

DOI: 10.13382/j.jemi.B2407656

基于随机森林的管道漏磁缺陷检测与量化

石晴¹ 张国山¹ 刚蓓² 李志华² 胡家铨² 刘思娇²

(1. 天津大学电气自动化与信息工程学院 天津 300072; 2. 北京华航海鹰新技术开发有限责任公司 北京 100013)

摘要: 油气管道的缺陷尺寸量化是管道检测的关键问题和最终目标。传统的缺陷检测方法往往停留在缺陷分类的阶段, 数据处理不具体给后续结果分析增加了难度; 智能识别方法又对漏磁数据的质量有更高要求。因此提出一种粒子群优化算法结合随机森林(PSO-RF), 实现管道缺陷长、宽、深的自动量化。首先对一组缺陷漏磁数据进行多维度的特征提取, 然后利用随机森林算法进行回归预测; 针对随机森林算法最佳参数不宜获得的难点, 使用粒子群优化算法进行超参数调优, 最终获得比较准确的缺陷长、宽、深预测数据。将 PSO-RF 算法与经典的卷积神经网络和粒子群算法-支持向量回归(PSO-SVR) 训练算法进行对比, 对长、宽、深的量化精度分别提高了 28%、32%、68%, 验证了 PSO-RF 算法的有效性与优越性。最后使用一组带标签的管道缺陷数据对算法进行验证, 长、宽、深量化误差在 20% 以内的数据分别达到 80.3%、88.5% 和 95.9%。

关键词: 漏磁检测; 特征提取; 随机森林; 粒子群优化; 缺陷量化

中图分类号: TN081 **文献标识码:** A **国家标准学科分类代码:** 510.4010

Detection and quantification of pipeline magnetic flux leakage defects based on random forest

Shi Qing¹ Zhang Guoshan¹ Gang Bei² Li Zhihua² Hu Jiacheng² Liu Sijiao²

(1. School of Electrical Automation and Information Engineering, Tianjin University, Tianjin 300072, China;

2. Beijing Huahang Haiying New Technology Development Co., Ltd., Beijing 100013, China)

Abstract: The quantification of defect size in oil and gas pipelines is a key issue and ultimate goal of pipeline inspection. Traditional defect detection methods often remain in the stage of defect classification, and the lack of detailed data increases the difficulty of subsequent processing; Intelligent recognition methods have higher requirements for the quality of magnetic leakage data however. Therefore, a combining particle swarm optimization and random forest (PSO-RF) is proposed to quantify the length, width, and depth of pipeline defects. Firstly, multi-dimensional feature extraction is performed on a set of defect magnetic leakage data, and then the random forest algorithm is used for regression prediction; In view of the difficulty of obtaining the best parameters of random forest algorithm, particle swarm optimization algorithm is used to optimize the hyperparameters, and finally more accurate prediction data of defect length, width and depth are obtained. The PSO-RF algorithm was compared with classical CNN and PSO-SVR training algorithms. The quantization accuracy of length, width and depth was improved by 28%, 32% and 68% respectively, verifying the effectiveness and superiority of the PSO-RF algorithm. Finally, a set of labeled pipeline defect data was used to validate the algorithm, and the data with quantization errors of length, width and depth within 20% achieved 80.3%, 88.5% and 95.9% respectively.

Keywords: magnetic flux leakage detection; feature extraction; random forest; particle swarm optimization; defect quantification

0 引言

管道作为运输物料的一种经济高效的媒介, 被广泛用于输送水、石油和天然气等人民生活所需物资^[1]。

在长期服役过程中, 由于恶劣的外部工作环境和内部传输介质的共同作用^[2], 管道壁容易产生变形和腐蚀, 可能导致油气泄漏, 引发安全事故。因此, 需要对油气管道进行定期检测, 确保管道安全运行。然而, 传统的检测方法往往依赖于人力或简单的管道压力监测, 灵敏性较差, 也

无法准确定位泄漏位置和大小,因此无损检测凭借其对于管道的非破坏性以及缺陷检测的高效性和安全性,在检测管道泄漏问题时得到广泛应用。在各种无损检测技术中,漏磁检测适用于各种金属材料,因而在油气行业迅速发展并趋于成熟。

漏磁缺陷检测主要包括两方面的问题,1)缺陷识别和定位;2)缺陷的量化问题。

缺陷识别方法主要包括特征提取和分类器训练。如刘金海等^[3]提出一种基于漏磁内检测和随机森林的缺陷识别方法,分别对小缺陷和大缺陷进行分类,相比常规分类方法取得更高的准确率;杨理践等^[4]提出一种基于深度学习的智能识别方法,跳过人工选择特征的过程,利用漏磁检测图像进行缺陷的分类和定位,获得了较高的识别精度;Chen 等^[5]建立了管道机械损伤凿槽的瞬态力学模拟模型,实现对机械损伤中凿痕的检测和识别;赵春华等^[6]引入特征融合网络增强模型改进了 YOLOv7-tiny 网络,降低钢材表面缺陷模型尺寸的同时,提高了检测的准确率和平均精度均值。

缺陷量化是指通过漏磁信号数据定量确定缺陷轮廓的大小。Ravan 等^[7]使用 Canny 边缘检测算法对二维测量信号的彩色图进行处理,构造出缺陷开口形状的粗模型,并基于该模型对缺陷深度进行预测;Xu 等^[8]提出了一种基于深度学习中 YOLOv5 和卷积神经网络(CNN)模型的目标检测框架,实现了漏磁伪彩图的目标分类和漏磁深度的同步识别;秦浩东等^[9]建立小管径弯头不同位置缺陷内检测模型,并结合改进 Canny 算子的缺陷量化方法提高了缺陷量化精度;Sun 等^[10]提出一种基于深度学习的双馈交叉残差网络(dual residual networks, DRN)自动提取缺陷的深层特征,实现了对缺陷长度、宽度和深度的同时高精度量化。

漏磁缺陷检测技术已在实际广泛应用,但是纵览前人工作,可以发现两个问题。1)传统算法和人工智能算法各有优劣,很难抉择。传统算法首先需要根据经验对原始数据进行特征提取,因而可解释性强,但往往停留在缺陷分类的阶段,在缺陷量化领域的应用不足。基于机器学习的智能识别算法自动对漏磁数据集进行特征提取,可信度高,但深度网络需要大量的缺陷数据优化网络参数^[11],并且学习过程未知,量化结果不易分析;此外,其处理对象一般是图片,需要对漏磁数据进行二次处理,增加了工作量,也对检测环境(如光线等外部干扰)提出更高要求^[12]。2)缺陷的量化没有统一的标准,也较难获得实际工程数据,大部分文献通过仿真或实验室搭建的管道进行试验,获得的缺陷数据比较单一,数据集也较小,具体的实用性很难界定。

考虑到实际工程的应用要求和条件,本文决定应用传统算法解决缺陷量化问题,兼顾量化结果的可解释性

和原始数据的完整度。采用一组三轴高精内检测器在环路试验中检测到的缺陷数据为样本集(610个),对缺陷的长、宽、深进行预测。

1 漏磁检测实验

1.1 漏磁缺陷检测原理

内检测器在管道内运动时,利用相连的钢刷对管道内壁进行磁化,并通过磁传感器检测管壁表面的漏磁信号,磁场通过磁体、磁轭和管壁形成闭合回路。缺陷附近的壁厚变化或其他异常现象,会造成磁力线畸变,形成漏磁^[13]。漏磁缺陷检测原理如图 1 所示。

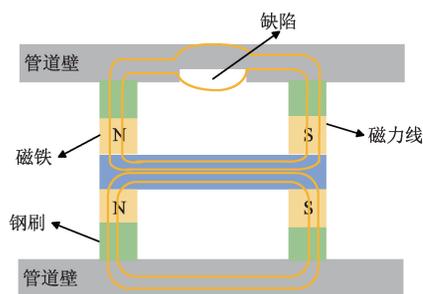


图 1 漏磁缺陷检测原理

Fig. 1 Schematic diagram of magnetic leakage defect detection

漏磁检测装置主要由检测元件与磁化装置两部分组成^[14],检测元件包括轴向、径向和周向传感器,磁化装置由分布在检测器表面的钢刷阵列组成。机器在管道内匀速移动,对某条管线进行漏磁缺陷检测的现场如图 2 所示。



图 2 某管道漏磁检测现场

Fig. 2 Magnetic flux leakage testing site for a pipeline

1.2 漏磁曲线特点

利用磁感应传感器获取漏磁场信号,然后送入计算机进行信号处理,对漏磁场磁通密度分量进行分析能进一步了解相应缺陷特征。可视化后的漏磁曲线横坐标轴是里程(也就是采样点),纵坐标轴表示不同通道测得的当前位置的磁场信号。理想状态下的管道磁化后,磁感线被约束在管道内部,几乎没有磁场泄漏,因而漏磁曲线是一条几乎没有起伏的直线;如果管道内部有缺陷,其

磁导率很小,磁阻很大,磁路中的磁通发生畸变,在缺陷处形成漏磁场,漏磁曲线会出现明显起伏,如图3所示。

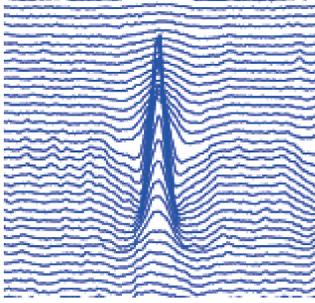


图3 可视化后的缺陷漏磁信号

Fig. 3 Visualized defect magnetic leakage signal

1.3 特征提取

管道缺陷处的漏磁信号是由多个通道的传感器检测到的波形变化组合而成的,不同尺寸大小的缺陷影响的传感器通道数不同^[12],但是漏磁变化情况和缺陷尺寸之间并没有精确的模型,不同的特征选择和提取方法对识别结果有很大影响,因此选择合适的特征对漏磁曲线进行特征提取是缺陷尺寸量化的关键所在。目前常见的特征提取种类主要有以下几方面:信号峰谷值、平均信号强度、信号上升速率、谷宽等,这些特征可以大致涵盖缺陷水平,但实际检测中缺陷形状多样复杂,依旧缺少一些细节信息。某一缺陷的轴向信号如图4所示,该信号由两个较小负峰以及一个较大正峰组成,除上述提到的特征外,缺陷起伏的坡度、缺陷中心采样点周围的拐点间距和特殊点间距等也会对缺陷的量化结果有一定影响。

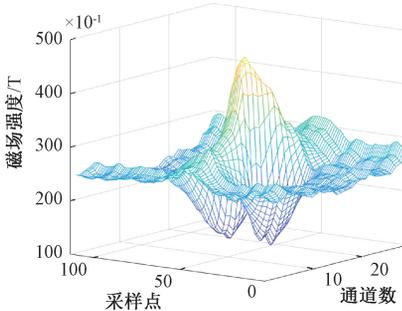


图4 缺陷轴向信号

Fig. 4 Defect axial signal

对漏磁缺陷信号进行分析,发现与周向和径向数据相比,轴向数据的变化更明显、更准确,因此在三轴信号中,着重对缺陷的轴向数据特征进行提取。经过对缺陷信号曲线的具体分析,最终提取信号曲线缺陷体积和表面积、谷谷间距、峰峰间距、面能量等31个特征,如表1所示。

表1 特征提取名称及含义

Table 1 Name and meaning of feature extraction

特征序号	特征名称	解释
1	轴向缺陷体积	整个缺陷区域的体积
2	轴向缺陷表面积	整个缺陷区域的表面积
3	轴向谷谷间距	轴向上左谷与右谷的距离
4	轴向谷谷均值	轴向上左谷与右谷值的平均
5	轴向峰峰间距	轴向上左峰与右峰的距离
6	轴向拐点间距	轴向上左右谷之间的左右拐点距离
7	轴向特殊点间距	轴向上左右谷之间的左右特殊点距离
8	轴向左峰谷值	轴向上左峰到左谷的值的差
9	轴向右峰谷值	轴向上右峰到右谷的值的差
10	轴向面积1	轴向上,边界为左右谷的面积
11	轴向面积2	轴向上,边界为特殊点的面积
12	轴向面积3	轴向上,边界为拐点的面积
13	轴向面能量1	轴向上,边界为左右谷的面能量
14	轴向面能量2	轴向上,边界为特殊点的面能量
15	轴向面能量3	轴向上,边界为拐点的面能量
16	周向截面面积1	在周向截面中的面积
17	周向截面面积2	在周向截面中边界为特殊点的面积
18	周向截面面积3	在周向截面中边界为拐点的面积
19	周向截面面能量1	在周向截面中的面能量
20	周向截面面能量2	在周向截面中边界为特殊点的面能量
21	周向截面面能量3	在周向截面中边界为拐点的面能量
22	周向特殊点距离	在周向中特殊点的距离
23	周向拐点距离	在周向中拐点的距离
24	径向截面面积1	在径向截面中的面积
25	径向截面面积2	在径向截面中边界为特殊点的面积
26	径向截面面积3	在径向截面中边界为拐点的面积
27	径向截面面能量1	在径向截面中的面能量
28	径向截面面能量2	在径向截面中边界为特殊点的面能量
29	径向截面面能量3	在径向截面中边界为拐点的面能量
30	径向特殊点距离	在径向中特殊点的距离
31	径向拐点距离	在径向中拐点的距离

上述31个特征分别对轴向、周向、径向漏磁检测数据进行特征提取,其中按照式(1)和(2)计算特殊点。

$$Y_{left} = Y_{vleft} + \alpha D_{vleft} \quad (1)$$

$$Y_{right} = Y_{vright} + \alpha D_{vright} \quad (2)$$

式中: α 为特殊点系数; D_{vleft} 、 D_{vright} 为双边峰谷差值; Y_{vleft} 、 Y_{vright} 为双边谷值; Y_{left} 、 Y_{right} 为双边特殊点值。

2 基于粒子群优化 PSO 的随机森林 RF 算法

2.1 随机森林算法

随机森林是一种基于决策树的集成学习技术^[15],训练过程中每棵树随机选择训练集和训练特征进行训练,

以避免过拟合问题^[16],在复杂问题预测研究中具备较强的适用性,算法具体步骤如下^[17-18]。

步骤 1) 在总样本集 N 中采取有放回随机抽样, 获得训练集 $1, 2, \dots, m$, 对每个训练集, 生成一棵决策树, 共 m 棵; m 的设置与训练集的规模和模型的复杂度有关, 一般来说, m 越大, 模型越精确, 但过大的 m 也会造成资源浪费和训练时间过长, 综合考虑本文将 m 的初始值设置为 50, RF 其他参数的详细设置在 3.2 节具体说明。

步骤 2) 在步骤 1) 中, 有 $1/3$ 的数据未被选中, 这一部分称为袋外数据, 将其作为测试集。对于测试集中的每个样本, 每棵回归树都给出一个预测结果。

步骤 3) 取所有树对某样本预测结果的平均值作为该样本的最终预测结果。

算法的训练流程如图 5 所示。

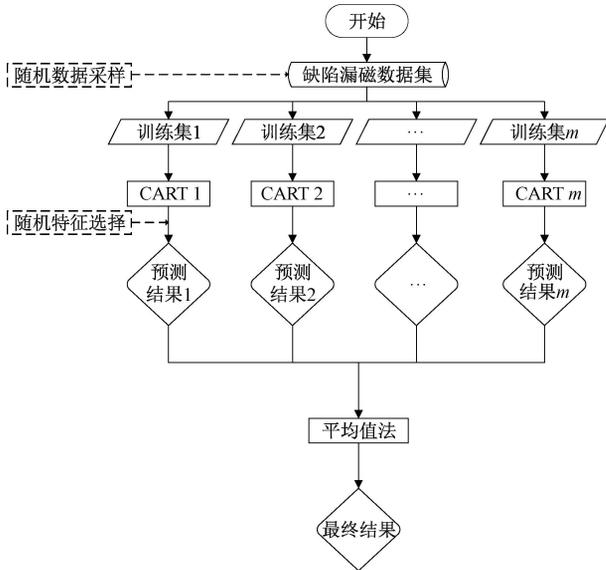


图 5 随机森林运算流程

Fig. 5 Random forest operation process

2.2 粒子群优化算法

粒子群优化算法是一种可靠的机器学习模型调整方法^[19], 通过初始化一群随机粒子, 反复迭代计算找到全局最优解^[20-21]。

步骤 1) 粒子群初始化, 设定粒子种群大小、适应值函数及搜索空间等参数, 本文设置粒子种群大小为 20。

步骤 2) 计算每个粒子的适应度。

步骤 3) 更新每个粒子的速度和位置, 对于第 i 个粒子, 分别计算其个体和全局最优解的影响, 然后按照式(3)和(4)更新该粒子的速度和位置。

$$v_{id}^k = wv_{id}^{k-1} + c_1r_1(P_{best} - x_{id}^{k-1}) + c_2r_2(G_{best} - x_{id}^{k-1}) \quad (3)$$

$$x_{id}^k = x_{id}^{k-1} + v_{id}^{k-1} \quad (4)$$

式中: w 为非负的惯性权重; v_{id}^{k-1} 和 x_{id}^{k-1} 分别为粒子 i 经历第 k 次迭代后在第 d 维搜索空间的速度和位置; P_{best} 和 G_{best} 为粒子在当下时刻的个体最优解和全局最优解; c_1, c_2 为非负的加速度常数; r_1, r_2 为 $(0, 1)$ 的随机数。

步骤 4) 判断终止条件, 若达到最大迭代次数或满足终止条件, 则算法结束; 否则返回步骤 2)。

步骤 5) 输出最优解。

算法流程如图 6 所示。

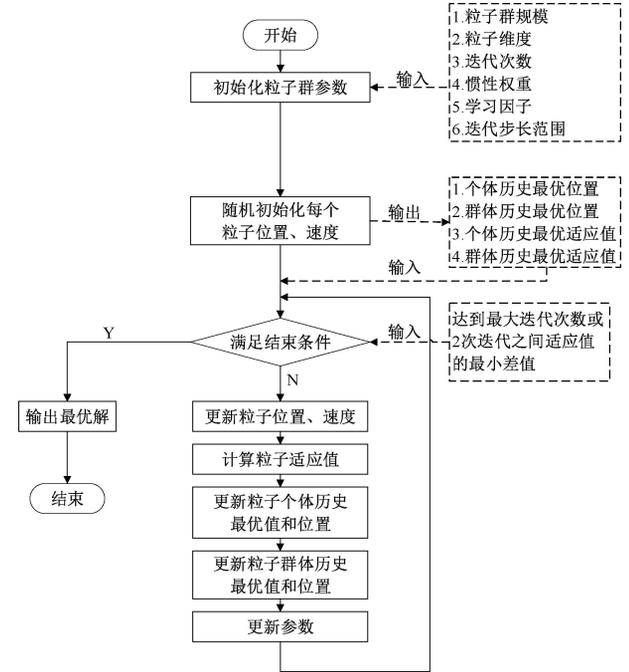


图 6 粒子群优化算法流程

Fig. 6 Particle swarm optimization algorithm flowchart

3 结果分析

3.1 PSO-RF 算法设计

RF 参数较多, 包括树的数量和最大深度、叶子数量、节点可分的最小样本数、每次分裂考虑的最大特征数等, 但研究表明, 树的数量和最大深度、叶子数量对模型训练的效果有较大影响, 因此使用 PSO 算法对上述参数组合进行寻优。

步骤 1) 输入数据集和参数, 设置粒子群规模为 20, 粒子维度为 3, 最大迭代次数为 50 次。

步骤 2) 设计适应值函数, 根据当前粒子群的解构造 RF 模型, 取袋外误差均值作为适应度函数。

步骤 3) 根据 PSO 最终寻优结果构建 RF 模型。

步骤 4) 对数据集进行筛选, 剔除样本集中的异常点。首先使用步骤 3) 中构建的模型进行第 1 次训练, 并使用训练好的模型对样本集进行预测, 按照式(5)选择

与真实值误差在 10% 之内的样本称为正确集, 作为下一次训练的样本集, 直到达到最大迭代次数或连续两次预测的结果相同(即正确集的样本数量不变), 筛选流程如图 7 所示。

$$error = \frac{|y_{pred} - y_{true}|}{y_{true}} \times 100\% \quad (5)$$

式中: y_{pred} 和 y_{true} 分别是样本集的预测值和真实值; $error$ 为预测值的相对误差。

步骤 5) 利用袋外数据集进行验证, 并以预测值与实际值的均方误差 (MSE) [22] 为指标, 评价模型的优劣, 如式 (6) 所示。

$$MSE = \left(\frac{1}{n}\right) \times \sum (y_{true} - y_{pred})^2 \quad (6)$$

步骤 6) 输出特征重要性排序并进行特征筛选和重组。

步骤 7) 使用新的特征进行训练。

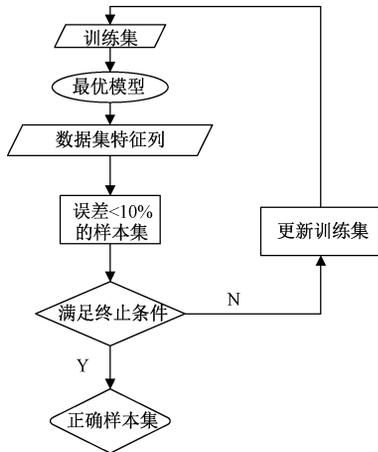


图 7 训练集筛选流程

Fig. 7 Training set selection process

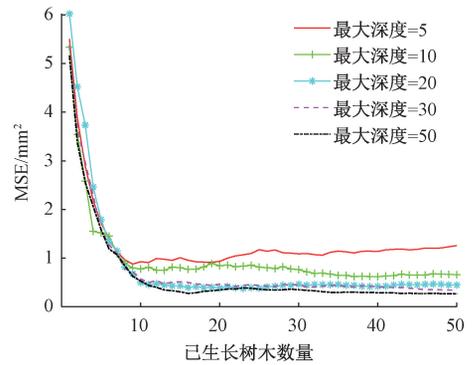
3.2 训练结果

为了节约运算时间, 需要预先确定 PSO 算法进行参数寻优的大致范围。以缺陷深度为例, 分别改变树木最大深度和叶子数量, 训练结果如图 8(a) 和 (b) 所示。

由图 8 可知, 其他参数一定时, 均方误差随着树木数量的增加而减小, 树木数量超过 30 后, 均方误差的变化趋于平缓; 树木数量一定时, 均方误差随树木最大深度的增大而减小, 随叶子数量的减小而减小。总体来说, 树木越多, 树木深度越大, 树叶越少, 最终预测的误差越小。但这个结论并不是完全正确的, 图 8(a) 中, 树木数量为 10 时, 树木最大深度为 20 时, 均方误差最小; 图 8(b) 中, 树木数量为 10 时, 树叶数量为 10 的时候均方误差最小, 树叶数量为 20 的时候次之; 树木最大深度和叶子数量一定时, 随树木数量的增加, 均方误差并非严格单调, 而是在振荡中逐渐减小。因此, RF 的不同参数之间有一定的

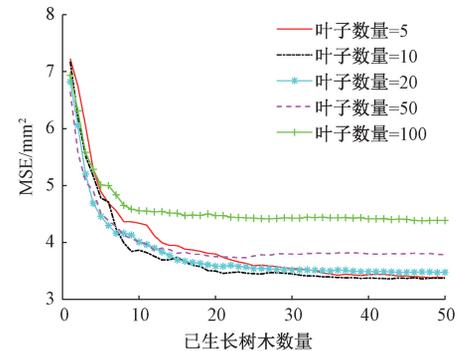
相关性, 最优参数不易选择。基于此问题, 使用 PSO 算法对 RF 进行超参数优化是有必要的。

此外, 随着树木数量的增加, 均方误差的大小也趋于稳定。图 8(a) 中, 树木数量为 50 时, 最大深度由 5 变为 10 的时候均方误差减小得最明显, 而由 20 增长到 50 的过程中, 虽然最大深度的变化更大了, 但均方误差的减小幅度却降低了。由此可见, 最终 RF 的规模不必过分大, 否则计算量会迅速增加而准确度变化不大。经过综合考虑, 将 PSO 寻优时树木数量、叶子数量和最大深度的上限分别定为 50、20 和 20, 最终对缺陷长、宽、深进行预测的最佳参数如表 2 所示。



(a) 改变树木最大深度对缺陷深度预测结果

(a) Prediction of defect depth by changing the Max depth



(b) 改变叶子数量对缺陷深度预测结果

(b) Prediction of defect depth by changing Min leaf size

图 8 不同树木深度和叶子数量对缺陷深度预测结果的影响

Fig. 8 Influence of different tree depth and leaf number on defect depth prediction results

表 2 长、宽、深预测最优参数

Table 2 Optimal parameters for predicting length, width, and depth

PSO 寻优	最佳树木	最佳树叶	最佳深度
长	50	2	20
宽	45	5	12
深	46	9	16

在 RF 模型训练过程中, 不同特征对预测结果的贡献度和重要性不同, 重要性越大的特征量与预测结果的

相关性越大,使用强相关特征训练模型,在一定程度上可以提高预测精度^[23]。第 j 个特征贡献度的计算方法如式(7)所示。

$$MDA = \frac{1}{m} \sum_t \left[\frac{1}{|D_t|} \left(\sum_{X_i^j \in D_t^j} (R_k(X_i^j) - y_i^k)^2 - \sum_{X_i \in D_t} \sum_k (R_k(X_i) - y_i^k)^2 \right) \right] \quad (7)$$

式中: m 为 RF 中树的数量; (X_i, y_i) 为样本; X_i^j 为 X_i 的第 j 个特征随机交换后的样本; D_t 为随机树 t 的袋外样本集; D_t^j 为第 j 维特征交换后形成的样本集; $R_k(X_i)$ 为样本 X_i 的预测输出。

使用表 2 中的最优参数进行训练,按照特征对训练的贡献进行重要性排序结果如图 9(a)~(c)所示。

本文对漏磁数据总共提取缺陷区域体积、表面积、面能量等 31 个特征,选取贡献度前 5 名的特征进行训练。在对缺陷长度进行预测时,特征 3、7、4、6、11 的影响较大;在对缺陷宽度进行预测时,特征 4、31、23、26、11 的影响较大;在对缺陷深度进行预测时,特征 23、5、8、31、9 的影响较大。

通过图 9 发现,在对缺陷的尺寸进行量化时,并不是所有的特征都与之相关。如特征 5,在进行长度和深度预测时比较重要,在进行宽度预测时贡献度则较低,因此对于宽度预测来说,该特征属于需要剔除的无效量。筛选特征贡献度前 5 名进行训练,特征筛选前后的训练结果如图 10(a)~(c)所示。

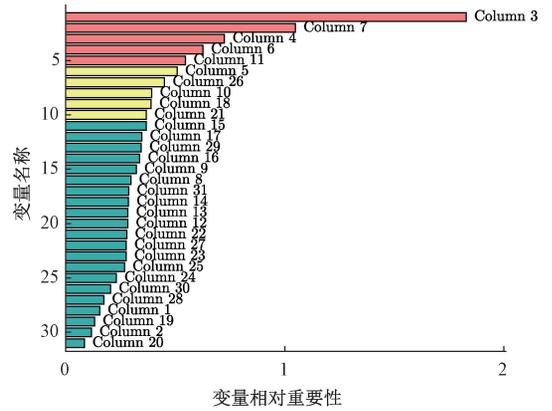
特征筛选后的预测结果相比筛选前普遍获得了更小的均方误差。由图 10(a)~(c)所示,特征筛选后,模型对长、宽、深预测结果的均方误差分别缩小了 7%、4%、5.5%左右,由此证明了特征筛选的必要性。

3.3 性能优化验证

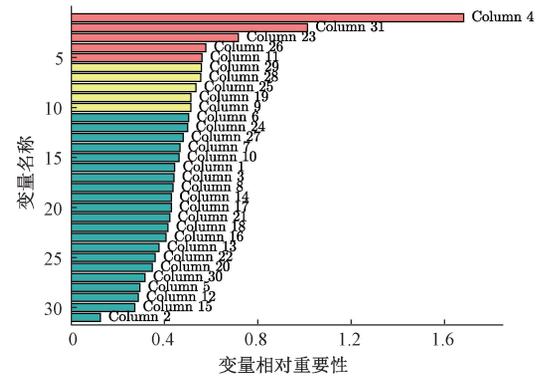
为了验证 PSO-RF 算法的有效性,将其分别与基于 PSO 的支持向量回归(SVR)算法和卷积神经网络(CNN)进行对比。选取原始数据集的前 480 个样本作为训练集,后 130 个样本作为验证集,最终对比结果如表 3 和 4 所示。

由表 3 所示,相比 CNN 算法,PSO-RF 算法在对长、宽、深进行预测时,均方误差分别缩小了 28%、32%、68%;而在与 PSO-SVR 对比时,长度和深度的均方误差分别降低 12% 和 55%左右,宽度的均方误差增大了 9%左右。虽然在与 PSO-SVR 的对比中宽度预测表现差一些,但实际工程检测对缺陷深度的关注度远高于长度和宽度,此外 PSO-RF 相比 PSO-SVR 结构更简单,综合考虑下,PSO-RF 的训练效果优于 PSO-SVR。

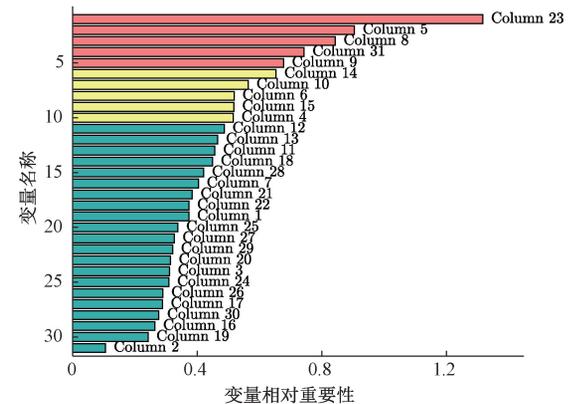
由表 4 所示,PSO-RF 算法在 3 种尺寸预测中运行时间均远远小于 CNN 算法,说明利用 PSO-RF 算法进行缺陷量化的准确度和效率都得到了保证。



(a) 长度预测特征重要性排序
(a) Importance ranking of length prediction features



(b) 宽度预测特征重要性排序
(b) Importance ranking of width prediction features



(c) 深度预测特征重要性排序
(c) Importance ranking of depth prediction features

图 9 缺陷量化特征重要性排序

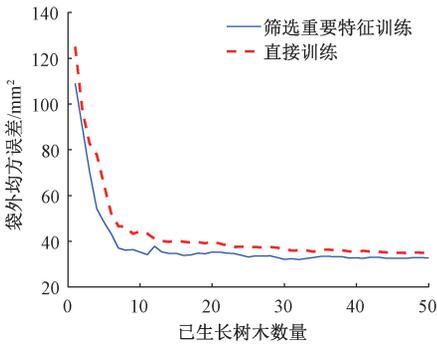
Fig. 9 Importance ranking of defect quantification features

表 3 PSO-RF 与 PSO-SVR、CNN 均方误差对比

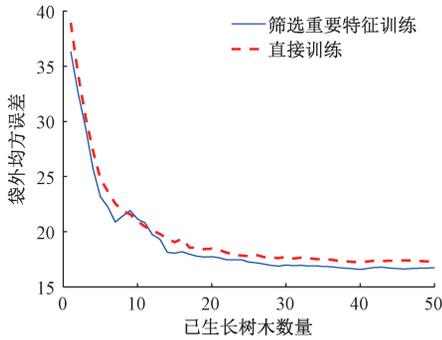
Table 3 Comparison of MSE between

PSO-RF, PSO-SVR and CNN (mm²)

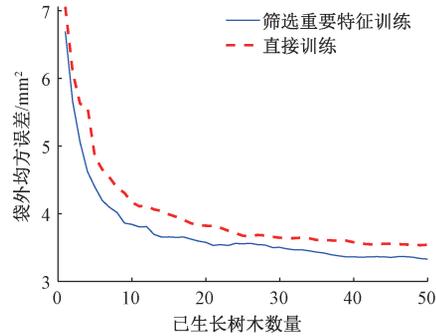
种类	PSO-RF	PSO-SVR	CNN
长	8.841 62	9.775 0	12.258 8
宽	12.749 80	11.750 9	18.680 1
深	0.722 56	1.799 4	2.286 6



(a) 特征筛选前后长度预测结果
(a) Length prediction before and after feature screening



(b) 特征筛选前后宽度预测结果
(b) Width prediction before and after feature screening



(c) 特征筛选前后深度预测结果
(c) Depth prediction before and after feature screening

图 10 特征筛选前后缺陷量化结果

Fig. 10 Defect quantization results before and after feature screening

表 4 PSO-RF 与 CNN 运算时间对比

Table 4 Comparison of PSO-RF and CNN operation times (s)

种类	PSO-RF	CNN
长	2. 909 499	17. 833 182
宽	2. 346 373	17. 578 288
深	3. 141 174	18. 284 931

3.4 实例应用

为了验证 PSO-RF 算法在实际工程中的应用性,选取一组带标签的管道缺陷数据,包含样本 1 047 个。这

组数据同样使用 RF 模型进行训练和预测,且随机选取若干缺陷样本作为待验证点,在相对应的管道挖开进行人工测量,并与预测结果进行对比,被证明是合格的。使用本文训练的模型对该样本集进行预测,结果如表 5 所示。

表 5 PSO-RF 预测与第三方预测结果对比

Table 5 Comparison of PSO-RF prediction and third party prediction results

种类	<10%	10%~20%	20%~30%	>30%
长	212	629	132	74
宽	306	621	47	73
深	562	442	43	0

由表 5 可知,经过 PSO-RF 算法训练出的模型,在对长度尺寸的预测上,约 80.3% 的数据误差在 20% 以内;在对宽度尺寸的预测上,约 88.5% 的数据误差在 20% 以内;在对深度尺寸的预测上,约 95.9% 的数据误差在 20% 以内。在实际工程应用中,关于深度的尺寸量化是三维缺陷尺寸量化中最重要的一步,表 5 说明 PSO-RF 算法在对缺陷的三维尺寸量化尤其是对深度的预测上取得了良好效果。

4 结 论

本文对漏磁信号曲线进行分析,提取缺陷区域体积、表面积、面能量等多个特征;将 RF 算法和 PSO 算法结合,构造出一种无需人为调参的 PSO-RF 算法,并以环路实验中所获得的缺陷漏磁数据为样本,实现了缺陷的尺寸量化,与经典的 CNN 算法和 PSO-SVR 算法相比较,证明了 PSO-RF 算法的优越性,并使用一组真实缺陷数据进行预测,在缺陷量化问题上取得了良好的效果。

缺陷的尺寸量化没有统一的标准,因此使用不同的模型训练获得的预测结果往往大不相同,而不同机构使用的模型往往是机密所在,因此在不进行实地测量的情况下,不同训练模型处理相同数据获得的预测结果的质量和真实度往往有很大争论。因此后续将在未知模型的基础上,通过漏磁检测数据和缺陷尺寸量化结果进行质量评价方面的进一步研究。

参考文献

[1] 王国涛,郭天昊. 油气管道特殊缺陷的漏磁信号识别方法 [J]. 沈阳工业大学学报, 2019, 41 (4): 401-405.

WANG G T, GUO T H. Recognition method of magnetic flux leakage signal for special defect of oil and gas pipeline [J]. Journal of Shenyang University of

- Technology, 2019, 41(4): 401-405.
- [2] 焦敬品, 常予, 李光海, 等. 金属包覆层管道外壁损伤低频漏磁检测技术[J]. 北京工业大学学报, 2018, 44(12): 1471-1477.
JIAO J P, CHANG Y, LI G H, et al. Outer wall damage of cladding tube based on low frequency magnetic flux leakage technology[J]. Journal of Beijing University of Technology, 2018, 44(12): 1471-1477.
- [3] 刘金海, 付明芮, 唐建华. 基于漏磁内检测的缺陷识别方法[J]. 仪器仪表学报, 2016, 37(11): 2572-2581.
LIU J H, FU M R, TANG J H. MFL inner detection based defect recognition method[J]. Chinese Journal of Scientific Instrument, 2016, 37(11): 2572-2581.
- [4] 杨理践, 曹辉. 基于深度学习的管道焊缝法兰组件识别方法[J]. 仪器仪表学报, 2018, 39(2): 193-202.
YANG L J, CAO H. Deep learning based weld and flange identification in pipeline[J]. Chinese Journal of Scientific Instrument, 2018, 39(2): 193-202.
- [5] CHEN J ZH, KANG X W, ZHANG X W, et al. Research on unsaturated magnetization mfl detection of gouge in oil and gas pipeline[C]. Advances in Precision Instruments and Optical Engineering: Proceedings of the International Conference on Precision Instruments and Optical Engineering, 2021. Singapore: Springer Nature Singapore, 2022: 35-45.
- [6] 赵春华, 罗顺, 谭金铃, 等. 基于 PC-YOLOv7 算法钢材表面缺陷检测[J]. 国外电子测量技术, 2023, 42(9): 137-145.
ZHAO CH H, LUO SH, TAN J L, et al. Steel surface defect detection based on PC-YOLOv7 algorithm [J]. Foreign Electronic Measurement Technology, 2023, 42(9): 137-145.
- [7] RAVAN M, AMINEH R K, KOZIEL S, et al. Sizing of 3-D arbitrary defects using magnetic flux leakage measurements [J]. IEEE Transactions on Magnetics, 2010, 46(4): 1024-1033.
- [8] XU ZH CH, LIU K R, GU B, et al. Image recognition model of pipeline magnetic flux leakage detection based on deep learning[J]. Corrosion Reviews, 2023, 41(6): 689-701.
- [9] 秦浩东, 张颖, 赵鹏程. 改进 Canny 算子的小管径弯头漏磁缺陷图像量化方法[J]. 电子测量技术, 2024, 47(5): 150-157.
QIN H D, ZHANG Y, ZHAO P CH. Image quantization method of magnetic leakage defect of small diameter elbow with improved Canny operator [J]. Electronic Measurement Technique, 2024, 47(5): 150-157.
- [10] SUN H Y, PENG L SH, HUANG S L, et al. Development of a physics-informed doubly fed cross-residual deep neural network for high-precision magnetic flux leakage defect size estimation [J]. IEEE Transactions on Industrial Informatics, 2021, 18(3): 1629-1640.
- [11] 刘金海, 赵真, 付明芮, 等. 基于主动小样本学习的管道焊缝缺陷检测方法[J]. 仪器仪表学报, 2022, 43(11): 252-261.
LIU J H, ZHAO ZH, FU M R, et al. Pipeline weld defect detection method based on active small sample learning [J]. Chinese Journal of Scientific Instrument, 2022, 43(11): 252-261.
- [12] 赵治. 基于机器学习的管道缺陷识别量化方法研究[D]. 沈阳: 沈阳工业大学, 2023.
ZHAO ZH. Research on quantitative method for pipeline defect recognition based on machine learning [D]. Shenyang: Shenyang University of Technology, 2023.
- [13] HUANG S L, PENG L SH, SUN H Y, et al. Deep learning for magnetic flux leakage detection and evaluation of oil & gas pipelines: A review [J]. Energies, 2023, 16:1372.
- [14] 刘杰. 速度可控的管内漏磁检测机器人研究[D]. 北京: 中国石油大学(北京), 2023.
LIU J. Research on a speed controllable magnetic leakage detection robot inside a pipe [D]. Beijing: China University of Petroleum (Beijing), 2023.
- [15] 常梦容, 王海瑞, 肖杨. mRMR 特征筛选和随机森林的故障诊断方法研究[J]. 电子测量与仪器学报, 2022, 36(3): 175-183.
CHANG M R, WANG H R, XIAO Y. Research on fault diagnosis method based on mRMR feature screening and random forest[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(3): 175-183.
- [16] SHI Y, CUI L M, QI ZH Q, et al. Automatic road crack detection using random structured forests [J]. IEEE transactions on intelligent transportation systems, 2016, 17(12): 3434-3445.
- [17] GEORGANOS S, GRIPPA T, NIANG GADIAGA A, et al. Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling [J]. Geocarto International, 2019, 36(2): 121-136.
- [18] SPEISER J L, MILLER M E, TOOZE J, et al. A comparison of random forest variable selection methods for classification prediction modeling [J]. Expert Systems

with Application, 2020, 134: 93-101.

- [19] GONG Y J, LI J J, ZHOU Y, et al. Genetic learning particle swarm optimization [J]. IEEE Transactions on Cybernetics, 2015, 46(10): 2277-2290.
- [20] HAJIHASSANI M, JAHED ARMAGHANI D, KALATEHJARI R. Applications of particle swarm optimization in geotechnical engineering: A comprehensive review[J]. Geotechnical and Geological Engineering, 2018, 36: 705-722.
- [21] BONYADI M R, MICHALEWICZ Z. Particle swarm optimization for single objective continuous space problems: A review [J]. Evolutionary Computation, 2017, 25(1): 1-54.
- [22] HAN Q H, GUI CH Q, XU J, et al. A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm [J]. Construction and Building Materials, 2019, 226: 734-742.
- [23] 李郅琴, 杜建强, 聂斌, 等. 特征选择方法综述[J]. 计算机工程与应用, 2019, 55(24): 10-19.
- LI ZH Q, DU J Q, NIE B, et al. Summary of feature selection methods [J]. Computer Engineering and Applications, 2019, 55(24): 10-19.

作者简介



石晴 (通信作者), 2022 年于天津大学获得学士学位, 现为天津大学硕士研究生, 主要研究方向为漏磁检测和智能控制。

E-mail: shiqing@tju.edu.cn

Shi Qing (Corresponding author)

received her B. Sc. degree from Tianjin University in 2022. Now she is a M. Sc. candidate in Tianjin University. Her main research interests include magnetic leakage detection and intelligent control.



张国山, 1983 年于东北师范大学获理学学士学位, 1989 年毕业于东北工学院获理学硕士学位, 1996 年于东北大学获工学博士学位。现任天津大学教授、博士生导师, 目前研究方向为非线性系统控制和智能控制。

E-mail: zhanggs@tju.edu.cn

Zhang Guoshan received his B. Sc. degree from Northeast Normal University in 1983, M. Sc. degree from Northeast Institute of Technology in 1989, and Ph. D. degree from Northeastern University in 1996, respectively. Now he is a professor and Ph. D. supervisor at Tianjin University. His main research interests include nonlinear system control and intelligent control.