

引文: 杨阳, 李成志, 杜选, 等. 基于 KNN 和随机森林算法的腐蚀泄漏风险软检测模型[J]. 油气储运, 2024, 43(9): 1064–1072.

YANG Yang, LI Chengzhi, DU Xuan, et al. Soft detection model of corrosion leakage risk based on KNN and random forest algorithms[J]. Oil & Gas Storage and Transportation, 2024, 43(9): 1064–1072.

基于 KNN 和随机森林算法的腐蚀泄漏风险软检测模型

杨阳^{1,2} 李成志¹ 杜选¹ 于晓³ 董绍华³

1. 昆仑数智科技有限责任公司; 2. 华北电力大学能源动力与机械工程学院; 3. 中国石油大学(北京)安全与海洋工程学院

摘要:【目的】城镇燃气管网完整性管理需要有效的风险评价方法, 腐蚀泄漏风险评价需要将风险评价因子充分与各项检测业务相结合, 然而当前检测数据繁杂且缺失严重, 亟需一种可预测并评价腐蚀泄漏风险的方法。【方法】通过相关性分析, 筛选出与腐蚀泄漏风险相关的关键指标, 结合管道本体数据与周围环境数据, 采用KNN(K-Nearest Neighbor)与随机森林算法, 建立智能软检测模型。【结果】该模型能够对缺失检测数据进行预测, 实现关键指标的间接测量, 模型预测值与真实测量值的相对误差小于25%, 达到合格水平。该模型可在数据缺失情况下有效预测管道腐蚀泄漏风险, 为定量评价奠定基础。与前人研究相比, 模型在多因素耦合关系提取与算法选择上进行创新, 提高了预测的准确性和可靠性。然而, 部分异常数据表明该模型在某些条件下的预测能力有限, 且模型依赖于数据完整性和准确性。设法提高检测数据数量和质量, 优化关键风险指标特征提取方法, 可以进一步提高模型精度。【结论】丰富了燃气管道腐蚀泄漏风险预测理论, 在提高管道运行安全性与可靠性方面具有实用价值, 未来应着重改进数据采集和分析技术, 进一步优化模型结构, 提升其在不同应用场景下的适应性与准确性。(图10, 表6, 参25)

关键词: 城镇燃气管网; KNN; 随机森林; 泄漏风险; 软检测; 定量风险评价

中图分类号: TE988

文献标识码: A

文章编号: 1000-8241(2024)09-1064-09

DOI: 10.6047/j.issn.1000-8241.2024.09.012

Soft detection model of corrosion leakage risk based on KNN and random forest algorithms

YANG Yang^{1,2}, LI Chengzhi¹, DU Xuan¹, YU Xiao³, DONG Shaohua³

1. Kunlun Digital Technology Co. Ltd.; 2. School of Energy Power and Mechanical Engineering, North China Electric Power University;
3. College of Safety and Ocean Engineering, China University of Petroleum (Beijing)

Abstract: [Objective] The integrity management of urban gas pipeline networks demands effective risk assessment methods. Corrosion leakage risk assessment necessitates the comprehensive integration of risk assessment factors with various detection operations. Current detection tasks face challenges due to data complexities and significant data deficiencies. Therefore, it is vital to develop a method for predicting and evaluating corrosion leakage risks. [Methods] Key indicators associated with corrosion leakage risks were selected through a correlation analysis. These identified indicators were then employed to develop an intelligent soft detection model that integrates pipeline and environmental data, based on the K-Nearest Neighbor (KNN) and Random Forest algorithms. [Results] The model conducted predictions on missing detection data and achieved indirect measurements of key indicators, with a relative error between predicted and measured values staying below 25%, meeting acceptable standards. It effectively forecasts pipeline corrosion leakage risks in instances of missing data, paving the way for additional quantitative assessments. In comparison to prior research, the model displayed enhanced prediction accuracy and reliability, attributed to innovations in extracting multi-factor coupling relationships and algorithm choices. Nonetheless, the emergence of some abnormal data suggested constraints on its predictive capacity under specific circumstances and its dependence on complete and precise data. Consequently, enhancing both the quantity and quality of detection data, along with refining the feature extraction approach for key risk indicators, is anticipated to further boost the accuracy of the model. [Conclusion] This research enriches the risk prediction theory concerning corrosion leakage in gas pipelines and offers practical benefits in enhancing

pipeline operation safety and reliability. Future research efforts should focus on enhancing data acquisition and analysis techniques, optimizing the model structure, and improving the model adaptability and accuracy across various application scenarios. (10 Figures, 6 Tables, 25 References)

Key words: urban gas pipeline network, K-Nearest Neighbor (KNN), random forest, leakage risk, soft detection, quantitative risk assessment

随着社会经济水平不断发展,燃气作为主要消耗能源的比例持续增长,燃气管道作为燃气的传输载体,已成为重要的城市基础设施之一。埋地钢质管道长期埋于地下,随着时间推移,在施工、地形沉降、土壤腐蚀环境等因素影响下,存在发生泄漏的风险,可能给燃气公司与人民群众带来严重损失^[1]。定期对燃气管道实施腐蚀泄漏检测,定量评价燃气管道腐蚀泄漏风险,可为埋地管道的维护管理提供科学依据,减少潜在的不安全隐患,对于保证燃气管网安全运行具有重大意义^[2]。

由于城镇燃气埋地管道具有复杂性与隐蔽性,目前缺乏适用于燃气管道的腐蚀泄漏检测技术^[3]。近年来,基于数据驱动的管道腐蚀泄漏检测算法得到了广泛研究与应用,该类方法通常使用压力、流量、声音等时域信号数据检测管道的腐蚀泄漏^[4-5],仅针对已发生泄漏的管段。从风险管理角度考虑,如设法利用管道检测数据预测泄漏风险,可将风险前置,提前采取风险防控手段。但受管道压力检测与防腐检测的检测成本、人力与设备资源、技术可行性等条件限制,实际开展检测的管道仅占燃气管网约 15%^[6],且有大量数据未被很好地开发利用。机器学习技术的发展为实现燃气管道腐蚀泄漏软检测提供了可能。目前, KNN (K-Nearest Neighbor) 算法与随机森林算法在故障诊断与预测方面得到了广泛应用,其可靠性与准确性已得到验证^[7],相对误差基本小于 0.25,但上述算法尚未在管道腐蚀泄漏风险预测领域投入应用。在此,提出一种燃气腐蚀泄漏软检测模型,利用相关系数法从关键腐蚀泄漏风险指标的影响特征集中剔除不相关及冗余的影响特征,提取出最能代表某种腐蚀泄漏风险的低维主特征,即将大量腐蚀泄漏检测数据及管道内外部数据降维,导入 KNN 或随机森林中,作为软检测模型的输入,得到与腐蚀泄漏风险强相关但未检出数据的预测值,进一步进行腐蚀泄漏风险指标的风险值计算,以期为泄漏风险定量评价提供有效补充。

1 模型建立

1.1 指标选取

泄漏风险指标的数量会影响风险评估结果,过多将造成风险信息冗余,过少则缺失风险信息^[8]。通过对燃气管道泄漏风险影响因素与检测项目进行双向整合,结合事故数据库,对燃气管道泄漏事故进行原因分析,将燃气管道泄漏风险影响因素概括成 11 个风险强度指标,以表征管道泄漏影响的 11 个方面,即土壤腐蚀性风险强度、杂散电流干扰风险强度、防腐层风险强度、阴极保护风险强度、管道本体风险强度、庭院线环境风险强度、地质沉降风险强度、地质扰动风险强度、天气风险强度、施工风险强度、穿跨越风险强度。对指标进行分类时,通常需考虑必要性、重要性、测量难度^[9]。其中,必要性指国家规定检测的指标,重要性指与风险相关的重要指标,测量难度考虑检测的技术难度与成本要求。必要性通常是无法改变的,因此,从指标的重要性和测量难度对指标进行分类(图 1),可见土壤腐蚀性风险强度、杂散电流干扰风险强度、防腐层风险强度指标的重要性相对较高、测量难度相对较低,最适合作为泄漏风险关键指标。在确定风险强度关键指标后,针对关键指标对应的检测项目数据、本体数据及周围环境数据,利用皮尔逊相关系数进行多因素耦合研究,精简软检测模型的数据输入内容,并通过分别构建关键指标的影响因素体系,利用 KNN 算法和随机森林算法搭建软检测模型,对未能检测的重要数据进行预测。

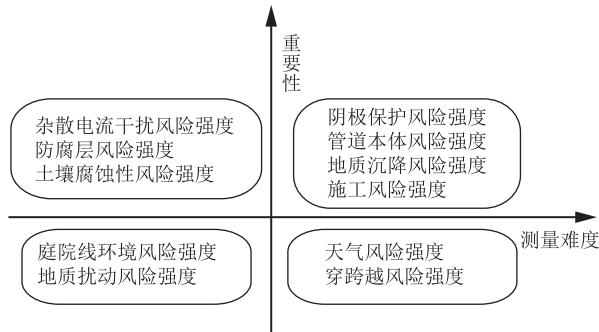


图 1 腐蚀泄漏风险强度指标分类示意图
Fig. 1 Classification of corrosion leakage risk intensity indicators

1.2 多因素耦合关系

相关系数最早由统计学家卡尔·皮尔逊设计,是研究变量之间线性相关程度的量。由于研究对象的不同,相关系数有多种定义方式^[10],统计学中常用的3种相关系数为皮尔逊相关系数(Pearson Correlation Coefficient, PCC)、斯皮尔曼秩相关系数(Spearman's Rank Correlation Coefficient, SRCC)及肯德尔相关系数(Kendall Rank Correlation Coefficient, KRCC)^[11]。PCC适用于连续数据、服从正态分布、线性关系的情况;SRCC利用两变量的秩次大小作线性相关分析,适用范围较广,但对于线性关系的数据统计效能较低;KRCC则适用于2个分类变量均为有序分类的情况^[12-14]。利用线性判别分别分析腐蚀检测数据、管道本体数据、周围环境的关联度,腐蚀泄漏相关数据更符合线性关系^[15],因此选择PCC,其表达式为:

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}} \quad (1)$$

式中: r 为PCC; σ_{XY} 为变量 X 与变量 Y 的协方差; σ_X^2, σ_Y^2 分别为变量 X 、变量 Y 的方差。

当 $r>0$ 时,两个变量正相关;当 $r<0$ 时,两个变量负相关;当 $r=0$ 时,两个变量不相关。 r 的绝对值越大,相关性越强(表1)^[16]。

表1 皮尔逊相关系数绝对值与相关性对应关系表
Table 1 Correspondence between Pearson correlation coefficients (absolute values) and correlation levels

$ r $	相关性
(0.8, 1.0]	非常强相关
(0.6, 0.8]	强相关
(0.4, 0.6]	中等相关
(0.2, 0.4]	弱相关
[0, 0.2]	非常弱相关或不相关

利用皮尔逊相关系数研究检测数据、管道本体数据、周围环境数据的多因素耦合关系,可以了解各部分数据的相关性,剔除相关性较大的外部数据,关注与关键风险指标及其对应检测内容最强相关的外部数据。根据相关性,从关键风险指标的影响特征集中剔除不相关或冗余的影响特征,提取出易于识别的低维主特征。

1.3 软检测模型

1.3.1 算法理论基础

在建立软检测模型过程中,涉及到KNN算法与随机森林算法。KNN算法通过测量不同特征值之间

的距离对其进行分类,如果一个样本在特征空间中的 k 个最相似(即特征空间中最邻近)样本中的大多数属于某一个类别,则该样本也属于这个类别,其中 k 通常是不大于20的整数^[17]。KNN算法中所选择的临近样本均是已正确分类的对象,该方法在定类决策上只依据最邻近的一个或者几个样本的类别决定待分样本所属的类别。在KNN中,通过计算对象间距离作为各个对象之间的非相似性指标,避免了对象之间的匹配问题^[18]。同时,KNN通过依据 k 个对象中占优的类别进行决策,而不是单一的对象类别决策。这两点就是KNN算法的优势^[19-21]。

随机森林是一种有监督学习算法,具有操作简单、预测精度高且能够对评价指标重要性进行识别等特点^[22]。该算法在应用时创建一个包含多个决策树的分类器,使其拥有某种方式的随机性,其输出的类别由个别树输出类别的众数决定^[23-24]。随机森林算法的本质是对决策树算法的优化,将多棵决策树组合在一起,每棵树的形成均取决于一个独立抽出的样本。每棵树均具有相同的分布,分类误差依赖于每棵决策树对类别划分的能力及决策树之间的关联性^[25]。

1.3.2 模型结构

燃气管道腐蚀泄漏软检测模型利用相关系数法,从关键腐蚀泄漏风险指标的影响特征集中剔除不相关及冗余的影响特征,提取出最能代表腐蚀泄漏风险的低维主特征,然后根据风险强度指标对提取的特征进行匹配选择,利用典型的机器学习算法,对与腐蚀泄漏风险强相关但未检测出的数据进行预测,最终用于燃气泄漏风险评价(图2)。

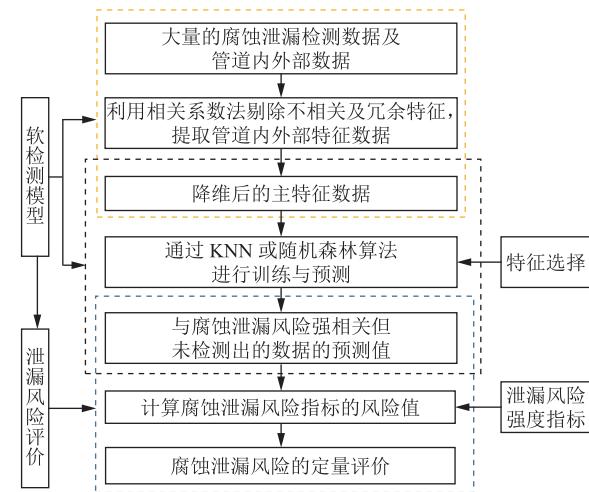


图2 燃气管道腐蚀泄漏软检测模型结构图
Fig. 2 Structure of soft detection model for corrosion leakage of gas pipeline

2 案例分析

2.1 数据预处理

2.1.1 数据收集

在对管道腐蚀泄漏风险进行预测时,需要收集的数据包括:①管道本体数据,指管道投入生产运行时的初始状态信息,包括管道的投运年限、管径、管材、压力级别、管理单位及埋深。②管道检测数据,包括防腐检测、压力管道检测、动态杂散电流专项检测数据及日常检测数据。其中,防腐检测是针对中低压管道的检测,压力管道检测是针对次高压及以上管道的检测,其数据包括土壤腐蚀性检测数据、静态杂散电流检测数据、防腐层检测数据、阴保设施检测数据等。③周围环境数据,包括地铁运行与分布、铁路分布、水系面、无轨电车及充电桩等电气化设备数据、地质沉降、地质扰动、管道穿跨越、天气、施工、庭院线所对应的小区属性数据。

2.1.2 数据相关性分析

2.1.2.1 本体数据

本体数据可用特征共有 6 个,分别为年份、管材、管径、压力水平、管理单位、管道埋深。其中,管理单位按照公司实际行政级别进行等级划分,并对等级进行数字化处理。由管道本体数据相关系数(图 3)可见,压力水平与管径的相关系数最大,为 0.332 621,可以判定本体数据各特征之间存在较弱的相关性,全部保留。

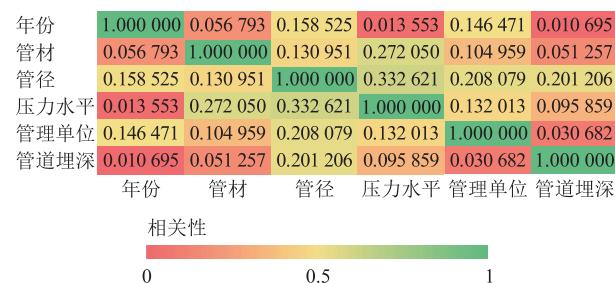


图 3 管道本体数据相关系数图
Fig. 3 Correlation coefficients of pipeline data

2.1.2.2 检测数据

对于土壤腐蚀性检测与静态杂散电流检测,相关可用检测项有仪器显示值、土壤电阻率、直流电位径向梯度、直流电位纵向梯度。分析各检测项之间的相关性(图 4)可见,仪器显示值与土壤电阻率的相关系数为 0.667 109,存在强相关。由于土壤电阻率是基于仪器显示值、地表至某土层深度、测量仪电极的距离等计算出的一段管段附近土壤电阻率平均值,相对于仪器显示值更具有代表性,因而在后续研究中仅关注土壤电阻率。

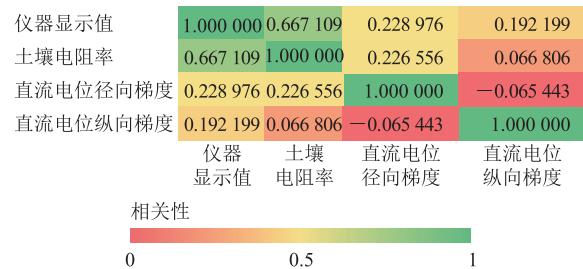


图 4 土壤腐蚀性检测项相关系数图
Fig. 4 Correlation coefficients of soil corrosivity detection data

检测防腐层时,可判断检测点为破损点的检测项有:开始电位、衰减后电位、电位差、绝缘电阻值、破损统计。其中,电位差为开始电位与衰减后电位之差,绝缘电阻值由工作人员结合电位差与工作经验估计得出。分析各检测项之间的相关性(图 5)可见,各检测项之间的相关性很强,特别是开始电位与衰减后电位的相关系数达 0.963 467。由于电位差在不同时刻的测量值不同,在后续研究中选择能够清楚反映管道防腐层破损情况的绝缘电阻值作为研究因素。

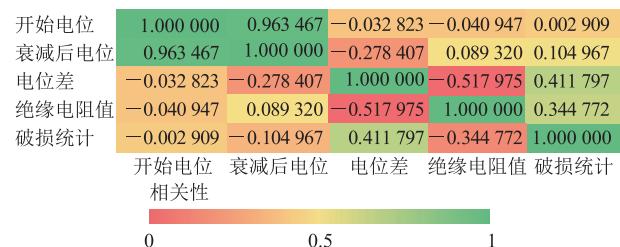


图 5 防腐层检测项相关系数图
Fig. 5 Correlation coefficients of anti-corrosive coating detection data

评价阴保设施状态的检测项有:保护电位、阳极开路电位、阳极输出电流。分析各检测项之间的相关性(图 6)可见,保护电位与阳极开路电位之间的相关性系数为 0.78,二者强相关,选取其中 1 个作为阴保设施检测的相关因素。

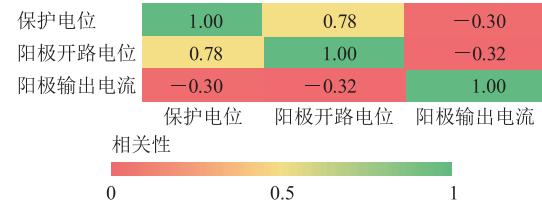


图 6 阴保设施检测项相关系数图
Fig. 6 Correlation coefficients of cathodic protection detection data

2.1.2.3 周围环境数据

周围环境数据可用特征共有 14 个,分别为距离样本点最近的 5 个充电桩距离 cd1、cd2、cd3、cd4、cd5,距离样本点最近的地铁距离 ddis,500 m、1 000 m 范围内的地铁数量 d500、d1000,距离样本点最近的铁路距离 tdis,700 m 内是否有铁路 d700(无、有时分别取值 0、1),

2 000 m 范围内的铁路数量 d2000, 离样本点最近的水系的距离 dis、面积 are、长度 long。分析上述特征之间的相关性(图 7)可见, 距离样本点最近的 5 个充电桩距离之间存在极强的相关性, 因此仅保留其中 1 个, 即

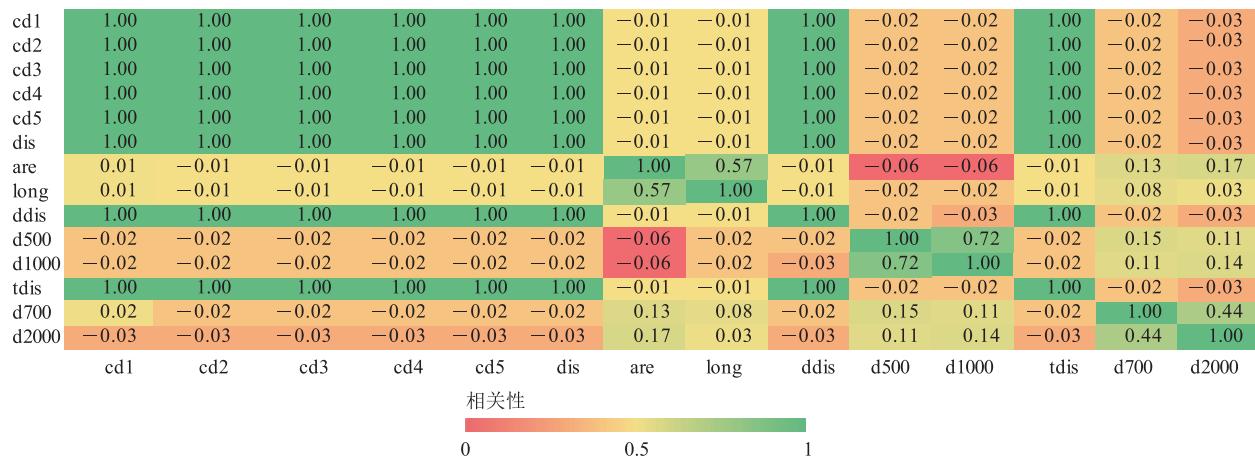


图 7 周围环境特征相关系数图
Fig. 7 Correlation coefficients of environmental data

2.2 构造关键指标软检测模型

2.2.1 土壤腐蚀性风险强度

土壤腐蚀性风险强度数据包括土壤电阻率、最近水系面距离、水系面积及水系面长度。其中, 最近水系面距离、水系面积、水系面长度是完整且客观的, 无需对数据进行额外补充处理; 土壤电阻率在实际检测中是通过不连续采集获得的, 并未覆盖全部管道, 因此需要通过建立土壤电阻率软检测模型的方式, 对未检测土壤电阻率的管段进行补充。对目前获取的 2 912 条原始土壤电阻率数据进行整合, 由于数据的来源不尽相同, 格式包括各种形式的电子文档与纸质检测报告, 在使用数据之前, 需对原始检测数据进行数据整合与正确性校验, 分析异常值, 并利用正则表达式将其转换为正确格式。通过数据整合, 最终得到可用的土壤电阻率检测数据 2 912 条, 数据输出为 Excel 格式(表 2)。

由于土壤电阻率与其位置信息有着密切的关联, 在对土壤电阻率进行预测时, 采用 KNN 算法搭建土壤电阻率软检测模型, 采用最近邻算法, 即 $k=1$, 对未检测点的土壤电阻率进行预测并补充数据。根据 GB/T 21447—2018《钢质管道外腐蚀控制规范》中的定义, 对土壤电阻率进行分级(表 3), 同时基于该标准统计各检测点分级情况, 等级为 1、2、3 的监测点数目分别为 43、375、2 494。

在建立土壤电阻率软检测模型时, 针对上述 2 912 个检测样本, 利用其中的 2 184 个样本进行训练及交叉验证, 构建土壤电阻率软检测模型。之后, 利用剩余的

最近充电桩距离; 离样本点最近的一条水系的面积与长度之间的相关系数为 0.57, 存在中等程度的相关性, 暂且保留; 500 m 内地铁数量与 1 000 m 内地铁数量的相关系数为 0.72, 存在强相关, 保留其中 1 个。

表 2 土壤电阻率检测数据表(部分)
Table 2 Soil resistivity detection data (partial)

管段 编号	土壤电阻率/ ($\Omega \cdot m$)	坐标		数据来源或 数据库名称
		x	y	
1	71.000	503 761.59	307 201.66	word
2	73.000	504 519.49	304 619.07	word
3	60.000	505 102.96	303 960.47	word
...
330	121.832	517 586.76	289 498.04	n1516
331	128.112	516 760.24	289 207.53	n1516
332	133.136	515 538.46	290 712.47	n1516
...
1 876	91.688	529 160.72	283 309.18	hbase
1 877	89.678	529 668.74	280 087.98	hbase
1 878	115.552	529 666.49	280 702.08	hbase
...
2 299	8.792	504 303.56	284 174.05	动态杂散电流检测
2 300	10.676	530 326.01	302 429.88	动态杂散电流检测
2 301	13.816	513 358.47	315 916.48	动态杂散电流检测
...

表 3 土壤电阻率及腐蚀性分级表
Table 3 Classification of soil resistivity and corrosive property

等级	腐蚀性	土壤电阻率/($\Omega \cdot m$)
1	弱	大于 50
2	中等	20~50
3	强	小于 20

728 个样本数据对软检测模型预测结果进行验证, 准确率为 0.78。

2.2.2 杂散电流干扰风险强度

在对未检测点所在管段进行杂散电流评分时, 其基础是已检测点的杂散电流特征, 因此建立杂散电流

直流干扰风险强度软检测模型是评价杂散电流风险强度的关键。首先对检测数据进行数据预处理。在杂散电流检测中,土壤电阻率数据是缺失的,基于上述土壤电阻率KNN软检测模型,对缺失数据进行填补(表4)。

表4 杂散电流检测土壤电阻率KNN填补表

Table 4 KNN-based complement of soil resistivity for stray current detection

年份	位置	最近充电桩距离/km	土壤电阻率/(Ω·m)	数据源	直流干扰风险
2014	地下	682.466 738 20	35.168	实际检测	弱
2012	地下	379.442 907 00	70.336	实际检测	欠保护
2012	地下	379.498 336 70	70.336	实际检测	欠保护
2012	地下	216.817 922 50	91.688	实际检测	欠保护
2012	地下	58.688 686 64	75.360	实际检测	弱
2007	地下	218.400 991 20	64.000	软检测	弱
2007	地下	898.149 822 60	111.784	软检测	欠保护
1971	地下	12.449 029 21	45.216	实际检测	弱
1971	地下	496.895 710 50	138.160	实际检测	弱
1971	地下	156.543 996 60	27.632	实际检测	弱
1971	地下	640.446 737 80	20.096	实际检测	欠保护

在搭建杂散电流直流干扰风险强度软检测模型前,先对其进行特征分析,可以发现高压/次高压管道与中/低压管道的杂散电流干扰差异明显,因此在搭建直流干扰风险强度软检测模型时,应分别基于二者数据建立模型。考虑到特征之间存在高度相关性,在建模前进行特征选取工作(表5、表6)。

表5 中高压/次高压管道直流干扰风险强度软检测模型所用特征表

Table 5 Features in soft detection model for DC interference risk intensity of MP and HP/Sub-HP pipelines

特征名称	含义	特征名称	含义
x	横坐标	Subway line	地铁线路
y	纵坐标	flow	载客量
dis	最近水系面距离	date	地铁投运年限
are	水系面积	R	土壤电阻率
ddis	最近地铁距离	place	纵向位置
d500	500 m 内地铁个数	cd	最近充电桩距离
tdis	最近铁路距离	management	管理单位
d700	700 m 内是否有铁路		

表6 中压/低压管道直流干扰风险强度软检测模型所用特征表

Table 6 Features in soft detection model for DC interference risk intensity of MP/LP pipelines

特征名称	含义	特征名称	含义
x	横坐标	d700	700 m 内是否有铁路
y	纵坐标	flow	载客量
dis	最近水系面距离	date	地铁投运年限
are	水系面积	R	土壤电阻率
ddis	最近地铁距离	place	纵向位置
d500	500 m 内地铁个数	cd	最近充电桩距离
tdis	最近铁路距离		

基于上述特征,经过多次模型调参迭代,搭建的高压/次高压管道直流干扰风险强度软检测模型为随机森林模型,设置树个数为5(Tree 0~Tree 4),将特征投射到二维平面,为更好地解释该模型的分类过程,对模型的过程与结果进行可视化呈现(图8)。分类结果图的划分是一个综合判断过程,其划分依据由的划分结果投票判断所得。已检测次高压/高压管道共313个样本,在建模时,用其中250个样本进行训练及交叉验证,利用剩余63个样本对搭建的软检测模型预测结果进行验证,准确率为0.82。

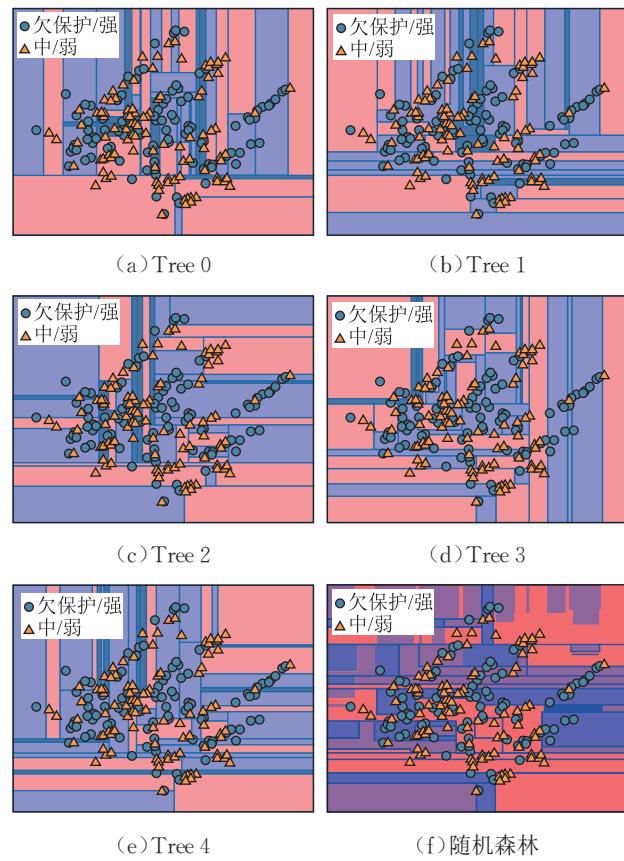


图8 高压/次高压管道直流干扰风险强度软检测模型分类可视化图

Fig. 8 Visualized classification of soft detection model for DC interference risk intensity of HP/Sub-HP pipelines

中压/低压管道直流干扰风险强度软检测模型为随机森林模型,设置树个数为5,将特征投射到二维平面可视化呈现(图9),可见,与高压/次高压管道直流干扰风险强度软检测模型类似。已检测中低压管道共341个样本,利用其中272个样本数据进行训练与交叉验证,其他69个样本对搭建的软检测模型预测结果进行验证,准确率为0.753 6。

2.2.3 防腐层风险强度

在防腐层风险强度指标计算相关影响因素中,防

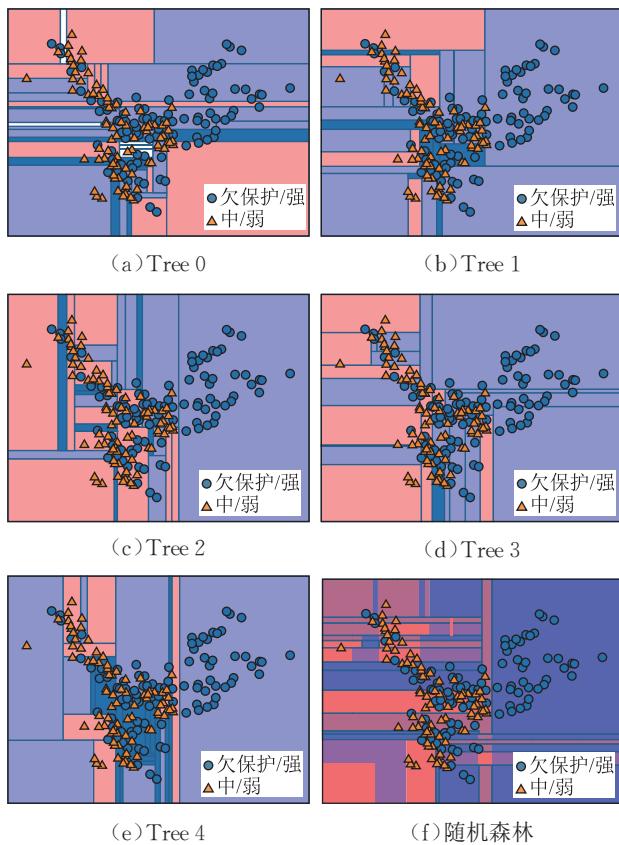


图9 中压/低压管道软检测模型直流干扰风险强度分类可视化图

Fig. 9 Visualized classification of soft detection model for DC interference risk intensity of MP/LP pipelines

防腐层绝缘电阻起到决定性的作用。而防腐层绝缘电阻这一检测指标在实际检测中并未覆盖到全部管道,因此需要对防腐层绝缘电阻建立软检测模型,从而对未检测防腐层绝缘电阻的管段进行补充。对已检测的防腐层绝缘电阻数据进行整合,最终得到9 972条数据。基于上述数据,收集与燃气管道防腐层绝缘电阻相关的特征数据,包括以管道压力级别、管径、投运年限、管理单位等为代表的管道本体属性数据与以管道与最近水系面、附近地铁、电气化铁路的位置关系等为代表的管道周边环境数据。选取原则包括:①选择对燃气管道腐蚀泄漏风险影响大的数据和管道泄漏后显著影响周围环境的数据;②遵守燃气公司管道腐蚀泄漏风险管控的管理规范、技术规程与数据规范;③参考管道泄漏失效领域的专家意见。

基于燃气管道内、外部数据,通过随机森林算法搭建软检测模型。针对已检测管道共计9 972个检测样本,其中8 972个样本用于进行训练及验证,构建防腐层绝缘电阻软检测模型;剩余1 000个样本用于对搭建的软检测模型预测结果进行验证,结果显示,相关系数为0.93,一致性系数为0.829 3(图10)。

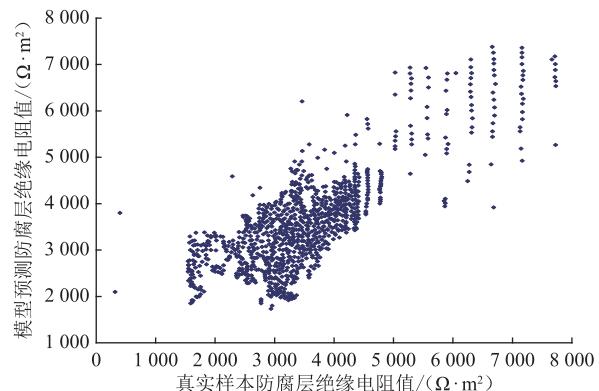


图10 防腐层绝缘电阻软检测模型预测效果验证图
Fig. 10 Verification of prediction effect of soft detection model for insulation resistance of anti-corrosive coating

3 结论

1)为实现燃气管道腐蚀泄漏风险的预测,提出了一种基于多因素耦合关系提取、KNN 算法与随机森林算法的腐蚀泄漏风险软检测模型。该模型利用管道腐蚀检测数据,对土壤腐蚀风险强度、杂散电流干扰风险强度及防腐层风险强度等重点检测内容进行预测,实现了难测量要素值的数学预测,预测值与真实测量值的相对误差小于25%,精度达到了合格水平,显示了模型在燃气管道腐蚀泄漏风险预测中的普遍适用性。然而,部分异常数据表明模型在某些条件下的预测能力有限,且模型性能依赖于检测数据的完整性和准确性。

2)与以往研究相比,该模型在构建方法与算法选择上有所创新,结合了多因素耦合关系提取与先进的机器学习算法,在模型构建过程中强调了关键风险指标影响特征选取的重要性,提高了预测的准确性和可靠性。该模型在理论上丰富了燃气管道腐蚀泄漏风险预测的理论框架,在应用方面提高了管道运行的安全性和可靠性,具有显著的实用价值。在未来进一步研究中,建议增加并完善腐蚀检测数据,探索更多关键风险指标的影响特征,优化特征提取方法,并结合其他先进机器学习算法以提升模型的预测精度及稳定性,从而提升燃气管道腐蚀泄漏风险预测的整体水平,确保管道的安全运行。

参考文献:

- [1] 金龙,陈樑,王海燕,张昇.不同土壤对输气管道泄漏扩散影响模拟分析[J].石油与天然气化工,2020,49(4):101-108. DOI: 10.3969/j.issn.1006-1390.2020.04.016.

- 10.3969/j.issn.1007-3426.2020.04.018.
- JIN L, CHEN L, WANG H Y, ZHANG S. Simulation analysis of the influence of different soil on gas pipeline leakage and diffusion[J]. Chemical Engineering of Oil and Gas, 2020, 49(4): 101-108.
- [2] 徐鹏,潘丹丹,袁勋,张正雄,熊亚选. 燃气管道泄漏扩散研究进展[J]. 油气储运,2022,41(1):21-28. DOI: 10.6047/j.issn.1000-8241.2022.01.003.
- XU P, PAN D D, YUAN X, ZHANG Z X, XIONG Y X. Progress of research on leakage diffusion of gas pipelines[J]. Oil & Gas Storage and Transportation, 2022, 41(1): 21-28.
- [3] 高琳,曹建国. 管道泄漏检测方法研究综述[J]. 现代制造工程, 2022(2): 154-162. DOI: 10.16731/j.cnki.1671-3133.2022.02.024.
- GAO L, CAO J G. Review of diagnostic technique for pipe leakage[J]. Modern Manufacturing Engineering, 2022(2): 154-162.
- [4] 李睿. 基于龙格-库塔法的天然气管道泄漏检测与定位[J]. 油气储运,2019,38(7):816-821. DOI: 10.6047/j.issn.1000-8241. 2019.07.016.
- LI R. Gas pipeline leakage detection and location based on Runge-Kutta method[J]. Oil & Gas Storage and Transportation, 2019, 38(7): 816-821.
- [5] 谭震,郭新蕾,李甲振,郭永鑫,潘佳佳. 基于多尺度卷积神经网络的管道泄漏检测模型研究[J]. 水利学报,2023,54(2): 220-231. DOI: 10.13243/j.cnki.slxb.20220644.
- TAN Z, GUO X L, LI J Z, GUO Y X, PAN J J. Multi-scale convolutional neural network model for pipeline leak detection[J]. Journal of Hydraulic Engineering, 2023, 54(2): 220-231.
- [6] 罗宗林. 城市直埋燃气管道泄漏扩散对环境影响的研究[D]. 哈尔滨: 哈尔滨工业大学,2021.
- LUO Z L. Study on environmental impact of leakage and diffusion of urban directly buried gas pipeline[D]. Harbin: Harbin Institute of Technology, 2021.
- [7] HU Z K, CHEN B Q, CHEN W L, TAN D B, SHEN D T. Review of model-based and data-driven approaches for leak detection and location in water distribution systems[J]. Water Supply, 2021, 21(7): 3282-3306. DOI: 10.2166/ws.2021.101.
- [8] 戴海霞,李文. 基于 Spearman 秩相关系数与 AHP 法的天然气管道运行风险绩效评估[J]. 山东化工,2023,52(2):178-181. DOI: 10.3969/j.issn.1008-021X.2023.02.050.
- DAI H X, LI W. Natural gas pipeline operation risk performance evaluation based on Spearman's rank correlation coefficient and AHP method[J]. Shandong Chemical Industry, 2023, 52(2): 178-181.
- [9] 姜礼涛,周爱红,刘超. 基于灰色关联度的隧洞围岩分类研究[J]. 河北地质大学学报,2022,45(4):58-64. DOI: 10.13937/j.cnki.hbdzxxb.2022.04.009.
- JIANG L T, ZHOU A H, LIU C. Study on classification of tunnel surrounding rock based on grey correlation degree[J]. Journal of Hebei GEO University, 2022, 45(4): 58-64.
- [10] 赵源上,林伟芳. 基于皮尔逊相关系数融合密度峰值和熵权法典型场景研究[J]. 中国电力,2023,56(5):193-202. DOI: 10.11930/j.issn.1004-9649.202212032.
- ZHAO Y S, LIN W F. Research on typical scenarios based on fusion density peak value and entropy weight method of Pearson's correlation coefficient[J]. Electric Power, 2023, 56(5): 193-202.
- [11] 纪德洋,金锋,冬雷,张姗,于坤洋. 基于皮尔逊相关系数的光伏电站数据修复[J]. 中国电机工程学报,2022,42(4):1514-1522. DOI: 10.13334/j.0258-8013.pcsee.211172.
- JI D Y, JIN F, DONG L, ZHANG S, YU K Y. Data repairing of photovoltaic power plant based on Pearson correlation coefficient[J]. Proceedings of the CSEE, 2022, 42(4): 1514-1522.
- [12] SELVY P T, RAGUL M, VIGNESH G N, ANITHA M. Real time efficient accident predictor system using machine learning techniques (kNN, RF, LR, DT)[J]. International Journal of Engineering and Advanced Technology, 2020, 10(2): 108-111. DOI: 100.1/ijeat.D6910049420.
- [13] 宫彦双,谷坛,安超,张爱良,廖柯熹. 基于 Pearson 相关系数的集输管道流动腐蚀主控因素分析[J]. 石油与天然气化工, 2021,50(2):93-99. DOI: 10.3969/j.issn.1007-3426.2021.02.016.
- GONG Y S, GUT, ANC, ZHANG A L, LIAO K X. Analysis of main control factors for flow corrosion of gathering and transportation pipeline based on Pearson correlation coefficient[J]. Chemical Engineering of Oil and Gas, 2021, 50(2): 93-99.
- [14] 陈世群,高伟,陈孝琪,涂彦昭,杨艳. 一种基于极限学习机和皮尔逊相关系数的光伏阵列故障快速诊断方法[J]. 电气技术, 2021,22(10):57-64. DOI: 10.3969/j.issn.1673-3800.2021.10.010.
- CHEN S Q, GAO W, CHEN X Q, TU Y Z, YANG Y. A fast fault diagnosis method for photovoltaic array via extreme learning machine and Pearson's correlation coefficient[J].

- Electrical Engineering, 2021, 22(10): 57–64.
- [15] 姚梦月. 埋地油气管道外腐蚀速率及剩余寿命预测模型优化研究[D]. 西安: 西安建筑科技大学, 2020.
- YAO M Y. Study on the optimization of prediction model for the external corrosion rate and residual life of buried oil and gas pipelines[D]. Xi'an: Xi'an University of Architecture and Technology, 2020.
- [16] XUE X M, ZHOU J Z. A hybrid fault diagnosis approach based on mixed-domain state features for rotating machinery[J]. ISA Transactions, 2017, 66: 284–295. DOI: 10.1016/j.isatra.2016.10.014.
- [17] ZHANG J, BIAN Z K, WANG S T. Self-paced and Bayes-decision-rule linear KNN prediction[J]. International Journal of Machine Learning and Cybernetics, 2022, 13(11): 3267–3283. DOI: 10.1007/s13042-022-01593-9.
- [18] 郭小萍, 滕佳岐, 李元. 基于EEMD-kNN的工业过程微小故障检测[J]. 计算机应用与软件, 2023, 40(4): 36–40, 53. DOI: 10.3969/j.issn.1000-386x.2023.04.007.
- GUO X P, TENG J Q, LI Y. Incipient fault detection for industry process based on EEMD-kNN[J]. Computer Applications and Software, 2023, 40(4): 36–40, 53.
- [19] 刘培栋, 刘立巍, 李明, 赵光, 刘书元, 邵壮, 等. 基于KNN和RF结合的供热机组热负荷预测研究[J]. 全球能源互联网, 2021, 4(6): 595–601. DOI: 10.19705/j.cnki.issn2096-5125.2021.06.009.
- LIU P D, LIU L W, LI M, ZHAO G, LIU S Y, SHAO Z, et al. Research on heat load forecasting of heating unit based on combination of KNN and RF[J]. Journal of Global Energy Interconnection, 2021, 4(6): 595–601.
- [20] 陈智鹏. 基于遗传KNN聚类的机械加工故障趋势预测模型[J]. 机械与电子, 2019, 37(5): 59–62, 70. DOI: 10.3969/j.issn.1001-2257.2019.05.014.
- CHEN Z P. Fault trend prediction model of machining based on genetic KNN clustering[J]. Machinery & Electronics, 2019, 37(5): 59–62, 70.
- [21] 李洪奇, 杨中国, 朱丽萍, 刘蔷. 基于数据集特征的KNN最优K值预测方法[J]. 计算机应用与软件, 2016, 33(6): 54–58, 83. DOI: 10.3969/j.issn.1000-386x.2016.06.014.
- LI H Q, YANG Z G, ZHU L P, LIU Q. Prediction method of optimal K value in KNN based on dataset features[J]. Computer Applications and Software, 2016, 33(6): 54–58, 83.
- [22] 林艳竹, 韩忠, 黄林显, 邢立亭, 梁浩, 侯金霄. 基于随机森林法的弥河—潍河流域地下水质量评价研究[J]. 水文, 2023, 43(3): 60–64, 70. DOI: 10.19797/j.cnki.1000-0852.20210513.
- LIN Y Z, HAN Z, HUANG L X, XING L T, LIANG H, HOU J X. Research on groundwater quality assessment of Mihe–Weihe River Basin based on random forest algorithm[J]. Journal of China Hydrology, 2023, 43(3): 60–64, 70.
- [23] MADANI A, HAGAGE M, ELBEIH S F. Random forest and logistic regression algorithms for prediction of groundwater contamination using ammonia concentration[J]. Arabian Journal of Geosciences, 2022, 15(20): 1619. DOI: 10.1007/s12517-022-10872-2.
- [24] ZHANG H Z, ZIMMERMAN J, NETTLETON D, NORDMAN D J. Random forest prediction intervals[J]. The American Statistician, 2020, 74(4): 392–406. DOI: 10.1080/00031305.2019.1585288.
- [25] 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.

(编辑:张腾)

基金项目: 中国石油天然气集团有限公司项目“天然气管网及站场智能管控系统研发”, 2021DJ7304。

作者简介: 杨阳, 男, 1986年生, 工程师, 2013年硕士毕业于西南石油大学油气田材料与应用专业, 现主要从事油气管道完整性管理、材料科学与工程等专业方向的研究工作。地址: 北京市昌平区沙河西沙屯中石油科技园12地块B1座, 102206。电话: 18210920815。Email: yangyang2021@cnpc.com.cn

- Received: 2023-08-29
- Revised: 2023-10-17
- Online: 2024-06-14

