

引文: 宋福霖, 赵弘, 苗兴园. 基于 HMOGWO-RF 的埋地管道点蚀深度机理-学习预测模型[J]. 油气储运, 2024, 43(11): 1249-1259.

SONG Fulin, ZHAO Hong, MIAO Xingyuan. Mechanism-learning prediction model for pitting depth of buried pipeline based on HMOGWO-RF[J]. Oil & Gas Storage and Transportation, 2024, 43(11): 1249-1259.

基于 HMOGWO-RF 的埋地管道点蚀深度机理-学习预测模型

宋福霖 赵弘 苗兴园

中国石油大学(北京)机械与储运工程学院

摘要:【目的】截至 2025 年,中国油气管网规模将达到 24×10^4 km,管道输送已然成为中国油气运输的重要方式之一。受管道周围土壤环境等因素影响,管道腐蚀现象时有发生,降低了管道的使用年限。为保障埋地管道的安全运行,需有效预测其所受腐蚀程度。【方法】将随机森林(Random Forest, RF)算法与多目标优化方法相结合,提出腐蚀机理引导下的埋地管道点蚀深度预测模型,将管道腐蚀的腐蚀机理知识引入机器学习模型中,提高模型的可解释性。根据特征变量之间的交互作用机制,构建新的特征变量,以更好地反映管道周围土壤环境的影响因素。通过随机森林算法中的基尼系数计算新特征空间中所有特征的重要性,利用混合多目标灰狼优化(Hybrid Multi-Objective Grey Wolf Optimization, HMOGWO)算法求解随机森林算法的最优超参数,并将特征选择融入多目标优化中。在多目标优化的过程中,综合考虑特征数量、预测准确率、模型稳定性 3 个优化目标,并设计综合评价指标,对比分析 Pareto 解集,以获取特征子集与最优超参数组合,得到最具代表性、优化性能最佳的特征子集,提高模型稳定性与预测准确性。【结果】模型设计完成后,采用实际埋地管道的点蚀数据集对模型进行验证,将三目标 HMOGWO 算法与 RF 模型相结合,模型的预测性能及稳定性远超三目标灰狼优化算法、双目标 HMOGWO 算法、双目标灰狼优化算法、单目标灰狼优化算法及单目标粒子群优化算法。【结论】该模型可以实现埋地管道最大点蚀深度的准确预测,所提出的腐蚀机理引导下的埋地管道点蚀深度预测模型可以提高管道腐蚀预测的可解释性与准确性,有助于延长管道的使用寿命,对于石油和天然气运输行业具有重要的实际应用意义。(图 12,表 2,参 22)

关键词: 埋地管道; 点蚀深度; 腐蚀机理引导; 随机森林算法; 多目标灰狼优化算法; 特征重要性

中图分类号: TE832

文献标识码: A

文章编号: 1000-8241(2024)11-1249-11

DOI: 10.6047/j.issn.1000-8241.2024.11.006

Mechanism-learning prediction model for pitting depth of buried pipeline based on HMOGWO-RF

SONG Fulin, ZHAO Hong, MIAO Xingyuan

College of Mechanical and Transportation Engineering, China University of Petroleum (Beijing)

Abstract: 【Objective】 China's oil and gas pipeline networks are expected to reach 24×10^4 km by 2025. Pipeline transportation has become one of the key means of transportation in the country. However, these pipelines are vulnerable to corrosion caused by the surrounding soil environment and other factors, which shortens their life in service. To ensure the safe operation of buried pipelines, accurately predicting the degree of corrosion is crucial. 【Methods】 This paper presents a prediction model for the pitting depth of buried pipelines, guided by the corrosion mechanism and combining a Random Forest (RF) algorithm with a Multi-Objective Optimization process. The incorporation of knowledge about the pipeline corrosion mechanism enhances the interpretability of the machine-learning (ML) model. By building on the interaction mechanisms among characteristic variables, new variables were created to better reflect the influencing factors of the surrounding soil environment. The Gini coefficients in the Random Forest algorithm were used to evaluate the importance of all features in the new characteristic space through calculations. Additionally, a Hybrid Multi-Objective Grey Wolf Optimization (HMOGWO) algorithm was adopted to determine the optimal hyperparameters of the RF algorithm. This feature selection approach was integrated with the multi-objective optimization process, considering three comprehensive optimization objectives: the number of features, prediction accuracy, and

model stability. Using a defined comprehensive evaluation index, a comparative analysis of the Pareto solution set was conducted to obtain the optimal combination of feature subsets and hyperparameters. The resulting feature subsets, which are both representative and optimized for performance, contribute to improvements in model stability and prediction accuracy. **[Results]** The designed model was validated using a pitting dataset of real-world buried pipelines. By leveraging the combination of the three-objective HMOGWO algorithm and the RF model, it significantly outperformed the three-objective MOGWO algorithm, the two-objective HMOGWO algorithm, the two-objective MOGWO algorithm, as well as both the single-objective GWO algorithm and the single-objective PSO algorithm in terms of prediction performance and stability. **[Conclusion]** The proposed model has proven effective in accurately predicting the maximum pitting depth of buried pipelines. It is more interpretable and accurate in pipeline corrosion prediction, guided by the corrosion mechanism. This model is shown to be valuable in prolonging the service life of pipelines, highlighting its significance for practical applications in the oil and gas transportation sector. (12 Figures, 2 Tables, 22 References)

Key words: buried pipeline, pitting depth, guided by corrosion mechanism, Random Forest (RF) algorithm, Hybrid Multi-Objective Grey Wolf Optimization (HMOGWO) algorithm, feature importance

根据《中长期油气管网规划》，截至 2025 年，中国油气管网规模将达到 24×10^4 km，将构建布局合理、覆盖广泛、外通内畅、安全高效的现代油气管网^[1]。然而，随着服役时间的增长，油气管道不可避免面临腐蚀风险，腐蚀导致的管道事故时有发生，造成了重大的经济损失及环境污染。据统计，当前 20%~30% 的管道使用年限已超过 20 年，在周围土壤环境的作用下，点蚀现象时有发生，可导致管道泄漏失效。因此，准确预测埋地管道的点蚀对于保证管道可靠性及安全性具有重要意义^[2]。

近年来，国内外对于管道点蚀预测建立了多种模型，建立时大多进行室内试验或以某种恒定不变的腐蚀速率及环境条件进行模拟，但未考虑埋地管道土壤特性与腐蚀速率在实际工况中存在随机变化的情况。骆正山等^[3]提出粗糙集 (Rough Set, RS) 结合多策略改进的麻雀搜索算法 (Multi-strategy Improved Sparrow Search Algorithm, MISSA) 与最小二乘支持向量机 (Least Squares Support Vector Machine, LSSVM) 建立预测模型，利用 MISSA 求解出 LSSVM 中核函数参数及惩罚因子的最优解，同时选取径向基核函数，使其预测性能达到最优。Sim 等^[4]提出一种通过幂律函数计算点蚀深度的方法，将点蚀看作一个与暴露时间相关的过程。Caleyo 等^[5]提出了一种点蚀预测模型，该模型考虑了土壤、管道的化学及物理特性。上述模型、实验研究以及仿真模拟均需在某一限定条件下方可得到相应结论，受腐蚀过程中各种条件的限制以及模型中各参数简单的线性关系影响，无法精确预测真实工况下埋地管道的点蚀过程。

随着科技迅猛发展，机器学习 (Machine Learning, ML) 方法受到深入关注。史航^[6]提出了管道随机点蚀模型的构建方法，对不同点蚀参数和管道因素下的管道悬空沉降进行了模拟，得到点蚀深度、数量、半径、管道内压、点蚀区长度等参数对埋地悬空管道力学性能的影响规律。Li 等^[7]提出了一种基于集成的麻雀搜索算法 (Sparrow Search Algorithm, SSA) 与长-短期记忆 (Long Short-Term Memory, LSTM) 的最大点蚀深度预测数据驱动模型，测试集的均方根误差 (RMSE) 为 0.060 7。吕林林等^[8]采用核主成分分析法 (Kernel Principal Component Analysis, KPCA) 进行降维操作，随后利用改进的蝗虫优化算法 (Improved Grasshopper Optimization Algorithm, IGOA) 对极限学习机 (Extreme Learning Machine, ELM) 进行优化，确定最优网络结构和激励函数，提出了 KPCA-IGOA-ELM 组合预测模型。Peng 等^[9]提出了一种多相管道内腐蚀速率预测的混合模型，该模型结合了主成分分析 (Principal Component Analysis, PCA)、混沌粒子群优化 (Chaos Particle Swarm Optimization, CPSO) 与支持向量回归 (Support Vector Regression, SVR)，其平均绝对误差 (MAE) 仅为 0.083。

已有的研究在利用 ML 方法进行腐蚀预测方面已取得很大进展，但在 ML 模型中集成的腐蚀机理较少，导致模型的可解释性较差。ML 模型的超参数对预测性能具有显著影响，上述研究建立了具有智能优化算法的混合模型，但仅考虑了预测精度，未考虑模型的稳定性。同时，上述模型大多通过原始数据集或数据降维进行训练，特征选择未参与到优化过程中。为解决

上述问题,将腐蚀机理与领域知识融入 ML 模型中,考虑土壤特性、管道特性、时间变化进行最大点蚀深度预测。通过原始特征与自变量之间的相互作用构建新的特征空间。由于各输入变量对最大点蚀深度的影响程度不同,需要进行特征选择并确定特征子集。使用随机森林(Random Forest, RF)的基尼系数对所有变量的特征重要性进行排序,提出了一种基于非线性收敛因子与天牛须搜索(Beetle Antennae Search, BAS)算法的混合多目标灰狼优化(Hybrid Multi-Objective Grey Wolf Optimization, HMOGWO)算法,用于优化 RF 模型的超参数。此外,通过最小化预测误差与输入变量的数量,综合考虑特征数、预测精度及模型稳定性,设计综合评价指标,用于确定最优的超参数及最合适的特征子集。

1 理论基础

1.1 腐蚀机理

管道腐蚀可分为全面腐蚀与局部腐蚀。全面腐蚀指管道管壁整体以一种缓慢、均匀的速度发生腐蚀减薄;局部腐蚀则是管道表面某个或某些局部位置发生腐蚀的现象^[10]。由于埋地管道沿线各处所处的环境存在差异,其腐蚀速率与腐蚀程度相应存在明显差异,发生局部腐蚀的可能性远大于全面腐蚀,且局部腐蚀向纵深发展速度快、检测难度大,更易对管道造成重大危害。局部腐蚀包含点蚀、缝隙腐蚀、电偶腐蚀、应力腐蚀等,其中点蚀是管道最常见的腐蚀形式,其蚀坑直径较小、深度较深,具有很强的隐蔽性。同时,由于土壤微生物作用、腐蚀物质、湿度与温度变化、电化学作用均对点蚀具有影响^[11],其机制相对复杂,难以有效防控。

1.2 随机森林理论

集成学习是通过对多个弱学习器进行结合获得一个强学习器来提高预测结果的一种方法。RF 算法是一种常用的集成学习方法,其引入了 Bagging 框架与随机属性选择,计算方便、准确率高。首先,采用 Bootstrap 采样从训练数据集中提取 n 个样本,建立 n 个决策树。在这些决策树的每个分支上,节点上的均方误差(MSE)不断更新,直到得出最优结果为止(图 1),将所有决策树的平均预测值作为 RF 算法的最终结果^[12]:

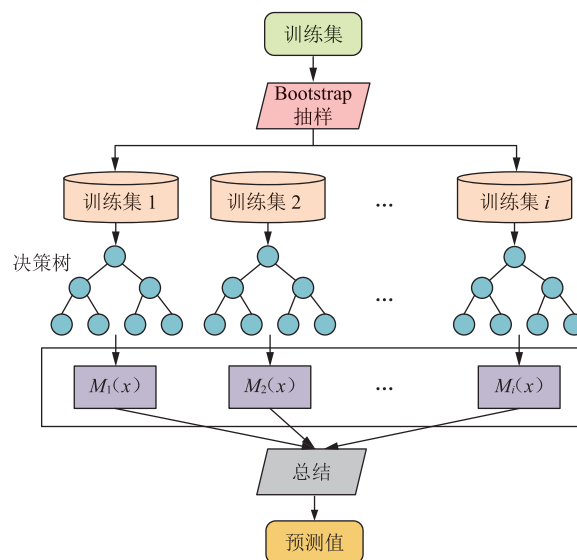


图 1 RF 算法流程图

Fig. 1 Architecture of RF algorithm

$$f_{\text{RF}}(x) = \frac{1}{n} \sum_{i=1}^n M_i(x) \quad (1)$$

式中: $f_{\text{RF}}(x)$ 为 RF 模型的最终结果; $M_i(x)$ 为第 i 个决策树模型的结果。

1.3 多目标灰狼优化算法

多目标灰狼优化(Multi-Object Grey Wolf Optimization, MOGWO)算法为一种基于多目标优化的群体智能算法,是灰狼优化算法的改进与扩展^[13]。该算法通过灰狼个体之间的协作与竞争,不断搜索、优化多个目标函数的解向量,从而达到多目标优化的目的。相比于传统的单目标优化算法,MOGWO 算法能够在有限的搜索空间中找到多个最优解,提供更多选择及更好的决策支持^[14]。

2 机理引导下的点蚀深度预测模型

当前,国内外的最大点蚀深度预测经验模型缺乏对管道周围环境土壤特性以及管道各方面性能因素的考虑,无法准确描述点蚀过程。此外,传统的 ML 模型缺乏腐蚀机理引导,使得其在预测性上也有所欠缺。为解决上述问题,首先基于腐蚀机理不同变量之间的相互作用,在腐蚀数据集原始特征基础上构建新的特征参数,然后通过 RF 模型的基尼系数计算出特征参数的重要性,最后利用 HMOGWO 算法对 RF 模型超参数进行优化,综合考虑特征数量、预测精度、模型稳定性等多个目标,实现多因素下的埋地管道最大点蚀深度预测(图 2)。HMOGWO 算法集成了

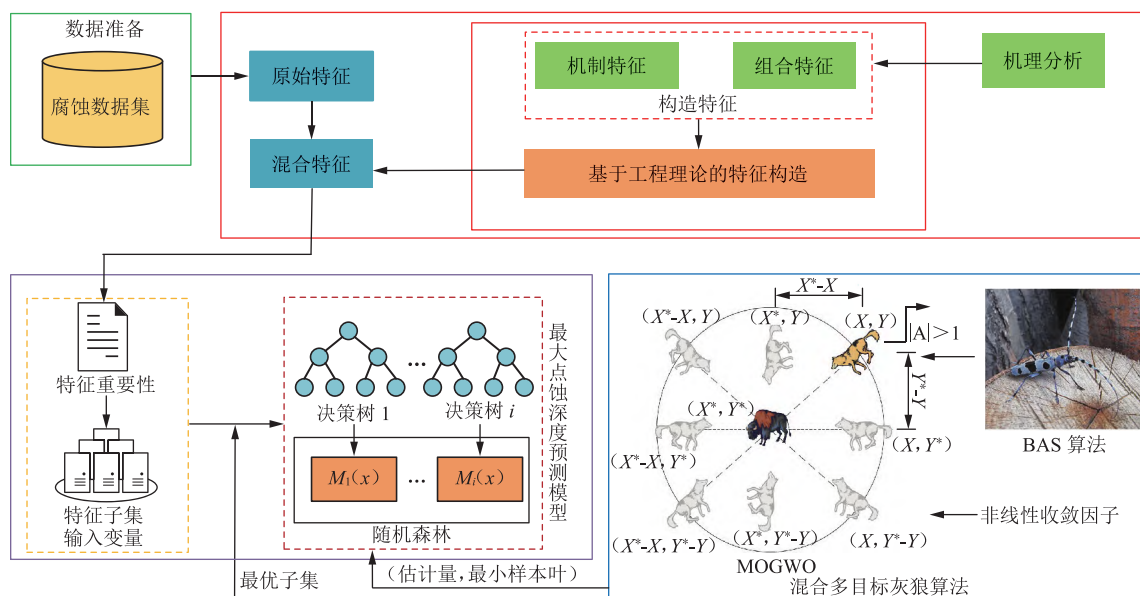


图 2 腐蚀机理引导的 ML 模型整体框架示意图

Fig. 2 Overall framework of ML model guided by corrosion mechanism

MOGWO 算法、BAS 算法及非线性收敛因子,使得优化不仅由灰狼指导,也由天牛须进行指导,收敛性能较佳。

2.1 腐蚀机理引导的特征空间

2.1.1 特征构建

埋地管道的最大点蚀深度与环境因素有关,如 pH 值 (pH)、暴露时间 (t)、氧化还原电位 (E_{RP})、涂层类型 (K_C)、管地电位 (E_{PP})、电阻率 (ω)、含水量 (C_W)、土壤容积密度 (ρ_s)、氯离子浓度 (C_C)、碳酸氢盐浓度 (C_B) 及硫酸盐离子浓度 (C_S),各参数设置为预测模型的独立输入变量,各变量之间存在相互作用关系,其中, ω 与 ρ_s 之间的相互作用关系为^[15]:

$$\log \omega = a + \beta_{D_C} D_C + \beta_{D_S} D_S \quad (2)$$

$$D_C = \frac{\rho_s}{\rho_{s,r}} \times 100 \quad (3)$$

式中: a 、 β_{D_C} 、 β_{D_S} 为常数系数; D_C 为压实度; D_S 为饱和度; $\rho_{s,r}$ 为同种土壤的参考容积密度, kg/m^3 。

C_W 与 ω 之间的相互作用关系为^[16]:

$$C_W = a\omega^b \quad (4)$$

式中: a 、 b 为常数系数。

E_{RP} 与 pH 值之间的相互作用关系为^[17]:

$$E_{RP} = \frac{k_B T \ln 10}{2e} [r(\text{H}_2) - 2 \text{pH}] \quad (5)$$

式中: k_B 为玻尔兹曼常数; e 为基本电荷, C ; T 为热力学温度, K ; $r(\text{H}_2)$ 为氢分子热力学活度的倒数。

考虑输入变量之间的相互作用,设计腐蚀机理引

导的特征空间。为了更好地预测埋地管道最大点蚀深度,基于各输入变量之间的相互作用关系构建新的特征参数, E_{RP}/pH 、 C_W/ω 、 $C_W/10^{\rho_s}$ 、 $\omega/10^{\rho_s}$ 。最后,构造 HMOGWO-RF 模型预测最大点蚀深度 (图 3)。

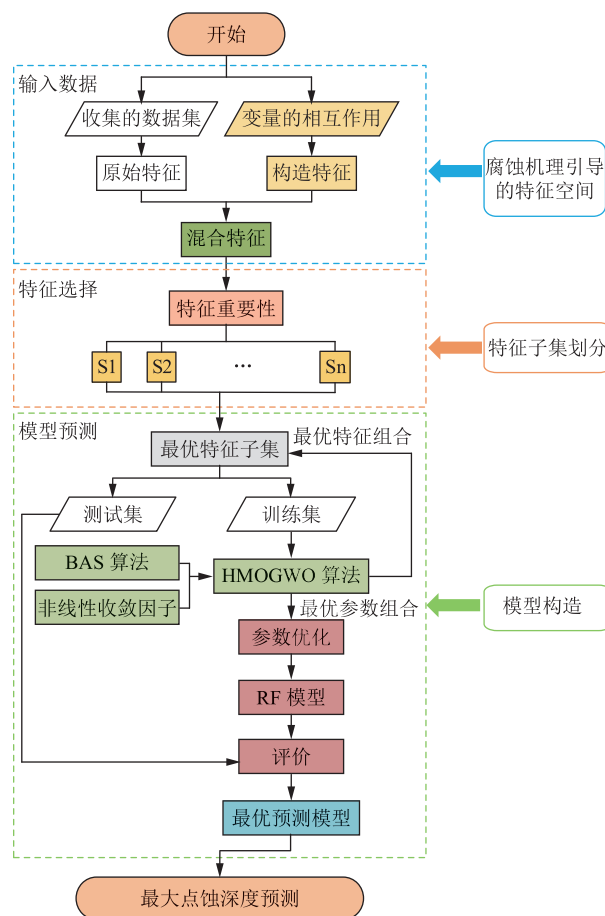


图 3 最大点蚀深度预测流程图

Fig. 3 Flow chart for prediction process of maximum pitting depth

2.1.2 特征重要性

在构造新特征之后,输入变量中包含很多重要性未知的特征,因此采用 RF 算法的基尼系数 G_F 选择具有代表性的特征:

$$G_F = 1 - \sum_{v=1}^m \lambda_{sv}^2 \quad (6)$$

式中: m 为训练集中的特征数; λ_{sv} 为第 v 个特征在节点 s 中的比例。

节点 s 的特征 F_i 的重要性为 I_{F_i} :

$$I_{F_i} = G_F - G_1 - G_2 \quad (7)$$

式中: G_1 、 G_2 分别为分枝后两个新节点的基尼指数。

特征 F_i 在第 k 棵树中的重要性为 $I_{F_i,k}$:

$$I_{F_i,k} = \sum_{s \in M} I_{F_i,s} \quad (8)$$

式中: M 为特征 F_i 在第 k 棵树中出现的次数。

第 i 个特征的归一化权重为 $\overline{I_{F_i}}$ 为:

$$\overline{I_{F_i}} = \frac{I_{F_i}}{\sum_{i=1}^h I_{F_i}} \quad (9)$$

式中: h 为特征的数量。

2.2 腐蚀机理引导的 HOMOGWO-RF 模型

2.2.1 混合多目标灰狼优化算法

在 HOMOGWO 算法中,灰狼可以分为 4 个层次,定义最优解决方案为 α ,第二优、第三优的解决方案分别为 β 、 δ ,其余的解为 ω 。灰狼的捕猎(优化)过程由方案 α 、 β 、 δ 进行指导判别。灰狼的位置代表一个可能的解决方案,灰狼 ω (候选解)将跟随灰狼 α 、 β 、 δ (最优解)去寻找最优解决方案^[18-19]。HOMOGWO 算法在灰狼捕猎过程中关注的是整个种群,忽略了个体判断,使其容易陷入局部最优。为解决此问题,引入 BAS 算法对其进行优化^[20-21]。天牛左、右两侧触角的位置 X_R 、 X_L 定义为:

$$\begin{cases} X_R = X + l\mathbf{d} \\ X_L = X - l\mathbf{d} \end{cases} \quad (10)$$

式中: l 为左、右触角之间的距离; \mathbf{d} 为随机单位向量。

根据左、右触角所感知气味浓度的差异,在第 j 次迭代天牛位置 X_j 的基础上,计算出第 $j+1$ 次迭代天牛位置 X_{j+1} :

$$X_{j+1} = X_j + \delta_j \mathbf{d} \text{sign}[f(X_R) - f(X_L)] \quad (11)$$

式中: δ_j 为第 j 次迭代的搜索步长; $\text{sign}(\cdot)$ 为符号函数; $f(\cdot)$ 为适应度函数。

基于 BAS 算法,在每次迭代中都考虑了个体对环境的判断。 α 、 β 、 δ 灰狼为搜索过程中的天牛,灰狼的位置更新不仅由 α 、 β 、 δ 灰狼指导,也由天牛须指导,精英个体在每次迭代中对比其左右两侧的适应度函数值,并以此更新狼的位置。

2.2.2 混合多目标灰狼优化算法-随机森林模型

采用 HMOGWO 与 RF 的混合模型预测埋地管道最大点蚀深度(图 4)。由于决策树的超参数会影响预测精度,在此提出通过 HMOGWO 优化射频模型的超参数。同时,为使预测误差和计算成本最小化构建三目标优化模型,使用输入特征数量、测试集的平均绝对百分比误差(MAPE)与标准差(STD)作为适应度函数,并采用 HMOGWO 算法对其进行最小化。特征数越小,计算效率越高;MAPE 越小,预测误差越小;STD 越小,模型稳定性越好。

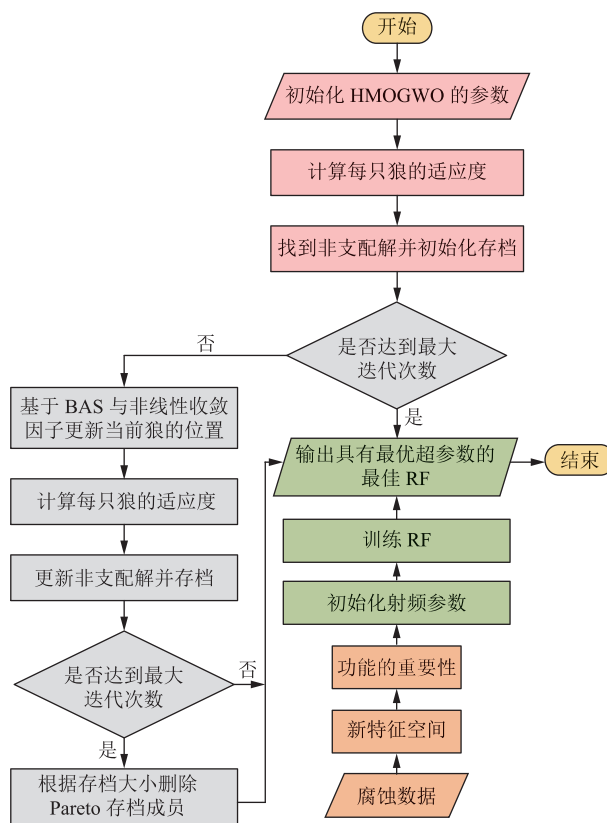


图 4 HMOGWO-RF 模型流程图
Fig. 4 Schematic diagram of HMOGWO-RF model

3 结果与讨论

3.1 数据采集

采用 Velázquez 等^[22]收集的油气管道最大点蚀深度数据集评估所提出模型的预测精度,将影响因

素作为输入变量(表 1)。将构造的新变量 E_{RP}/pH 、 C_W/ρ 、 $C_W/10^{\rho_s}$ 、 $\rho/10^{\rho_s}$ ，并对数据进行归一化处理，将数据集中的 80% 作为训练集，数据集中的 20% 作为测试集。

表 1 预测性能评价统计度量表
Table 1 Statistics for evaluation of prediction performance

参数	t/a	pH	E_{PP}/V	$\omega/(W \cdot m)$	C_W	$\rho_s/(g \cdot mL^{-1})$	C_C	C_B	C_S	E_{RP}/mV	K_C
最小值	5.00	4.140	-1.97	1.90	8.80%	1.10	0.99×10^{-6}	0.99×10^{-6}	0.99×10^{-6}	2.10	0.30
最大值	50.00	9.880	-0.42	399.50	66.00%	1.56	672.70×10^{-6}	195.20×10^{-6}	$1\,370.20 \times 10^{-6}$	348.00	1.00
平均值	22.99	6.139	-0.88	50.15	23.90%	1.30	47.73×10^{-6}	19.67×10^{-6}	152.97×10^{-6}	167.04	0.77
标准差	9.12	0.930	0.24	55.92	6.66%	0.09	75.16×10^{-6}	25.33×10^{-6}	168.18×10^{-6}	85.48	0.13

3.2 特征选择与性能对比

利用 RF 的基尼系数评估新特征空间中所有输入变量的特征重要性(图 5)，可见对于管道点蚀， E_{RP}/pH 是最关键的影响因素，而 K_C 的影响程度最低。

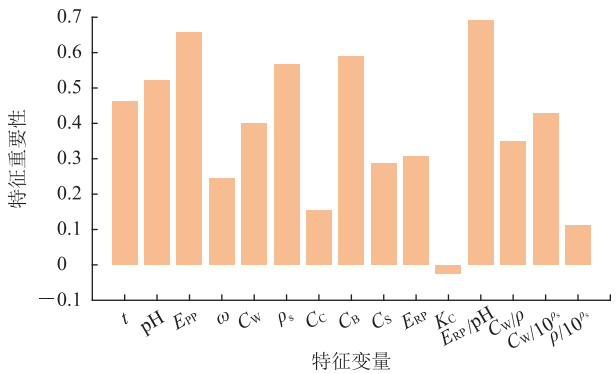


图 5 各输入变量特征重要性统计图
Fig. 5 Statistics of feature importance for input variables

采用 MAE、MAPE、MSE、RMSE、STD 等统计指标及相关系数 R^2 评价模型预测性能。将包含所有输入变量的数据集根据重要性划为 D1~D15 共 15 个特征子集(pH 、 t 、 E_{RP} 、 T_C 、 E_{PP} 、 ω 、 C_W 、 ρ_s 、 C_C 、 C_B 、 C_S 、 E_{RP}/pH 、 C_W/ω 、 $C_W/10^{\rho_s}$ 、 $\omega/10^{\rho_s}$)，通过 HMOGWO-RF 模型选择最合适的特征子集，在多目标优化过程中评估每个特征子集的性能。HMOGWO-RF 模型的多目标由特征数数量、MAPE、STD 组成。优化结果包含 20 个 Pareto 最优解(图 6)，特征数在 1~9 之间，MAPE 为 0.004~0.588，STD 为 0.032~2.194，超参数与特征子集的不同组合对整体预测性能有较大影响。

在得到 Pareto 最优解后，设计综合评价指标 K 用于选择最优解，确定最优超参数和特征子集，使得预测误差和特征数最小：

$$K=0.2 \overline{\mu}+0.4 \overline{E_{MAP}}+0.4 \overline{D_{ST}} \tag{12}$$

$$\mu=\frac{q}{h} \tag{13}$$

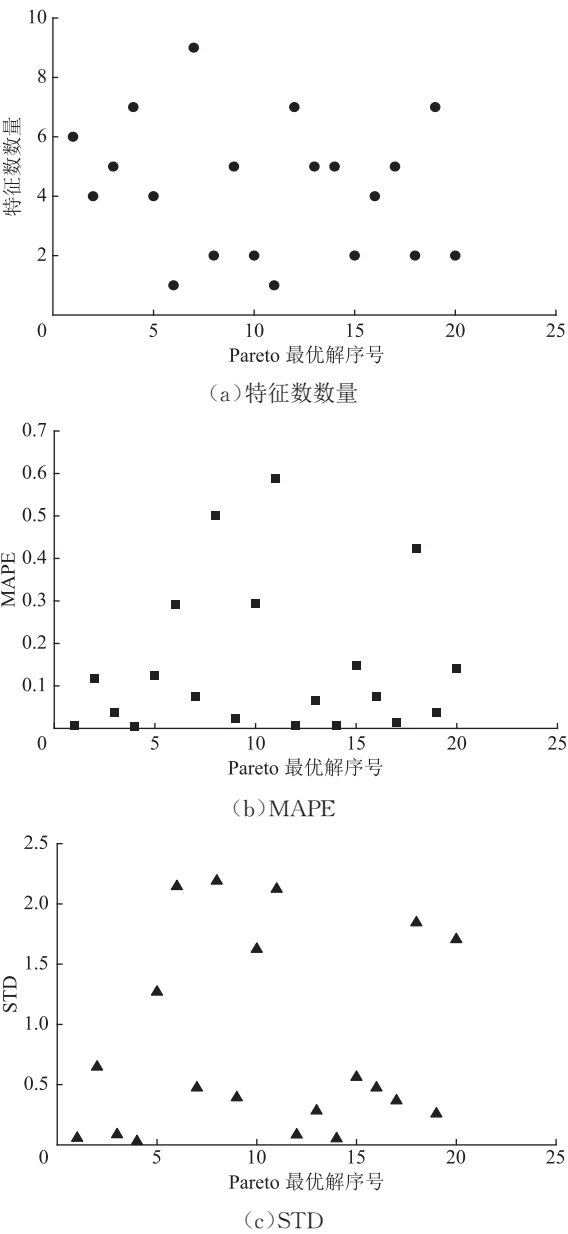


图 6 HMOGWO-RF 模型优化结果图
Fig. 6 Optimization results of HMOGWO-RF model

式中： q 为特征子集的特征数； $\overline{E_{MAP}}$ 、 $\overline{D_{ST}}$ 分别为归一化后平均绝对百分比误差、标准差； $\overline{\mu}$ 为归一化的 μ 。
 K 值越小，整体模型预测性能越好。将数据代入

式(12),可知最优特征数为 5,决策树数为 28,最小样本叶为 1(图 7)。对于所选的特征子集与 HMOGWO-RF 模型,MAPE、STD 分别为 0.006、0.055。

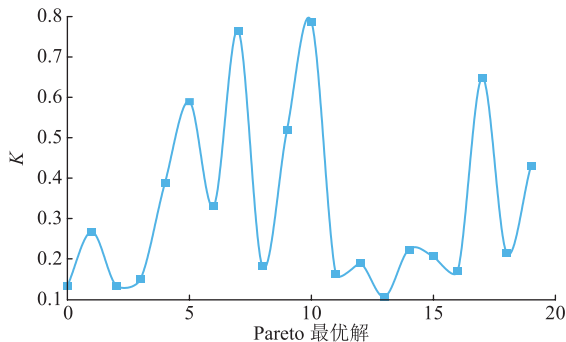


图 7 最优解综合评价指标随 Pareto 最优解变化曲线
Fig. 7 Variation curve of comprehensive evaluation index for optimal solution with Pareto optimal solution

根据 Pareto 最优解结果,选择 S1~S8 共 8 个特征子集进行对比(表 2),并统计不同特征子集的预测性能(图 8)。可见, S8 子集缺乏腐蚀机理,导致其具有较高的预测误差;S3、S4、S5 子集引入了腐蚀机理,预测结果更加准确;S1、S7 子集由于信息的缺乏或冗余导致其预测性能较差。这表明特征选择结合多目标优化和自变量之间的交互作用对模型的改进具有

表 2 HMOGWO-RF 模型特征子集列表
Table 2 List of feature subsets for HMOGWO-RF model

数据集	特征变量	特征数量
S1	$E_{pp}, E_{RP}/pH$	2
S2	$E_{pp}, \rho_s, C_B, E_{RP}/pH$	4
S3	$pH, E_{pp}, \rho_s, C_B, E_{RP}/pH$	5
S4	$t, pH, E_{pp}, \rho_s, C_B, E_{RP}/pH$	6
S5	$T, pH, E_{pp}, \rho_s, C_B, E_{RP}/pH, C_w/(10^{\rho_s})$	7
S6	$T, pH, E_{pp}, C_w, \rho_s, C_B, E_{RP}/pH, C_w/\rho, C_w/(10^{\rho_s})$	9
S7	所有特征变量	15
S8	原始数据	11

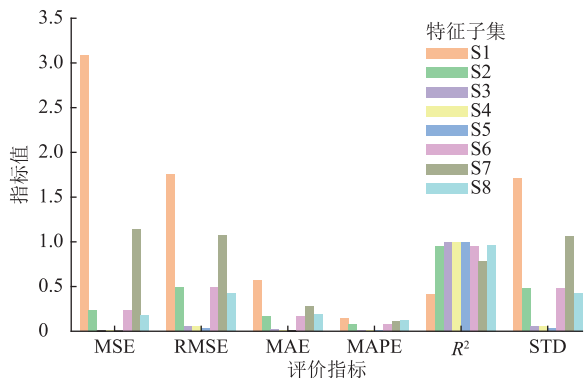
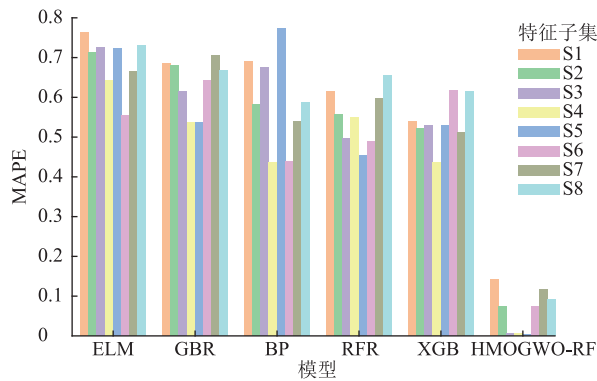


图 8 不同特征子集的预测性能指标对比图
Fig. 8 Comparison of prediction performance indexes among different feature subsets

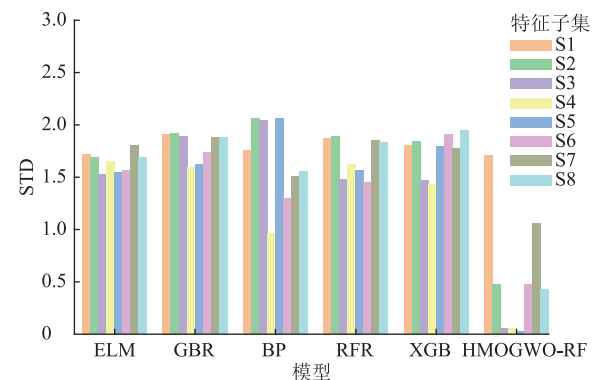
重要意义。同时,考虑预测性能和特征数,通过所提出的 HMOGWO-RF 模型,选择 S3 子集作为最适特征子集。

3.3 模型验证

为验证所提出模型的优越性,与采用集成学习模型 的单随机森林回归(Random Forest Regressor, RFR)模型、梯度提升回归树(Gradient Boosting Regression Tree, GBRT)模型、极端梯度提升(eXtreme Gradient Boosting, XGB)模型、反向传播神经网络(Back Propagation Neural Network, BPNN)、极限学习机(Extreme Learning Machines, ELM)进行对比(图 9、图 10)。可见,所提出 HMOGWO-RF 模型对埋地管道最大点蚀深度的预测值与实际值最接近,引入腐蚀机理以及采用混合模型具有显著优势,在 S3 子集上,其误差最低, R^2 为 0.999。



(a)MAPE



(b)STD

图 9 在不同特征子集上各模型预测精度对比图
Fig. 9 Comparison of prediction accuracy across models for different feature subsets

选取若干以 RF 模型为基础的混合模型与所提出模型进行对比,并采用传统三目标 MOGWO 算法验证 HMOGWO 算法的改进效果(图 11)。同时,为进一步验证所提出模型的优越性,采用双目标 HMOGWO 算法、双目标 MOGWO 算法、单目标 GWO 算法、单目标 PSO 算法与所提出模型进行性能对比。双目标

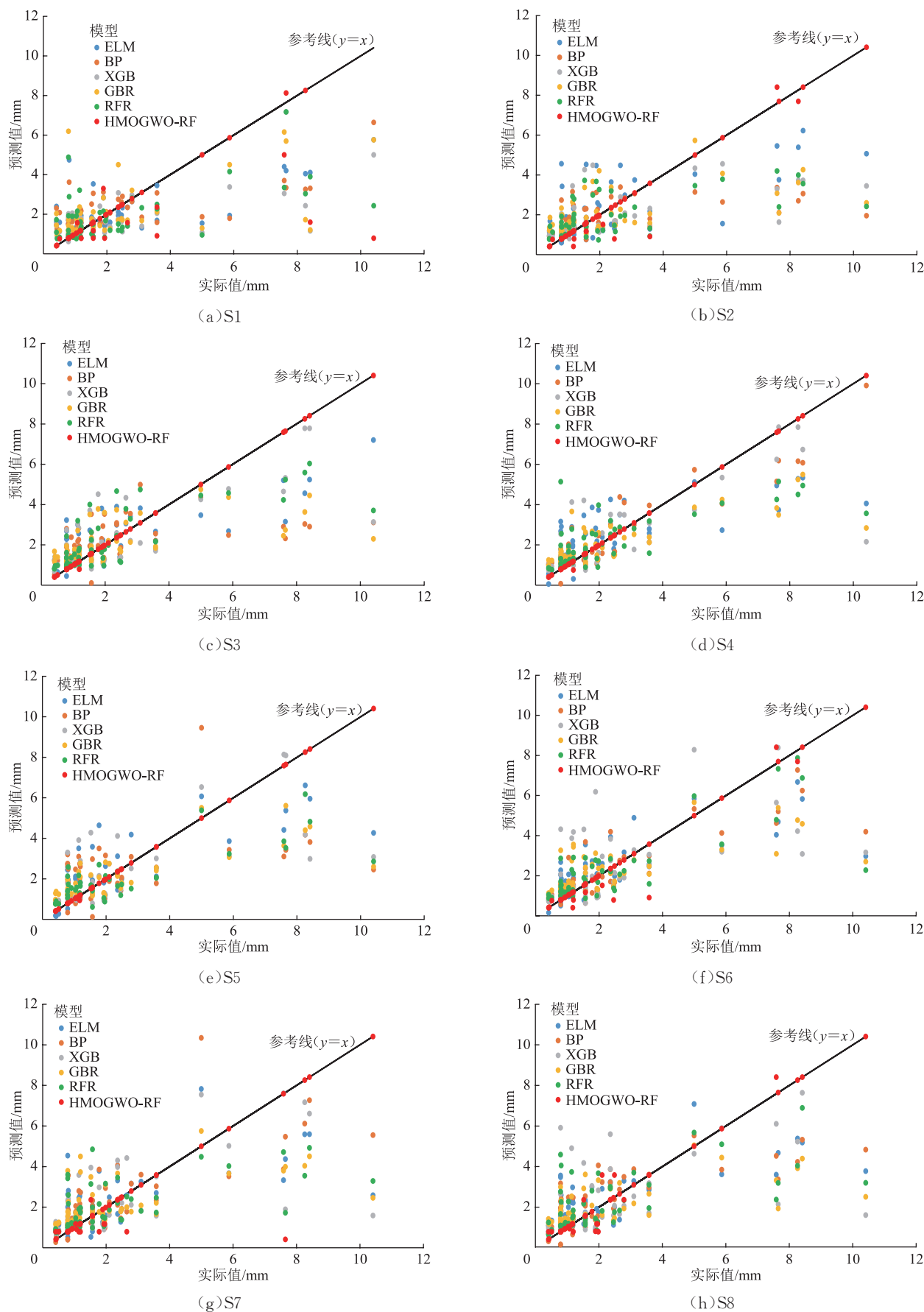


图 10 各模型在不同特征子集上的预测值与实际值对比图

Fig. 10 Comparison between predicted values from models and actual values for different feature subsets

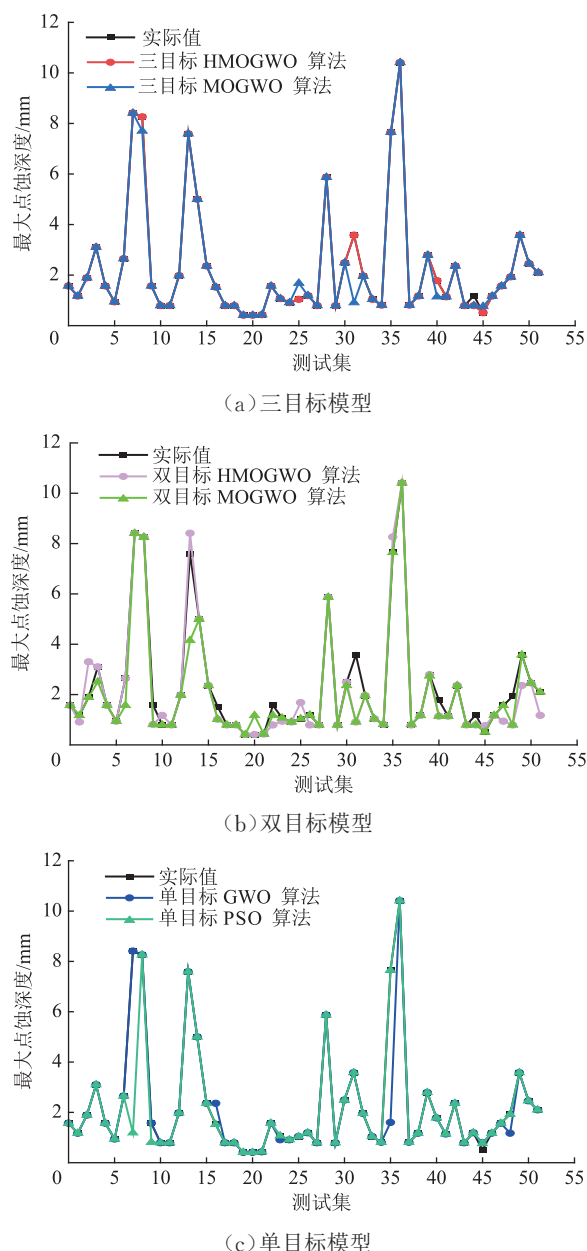


图 11 不同混合模型预测结果与实际值对比图

Fig. 11 Comparison between predicted values from different hybrid models and actual values

算法的优化目标是 MAPE、STD, 不考虑特征数。与双目标算法相比, 传统的单目标算法 (GWO、PSO) 的优化目标仅为 MAPE, 不考虑模型的稳定性。三目标 HMOGWO 算法的最大预测误差值为 0.400 mm, 平均预测误差值为 0.008 mm; 双目标 HMOGWO 算法的最大预测误差值为 2.670 mm, 平均预测误差值为 0.280 mm; 单目标 GWO 算法的最大预测误差值为 6.050 mm, 平均预测误差值为 1.555 mm。可见, 三目标模型的预测结果最贴近实际值, 预测效果更佳。

统计各混合模型的评价指标 (图 12), 可见相比于多目标 PSO, 单目标 PSO 的误差较大、 R^2 较小, 其整体

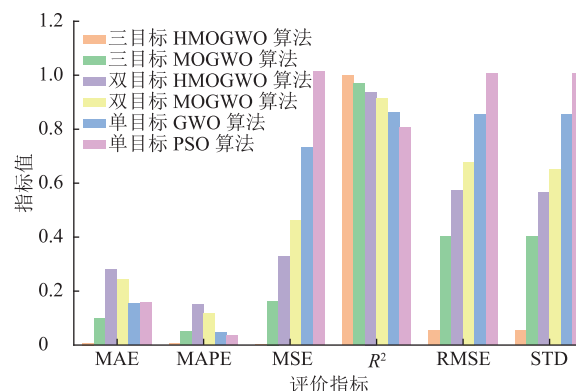


图 12 不同混合模型的评估指标对比图

Fig. 12 Comparison of evaluation indexes for different hybrid models

预测性能及稳定性最差。相比于双目标 MOGWO 算法, 双目标 HMOGWO 算法的 MAE、MAPE 与其相当、 R^2 较大、MSE、RMSE、STD 较低; 与双目标 MOGWO 算法相比, 三目标 HMOGWO 算法的各项误差均较低、 R^2 值较大。结合 HMOGWO 算法的混合模型的性能明显优于采用 MOGWO 算法的混合模型。

4 结论

1) 为准确预测管道最大点蚀深度, 提出了一种将腐蚀机理与多目标优化相结合的腐蚀机理引导 HMOGWO-RF 模型, 用于预测埋地管道的最大点蚀深度。根据不同自变量之间的相互作用构造新的特征变量来建立新的特征空间, 然后对特征重要性进行分析, 研究各输入变量的影响程度, 并根据重要性排序形成不同的特征子集。在多目标优化的过程中, 该模型综合考虑了特征数量、预测准确率、模型稳定性 3 个优化目标, 利用油气埋地管道的点蚀数据集验证预测效果, 选择最合适的子集。

2) 通过设计综合评价指标的形式评价 Pareto 最优解的预测性能, 发现将工程理论、领域知识集成到机器学习模型中是改进模型的必要条件。与传统 ML 模型及其他混合模型相比, HMOGWO-RF 模型的整体预测性能优于传统 ML 模型, 该模型的 MAE、MAPE、MSE、RMSE 和 STD 的最小值分别为 0.008、0.006、0.003、0.055、0.055, R^2 最高可达 0.999。基于所建立的数据驱动模型, 可以预测所研究管道的最大点蚀深度。该模型有助于提高管道腐蚀预测的可解释性与准确性, 延长管道使用寿命, 可为石油和天然气运输行业的发展提供技术参考。

参考文献:

- [1] 李秋扬, 赵明华, 任学军, 王乐乐, 冯学书, 牛亚琨. 中国油气管道建设现状及发展趋势[J]. 油气田地面工程, 2019, 38(增刊 1): 14-17. DOI: 10.3969/j.issn.1006-6896.2019.z1.004.
- LI Q Y, ZHAO M H, REN X J, WANG L L, FENG X S, NIU Y K. Construction status and development trend of Chinese oil & gas pipeline[J]. Oil-Gas Field Surface Engineering, 2019, 38(S1): 14-17.
- [2] 吴志平, 陈振华, 戴联双, 胡亚博, 毕武喜. 油气管道腐蚀检测技术发展现状与思考[J]. 油气储运, 2020, 39(8): 851-860. DOI: 10.6047/j.issn.1000-8241.2020.08.002.
- WU Z P, CHEN Z H, DAI L S, HU Y B, BI W X. Development status and thinking of oil and gas pipeline corrosion detection technology[J]. Oil & Gas Storage and Transportation, 2020, 39(8): 851-860.
- [3] 骆正山, 徐龙寅, 骆济豪, 王小完. 改进 SSA-LSSVM 模型在埋地管道点蚀深度预测中的应用[J]. 安全与环境学报, 2023, 23(9): 3115-3122. DOI: 10.13637/j.issn.1009-6094.2022.0395.
- LUO Z S, XU L Y, LUO J H, WANG X W. Application of improved SSA-LSSVM model in prediction of pitting depth of buried pipelines[J]. Journal of Safety and Environment, 2023, 23(9): 3115-3122.
- [4] SIM S, COLE I S, CHOI Y S, BIRBILIS N. A review of the protection strategies against internal corrosion for the safe transport of supercritical CO₂ via steel pipelines for CCS purposes[J]. International Journal of Greenhouse Gas Control, 2014, 29: 185-199. DOI: 10.1016/j.ijggc.2014.08.010.
- [5] CALEYO F, VELÁZQUEZ J C, VALOR A, HALLEN J M. Probability distribution of pitting corrosion depth and rate in underground pipelines: a Monte Carlo study[J]. Corrosion Science, 2009, 51(9): 1925-1934. DOI: 10.1016/j.corsci.2009.05.019.
- [6] 史航. 含点蚀埋地长输管道悬空沉降失效评估方法研究[D]. 大庆: 东北石油大学, 2023.
- SHI H. Study on failure evaluation method of suspended settlement of long-distance pipeline with pitting corrosion[D]. Daqing: Northeast Petroleum University, 2023.
- [7] LI X H, GUO M M, ZHANG R R, CHEN G M. A data-driven prediction model for maximum pitting corrosion depth of subsea oil pipelines using SSA-LSTM approach[J]. Ocean Engineering, 2022, 261: 112062. DOI: 10.1016/j.oceaneng.2022.112062.
- [8] 吕林林, 王杰, 祁庆芳, 郭策, 贺蓉蓉, 孙小伟. 基于 KPCA-IGOA-ELM 的油气混输管道腐蚀速率预测模型[J]. 油气储运, 2023, 42(7): 785-792. DOI: 10.6047/j.issn.1000-8241.2023.07.007.
- LYU L L, WANG J, QI Q F, GUO C, HE R R, SUN X W. Corrosion rate prediction model of oil-gas mixed transportation pipelines based on KPCA-IGOA-ELM[J]. Oil & Gas Storage and Transportation, 2023, 42(7): 785-792.
- [9] PENG S B, ZHANG Z, LIU E B, LIU W, QIAO W B. A new hybrid algorithm model for prediction of internal corrosion rate of multiphase pipeline[J]. Journal of Natural Gas Science and Engineering, 2021, 85: 103716.
- [10] 闻亚星, 吕坦, 国滨, 王锋, 陈金忠, 马义来. 基于 CART-熵权法的管道腐蚀状态评估及其应用[J]. 腐蚀与防护, 2023, 44(9): 16-21, 100. DOI: 10.11973/fsyfh-202309003.
- WEN Y X, LYU T, GUO B, WANG F, CHEN J Z, MA Y L. Evaluation and its application of pipeline corrosion status based on CART-entropy weight method[J]. Corrosion and Protection, 2023, 44(9): 16-21, 100.
- [11] 吴明, 谢飞, 陈旭, 王丹, 孙东旭. 埋地油气管道腐蚀失效研究进展及思考[J]. 油气储运, 2022, 41(6): 712-722. DOI: 10.6047/j.issn.1000-8241.2022.06.013.
- WU M, XIE F, CHEN X, WANG D, SUN D X. Research progress and thinking on corrosion failure of buried oil and gas pipelines[J]. Oil & Gas Storage and Transportation, 2022, 41(6): 712-722.
- [12] 陈一诺, 董绍华. 基于随机森林算法的管道缺陷预测方法[J]. 油气储运, 2018, 37(9): 975-979. DOI: 10.6047/j.issn.1000-8241.2018.09.003.
- CHEN Y N, DONG S H. Pipeline defect prediction method based on Random Forests algorithm[J]. Oil & Gas Storage and Transportation, 2018, 37(9): 975-979.
- [13] MIRJALILI S, SAREMI S, MIRJALILI S M, DOS S. COELHO L. Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization[J]. Expert Systems with Applications, 2016, 47: 106-119. DOI: 10.1016/j.eswa.2015.10.039.
- [14] 黄文聪, 张凤顺, 胡滢, 余文锦, 常雨芳. 基于 IMOGWO 算法的无线电能传输系统参数优化[J/OL]. 控制工程: 1-9[2023-10-20]. <https://doi.org/10.14107/j.cnki.kzgc.20220596>. DOI: 10.14107/j.cnki.kzgc.20220596.
- HUANG W C, ZHANG F S, HU Y, YU W J, CHANG Y

- F. Parameter optimization of wireless power transfer system based on IMOGWO algorithm[J/OL]. Control Engineering of China: 1-9[2023-10-20]. <https://doi.org/10.14107/j.cnki.kzgc.20220596>.
- [15] AL-KAYSSI A W. Use of water retention data and soil physical quality index S to quantify hard-setting and degree of soil compactness indices of gypsiferous soils[J]. Soil and Tillage Research, 2021, 206: 104805. DOI: 10.1016/j.still.2020.104805.
- [16] DE MELO L B B, SILVA B M, PEIXOTO D S, CHIARINI T P A, DE OLIVEIRA G C, CURI N. Effect of compaction on the relationship between electrical resistivity and soil water content in oxisol[J]. Soil and Tillage Research, 2021, 208: 104876. DOI: 10.1016/j.still.2020.104876.
- [17] HUSSON O, HUSSON B, BRUNET A, BABRE D, ALARY K, SARTHOU J P, et al. Practical improvements in soil redox potential (Eh) measurement for characterisation of soil properties. Application for comparison of conventional and conservation agriculture cropping systems[J]. Analytica Chimica Acta, 2016, 906: 98-109. DOI: 10.1016/j.aca.2015.11.052.
- [18] MIRJALILI S, MIRJALILI S M, LEWIS A. Grey wolf optimizer[J]. Advances in Engineering Software, 2014, 69: 46-61. DOI: 10.1016/j.advensoft.2013.12.007.
- [19] XIA Y Y, ZHANG C, WANG C X, LIU H J, SANG X X, LIU R, et al. Prediction of bending strength of glass fiber reinforced methacrylate-based pipeline UV-CIPP rehabilitation materials based on machine learning[J]. Tunnelling and Underground Space Technology, 2023, 140: 105319. DOI: 10.1016/j.tust.2023.105319.
- [20] 方晓彤, 张立涛, 闫秀霞, 董春肖. 基于 BAS-BP 神经网络的医药供应链协同绩效智能评价[J]. 计算机应用与软件, 2023, 40(8): 59-66, 113. DOI: 10.3969/j.issn.1000-386x.2023.08.010.
- FANG X T, ZHANG L T, YAN X X, DONG C X. Intelligent evaluation of pharmaceutical supply chain collaborative performance based on BAS-BP neural network[J]. Computer Applications and Software, 2023, 40(8): 59-66, 113.
- [21] 梁海波, 王怡, 贾武升. 基于机器学习的天然气钢质管道缺陷检测方法研究[J]. 安全与环境学报, 2023, 23(10): 3528-3537. DOI: 10.13637/j.issn.1009-6094.2022.1150.
- LIANG H B, WANG Y, JIA W S. Research on defect detection method for natural gas steel pipeline based on machine learning[J]. Journal of Safety and Environment, 2023, 23(10): 3528-3537.
- [22] VELÁZQUEZ J C, CALEYO F, VALOR A, HALLEN J M. Predictive model for pitting corrosion in buried oil and gas pipelines[J]. Corrosion, 2009, 65(5): 332-342. DOI: 10.5006/1.3319138.
- (编辑: 张腾)
-
- 基金项目:** 国家自然科学基金资助项目“海底油气管道封堵致振机理及气动减振控制方法研究”, 51575528; 中国石油大学(北京)前瞻导向及培育项目“油气管道及站场关键检测技术与实验方法研究”, 2462022QEDX011。
- 作者简介:** 宋福霖, 男, 1998 年生, 在读博士生, 2020 年毕业于辽宁石油化工大学过程装备与控制工程专业, 现主要从事无损检测方向的研究工作。地址: 北京市昌平区府学路 18 号, 102249。电话: 13342127149。Email: 2379991464@qq.com
- 通信作者:** 赵弘, 女, 1975 年生, 教授, 博士生导师, 2003 年博士毕业于西安交通大学机械电子工程专业, 现主要从事无损检测方向的研究工作。地址: 北京市昌平区府学路 18 号, 102249。电话: 13522390529。Email: hzhao_cn@163.com
- Received: 2023-11-22
 - Revised: 2024-01-04
 - Online: 2024-09-09

