器学习算法的钢材大气腐蚀深度预测方法。首先,收 集了我国 10 个大气暴露站的腐蚀检测数据、环境特征和材料特征,采用规范公式与 6 种机器学习算法预测钢材腐蚀深 度。然后分析预测误差,筛选适用于钢材大气腐蚀的预测模型。进一步分析材料与环境特征敏感性,揭示影响钢材大 气腐蚀的主要材料与环境因素。结果表明,相比于规范公式,应用随机森林(RF)和长短期记忆循环神经网络 (LSTM)算法的预测模型精度大幅提升;除了规范公式中的各类特征外,有关雨水酸碱性和雨水腐蚀性离子浓度的特 征对钢材腐蚀行为有较大影响,应予以考虑。

[关键词]钢材;大气暴露站;机器学习;特征敏感性

### 0 引言

钢材大气腐蚀普遍存在于制造业、建筑业、 交通运输业等领域。准确认识并预测钢材的腐 蚀行为和失厚速率对于预防腐蚀失效、降低腐 蚀成本具有重要意义。以往研究多采用室内加 速试验或室外暴露试验数据开展钢材腐蚀研究, 重点分析了钢材大气腐蚀的环境因素、材料因 素和时间因素<sup>[1]</sup>。为了进一步获得具有地方腐 蚀特性的长期室外暴露腐蚀数据,中国科技部 国家材料环境腐蚀野外科学观测研究平台自 1984年在全国多个具有典型气候特征的地方建 立大气腐蚀站点,并在每个站点开展了多种金 属材料的室外大气挂片试验,目前已经积累了 大量的腐蚀数据。

近年来,随着人工智能的兴起,机器学习、 深度学习等技术为处理高维数据、挖掘特征关 系、精准预测腐蚀速率提供了新的思路<sup>[2-4]</sup>。支 元杰<sup>[5]</sup> 以17种低合金钢在6个大气站共409项 大气腐蚀数据为样本,采用线性回归、人工神 经网络、随机森林 (random forest, RF)、支持 向量机 (support vector machines, SVM) 和基于 加权 K 近邻的随机森林动态集成选择算法 (RF -WKNNs)进行分析,给出了大气环境特征对 腐蚀速率影响的定量分析。进一步提出全连接 级联动集成选择森林算法 (DCCF-WKNNs) 并 预测腐蚀失厚率。结果表明,改进后的模型对 于大气环境下的小样本低合金钢腐蚀数据预测 具有较高准确性和泛化性,重要性较高的环境 特征为雨水 pH 值、SO, 浓度、Cl<sup>-</sup>浓度、年平 均相对湿度(RH)、年平均温度(T)和降雨 量。其另一篇论文<sup>[6]</sup>对 Q235 钢的大气腐蚀数 据分析表明贡献度较高的环境特征还包括 NO。 和自然降尘量 (水溶性和非水溶性)。但以上 研究中采用的大气站传统检测数据量较小,环 境和材料特征相对较少,未考虑具有特殊气候 大气站的腐蚀数据(如万宁等)。此外,对于 深度学习算法,如循环神经网络、卷积神经网 络 (convolutional neural networks, CNN) 的应 用尚未开展。

针对以上不足,本文收集了我国10个大气

# • 沩江贡盈千枝 •

<sup>(1.</sup> 浙江数智交院科技股份有限公司,杭州 310030;

<sup>2.</sup> 综合交通运输理论交通运输行业重点实验室, 杭州 310030;

收稿日期: 2023-10-30

作者简介: 沈坚(1969-), 男, 正高级工程师, 从事公路、水运交通工程建设与管理的研究。

**化**科技创新 Citchuangxin

站的腐蚀检测数据、环境特征以及材料组成特征,采用6种机器学习算法预测我国大气站环境腐蚀性等级与钢材腐蚀速率,分析大气腐蚀规范公式与各类机器学习算法对腐蚀深度预测的误差,明确泛化性更强、准确度更高的腐蚀深度预测模型,揭示影响腐蚀的主要环境与材料组成特征。

## 1 数据采集与处理

### 1.1 数据采集

采集我国 10 个大气站 1984-2016 年的腐蚀 数据与环境特征,数据来源为中国腐蚀与防护 网数据管理系统。其中 10 个大气站分别为万 宁、江津、青岛、琼海、广州、北京、武汉、 沈阳、敦煌、拉萨。

腐蚀数据共 1364 组,每组包括钢材牌号、 钢材类别、试验地点、试验开始时间、试验结 束时间、试验周期 t/month 和腐蚀失厚率 r<sub>corr</sub>/ (μm·a<sup>-1</sup>)。钢材牌号共 34 种,可分为 9 类, 分别为碳素结构钢 (273 组)、优质碳素结构钢 (103 组)、低合金钢 (411 组)、不锈钢 (107 组)、耐候钢 (348 组)、普通冷轧钢 (37 组)、 细晶粒结构钢 (37 组)、弹簧钢 (27 组)和稀 土结构钢 (21 组),其中低合金钢、耐候钢、 碳素结构钢约占总体的 76%。经统计,10 个大 气站的腐蚀数据箱型图见图 1。可以看出,腐蚀 失厚率变化幅度较大,万宁站、青岛站、琼海 站和沈阳站离群值较多,其中万宁站和沈阳站 腐蚀失厚率的变异系数达到 0.95 以上,不确定 性较高。

环境数据共 21768 个,包括试验地点、试 验时间、海拔 H/m、温度 T/℃、湿度 RH/%、 降水量  $P_{rain}$ /mm、海盐粒子沉积率  $D_{seasalt}$ /  $(10^{-2}mg \cdot cm^{-2} \cdot d^{-1})、硫酸盐化速率 <math>D_{sulfation}$ /  $(10^{-2}mg \cdot cm^{-2} \cdot d^{-1})$ 、NH<sub>3</sub> 沉积率  $D_{NH3}$ /  $(10^{-2}mg \cdot cm^{-2} \cdot d^{-1})$ 、NO<sub>2</sub> 沉积率  $D_{NO2}$ /  $(10^{-2}mg \cdot cm^{-2} \cdot d^{-1})$ 、H<sub>2</sub>S 沉积率  $D_{H2S}$ /  $(10^{-2}mg \cdot cm^{-2} \cdot d^{-1})$ 、雨水 pH 值  $R_{pH}$ 、雨水 SO<sub>4</sub><sup>2-</sup>浓度  $R_{SO4}$ /  $(mg \cdot m^{-3})$ 、雨水 CI<sup>-</sup>浓度  $R_{CI}$ /  $(mg \cdot m^{-3})$ 、自然降尘量  $D_{dust}/(g \cdot cm^{-2} \cdot month^{-1})$ 、日照时数  $t_{sun}/h$  和平均风速  $V_{wind}/(m \cdot s^{-1})$ 。环境数据中气候环境信息和各类污染物 沉积速率或浓度共 15 项。



#### 1.2 数据处理

处理腐蚀数据和环境数据,得到可用于机 器学习算法的数据集。首先提取腐蚀数据中的 钢材牌号,使用化学元素含量来表示各牌号钢 材之间的差异.化学成分包括.铁Fe、碳C、 磷P、硫S、硅Si、镍Ni、铬Cr、铜Cu、锰 Mn、钼 Mo、铝 Al、钒 V、钛 Ti 和铌 Nb。将腐 蚀失厚率乘以试验周期得到腐蚀深度 D\_\_\_\_/ (µm) 作为输出量进行训练和测试。进一步计 算皮尔逊相关系数矩阵以明确输入量与输出量 的相关关系,其中时间特征(试验周期)与输 出量的相关系数为 0.36、材料特征、环境特征 与输出量的相关系数矩阵热力图见图 2。从中可 以看出铁与铬、镍相关系数较高,可达 0.97, 降水量与温度、湿度相关系数较高,可达0.93。 为了提高输入量的质量,应剔除数据集中相关 性过高的特征,因此删除铁含量与降水量这两 项输入特征。

经过上述处理,可构建一个包含 1364 条腐 蚀样本的数据集,每个样本包含 1 项时间特征 (试验周期)、13 项材料特征(化学成分含量)、 14 项环境特征,共计 28 维特征作为输入量,1 项腐蚀深度作为输出量。

• 浙江委备册技 •









图 2 腐蚀深度与特征的相关系数矩阵热力图

## 2 预测模型与方法

## 2.1 规范公式

依据《金属和合金的腐蚀 大气腐蚀性 第1 部分:分类、测定和评估》(GB/T 19292. 1-2018)<sup>[7]</sup>中的规定,碳钢腐蚀深度与暴露时间、 化学成分含量、环境参数等 12 维数据有关。碳 钢大气腐蚀深度 *D*<sub>ear</sub> 为

 $D_{\rm corr} = r_{\rm corr, 1} \cdot t^{\left(0.569 + \sum b_i w_i + 0.0845 S_d^{0.36}\right)}$ (1)

 $r_{\text{corr, 1}} = 1.77 \cdot P_{\text{d}}^{0.52} \cdot \exp(0.020 \cdot \text{RH} + f_{\text{st}}) +$ 

0.102 · S<sup>0.62</sup> · exp(0.033 · RH + 0.040 · T) (2)
式中: D<sub>corr</sub> 为碳钢腐蚀深度, μm; t 为暴晒
时间, a; r<sub>corr,1</sub> 为第一年腐蚀速率, μm/a; T 为
年平均温度, C; RH 为年平均湿度,%; P<sub>d</sub> 为年

平均 SO<sub>2</sub> 沉积率, mg/m<sup>2</sup> · d;  $S_d$  为年平均 Cl<sup>-</sup>沉 积率, mg/m<sup>2</sup> · d;  $f_{st}$  为碳钢相关系数, 可采用 下式计算:

$$f_{\rm st} = \begin{cases} 0.150 \cdot (T - 10) & T \le 10 \ ^{\circ}{\rm C} \\ -0.054 \cdot (T - 10) & T > 10 \ ^{\circ}{\rm C} \end{cases}$$
(3)

式中: w<sub>i</sub> 为第 i 个化学元素的质量百分比; b<sub>i</sub> 为第 i 个化学元素的影响系数<sup>[8]</sup>。根据 GB/T 19292. 2-2018<sup>[8]</sup> 规定,暴露 20 年之后的腐蚀 速率是线性的,总腐蚀深度为

 $D_{\text{corr}}(t>20) = r_{\text{corr},1} \cdot [20^{b} + b(20^{b-1}) \cdot (t-20)]$ (4) 2.2 机器学习算法

本文涉及的机器学习算法为有监督学习算 法 (回归),包括 SVM、RF、径向基函数神经 网络 (radial basis function neural networks, RBFNN)、BP 神经网络 (back propagation neural networks, BPNN)、长短期记忆 (long short term memory, LSTM) 循环神经网络和 CNN。基本建 模步骤如下:首先,划分训练集与测试集。打 乱数据集,将数据划分为用于训练算法模型的 训练集和用于评估模型性能的测试集, 划分比 例取70%和30%。将测试集和训练集数据归一 化到区间 [0,1],避免数据的尺度效应。由于 万宁站腐蚀失厚率高,远大于其他大气站,因 此将万宁站与其他大气站数据集分开训练。其 次,训练模型。选择对应的机器学习算法,分 别确定万宁站和其他大气站数据的最优超参数。 采用训练集训练机器学习算法模型,得到训练 集的预测值。接下来,测试模型。采用测试集 仿真测试,反归一化后得到测试集的预测值。 计算回归问题的误差评价指标,即决定系数  $(R^2)$ 、均方根误差 (RMSE)、平均绝对误差 (MAE)、平均绝对误差率 (MAPE)。最后,分 析敏感性。采用平均影响值 (mean impact value, MIV) 法评估各类模型特征的敏感性。将归 一化后的特征数值增大20%,输入训练好的模 型中计算数据集的预测值。同时,将归一化后 的特征数值减小 20% 计算得到预测值, 两组预 测值之差的平均值即为 MIV。将 MIV 除以原预

A技创新 Lejichuangxin

测值得到平均相对影响值,进一步得到 28 维特征的平均相对影响值的百分比。

## 3 预测结果与分析

## 3.1 误差分析

规范公式和 6 种机器学习算法对钢材大气 腐蚀深度的预测结果见图 3,其中 *D*<sub>real</sub> 与 *D*<sub>pred</sub> 分别为腐蚀深度的真实值和预测值。训练集和 测试集的散点越接近于 *y* = *x* 直线,预测效果越 好。可以看出,规范公式对腐蚀深度较大的样 本预测结果并不理想,远小于真实值,尤其是 腐蚀深度离群样本差异极大。而机器学习算法 的预测效果展现出明显的优势,尤其以 FR、 LSTM、BPNN 的预测效果更优。





(e) BPNN

# • 浙江交易种技 •



图 3 不同方法腐蚀深度预测结果

按照规范公式, 万宁站数据预测误差指标 *R*<sup>2</sup>为0.23, RMSE为276.50, MAE为127.56, MAPE为15.01; 其他大气站数据预测误差指标 *R*<sup>2</sup>为0.24, RMSE为67.65, MAE为43.14, MAPE为8.45。表1给出了应用不同机器学习 算法对大气腐蚀深度预测结果的误差指标。图4 (a)描述了规范公式与不同机器学习算法在不 同地区数据集上的 RMSE。通过对比可以看出, 6种机器学习算法腐蚀深度预测结果的误差远小 于规范公式的预测误差,可见采用数据驱动的 机器学习算法可以提升腐蚀深度预测的准确性。 但并非所有算法都适用于钢材腐蚀深度预测问 题, CNN 在万宁站数据集上的性能极差,并不 适用于此类回归问题; RBFNN 在万宁站数据集 上的预测性能相对较差,误差远大于 RF 与 LSTM 的预测结果,因此应避免选择这两类算法 进行腐蚀预测; BPNN 与采用 RBF 核函数的 SVM 在训练万宁站数据集的过程中不确定性较 高,输出结果与数据集的划分有关,模型鲁棒 性较差; RF 与 LSTM 在所有大气站数据集上预 测结果的误差相对较低,模型具有较好的鲁棒 性,其中 LSTM 在万宁站数据集上的表现更好, 能够准确预测极端值,而 RF 在其他大气站数据 集上的表现更好,误差最低。

Keiichu

图4(b)进一步描述了规范公式与不同机 器学习算法在不同试验周期上的 RMSE。整体 上, RMSE 随着试验周期的增长而逐步增大,规 范公式在腐蚀初期(3~24个月)的预测精度较 高,而在之后误差快速增长,预测精度小于机 器学习算法的精度。当试验周期在 60 个月以内 时, RF 的预测效果更好,而当试验周期在 60 个月以上时,LSTM 的预测效果更好。综合考虑 以上内容,针对钢材大气腐蚀深度预测这一回 归问题,推荐选择 RF 或LSTM 进行模型训练与 预测,对于需要考虑较多离群样本的时间序列 问题,采用LSTM 更优。



浙江交通科技 •





3.2 敏感性分析

采用 MIV 法计算规范公式和机器学习算法 对 28 维特征的平均相对影响值百分比,结果见 表 2。可以看出,不同预测方法的敏感性分析结 果存在较大差异。若依据规范公式计算得到的 敏感性排序,则前 3 项重要材料特征为:硫>硅 >磷,前 3 项重要环境特征为:平均相对湿度> 硫酸盐化速率>海盐粒子。若依据 RF 计算得到 的敏感性排序,则前 3 项重要材料特征为:硅> 碳>磷,前 3 项重要环境特征为:雨水 pH 值>日 照时数>海盐粒子。若依据 LSTM 计算得到的敏 感性排序,则前 3 项重要材料特征为:铬>硅> 磷,前 3 项重要环境特征为:雨水 pH 值>平均 相对湿度>平均温度。

表1 不同机器学习模型预测结果的误差指标

试验地点	误差指标	样本集	SVM	RF	RBFNN	BPNN	LSTM	CNN
万宁	$R^2$	训练集	0.81	0. 91	0.70	0.95	0.91	0.33
		测试集	0.74	0. 65	0.48	0.85	0.87	0.38
	RMSE	训练集	152	105	178	68	100	287
	MAE	测试集	100	109	206	137	92	163
	MAPE	训练集	45	52	103	43	61	131
		测试集	59	69	140	75	58	94
		训练集	265%	252%	521%	264%	254%	1511%
		测试集	263%	272%	2220%	127%	1454%	304%
其他	$R^2$	训练集	0.90	0. 94	0.80	0.85	0.85	0.49
		测试集	0.86	0. 84	0.76	0.77	0.82	0.65
	RMSE	训练集	27	21	36	33	31	59
	MAE	测试集	21	23	33	29	31	38
	MAPE	训练集	10	7	20	19	15	28
		测试集	15	13	24	20	18	23
		训练集	130%	29%	202%	215%	60%	569%
		测试集	141%	57%	267%	156%	133%	237%

## • 浙江交易种技 •

Kcjichuangxin

特征	规范公式	SVM	RF	RBFNN	BPNN	LSTM	CNN	RF 重要性百分比
试验周期	16. 1%	6.8%	18.1%	5.5%	7.4%	17.6%	31.2%	15.0%
雨水 pH	-	18.1%	27.8%	22.2%	52.1%	14. 5%	0.1%	9.6%
铬	0.1%	6. 5%	0.3%	2.5%	3.2%	8.1%	4. 5%	8.1%
雨水 SO42-	-	3.9%	1.4%	5.5%	1.4%	1.5%	20.1%	5.6%
日照时数	_	2.7%	5.5%	3.2%	3.9%	0.5%	0.5%	5.5%
自然降尘量	-	1.5%	2.4%	0.5%	0.6%	1.8%	1.0%	5.2%
硫酸盐化速率	10. 3%	3.1%	0.1%	2.1%	0.4%	2.0%	0.2%	4.6%
平均温度	4.7%	3.3%	2.5%	3.8%	3.9%	4.2%	9.8%	4.5%
$NO_2$	-	2.2%	0.9%	3.1%	0.5%	3.0%	0%	4.4%
碳	0%	5.6%	6.7%	1.5%	0.5%	1.1%	0.1%	4.3%
铜	0.1%	3.0%	0.6%	10. 7%	1.9%	2.6%	0%	3.9%
硅	0.5%	1.0%	7.9%	4. 5%	1.9%	5.5%	0.4%	3.3%
海盐粒子	7.1%	0.1%	4.9%	1.5%	2.0%	3.8%	0.1%	3.2%
NH <sub>3</sub>	-	1.5%	1.2%	2.3%	2.9%	1.0%	0%	3.2%
平均相对湿度	59.5%	13.8%	3.3%	6.5%	5.6%	13.6%	7.1%	2.9%
海拔	_	0.1%	0.5%	0.6%	0.4%	1.3%	13.2%	2.9%
锰	-	2.5%	2.8%	6.7%	0.6%	1.4%	0.7%	2.7%
硫	1.0%	0.6%	3.5%	6.9%	2.3%	2.3%	0%	2.3%
$H_2S$		1.3%	1.0%	0.8%	0.2%	2.5%	0%	2.3%
磷	0.4%	7.1%	6.3%	1.5%	3.6%	5.4%	0%	2.1%
雨水 Cl⁻	_	0.6%	1.6%	4.7%	2.6%	2.5%	10. 8%	1.7%
平均风速	_	1.3%	0.3%	1.6%	0.8%	1.8%	0%	1.7%
镍	0.1%	2.8%	0%	0.1%	0%	0.1%	0.1%	1.5%
钼	-	0.1%	0.2%	0.1%	1.1%	0.4%	0%	0.7%
钛	-	5.5%	0%	0.2%	0.2%	0.1%	0%	0%
钒	-	3.1%	0%	1.0%	0.1%	1.0%	0%	0%
铌	-	1.2%	0%	0.2%	0%	0.6%	0%	0%
铝	_	0.8%	0.1%	0.2%	0.1%	0%	0%	0%

表 2 不同机器学习模型的特征敏感性指标

注:加粗的为特征敏感性或重要性降序排序的前10项。

此外,还计算了 RF 重要性指标,若依据 为: 铬>碳>铜,前 3 项重要环境特征为:雨水 RF 重要性指标排序,则前 3 项重要材料特征 pH 值>雨水 SO<sub>4</sub><sup>2-</sup>>日照时数。综合各类机器学

# • 浙江委益科技 •

习算法排序,相对更重要的特征前10项分别为:雨水pH值、试验周期、平均相对湿度、雨水SO<sub>4</sub><sup>2-</sup>、平均温度、铬、磷、硅、雨水Cl<sup>-</sup>、铜。其中有关雨水酸碱性和雨水腐蚀性离子浓度的指标未能被规范公式考虑。

将28 维特征归类为时间、材料和环境三类特征,对每类特征平均相对影响值百分比求和得到图 5。规范公式中材料特征的敏感性仅为2%,而机器学习算法大幅提升了材料特征的敏感性。观察 RF、LSTM 敏感性和 RF 重要性百分比,对钢材大气腐蚀深度影响的重要性排序为:环境特征>材料特征>时间特征。



4 结论

1)针对钢材大气腐蚀深度预测问题,应用 机器学习算法的预测精度高于规范公式,推荐 选择 RF 或 LSTM 进行模型训练与预测。对于试 验周期不超过 60 个月的预测问题,采用 RF 更 优;而对于试验周期超过 60 个月或需要考虑极 端值的时间序列问题的预测问题,采用 LSTM 更 优。

2) 通过对所有预测模型的 MIV 分析和 RF

2023年第4期

模型的重要性分析,钢材大气腐蚀深度影响的 特征排序为:环境特征>材料特征>时间特征, 其中重要的环境特征为:雨水 pH 值、平均相对 湿度、雨水 SO<sub>4</sub><sup>2-</sup>浓度、平均温度和雨水 Cl<sup>-</sup>浓 度;重要的材料特征为:铬、磷、硅和铜。有 关雨水酸碱性和雨水腐蚀性离子浓度的指标未 能被规范公式考虑。

#### 参考文献:

- [1] 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 (03):441-451.
- [2] Fang S F, Wang M P, Qi W H, et al. Hybrid genetic algorithms and support vector regression in forecasting atmospheric corrosion of metallic materials [J]. Computational Materials Science, 2008, 44 (2): 647–655.
- [3] Jiménez-Come M J, Turias I J, Ruiz-Aguilar J J. A twostage model based on artificial neural networks to determine pitting corrosion status of 316L stainless steel [J]. Corrosion Reviews, 2016, 34 (1-2); 113-125.
- [4] Pintos S, Queipo NV, de Rincon O T, et al. Artificial neural network modeling of atmospheric corrosion in the MICAT project [J]. Corrosion. Science. 2000, 42 (1): 35 - 52.
- [5] 支元杰.大气环境下小样本金属材料腐蚀的数据驱动预 测模型 [D].北京:北京科技大学,2019.
- [6] Zhi Y J, Jin Z H, Lu L, et al. Improving atmospheric corrosion prediction through key environmental factor identification by randomforest-based model [J]. Corrosion Science, 2021, 178: 109084.
- [7] 金属和合金的腐蚀 大气腐蚀性 第1部分:分类、测定和 评估:GB/T 19292.1-2018 [S].北京:中国标准出版 社,2018.
- [8] 金属和合金的腐蚀 大气腐蚀性 第2部分:腐蚀等级的指导值:GB/T 19292.2-2018 [S].北京:中国标准出版 社,2018.

• 浙江交益种技 •