

# 基于优化随机森林的 $H_2S$ 腐蚀产物类型 及腐蚀速率预测

曲志豪<sup>1</sup>, 唐德志<sup>2</sup>, 胡丽华<sup>3</sup>, 陈宏健<sup>2</sup>, 李慧心<sup>3</sup>, 贾海韵<sup>1</sup>, 王竹<sup>1</sup>, 张雷<sup>1</sup>

(1.北京科技大学, 北京 100083; 2.中国石油天然气股份有限公司规划总院, 北京 100083;  
3.中海油研究总院, 北京 100028)

**摘要:** **目的** 研究  $H_2S$  环境下碳钢腐蚀产物类型及失重腐蚀速率预测模型, 为含硫油气田管道腐蚀防护设计与选材提供依据。**方法** 整合  $H_2S$  腐蚀模拟实验数据, 采用随机森林算法对各腐蚀因素重要性进行排序, 一方面以腐蚀产物类型为输出量, 通过随机森林分类算法建立硫铁腐蚀产物类别预测模型, 另一方面以腐蚀速率为输出量, 通过随机森林回归算法建立腐蚀速率预测模型, 并与其他模型进行比较。运用网格搜索方法对各类算法的超参数进行优选, 以提高预测可靠性。**结果** 随机森林算法得出的影响  $H_2S$  腐蚀产物类型的因素重要性排序为:  $H_2S$  分压、温度、pH 值、实验周期、总压、 $CO_2$  分压。基于网格搜索优化的随机森林分类模型交叉验证得分超过 0.9,  $f_1$  得分达到 0.96, 优于其他三种常用分类模型。采用网格搜索优化的随机森林回归模型预测结果与实际值的均方误差为 0.86%。相关系数  $R$  值为 0.979, 优于其他两个回归模型。**结论** 网格搜索优化后的随机森林分类、回归模型对含  $H_2S$  复杂环境下的碳钢腐蚀产物类型及腐蚀速率预测准确性较高, 能够为油气田管道腐蚀防护提供参考。

**关键词:** 机器学习; 随机森林;  $H_2S$  腐蚀; 腐蚀产物; 腐蚀速率; 预测模型

**中图分类号:** TG172 **文献标识码:** A **文章编号:** 1001-3660(2020)03-0042-08

**DOI:** 10.16490/j.cnki.issn.1001-3660.2020.03.006

## Prediction of $H_2S$ Corrosion Products and Corrosion Rate Based on Optimized Random Forest

QU Zhi-hao<sup>1</sup>, TANG De-zhi<sup>2</sup>, HU Li-hua<sup>3</sup>, CHEN Hong-jian<sup>2</sup>, LI Hui-xin<sup>3</sup>,  
JIA Hai-yun<sup>1</sup>, WANG Zhu<sup>1</sup>, ZHANG Lei<sup>1</sup>

(1.University of Science and Technology Beijing, Beijing 100083, China; 2.Petro China Planning & Engineering Institute, Beijing 100083, China; 3.CNOOC Research Institute, Beijing 100028, China)

**ABSTRACT:** The work aims to investigate the prediction model of iron sulfide corrosion products and corrosion rate under the

收稿日期: 2019-12-19; 修订日期: 2020-03-08

Received: 2019-12-19; Revised: 2020-03-08

基金项目: 中央高校基本科研业务费专项资金资助项目 (FRF-IC-19-008); 中国博士后科学基金资助项目 (2019M650487)

Fund: Supported by Fundamental Research Funds for the Central Universities (FRF-IC-18-008), Project Funded by China Postdoctoral Science Foundation (2019M650487)

作者简介: 曲志豪 (1994—), 男, 硕士, 主要研究方向为基于机器学习的海底管道腐蚀风险评估与预测。

Biography: QU Zhi-hao (1994—), Male, Master, Research focus: risk assessment and prediction of submarine pipeline corrosion based on machine learning.

通讯作者: 张雷 (1978—), 男, 博士, 教授, 主要研究方向为腐蚀与防护。邮箱: Zhanglei@ustb.edu.cn

Corresponding author: ZHANG Lei (1978—), Male, Ph. D., Professor, Research focus: corrosion and protection of materials. E-mail: Zhanglei@ustb.edu.cn

引文格式: 曲志豪, 唐德志, 胡丽华, 等. 基于优化随机森林的  $H_2S$  腐蚀产物类型及腐蚀速率预测[J]. 表面技术, 2020, 49(3): 42-49.

QU Zhi-hao, TANG De-zhi, HU Li-hua, et al. Prediction of  $H_2S$  corrosion products and corrosion rate based on optimized random forest[J]. Surface technology, 2020, 49(3): 42-49.

$\text{H}_2\text{S}$  environment, and provide basis for designing and selecting the corrosion protection of pipelines in gas fields with sour oil. Based on the collected experimental data of  $\text{H}_2\text{S}$  corrosion, the priority of each corrosion factor was ranked by the random forest algorithm. On the one hand, the prediction model of the corrosion product category was established by the random forest classification algorithm with the corrosion product types as output. On the other hand, the prediction model of the corrosion rate was established by random forest regression algorithm with corrosion rate as output. The grid search method was used to optimize the super parameters of various algorithms to improve the prediction performance. According to the random forest algorithm, the characteristic importance of  $\text{H}_2\text{S}$  corrosion product types was ranked as follows:  $\text{H}_2\text{S}$  partial pressure, temperature, pH, experimental period, total pressure and  $\text{CO}_2$  partial pressure. The cross-validation score of random forest classification model based on grid search optimization exceeded 0.9 and  $f_1$  score reached 0.96, which was better than other three common classification models. The mean square error between the prediction result and the actual value of the random forest regression model with grid search optimization was 0.86%. The  $R$  value of the correlation coefficient was 0.979, which was better than the other two regression models. The random forest classification and regression models optimized by grid search have high accuracy in predicting corrosion product types and corrosion rates in complex  $\text{H}_2\text{S}$  environment, which can provide reference for the corrosion protection pipelines in oil and gas fields.

**KEY WORDS:** machine learning; random forest;  $\text{H}_2\text{S}$  corrosion; corrosion product; corrosion rate; prediction model

随着国家能源需求的日益增长,含硫油气田开发力度逐年增大,由于腐蚀引起的管道穿孔泄漏现象屡有发生,不仅造成了巨大经济损失,并且严重威胁人员安全和周边环境。通过有效的腐蚀预测科学评估管道风险,是油气生产和运输系统安全保障的重要技术手段,国内外众多学者分别从机制模型、经验模型和半经验模型等角度持续开展了多年工作。含  $\text{H}_2\text{S}$  环境中形成的腐蚀产物十分复杂,且显著影响腐蚀发展,明确其腐蚀产物类型对探寻  $\text{H}_2\text{S}$  腐蚀机制及油气田管道腐蚀防护均有重要意义<sup>[1]</sup>。多数学者发现,金属表面腐蚀产物的类型和致密程度对金属腐蚀形态和速率有着重要影响<sup>[2-4]</sup>。特别对于  $\text{H}_2\text{S}$  腐蚀,钢铁表面形成的多种结构及不同保护性的硫化铁化合物,会对腐蚀进程产生重要影响<sup>[5-6]</sup>。Smith 等<sup>[7]</sup>报道了马基诺矿和磁黄铁矿等腐蚀产物的形成存在一条温度- $\text{H}_2\text{S}$  分压的负斜率环境边界。Li 等<sup>[8]</sup>通过研究热力学原理,以温度和  $\text{H}_2\text{S}$  分压两个条件建立了马基诺矿和磁黄铁矿腐蚀产物的热力学模型。Nesic 等<sup>[9]</sup>基于热力学原理和腐蚀电化学研究,考虑环境温度、 $\text{H}_2\text{S}$  分压等参数,结合电位-pH 图对  $\text{H}_2\text{S}$  液相环境中的腐蚀产物进行了初步预测。在腐蚀速率预测方面,含  $\text{H}_2\text{S}$  条件的腐蚀预测研究相对有限。Zhao 等<sup>[10]</sup>使用灰色理论方法预测储油罐的腐蚀速率,分别建立了腐蚀厚度、速率、最大厚度和最大速率的时间变化模型。Hatami 等<sup>[11]</sup>以温度、 $\text{CO}_2$  分压、流速和 pH 为输入,建立了石油管道的腐蚀 SVM 模型。由于管道内腐蚀的影响因素较为复杂,从单一角度对含硫环境下的腐蚀状况进行分析,难以满足实际需求,且当考虑的影响因素较多时,又难以采用数学解析方法建立有效的腐蚀预测模型,导致预测结果与实际现场情况存在较大的偏差。

为了探索利用新的方法建立温度、 $\text{H}_2\text{S}$  分压、 $\text{CO}_2$  分压等多因素下的管道腐蚀风险预测模型,本文通过收集  $\text{H}_2\text{S}$  相关腐蚀实验数据,运用机器学习的方法,

构建随机森林预测模型,同时采用网格搜索算法,对模型超参数进行优化,最终实现对  $\text{H}_2\text{S}$  环境下硫化铁腐蚀产物类型及腐蚀速率的预测,为油气田金属管道的腐蚀防护提供参考依据。

## 1 $\text{H}_2\text{S}$ 腐蚀产物及腐蚀速率的影响因素

$\text{H}_2\text{S}$  环境下钢铁表面形成的腐蚀产物主要由非化学计量比的多晶型硫化铁化合物组成,目前已发现可能存在多种硫化铁固相,其中马基诺矿 (Mackinawite, M)、陨硫铁矿 (Troilite, T) 和磁黄铁矿 (Pyrrhotite, P) 最为常见<sup>[12-14]</sup>,下文均用缩写表示。

不同类型的腐蚀产物会对腐蚀进程产生重要影响<sup>[15-17]</sup>。Sun 等<sup>[18]</sup>认为几种硫化铁化合物腐蚀产物对基体的保护性能各不相同,一般认同其顺序为:马基诺矿<陨硫铁矿<磁黄铁矿<黄铁矿。Shi 等<sup>[6]</sup>通过在 0.1~3 MPa  $\text{H}_2\text{S}$ 、30~120 °C 条件下的实验发现,陨硫铁矿是马基诺矿向磁黄铁矿转化的中间产物,磁黄铁矿的出现会使得均匀腐蚀速率明显降低。Liu 等<sup>[19]</sup>通过高温高压模拟实验研究表明,马基诺矿疏松多孔,对基体的保护性较差,而热力学稳定性更好的磁黄铁矿会减缓腐蚀。腐蚀产物的微观结构和稳定性对腐蚀速率有重要影响。Gao 等<sup>[20]</sup>研究发现,马基诺矿和陨硫铁矿存在时,主要发生均匀腐蚀,当磁黄铁矿及黄铁矿出现时,局部腐蚀迅速发展。Li 等<sup>[21]</sup>通过实验研究发现, $\text{H}_2\text{S}$  腐蚀初期形成的马基诺矿产物对低合金钢有保护作用,磁黄铁矿的生成可以减缓全面腐蚀,促进局部腐蚀的发生和发展。

在  $\text{H}_2\text{S}/\text{CO}_2$  共存条件下,腐蚀速率的影响因素主要有温度、 $\text{H}_2\text{S}$  分压、 $\text{CO}_2$  分压和 pH 值等。有研究表明<sup>[22]</sup>,温度对腐蚀速率的影响主要通过腐蚀产物膜的形成和状态来体现。Huang<sup>[23]</sup>研究发现, $\text{H}_2\text{S}$  和  $\text{CO}_2$  共存环境下, $\text{CO}_2$  分压为 1.2 MPa 时, $\text{H}_2\text{S}$  分压的增

加促进了含硫产物的生成,并降低了碳钢的腐蚀速率。 $\text{CO}_2$ 的存在也影响着 $\text{H}_2\text{S}$ 腐蚀,在相同浓度下, $\text{CO}_2$ 和 $\text{H}_2\text{S}$ 共存时比单独 $\text{H}_2\text{S}$ 时的腐蚀性更强。李涌泉等<sup>[24]</sup>研究发现,钢的腐蚀速率随pH值升高呈下降趋势,pH值直接影响 $\text{H}_2\text{S}$ 对碳钢的腐蚀速率。实际环境中, $\text{H}_2\text{S}$ 腐蚀速率会同时受到温度、 $\text{H}_2\text{S}$ 分压、 $\text{CO}_2$ 分压和

pH值等多个因素的综合影响,其腐蚀进程更为复杂。

本文主要考虑温度、 $\text{H}_2\text{S}$ 分压、 $\text{CO}_2$ 分压和pH值对 $\text{H}_2\text{S}$ 腐蚀产物类型及腐蚀速率的影响,结合文献中数据及本实验室腐蚀模拟实验数据,组成预测模型输入数据集合,具体数据集合见表1,这为腐蚀预测模型提供了数据基础。

表 1 模型输入数据集合  
Tab.1 Input data set of model

No.	Temperature /°C	Test duration /h	$P_{\text{H}_2\text{S}}$ /MPa	$P_{\text{CO}_2}$ /MPa	pH	Corrosion product type	Corrosion rate /(mm·a <sup>-1</sup> )	No.	Temperature /°C	Test duration /h	$P_{\text{H}_2\text{S}}$ /MPa	$P_{\text{CO}_2}$ /MPa	pH	Corrosion product type	Corrosion rate /(mm·a <sup>-1</sup> )
1	60	120	0.3	0.18	3.7	M	0.59	31	90	96	1.2	0.6	3.41	MTP	1.5
2	60	120	1	0.6	3.4	M	0.9	32	80	168	0.65	0	3.58	M	0.64
3	90	168	0.5	0	3.6	MT	0.48	33	120	168	0.1	1.4	3.68	M	0.39
4	30	168	0.1	0	4	M	0.16	34	100	24	1.5	0	3.41	MTP	0.35
5	60	168	0.1	0	3.98	M	0.36	35	110	168	0.56	0	3.63	MT	0.26
6	90	168	0.1	0	3.98	M	0.24	36	110	168	0.86	0	3.53	MT	0.46
7	80	168	0.1	0	3.98	M	0.28	37	65	172	1.5	0	3.42	MTP	0.84
8	30	168	0.5	0	3.66	M	0.45	38	35	120	1.5	0	3.4	M	1.1
9	60	168	0.5	0	3.63	M	0.59	39	90	96	1.61	0	3.39	MP	1.3
10	30	168	1.2	0	3.47	M	1.2	40	100	120	1.5	0	3.5	MP	1.35
11	120	168	1.2	0	3.43	MP	1.1	41	100	96	1.2	0.6	3.3	MP	1.5
12	55	240	1.5	0	3.42	MTP	0.91	42	70	96	1.2	0.6	3.6	MT	1.3
13	100	124	1.5	0	3.41	MTP	1.23	43	75	196	0.1	0.64	3.66	M	0.39
14	30	168	0.33	0.16	3.64	M	0.34	44	100	168	1.8	1.4	3.5	MP	1.6
15	60	168	0.33	0.16	3.66	M	0.49	45	40	240	1.5	0	3.4	MTP	0.72
16	60	120	0.15	0.09	3.82	M	0.23	46	35	72	1.5	0	3.45	MTP	0.49
17	60	120	2.5	1.47	3.97	MTP	1.56	47	35	240	1.5	0	3.3	MTP	0.67
18	150	96	1.2	0.6	3.5	MP	1.3	48	60	168	1.2	0	3.45	MT	1.2
19	120	96	1.2	0.6	3.45	MP	1.5	49	100	168	1.2	0	3.4	MP	0.98
20	120	168	0.2	1.4	3.64	MT	0.8	50	120	128	1.1	1.4	3.44	MP	1.4
21	120	168	0.35	0	3.74	MT	0.89	51	60	120	1	0	3.5	M	0.87
22	90	96	1.2	0.6	3.41	MTP	1.7	52	120	96	1	0.6	0.45	MP	1.7
23	25	168	0.1	0	4.01	M	0.12	53	110	48	0.6	0	3.5	MT	1.35
24	50	168	0.1	0	3.99	M	0.31	54	110	72	0.6	0	3.5	MT	1.64
25	75	168	0.1	0	3.98	M	0.25	55	60	96	3	0	4	MP	0.79
26	90	72	1.5	0	3.4	MP	0.32	56	80	120	1	0	3.6	MT	0.8
27	75	72	1.5	0	3.44	MP	0.29	57	80	96	1	0	3.6	MT	0.6
28	100	72	1.5	0	3.47	MP	0.86	58	25	168	0.1	0	4.0	M	0.12
29	160	72	1.5	0	3.72	MP	0.95	59	60	120	2	1.2	3.2	MTP	1.45
30	60	120	1.5	0.88	3.33	MTP	1.72	60	120	168	1.2	0	3.4	MP	1.1

## 2 腐蚀模拟实验

通过碳钢的腐蚀模拟实验,调整 $\text{H}_2\text{S}$ 分压、 $\text{CO}_2$ 分压、温度,调节溶液pH值,经过不同实验周期的浸泡,可以获得不同条件组合下碳钢的 $\text{H}_2\text{S}$ 腐蚀宏观形态、失重腐蚀速率,并可以根据产物晶体形态,确定腐蚀产物类型。

通过实验结果的整理分析,可以观察到碳钢在不同实验条件下形成了不同组合的马基诺矿、陨硫铁矿、磁黄铁矿腐蚀产物,图1显示了几种较为典型的 $\text{H}_2\text{S}$ 腐蚀产物形态,图2为碳钢表面腐蚀产物的XRD物相分析。综合图1和图2可以看出,在40℃、0.05 MPa  $\text{H}_2\text{S}$ 分压条件下,碳钢表面形成了单一马基诺矿腐蚀产物;当温度增加到60℃、 $\text{H}_2\text{S}$ 分压增加到1.2 MPa时,碳钢表面获得了马基诺矿与针状陨硫

铁矿组合的腐蚀产物；保持 H<sub>2</sub>S 分压不变，升高温度至 90 ℃，可以发现马基诺矿、针状陨硫铁矿和棱柱状磁黄铁矿相互簇拥在碳钢表面；温度为 120 ℃、H<sub>2</sub>S 分压达到 1.2 MPa 时，可以获得少量马基诺矿与六方盘状磁黄铁矿组合的腐蚀产物。

为了进一步丰富数据来源，图 3 综合了腐蚀模拟实验及更多学者的研究工作，显示了不同温度及 H<sub>2</sub>S

分压条件下形成的硫铁腐蚀产物类型。可以发现，马基诺矿是钢在 H<sub>2</sub>S 环境下最主要的腐蚀产物，低温低 H<sub>2</sub>S 含量时，基体表面腐蚀产物主要是马基诺矿，磁黄铁矿主要在高温高 H<sub>2</sub>S 环境下出现，而陨硫铁矿则出现在马基诺矿向磁黄铁矿转化的过程中。以上研究工作中，大多考虑了温度及 H<sub>2</sub>S 分压条件对腐蚀产物的影响，而硫铁腐蚀产物的形成是温度、H<sub>2</sub>S 分压、

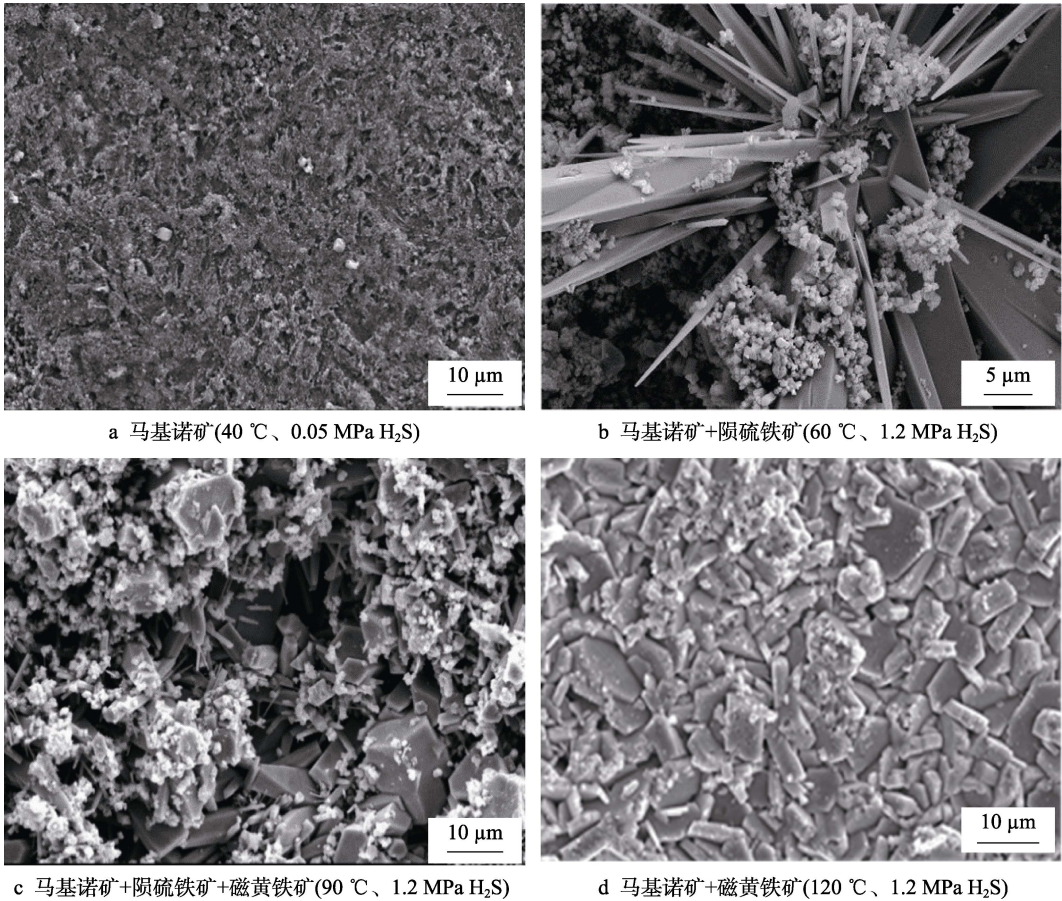


图 1 在不同实验条件下碳钢腐蚀产物膜微观形貌

Fig.1 Microstructure of corrosion product film of carbon steel under different experimental conditions:a) Mackinawite (40 ℃, 0.05 MPa H<sub>2</sub>S); b) Mackinawite+ Troilite (60 ℃, 1.2 MPa H<sub>2</sub>S); c) Mackinawite+Troilite+Pyrrhotite (90 ℃, 1.2 MPa H<sub>2</sub>S); d) Mackinawite+Pyrrhotite (120 ℃, 1.2 MPa H<sub>2</sub>S)

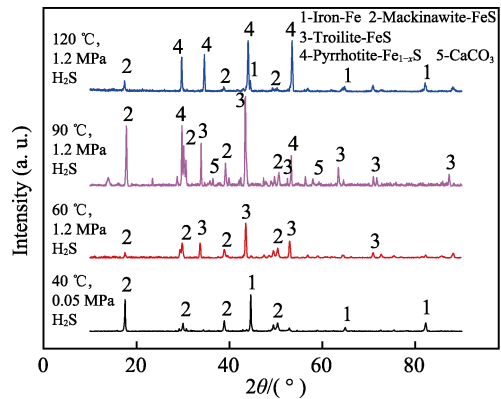


图 2 在不同实验条件下碳钢腐蚀产物的 XRD 物相分析

Fig.2 XRD for phase analysis of corrosion products formed on the surface of carbon steel under different experiment conditions

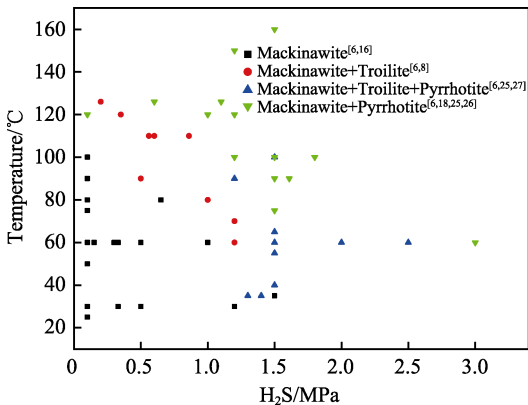


图 3 不同温度及 H<sub>2</sub>S 分压条件下的硫铁腐蚀产物类型

Fig.3 Type of ferrous sulfide corrosion products under different temperature and H<sub>2</sub>S partial pressure



CO<sub>2</sub> 分压、pH 等众多环境因素的综合作用，仅考虑少量因素存在一定局限性。因此本文通过选用机器学习模型，综合考虑多种环境因素，对 H<sub>2</sub>S 环境下的硫铁腐蚀产物类型进行预测。

### 3 随机森林预测模型及优化

随机森林 (Random Forest, RF) 是集群分类模型中的一种<sup>[28]</sup>，可应用于分类问题、回归问题以及特征选择问题<sup>[29]</sup>。其实质是对决策树算法的一种改进，森林中树的个数 ( $n\_estimators$ ) 及树的最大深度 ( $max\_depth$ ) 是随机森林算法中最重要的两个超参数，其对模型最终预测结果产生重要影响。本文通过穷举搜索的方式，将一定范围的超参数进行循环遍历，选取表现最佳的参数组合，即网格搜索 (grid search, GA)，此算法对模型超参数优化方面有着较好的效果<sup>[30]</sup>。图 4 展示了随机森林算法原理。

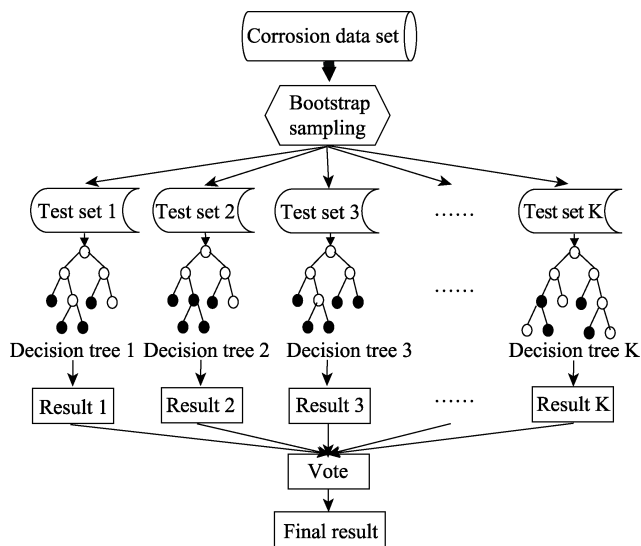


图 4 随机森林模型原理

Fig.4 Schematic diagram of random forest algorithm

### 4 归纳数据集及建模

本文基于 python 编程语言，应用 Spyder3.3.6 软件进行编程。所用数据集来源于本实验室 H<sub>2</sub>S 腐蚀模拟实验数据 (25 组) 及文献调研数据 (35 组)<sup>[6,8,16,18,25-27]</sup>。其中实验周期、温度、H<sub>2</sub>S 分压、CO<sub>2</sub> 分压及 pH 值作为预测模型的输入值，腐蚀产物类型及腐蚀速率作为预测模型的输出值，各自范围为：温度 25~160 °C，H<sub>2</sub>S 分压 0.1~3 MPa，CO<sub>2</sub> 分压低于 1 MPa，pH 值 3~5，实验周期 24~240 h。水化学成分也是内腐蚀的重要影响因素之一，但相关数据收集难度较大，本文暂未讨论。腐蚀产物类型用标签编码方法指定一个数字标签，以便模型识别和计算。

以 3 : 1 的比例将数据集划分为训练集和外部独立测试集，训练集用于对预测模型的构建，外部独立

测试集不用于模型的建立，仅用于对模型的评估。其中训练集数据 45 组，测试集数据 15 组。

针对分类模型适用性评估，本文采用  $K$  折交叉验证 (k-fold cross-validation) 方法进行评估，其中  $K$  取 5。采用查准率 (precision)、召回率 (recall) 和  $f_1$  得分对分类模型预测结果进行评价， $f_1$  得分为查准率与召回率的调和平均值，取值越接近 1，代表模型结果越好，计算公式如下：

$$f_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (1)$$

采用误差百分比、均方误差 (MSE)、相关性分析对回归模型预测结果进行评价。整体建模过程如图 5 所示。

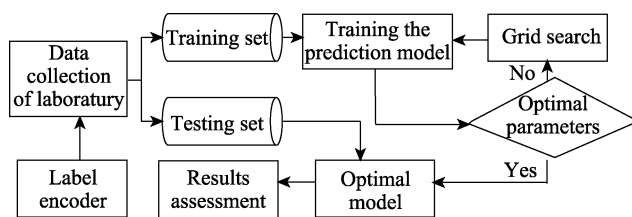


图 5 随机森林模型构建流程

Fig.5 Construction process of random forest model

### 5 模型预测结果与分析

#### 5.1 腐蚀产物类型预测

利用随机森林算法，对 5 个输入量与腐蚀产物类型的相关性进行排序，如图 6 所示。结果表明，H<sub>2</sub>S 分压对硫铁腐蚀产物类型预测影响最大，其次是温度及 pH，实验周期及 CO<sub>2</sub> 分压的影响最小。

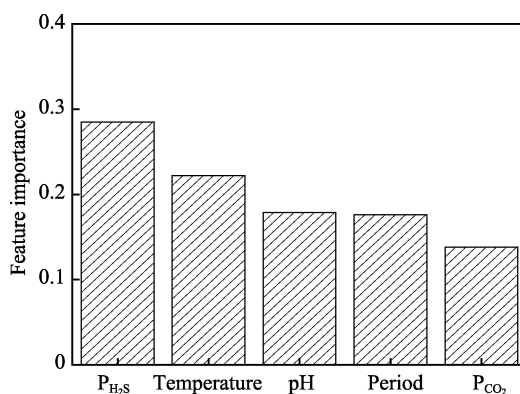


图 6 基于随机森林算法的输入量重要性排序

Fig.6 Importance ordering of input based on random forest algorithm

为了提高随机森林分类模型 (Random Forest Classification, RFC) 的预测精准度，采用网格搜索方法寻求参数最优值，各类参数选取范围如下：树的高度 ( $max\_depth$ ) 范围设置为 [3,4,5,6,7,8]，树的数量 ( $n\_estimators$ ) 范围设置为 [10,50,100,200]。通过网

格搜索算法多次寻优后，确定 RFC 算法最佳参数组合为[max\_depth: 6, n\_estimators: 50]。

为了进一步比较网格搜索优化后的随机森林模型（GA-RFC）的预测能力，本文另增加三个对比模型：基于线性核函数的支持向量机分类模型（SVC-linear）、网格搜索优化后的高斯径向基支持向量机模型（GA-SVC）、朴素贝叶斯分类模型（NB）。其中，GA-SVC 模型超参数设置为：惩罚函数 C=10, gamma=0.001, SVM-linear 及 NB 模型均使用默认参数。

表 2 为四种分类模型经过训练集数据训练后的五折交叉验证得分情况。从表中可知，SVC-linear 及 NB 模型得分较差，训练集及测试集得分均低于 0.87，说明模型对样本数据的分类适应性不好。GA-SVC 模型训练集得分为 0.924，测试集得分为 0.868，表明模型适应性有了一定提高。而 GA-RFC 模型无论针对训练集还是测试集，其得分均超过 0.9，模型适应性最佳。

表 2 不同分类模型的交叉验证得分  
Tab.2 Cross validation scores of different classification models

	GA-SVC		SVC-linear		GA-RFC		NB	
	Train	Test	Train	Test	Train	Test	Train	Test
1	0.92	0.85	0.87	0.77	0.93	0.89	0.89	0.69
2	0.96	0.83	0.83	0.75	0.96	0.91	0.83	0.92
3	0.90	0.92	0.96	0.75	0.97	0.91	0.83	0.92
4	0.90	0.92	0.85	0.67	0.98	0.87	0.90	0.67
5	0.94	0.82	0.82	0.91	0.92	1	0.84	0.91
Average	0.924	0.868	0.866	0.770	0.952	0.916	0.858	0.822

表 3 为不同分类模型预测结果的评价指标，从表中可以看出，针对 15 个测试数据，NB 模型预测结果

的查准率、召回率以及  $f_1$  得分均最低，分别为 0.64、0.63 和 0.63，且共出现了 5 个分类错误，对四类腐蚀产物分类均不理想。可能的原因是，NB 模型有分布独立的假设前提，当特征之间相关度较大时，其分类效果不理想。SVM-linear 模型预测结果的查准率为 0.75，召回率为 0.84， $f_1$  得分为 0.79，出现 3 个分类错误，分类结果不佳主要是本数据集具有非线性特点，SVM-linear 模型对此能力有限。GA-SVM 模型和 GA-RFC 模型的分类预测结果相对较好，仅出现了 1 个分类错误，查准率和召回率均较高， $f_1$  得分分别达到了 0.92 和 0.96，整体分类效果有了较大提升。

表 3 不同分类模型预测结果的评价指标  
Tab.3 Evaluation index of prediction results of different classification models

	GA-SVM	SVM-linear	GA-RFC	NB
Precision	0.88	0.75	0.95	0.64
Recall	0.97	0.84	0.97	0.63
$f_1$	0.92	0.79	0.96	0.63

5.2 腐蚀速率预测

腐蚀速率预测模型的输入数据集与腐蚀产物分类模型输入数据集相同。随机森林回归模型（Random Forest Regression, RFR）的参数优化过程同分类模型，最优参数组合为[max\_depth: 6, n\_estimators: 50]。本文特增加两个对比模型：基于线性核函数的支持向量机回归模型（SVR-linear）和优化后的高斯径向基支持向量机回归模型（GA-SVR）。表 4 为三种回归模型对 15 个目标腐蚀速率的预测结果，其中 SVR-linear 模型及 GA-SVR 模型整体预测误差较大，均方误差

表 4 不同回归模型预测结果  
Tab.4 Results predicted by different regression models

Experimental value /(mm·a <sup>-1</sup> )	GA-SVR		SVR-linear		GA-RFR	
	Predicted value	Error/%	Predicted value	Error/%	Predicted value	Error/%
0.24	0.24	0	0.5	108.33	0.27	12.5
0.91	0.9	-1.09	0.83	-8.79	0.88	-3.3
1.3	1.19	-8.46	1.2	-7.69	1.29	-0.78
0.36	0.22	-38.89	0.66	83.33	0.28	-22.22
0.91	0.85	-6.59	0.82	-9.89	0.91	0
0.86	1.07	24.42	0.69	-19.77	0.84	-2.33
0.59	0.52	-11.86	0.7	18.64	0.55	-6.78
1.56	1.9	21.79	1.6	2.56	1.5	-3.85
0.64	0.61	-4.69	0.44	-31.25	0.56	-12.5
1.3	1.11	-14.62	1.53	17.69	1.21	-6.92
0.34	0.48	41.18	0.36	5.88	0.48	41.18
0.9	1.12	24.44	0.85	-5.56	0.72	-20
0.49	0.5	2.04	0.7	42.86	0.5	2.04
0.23	0.4	73.91	0.6	160.87	0.38	65.22
1.1	0.97	-11.82	0.86	-21.82	0.94	-14.55
MSE/%	2.3		3.7		0.86	

分别达到了 3.7% 和 2.3%，而 GA-RFR 模型整体预测误差最小，均方误差仅有 0.86%。主要原因是支持向量机模型虽然可以建立小样本情况下的非线性映射，但是其本身属于浅层模型，对复杂结构的腐蚀数据集预测效果仍有待提高，而随机森林属于数据驱动的集成类模型，比一般浅层模型层数更多，更适用于腐蚀数据的预测。

三种模型预测值与实验测量值的相关性如图 7 所示，通过计算得知相关系数  $R$  值分别为： $R_{\text{SVR-linear}}=0.882$ ， $R_{\text{GA-SVR}}=0.937$ ， $R_{\text{GA-RFR}}=0.979$ 。GA-RFR 模型的  $R$  值更接近 1，表明其预测结果与实验测量值更接近，模型预测效果更好。

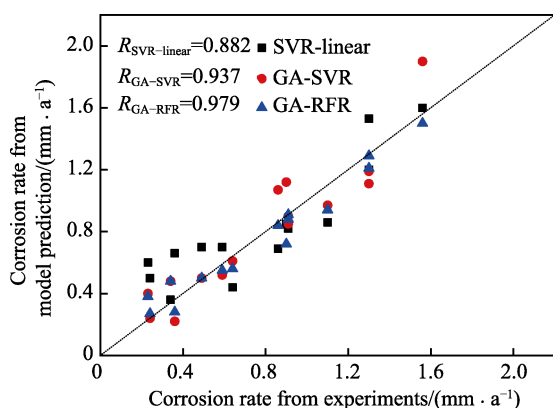


图 7 预测值与实验值的相关性

Fig.7 Correlation of predicted values and experiment values

机器学习模型可综合考虑多个环境因素对硫铁腐蚀产物类型及腐蚀速率进行预测，并且模型相对简单，预测效果较好。其中优化后的随机森林模型预测效果最好，主要是因为随机森林模型是层数较深的集成类模型，对于复杂的腐蚀预测有较好的适用性，这也证明了机器学习算法可应用于油气田管道的腐蚀风险预测，为含硫油气田管道的腐蚀防护提供依据。

## 6 结论

1) 影响  $\text{H}_2\text{S}$  腐蚀产物类型的因素重要性排序为： $\text{H}_2\text{S}$  分压、温度、pH 值、实验周期、 $\text{CO}_2$  分压。GA-RFC 模型对  $\text{H}_2\text{S}$  腐蚀产物分类预测效果较好，交叉验证得分超过 0.9， $f_1$  得分达到了 0.96，优于常见的 GA-SVC、SVC-linear 及 NB 模型，这为预测复杂环境下  $\text{H}_2\text{S}$  腐蚀产物类型提供了参考。

2) 针对复杂的腐蚀速率数据，GA-RFR 模型整体预测误差较小，均方误差仅有 0.86%，相关系数  $R$  值为 0.979，均优于 GA-SVR、SVR-linear 模型。属于集成类算法的 RFR 模型可用于复杂工况下的  $\text{H}_2\text{S}$  腐蚀速率预测，这对含硫油气田管道腐蚀防护有着重要意义。

## 参考文献:

- [1] SHOESMITH D W, TAYLOR P, BAILEY M G, et al. Electrochemical behaviour of iron in alkaline sulphide solutions[J]. *Electrochimica acta*, 1978, 23(9): 903-916.
- [2] ZHENG Shu-qi, ZHOU Cheng-shuang, CHEN Xing-yang, et al. Dependence of the abnormal protective property on the corrosion product film formed on  $\text{H}_2\text{S}$ -adjacent API-X52 pipeline steel[J]. *International journal of hydrogen energy*, 2014, 39(25): 13919-13925.
- [3] 李国敏, 刘烈伟, 郑家桑. 碳钢在含硫化氢及高压二氧化碳饱和的 NaCl 溶液中的腐蚀行为[J]. *中国腐蚀与防护学报*, 2000(4): 204-210.  
LI Guo-min, LIU Lie-wei, ZHENG Jia-shen. Corrosion behavior of carbon steel in high pressure carbon dioxide saturated NaCl solutions containing hydrogen sulfide[J]. *Journal of Chinese society for corrosion and protection*, 2000(4): 204-210.
- [4] MA Hou-yi, CHENG Xiao-liang, LI Gui-qiu, et al. The influence of hydrogen sulfide on corrosion of iron under different conditions[J]. *Corrosion science*, 2000, 42(10): 1669-1683.
- [5] SMITH S N, JOOSTEN M. Corrosion of carbon steel by  $\text{H}_2\text{S}$  in  $\text{CO}_2$ -containing oilfield environments[C]// NACE International. Houston: [s. n.], 2006: 06115.
- [6] SHI Feng-xian, ZHANG Lei, YANG Jian-wei, et al. Polymorphous FeS corrosion products of pipeline steel under highly sour conditions[J]. *Corrosion science*, 2016, 102: 103-113.
- [7] SMITH S N, PACHECO J L. Prediction of corrosion in slightly sour environments[C]// NACE International. Houston: [s. n.], 2002.
- [8] LI Hui-xin, ZHANG Lei, SHI Feng-xian, et al. The effect of corrosion product formation mode at the early stage of sour corrosion of pipeline steel[C]// NACE International. Houston: [s. n.], 2018.
- [9] NING Jing, ZHENG You-gui, BROWN Bruce, et al. A thermodynamic model for the prediction of mild steel corrosion products in an aqueous hydrogen sulfide environment[J]. *Corrosion*, 2015, 71(8): 945-960.
- [10] ZHAO Xiao-gang, ZHOU Yi, ZHAO Jian-yu, et al. Safety prediction of soleplate corrosion state in petroleum storage tank based on grey theory model[J]. *Chemical engineering transactions*, 2016, 51: 271-276.
- [11] HATAMI S, GHADERI-ARDAKANJ A, NIKNEJAD-KHOMAMI M, et al. On the prediction of  $\text{CO}_2$  corrosion in petroleum industry[J]. *The journal of supercritical fluids*, 2016, 117: 108-112.
- [12] SHOESMITH D W, TAYLOR P, BAILEY M G, et al. The formation of ferrous monosulfide polymorphs during the corrosion of iron by aqueous hydrogen sulfide at 21 °C[J]. *Journal of Electrochemical Society*, 1980, 127(5): 1007-

- 1015.
- [13] NESIC S. Key issues related to modelling of internal corrosion of oil and gas pipelines: a review[J]. Corrosion science, 2007, 49: 4308-4314.
- [14] NING Jing, ZHENG You-gui, YOUNG David, et al. A thermodynamic study of hydrogen sulfide corrosion of mild steel[J]. Corrosion, 2014, 70(4): 375-389.
- [15] 孙建波, 苏鑫, 张勇. 高温高压  $H_2S/CO_2$  腐蚀产物膜对低铬钢氢渗透行为的影响[J]. 表面技术, 2018(6): 17-23.  
SUN Jian-bo, SU Xin, ZHANG Yong. Effect of  $H_2S/CO_2$  Corrosion scales on the hydrogen permeation behavior of low chromium steels[J]. Surface technology, 2018(6): 17-23.
- [16] BAI Peng-peng, ZHENG Shu-qi, ZHAO Hui, et al. Investigations of the diverse corrosion products on steel in a hydrogen sulfide environment[J]. Corrosion science, 2014, 87: 397-406.
- [17] NING Jing, ZHENG You-gui, BROWN Bruce, et al. The role of iron sulfide polymorphism in localized  $H_2S$  corrosion of mild steel[J]. Corrosion, 2017, 72(2): 155-168.
- [18] SUN W, NESIC S, PAPAIVINASAM S. Kinetics of iron sulfide and mixed iron sulfide/carbonate scale precipitation in  $CO_2/H_2S$  corrosion[C]// NACE International. Houston: [s. n.], 2006.
- [19] LIU Ming, WANG Jian-qiu, KE Wei. Corrosion behavior of X52 pipeline steel in high  $H_2S$  concentration solutions at temperatures ranging from 25 °C to 140 °C[J]. Corrosion engineering science and technology, 2013, 48(5): 380-388.
- [20] GAO S J, BROWN B, YOUNG D, et al. Formation of iron oxide and iron sulfide at high temperature and their effect on corrosion[J]. Corrosion science, 2018, 135: 167-176.
- [21] 李慧心. 高含  $H_2S$  环境下低合金钢腐蚀产物演化及其对腐蚀行为的影响[D]. 北京: 北京科技大学, 2019.  
LI Hui-xin. Formation and evolution of  $H_2S$  corrosion products and the influence on corrosion mechanism under high  $H_2S$  partial pressure[D]. Beijing: University of Science and Technology Beijing, 2019.
- [22] 艾志久, 范钰伟, 赵乾坤, 等.  $H_2S$  对油气管材的腐蚀及防护研究综述[J]. 表面技术, 2015(9): 108-115.  
AI Zhi-jiu, FAN Yu-wei, ZHAO Qian-kun, et al. Review on  $H_2S$  corrosion of oil gas tubing and its protection[J]. Surface technology, 2015(9): 108-115.
- [23] HUANG X B. Corrosion of N80 tubing steel in brine at 1.2 MPa  $CO_2$  containing trace amounts of  $H_2S$ [J]. Corrosion engineering, science and technology, 2012, 47(1): 78-83.
- [24] 李涌泉, 田进. pH 值对 N80 钢固液两相冲刷腐蚀的影响[J]. 表面技术, 2013, 42(5): 29-31.  
LI Yong-quan, TIAN Jin. Influence of pH on erosion-corrosion of N80 steel in fluids with solid particles[J]. Surface technology, 2013, 42(5): 29-31.
- [25] WIKJORD A G, RUMMERY T E, DOERN F E, et al. Corrosion and deposition during the exposure of carbon steel to hydrogen sulphide-water solutions[J]. Corrosion science, 1980, 20(5): 651-671.
- [26] LIU Ming, WANG Jian-qiu, KE Wei, et al. Corrosion behavior of X52 anti- $H_2S$  pipeline steel exposed to high  $H_2S$  concentration solutions at 90 °C[J]. Journal of materials science & technology, 2014, 30(5): 504-510.
- [27] 杨建伟, 张雷, 丁睿明, 等. X60 管线钢在湿气和溶液介质中的  $H_2S/CO_2$  腐蚀行为[J]. 金属学报, 2008, 44(11): 1366-1371.  
YANG Jian-wei, ZHANG Lei, DING Rui-ming, et al.  $H_2S/CO_2$  corrosion behavior of X60 pipeline steel in wet gas and solution[J]. Acta metallurgica sinica, 2008, 44(11): 1366-1371.
- [28] JUREK-LOUGHREY A, BI Ya-xin, WU Sheng-li, et al. A survey of commonly used ensemble-based classification techniques[J]. The knowledge engineering review, 2014, 29(5): 551-581.
- [29] 王全才. 随机森林特征选择[D]. 大连: 大连理工大学, 2011.  
WANG Quan-cai. Random forest feature selection[D]. Dalian: Dalian University of Technology, 2011.
- [30] 郭超, 宋卫华, 魏威, 等. 基于网格搜索-支持向量机的采场顶板稳定性预测[J]. 中国安全科学学报, 2014, 24(8): 31-36.  
GUO Chao, SONG Wei-hua, WEI Wei, et al. Stope roof stability prediction based on both SVM and grid-search method[J]. China safety science journal, 2014, 24(8): 31-36.