

基于 XGBoost 的管道防腐层裂纹损伤识别算法研究

吕瑞宏, 王春云, 赵艺伟, 王晨丽
(沈阳工业大学 信息科学与工程学院, 沈阳 110870)

摘要: 埋地管道因为在长期服役后防腐层产生的裂纹易导致管道金属层受到损害, 为了避免损伤造成管道泄漏, 有必要建立管道防腐层裂纹损伤识别智能化平台; 通过引入集合经验模态分解, 提取管道防腐层结构在裂纹损伤状态下的频域特征值并构建裂纹损伤数据系统; 依次利用孤立森林、缺失森林和核主成分分析对损伤数据进行异常检测、数据补全和特征降维优化, 构建适于管道防腐层裂纹损伤识别的 extreme gradient boosting (XGBoost) 模型; 研究表明: 基于 XGBoost 的管道防腐层裂纹损伤识别模型可准确地对裂纹长度进行有效检测, 与 gradient boosting decision tree (GBDT) 和随机森林算法相比, 其决定系数最大且均方误差最小, 裂纹长度识别误差保持在 4.37 厘米以内, 为管道防腐层结构健康检测和安全运输提供了有效的识别方法。

关键词: 管道防腐层检测; 裂纹长度识别; 孤立森林; 缺失森林; 核主成分分析; 极端梯度提升树; 集合经验模态分解; 频域特征值

Research on Crack Damage Identification Algorithm of Pipeline Anticorrosion Coating Based on XGBoost

LÜ Ruihong, WANG Chunyun, ZHAO Yiwei, WANG Chenli

(School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China)

Abstract: Because buried pipeline services for a long time, cracks in the anti-corrosion coating are easy to damage the metal layer of pipelines, in order to avoid the damage caused by pipeline leakage, it is necessary to establish a pipeline anticorrosive coating crack damage identification intelligent platform. Ensemble empirical mode decomposition (EEMD) is introduced to extract the frequency domain characteristic values of pipeline anticorrosion layer structure in crack damage state, and build a crack damage data system. Isolation forest, Miss forest and Kernel principal component analysis are successively used to conduct anomaly the detection, data completion and feature dimension reduction optimization for damage data, and extreme gradient boosting (XGBoost) model is constructed for the crack damage identification of pipeline anticorrosion coating. Research shows that, based on XGBoost effectively for the crack length, the pipeline coating crack damage detection model can be accurately detected, compared with gradient boosting decision tree (GBDT) and random forest algorithms, the maximum coefficient and minimum mean square error are determined, the length of crack identification error keeps within 4.37 cm, which provides the effective identification method of pipeline anticorrosion layer structure for health detection and safety transportation.

Keywords: inspection of pipeline anticorrosion coating; crack length identification; isolation forest; miss forest; kernel principal component analysis; extreme gradient boosting; ensemble empirical mode decomposition; frequency domain eigenvalue

0 引言

由于土壤腐蚀和自然环境的影响, 长输埋地管道防腐层产生的裂纹易使管道金属层出现损伤, 从而导致管道泄漏、断裂甚至爆炸, 因此对管道防腐层进行裂纹损伤识别具有重要意义^[1-2]。管道防腐层损伤检测是保障管道安全运输的重要研究内容, 利用超声导波技术检测管道防腐层缺陷具有明显优势^[3-4]。

目前, 李强林等证明了超声导波检测技术可用于埋地管道的快速检测^[5]; 张广纯等应用超声导波检测技术识别了铝板缺陷的类型、大小及位置信息^[6]; 杨辰等将位移值和固有频率值相互融合成一种新的损伤特征向量矩阵, 该方法可以有效地对损伤位置以及程度进行识别和评估^[7]; 刘文光等提出以固有频率为损伤特征向量的裂缝识别方式; 王红君等提出一种基于集合经验模态分解 (EEMD, ensemble empirical mode decomposition) 小波阈值去噪和布谷鸟算法优化 BP 神

收稿日期: 2022-09-02; 修回日期: 2022-11-03。

基金项目: 辽宁省教育厅研究经费项目 (LQGD2020021)。

作者简介: 吕瑞宏 (1982-), 女, 山西河曲人, 博士, 副教授, 主要从事长输油气管道内检测技术、超声检测、弱信号处理方向的研究;

通讯作者: 王春云 (1997-), 女, 河北沧州人, 硕士研究生, 主要从事长输油气管道内检测技术、超声检测、弱信号处理方向的研究。

引用格式: 吕瑞宏, 王春云, 赵艺伟, 等. 基于 XGBoost 的管道防腐层裂纹损伤识别算法研究[J]. 计算机测量与控制, 2023, 31(5): 14-20, 27.

神经网络的方法, 实现了风电机组齿轮箱中故障的诊断^[9]; 杨晓晖等提出基于多粒度级联孤立森林 (IForest, isolation forest) 的检测模型, 提高了在高维数据中异常值检测的精确度^[10]; 吴小姣等^[11]将缺失森林 (MF, miss forest) 填补法与贝叶斯线性回归法、蒙特卡洛多重填补法、K 邻近填补法进行对比, 结果表明缺失森林在缺失数据中填补优势最大; 李仙琳^[12]等利用核主成分分析 (KPCA, kernel principal components analysis) 算法对钻井各参数进行信息整合, 结果表明基于 KPCA 优化的模型较其他模型具有较高的溢流诊断及模型泛化能力; 夏田等^[13]采用小波包分解与梯度提升树 (GBDT, gradient boosting decision tree) 算法相结合的方法对轴承故障进行损伤识别, 结果表明此方法在轴承损伤识别方面表现优异; 李超^[14]采用极端梯度提升 (XGBoost, extreme gradient boosting) 算法构建爆管风险评估模型, 结果证明基于 XGBoost 的管道爆管风险评估模型具有较高的精准性与稳定性。

基于以上研究, 针对管道防腐层结构的固有频率值对于不同损伤长度的裂纹的敏感程度不同, 本文以频域特征固有频率值作为管道防腐层裂纹损伤识别的特征向量, 利用孤立森林和缺失森林对频域特征值进行异常检测和数据库, 并应用 KPCA 对得到的损伤数据库进行特征降维, 获取优化后的裂纹损伤数据集, 构建基于 XGBoost 的管道防腐层裂纹损伤识别模型, 实现裂纹损伤长度的判定, 并通过有限元仿真与现场实验验证其有效性。

1 裂纹损伤数据系统的构建与优化

1.1 裂纹损伤数据系统的构建

在管道防腐层裂纹损伤识别中, 采集到的原始裂纹损伤信号 $s(t)$ 是非线性的, 设定 $s(t)$ 为:

$$s(t) = x(t) + n(t) \quad (1)$$

式中, $x(t)$ 为裂纹损伤信号, $n(t)$ 为噪声信号。为了从混叠的 $s(t)$ 中提取裂纹损伤信息, 消除噪声的干扰, 本文选用 EEMD 方法对 $s(t)$ 进行去噪分解, 在 EEMD 中引入噪声辅助信号分析, 可以有效解决经验模态分解 (EMD, empirical mode decomposition) 中可能出现的模态混叠现象。EMD 是将原始信号 $X(t)$ 分解为一系列本征模态函数 $m_i(t)$ 和一个残余量 $r_N(t)$ 之和, 表达式为:

$$X(t) = \sum_{i=1}^N m_i(t) + r_N(t) \quad (2)$$

式中, $m_i(t)$ 为本征模态函数, $r_N(t)$ 为残余量。

EEMD 在 $s(t)$ 中添加白噪声辅助分析, 将 $s(t)$ 分解为多个尺度分量之后继续进行总体平均运算, 去除白噪声的干扰, 获取最终的信号分解量^[15]。基本的实施流程如下: 首先选择集总平均次数 N 和标准高斯白噪声的标准差 σ 生成均值为 0、标准差为 $\sigma std(x)$ 的高斯白噪声并加入到 $s(t)$ 中, 然后对合成信号进行 EMD 处理并重复上述步骤共 N 次, 最后取 N 次分解得到的 IMF 分量的平均作为 EEMD 的 IMF 分量。

对采集的 $s(t)$ 进行 EEMD 时频分析, 得到 10 个 IMF

分量 $M_1(t), \dots, M_{10}(t)$ 和 1 个残余分量 $R_1(t)$, 由于前六阶 IMF 分量构成了 $s(t)$ 的主要瞬时信号, 后四阶 IMF 分量的能量分布较小, 其对于 $s(t)$ 的影响可忽略不计, 因此本文只取前 6 个 IMF 分量 $M_1(t), \dots, M_6(t)$ 进行分析。

对上述经过 EEMD 时频分析处理后获得的 $M_1(t), \dots, M_6(t)$ 进行 FFT 频域特征处理, 提取裂纹损伤信号的频域特征值 $C_i(k), i = 1, 2, \dots, 6$, 公式如式 (3) 所示:

$$FFT[ac_1(t) + bc_2(t) + \dots] = aC_1(k) + bC_2(k) + \dots \quad (3)$$

式中, a 与 b 均为常数, 通过 FFT 计算得到各个分量的频域特征 $C_1(k), \dots, C_6(k)$ 构成管道防腐层裂纹损伤数据集 S , 表达式为:

$$S = \{C_1, C_2, C_3, C_4, C_5, C_6\} \quad (4)$$

其中: C_1, \dots, C_6 为裂纹损伤信号的六阶频域特征值。

1.2 基于孤立森林的损伤数据异常检测

在对管道防腐层双层结构进行健康检测的过程中, 不可避免地会受到环境影响或人为因素的干扰, 可能导致采集到的管道损伤数据中存在异常值。管道损伤数据集中的异常值会影响后续进行管道损伤的类型与尺寸识别, 为减小管道损伤检测的误差, 需要对管道损伤数据集进行异常值识别, 检测出损伤数据集中的异常值并进行剔除。

孤立森林 (IForest) 异常值检测算法的基学习器是决策树^[16], IForest 的核心是将 S 中的异常值孤立出来^[17], 该算法在损伤数据异常识别中性能表现优异。IForest 定义异常样本为离群点, 离群点离分布密集的正常样本的距离较远, 而且离群样本通常呈现稀疏分布。IForest 针对异常值的孤立方法如图 1 所示, 对 S 进行多次样本分割, 因异常点分布稀疏的特性, 只需要极少次的分割就能够将其检测出来。

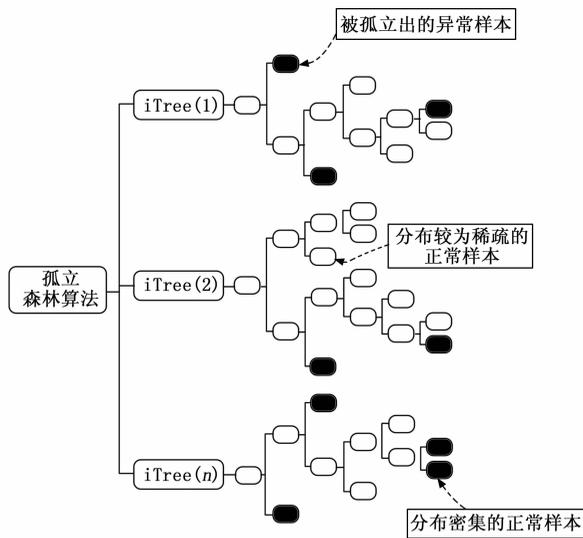


图 1 运用孤立森林进行异常检测

利用 IForest 进行异常检测时, 首先对 S 随机抽样并进行训练, 设定孤立树为 100 棵, 异常值比例为 0.05, 此时 S 为 6 维数据。应用 IForest 对 S 进行异常检测可视化时, 由

于结果图受 2 维平面的限制, 因此选择 S 中的两维特征进行可视化操作, 结果图如图 2 所示。

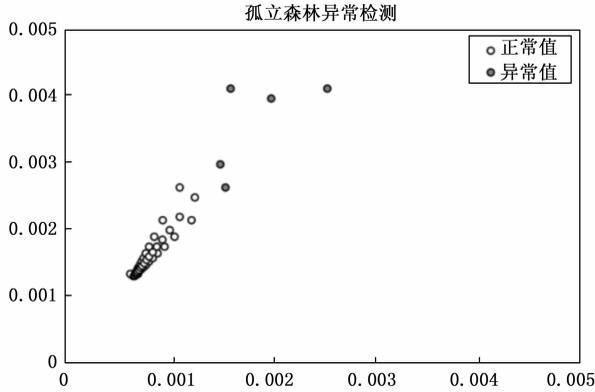


图 2 裂纹损伤数据集异常检测可视化

从统计学的角度分析, 裂纹损伤数据中的正常值分布在稀疏区域的机率很小, 因此能够确定分布在稀疏区域的裂纹损伤数据为异常值, 应用 IForest 将 S 中的异常值识别并剔除后构建新的裂纹损伤数据集 X 。

1.3 基于缺失森林的损伤数据补全

由于应用 IForest 对 S 进行异常值检测后剔除异常数据会导致数据缺失, 因此需要通过数据补全的方法补全 X 。传统的数据补全方式有均值填补、中位数填补以及众数填补等, 而采用传统的填补方式来补全 X 与 S 相比会产生较大的误差, 故本文选择缺失森林 (MF) 算法对 X 进行数据补全。MF 算法弥补了传统的数据补全方法的不足, MF 是基于随机森林的一种较为常用的数据补全算法, 首先通过完整部分的 X 训练出随机森林预测模型, 然后对 X 中缺失的损伤数据进行预测, 最后通过不断迭代来补全 X 中全部缺失的损伤数据。在缺失森林算法参数的设定中, 决策树数量为 500, 选择有放回采样, 并行决策树数量为 1。

裂纹损伤数据集 X 的数据补全过程如下:

- 1) 利用传统填补方式, 例如均值填补、众数填补对 X 进行填补;
- 2) X 的缺失列按缺失率进行排序, 缺失列的指标集记为 M ;
- 3) 当不满足停止准则 γ 时:
 - (1) 存储现有填补矩阵设为 R_{old}^{imp} ;
 - (2) 对于 $i \in M$, 将训练集 $y_{obs}^{(i)}$ 和 $r_{obs}^{(i)}$ 通过随机森林算法训练模型, 通过测试集 $r_{mis}^{(i)}$ 预测 $y_{mis}^{(i)}$;
 - (3) 利用模型预测的 $y_{mis}^{(i)}$ 来填补 X , 记为 X_{new}^{imp} ;
 - (4) 依次对 M 中缺失数据进行数据补全, 直到满足停止准则 γ 或已达到最大的迭代次数。
- 4) 最终对 X 进行数据补全, 构建补全后的裂纹损伤数据集为 D 。

1.4 基于 KPCA 的损伤数据特征降维

为了更全面反映裂纹损伤信号中包含的信息, 需要尽可能多的采集裂纹特征参数, 然而由于原始特征参数的维

数过大和相关性会为后续进行裂纹损伤长度识别造成影响, 因此本文利用 KPCA 对裂纹损伤数据集 D 进行有效性分析, 筛选出最能表征不同缺陷差异的裂纹特征参数, 并将高维原始特征空间降维到低维特征空间。

KPCA 是一种基于线性 PCA 的非线性扩展降维方法, KPCA 原理是将 D 的样本空间映射到更高维度的空间, 再进行 PCA 处理。假设裂纹损伤样本 $d_1, d_2 \dots d_M, \{d_i\}$ 表示裂纹损伤样本空间, 通过 KPCA 的映射为 Φ , 公式定义:

$$\begin{aligned} \Phi &= R^d \rightarrow F \\ da\xi &= \Phi(d) \end{aligned} \quad (5)$$

样本空间 d 到 F 的映射通过核函数实现, 由核函数映射的样本空间满足中心化条件为:

$$\sum_{\mu=1}^M \Phi(d_{\mu}) = 0 \quad (6)$$

则特征空间中的协方差矩阵为:

$$C = \frac{1}{M} \sum_{\mu=1}^M \Phi(d_{\mu}) \Phi(d_{\mu})^T \quad (7)$$

得到 C 的特征值 $\lambda (\lambda \geq 0)$ 和特征向量 v 为:

$$\begin{aligned} (\Phi(d_v) \cdot C_v) &= \lambda (\Phi(d_v) \cdot v), \\ v &= \sum_{i=1}^M \alpha_i \Phi(d_i) \end{aligned} \quad (8)$$

则有:

$$\begin{aligned} \frac{1}{M} \sum_{\mu=1}^M \alpha_{\mu} \left(\sum_{\omega=1}^M (\Phi(x_{\nu})) \cdot \Phi(x_{\omega}) \right. \\ \left. \Phi(d_{\omega}) \Phi(d_{\mu}) \right) &= \lambda \sum_{\mu=1}^M (\Phi(d_{\nu})) \cdot \Phi(d_{\mu}) \end{aligned} \quad (9)$$

其中: $\nu = 1, 2, \dots, M$ 。定义 $M \times M$ 维矩阵 K :

$$K_{\nu\omega} = (\Phi(d_{\nu}) \cdot \Phi(d_{\omega})) \quad (10)$$

则式 (9) 简化为:

$$M\lambda K\alpha = K^2\alpha \quad (11)$$

由此可得 λ 和 v , 对于裂纹损伤样本在特征向量空间 V^k 中映射为:

$$(v^k \cdot \Phi(d)) = \sum_{i=1}^M (\alpha_i)^k (\Phi(d_i), \Phi(d)) \quad (12)$$

将核函数替换内积, 则有:

$$(v^k \cdot \Phi(d)) = \sum_{i=1}^M (\alpha_i)^k K(d_i, d) \quad (13)$$

当式 (6) 不成立时, 需调整核矩阵为:

$$K_{\nu\omega} \rightarrow K_{\nu\omega} - \frac{1}{M} \left(\sum_{\omega=1}^M K_{\nu\omega} + \sum_{\omega=1}^M K_{\omega\nu} \right) + \frac{1}{M^2} \sum_{\omega, \tau=1}^M K_{\omega\tau} \quad (14)$$

本文采用 KPCA 方法对裂纹损伤数据集 D 进行特征提取, 处理过程如下:

首先获取 n 个指标, 其中每个指标均有 m 个裂纹损伤样本, 记为 $m \times n$ 维裂纹损伤数据矩阵, 然后选择高斯径向核函数并计算核矩阵 K , 通过公式 (14) 来调整核矩阵并计算裂纹损伤数据的特征值 $\lambda_1, \dots, \lambda_n$ 以及对应的特征向量 v_1, \dots, v_n ; 将特征值进行排序, 对损伤特征向量正交化得到 $\alpha_1, \dots, \alpha_n$, 计算损伤特征值贡献度 B_1, \dots, B_n 并对贡献度低的维度进行降维处理; 最后通过 K 计算裂纹损伤特征向量投影 $Y = KL \cdot \alpha$, 其中 $\alpha = (\alpha_1, \dots, \alpha_n)$, Y 即为经过 KPCA 降维优化

后的裂纹损伤数据集。

裂纹损伤数据集 D 中包含 6 维频域特征向量, 对其进行 KPCA 特征降维, 计算裂纹各个损伤特征的特征值和贡献度分别如图 3 和图 4 所示, 其中 PC1 到 PC6 对应 D 中的 6 维裂纹损伤特征。

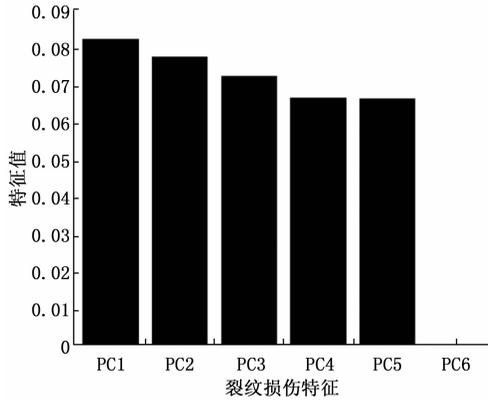


图 3 各损伤特征的特征值

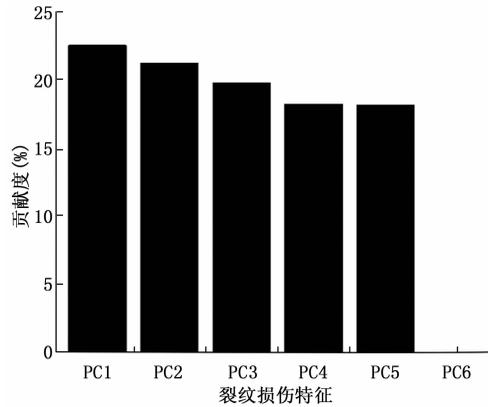


图 4 各损伤特征的贡献率

由图 3 可知, PC6 的特征值可以忽略不计, 由图 4 可知, PC1 到 PC5 的贡献率累加能够覆盖 99% 的裂纹损伤特征信息, 所以本文选取前 5 阶裂纹损伤特征向量构建新的裂纹损伤数据集 $Y = \{Y_1, \dots, Y_5\}$ 。

2 基于 XGBoost 的裂纹损伤识别

由于上述构建的裂纹损伤数据集 Y 是非线性的, 因此本文选择非线性逼近能力优秀、预测准确率高且训练时间短的极端梯度提升树 (XGBoost) 算法进行裂纹损伤长度识别。XGBoost 算法是在 GBDT 算法上进行优化, 通过多棵决策树来组合拟合上次回归预测反馈的残差, 是一种基于串行的集成学习算法^[18-21]。

XGBoost 的基学习器选择分类回归树 (CART), 模型公式如下:

$$\hat{y}_i = \varphi(x_i) = \sum_{k=1}^K f_k(x_i), f_k \in \Psi \quad (15)$$

其中: $\Psi = \{f(x) = \omega_{q(x)}\} (q: R^m \rightarrow T, \omega \in R^T)$, q 表示分类回归树映射损伤样本到叶节点的分数, $\omega_{q(x)}$ 表示所以分

类回归树映射分数的集合, T 代表分类回归树叶节点的数量, 则模型的目标函数为:

$$L(\Phi) = \sum_i l(\hat{y}_i, y_i) + \sum_K \Omega(f_k) \quad (16)$$

$$\text{where } \Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (17)$$

其中: $\sum_i l(\hat{y}_i, y_i)$ 与 $\sum_K \Omega(f_k)$ 分别代表目标函数的损失函数与目标函数的正则化, ω 代表叶子节点的分数。

决策树预测的残差值被生成的决策树拟合, 当 t 棵决策树生成后, 得到的公式为:

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (18)$$

将目标函数记为:

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (19)$$

提取第 t 棵分类回归树的预测值进行优化, 可以使用二阶泰勒展开近似表示为:

$$L^{(t)} \cong \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)} + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) \right] + \Omega(f_t) \quad (20)$$

公式 (20) 中 g_i 和 h_i 分别为 i 个样本预测值的一阶梯度和二阶梯度:

$$g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \quad (21)$$

$$h_i = \partial_{\hat{y}^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)}) \quad (22)$$

移除 $f_t(x_i)$ 无关项, 定义 $I_j = \{i | q(x_i) = j\}$ 为第 t 棵决策树中第 j 个叶子节点所在的损伤样本, 对损失函数进行简化得到:

$$L^{(t)} = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \quad (23)$$

计算得到第 t 次迭代中损失函数极值点的值为:

$$\omega_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (24)$$

将 ω_j^* 带入损伤函数中求出最优解目标函数为:

$$\tilde{L}^{(t)}(q) = - \frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (25)$$

由于原始裂纹损伤信号 $s(t)$ 是非线性的, 因此采用 EEMD 方法对 $s(t)$ 进行去噪分解, 得到多个 IMF 分量, IMF 分量中同时包含了裂纹损伤信息和噪声信息, 应用 FFT 提取频域特征固有频率值并构建裂纹损伤数据集 S , 利用 S 作为模型输入直接进行裂纹长度识别时会导致数据精度下降从而产生误判, 因此需要对 S 进行数据优化。由于 S 的复杂程度较高, 采集过程又存在环境影响和人工失误等因素, 可能会造成 S 中存在异常值。为了消除异常值, 采用 IForest 算法对 S 进行异常值检测。通过 IForest 识别并剔除异常值后, 构建的裂纹损伤数据集 X 中存在缺失值, 采用缺失森林算法对 X 进行数据补全, 构建裂纹损伤数据集 D 。KPCA 能够优化裂纹损伤数据, 提取主要损伤特征向量, 得到 5 阶裂纹损伤固有频率值 Y 作为裂纹损伤长度识

别的特征向量。因此，本文提出了 IF、MF、KPCA 优化 XGBoost 的管道防腐层裂纹损伤识别算法，由以上分析提出基于 XGBoost 的裂纹损伤长度识别模型的流程图如图 5 所示。

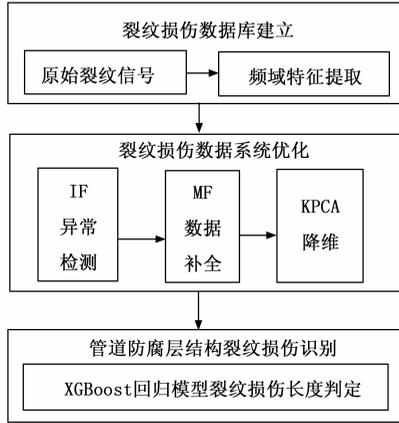


图 5 裂纹损伤长度识别模型流程图

3 含裂纹的管道防腐层有限元仿真

当管道的半径足够大时，可将管道防腐层双层结构看作管道—防腐层双层板状结构来进行研究，管道防腐层双层结构微体元模型如图 6 所示，其中钢制管道内径 R_1 、外径 R_2 ，防腐层厚度为 $h_v = (R - R_2)$ ，钢管厚度为 $h_E = (R_2 - R_1)$ ，取钢制管道中厚度为 $R - R_1$ 、宽度为 d 、体积为 V 的单元体可看作宽度为 d 的钢板—防腐层双层板状结构。

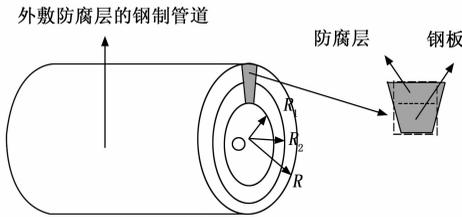


图 6 管道防腐层双层结构微体元分解图

工程中，3PE 防腐层因其电绝缘、防水性好、耐腐蚀等优点已成为管道防腐涂层的首选。由于环氧树脂的物理声学参数与 3PE 防腐层相近，因此本实验平台采用钢板与环氧树脂板结合形成双层板状结构来模拟管道防腐层双层结构进行损伤分析，其中含裂纹缺陷的钢板—环氧树脂板模型在仿真计算中的参数如表 1 所示。

表 1 钢板—环氧树脂板双层结构参数

模型结构参数	钢板	环氧树脂板
弹性模量/Pa	2.06×10^{11}	1×10^9
泊松比	0.3	0.38
密度/(kg/m ³)	7 800	980
纵波波速/(km/s)	5.95	2.73
横波波速/(km/s)	3.26	1.46

利用 COMSOL 仿真软件构建含裂纹缺陷的钢板—环氧

树脂板双层结构仿真模型，其中钢板的长×宽×高为 0.4 m×0.3 m×0.006 m，环氧树脂板的长×宽×高为 0.4 m×0.3 m×0.004 m。

利用 COMSOL 仿真软件加载幅值为 0.1 mm、频率 2.5 MHz 的正弦交变力作为钢板—环氧树脂板结构的激励源，超声波激励信号如公式 (26) 所示：

$$z(t) = \begin{cases} \sin(2\pi ft)\sin(2\pi ft/10), & t < 5f \\ 0, & t > 5f \end{cases} \quad (26)$$

通过 COMSOL 仿真依次均匀改变裂纹的长度，长度范围为 4 cm 到 40 cm，间隔 0.8 cm，对钢板—环氧树脂板双层结构进行动态仿真，通过超声导波进行缺陷检测获取损伤状态下的固有频率值，其动态仿真结构图如图 7 所示。

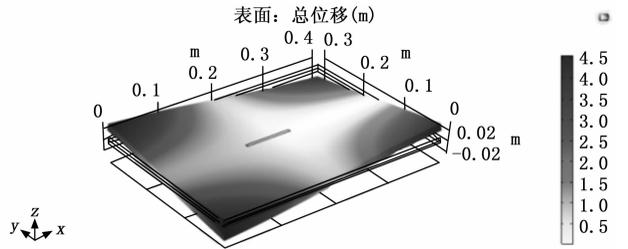


图 7 含裂纹缺陷的钢板—环氧树脂板动态仿真结构图

由于六维裂纹损伤特征向量的可视化在二维平面中受限，因此以裂纹损伤数据集中的第三阶固有频率值为例进行分析，得到含不同损伤长度的裂纹的钢板—环氧树脂板双层结构的固有频率值的变化如图 8 所示。

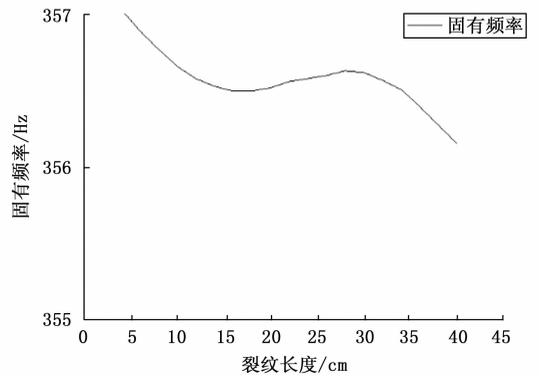


图 8 不同裂纹长度的管道—防腐层结构固有频率值

由图 8 可知，在钢板—环氧树脂板双层结构中，结构的固有频率值随着裂纹长度的增大呈现出逐渐减少的趋势，表明结构的固有频率值与裂纹损伤长度之间存在相关性，为后续使用 XGBoost 回归算法实现管道防腐层裂纹损伤长度判定提供了理论基础。

4 含裂纹的管道防腐层现场实验

含有裂纹缺陷的管道—防腐层双层板状结构设计为为双层板上层为钢板，下层为环氧树脂板，钢板的长×宽×高为 0.4 m×0.3 m×0.006 m，环氧树脂板的长×宽×高为 0.4 m×0.3 m×0.004 m。在钢板上放置发送和接收两个探

头, 现场实验中选取频率为 2.5 MHz, $K=2$ 的斜探头激发超声导波, 其中斜探头的指标如表 2 所示。

表 2 超声导波斜探头指标

斜探头指标	
频率/MHz	2.5
K 值	2
回波频率误差	$\leq 10\%$
空载始波占宽/mm	≤ 18
分辨率/dB	≥ 33
前沿距离/mm	13
灵敏度余量/dB	≥ 66

利用超声导波接收探头采集含裂纹缺陷的超声回波信号并使用示波器显示信号波形, 分别对采集到的裂纹回波信号进行分析与处理, 能够获取代表裂纹损伤特征的信息, 其中不同损伤长度的裂纹具有不同的损伤特征。

实验中保持环氧树脂板中裂纹的宽度为 4 mm, 高度为 3 mm, 通过依次均匀改变环氧树脂板上裂纹的长度, 分别采集裂纹长度从 4 cm 到 40 cm (间隔 0.8 cm) 的超声回波信号, 本文以裂纹长度为 8 cm 为例来分析, 接收到的原始裂纹损伤信号对其进行 EEMD 时频分析, 得到的结果图如图 9 所示。

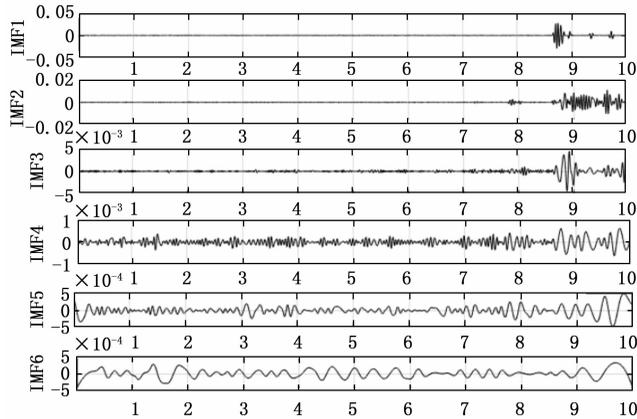


图 9 原始裂纹损伤信号 EEMD6 阶分解

在管道-防腐层双层结构裂纹损伤识别中, 首先分别对接收到的原始裂纹损伤信号进行时频分析, 然后通过快速傅里叶变换分别对其进行频域的损伤特征提取, 得到与损伤相对应的频域特征值。本文以裂纹长度为 8 cm 为例, 对接收到的回波信号分别进行 EEMD 时频分析并将前 6 个 IMF 分量进行 FFT 变换, 提取裂纹损伤信号频域特征量, 如图 10 所示为管道防腐层双层结构裂纹损伤前后的信号频域特征量的对比。

由图 10 可知, 当管道防腐层结构出现裂纹缺陷时, 结构的频域特征值减小, 验证了管道防腐层裂纹损伤动态仿真结论, 为后续使用 XGBoost 算法对裂纹进行长度识别提供了支持。

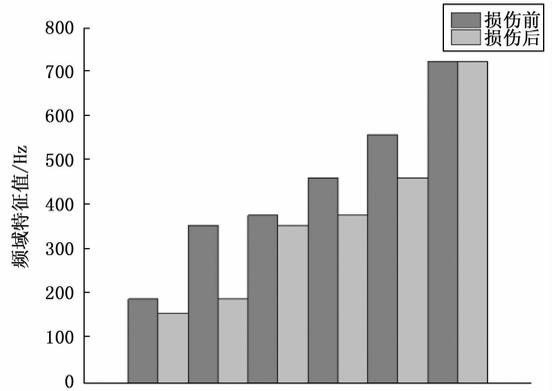


图 10 管道防腐层结构裂纹损伤前后的频域特征值

5 裂纹损伤长度识别算法对比

针对管道防腐层裂纹损伤长度识别, 采用裂纹损伤长度为 4~40 cm (间隔 0.8 cm) 的 46 组样本, 输入的频域特征值采用经过 IForest 异常检测、MF 数据补全、KPCA 降维优化后的损伤数据。

如图 11 和图 12 标准 XGBoost 分类算法出现相对较多的测试误差, 对管道防腐层进行损伤类型识别的精度相对较低, 而应用 IForest+MF+KPCA+XGBoost 分类算法的识别精度达到 90.9%, 相比只采用标准 XGBoost 分类算法的精度提高了 12.2%。其中, 横坐标为样本数量, 单位是 (个), 纵坐标 1, 2, 3 分别表示裂纹, 孔洞, 剥离三种损伤类型。

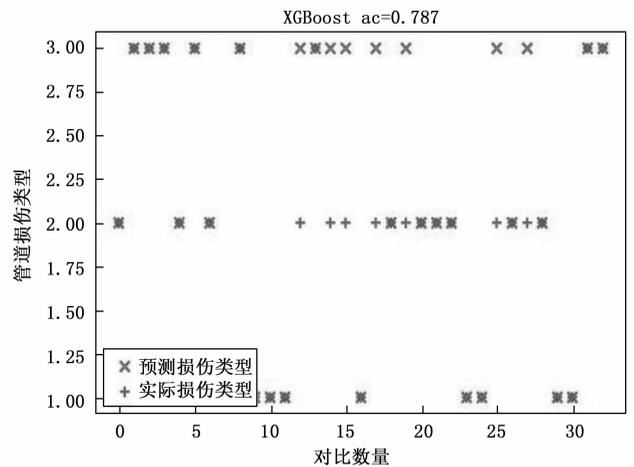


图 11 基于标准 XGBoost 分类算法的结构损伤类型识别结果

本文分别采用 XGBoost、GBDT、Random Forest 三种算法对管道防腐层裂纹损伤长度进行判定, 从上述构建的 46 组样本中选取 36 组样本作为训练样本, 剩余的 10 组样本作为测试样本, 选取均方误差和决定系数为性能度量。裂纹长度识别结果的均方误差 (MSE) 和决定系数分别如图 13 和图 14 所示。

由图 13 和图 14 可知, XGBoost 回归算法的均方误差最小且决定系数最大, 表明在管道防腐层裂纹长度识别中,

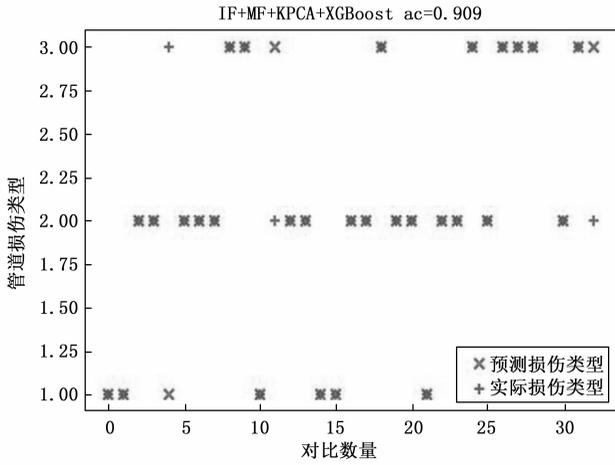


图 12 基于 IForest+MF+KPCA+XGBoost 分类算法的结构损伤类型识别结果

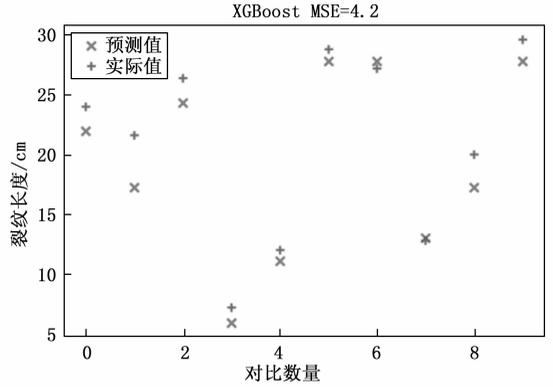


图 15 裂纹损伤长度识别结果

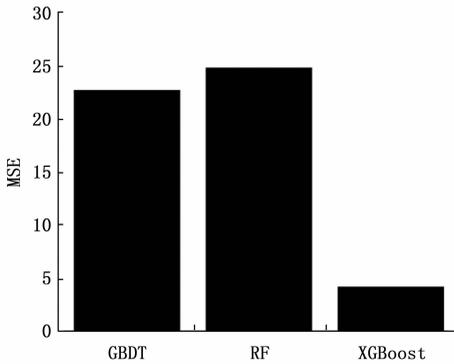


图 13 MSE 对比结果图

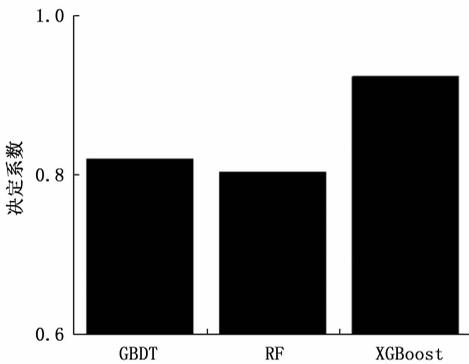


图 14 决定系数对比结果图

米表明使用 XGBoost 算法在管道防腐层裂纹长度识别方面表现良好。

表 3 裂纹损伤长度识别输出结果

裂纹长度/cm	识别输出/cm	相对误差/cm
21.6	17.234	4.366
26.4	24.319	2.081
7.2	5.902	1.298
12	11.1	0.9
28.8	27.792	1.008
27.2	27.792	0.592
12.8	12.997	0.197
20	17.234	2.766
29.6	27.792	1.808
24	21.961	2.039

6 结束语

本文引入集合经验模态分解、孤立森林、缺失森林、核主成分分析等方法，提出一种基于 XGBoost 的管道防腐层裂纹损伤长度识别方法。采用 IForest+MF+KPCA+XGBoost 分类算法的识别精度达到 90.9%，相比只采用标准 XGBoost 分类算法的精度提高了 12.2%，选取裂纹损伤信号的五阶固有频率作为管道防腐层裂纹损伤长度识别的特征向量，分别采用 XGBoost、GBDT、RF 三种回归算法对管道防腐层裂纹损伤长度进行判定。实验结果表明：应用 XGBoost 回归算法对裂纹损伤长度进行识别时，均方误差最小，决定系数最大，回归最大相对误差在 4.37 厘米以内，为管道防腐层结构裂纹智能化损伤检测提供了有效的识别方法。

参考文献：

[1] 高旭东, 邵永波, 谢丽媛, 等. X56 海底管道在腐蚀环境下疲劳裂纹扩展过程预测 [J]. 材料导报: 2020, 34 (2): 2123 - 2130.

[2] 杨理践, 吕瑞宏, 高松巍, 等. 有界双层结构 SH 导波频散特性分析 [J]. 沈阳工业大学学报: 2016, 38 (3): 286 - 292.

(下转第 27 页)

XGBoost 回归算法的效果最优，泛化能力最强。

XGBoost 回归算法对裂纹损伤长度识别的结果图如图 15 所示，横坐标为对比数量，单位是（个），在使用 XGBoost 回归算法对裂纹损伤长度进行判定时，设定 max_depth 为 5，初始学习率为 0.01，回归树的个数为 500 个。

XGBoost 回归算法对管道防腐层的裂纹进行损伤长度识别输出的具体结果如表 3 所示。

根据表 3 可知，裂纹损伤长度的识别输出值与测试结果的最大相对误差在 4.37 厘米以内，平均误差为 1.706 厘