

基于大数据分析的特种设备事故风险预测模型构建

蔺根宏

银川阅能天然气有限公司 宁夏 银川 750001

摘要: 本文探讨了基于大数据分析的特种设备事故风险预测模型构建。首先阐述特种设备数据类型、特点及大数据分析技术,接着介绍数据处理流程,包括数据采集、清洗与预处理。随后详细说明模型选择、构建流程及关键因素考量。最后通过交叉验证、留一法等方法验证模型,采用准确率、召回率等指标评估,并提出模型优化策略,为特种设备事故风险预测提供新思路。

关键词: 大数据分析; 特种设备; 事故风险预测; 模型构建; 数据处理

引言: 特种设备广泛应用于工业与民用领域,其安全运行至关重要。随着设备数量增加与运行复杂度提升,事故风险增大。大数据分析技术凭借强大的数据处理与挖掘能力,为特种设备事故风险预测提供新途径。通过整合多源数据,深入挖掘数据潜在价值,构建精准风险预测模型,可提前发现潜在风险,为设备安全管理提供科学依据,保障设备安全稳定运行。

1 大数据分析在特种设备事故风险预测中的应用基础

1.1 特种设备数据类型与特点

特种设备数据涵盖运行数据、维护数据、检验数据等多种类型。运行数据记录设备运行过程中的各项参数,包含设备的转速、压力、温度、位移等指标,反映设备实时工作状态。维护数据详细记载设备保养、维修、部件更换等信息,记录维护时间、维护内容、更换部件型号等,体现设备维护历史。检验数据包含定期检验、不定期检验结果,记录设备性能指标、缺陷状况、检验结论等,是设备安全性能的综合评估。这些数据具有多源性特点,来源于设备自身传感器、维护人员记录、检验机构报告等不同渠道。多源数据从不同角度反映设备状况,为风险预测提供全面信息。数据呈现动态性,设备运行参数随时间不断变化,维护和检验活动也持续产生新数据,数据持续更新,反映设备全生命周期状态演变。海量性也是显著特征,随着设备数量增加和运行时间累积,数据规模迅速扩大,庞大的数据量蕴含丰富设备运行规律和潜在风险信息,为精准风险预测提供可能。

1.2 大数据分析技术概述

数据挖掘技术通过关联分析、聚类分析、分类分析等方法,从大量特种设备数据中提取有价值信息。关联分析挖掘数据间潜在关联关系,发现设备运行参数、维护记录、检验结果之间的联系,找出影响设备安全的关

键因素组合。聚类分析依据数据特征将设备运行状态或故障模式分类,识别不同类型设备运行规律和潜在风险差异。分类分析构建分类模型,对设备状态进行分类预测,判断设备正常、异常或故障状态。机器学习技术在特种设备数据处理中发挥重要作用。决策树算法通过构建树形模型,依据数据特征进行决策分类,直观展现设备状态判断逻辑,用于设备故障初步诊断。支持向量机算法在高维数据空间寻找最优分类超平面,有效处理小样本、非线性数据,提高设备故障分类准确率。随机森林算法集成多个决策树模型,降低模型方差,提高预测稳定性和泛化能力,实现对设备故障概率的可靠预测。深度学习技术凭借多层神经网络强大特征提取能力,自动学习特种设备数据复杂特征。卷积神经网络善于处理图像、时间序列数据,可挖掘设备运行参数时间序列中的深层特征,捕捉设备运行趋势和异常变化模式^[1]。循环神经网络能够处理具有时间依赖关系的数据,记忆设备历史运行状态,更好地预测设备未来状态。这些技术在处理特种设备海量、动态、复杂数据时,展现出强大信息提取和分析能力,为事故风险预测提供技术支撑。

2 特种设备事故风险预测的数据处理

2.1 数据采集

特种设备数据采集来源广泛。设备自带传感器是关键数据源头,通过压力传感器、温度传感器、位移传感器等,实时监测设备运行时的压力、温度、位移等参数变化,持续输出设备运行状态信息。人工录入数据同样重要,维护人员在设备保养、维修过程中,详细记录维护时间、维修内容、更换部件型号等信息;操作人员也会记录设备启停、运行模式切换等操作细节,这些数据补充了传感器无法捕捉的人工操作信息。检验报告则是数据的重要补充,定期检验和不定期检验形成的报告,详细记载设备性能指标、缺陷状况、检验结论,为设备

安全性能评估提供依据。数据采集必须遵循完整性与准确性要求。完整性要求覆盖设备全生命周期数据,从安装调试、日常运行到维护检修、退役报废,每个阶段的数据都需完整采集。准确性要求确保采集数据真实反映设备实际状况,传感器需定期校准维护,保证数据测量精度;人工录入需建立严格审核机制,减少人为错误,保证数据真实可靠,为后续分析奠定坚实基础。

2.2 数据清洗

数据清洗是保障数据质量的必要步骤。原始数据中存在重复数据、缺失值与异常值,影响数据分析结果准确性。重复数据源于传感器故障、数据传输错误或人工重复录入,这些冗余数据干扰分析过程,需通过数据查重算法识别并删除重复记录。缺失值产生于传感器故障、数据传输中断或人工漏记等情况。处理缺失值有多种方法,对于少量缺失数据,可采用均值、中位数、众数等统计量填充;对于具有时间序列特征的数据,可利用相邻时间点数据进行插值填充;若缺失比例较大且无规律,可通过模型预测等方法估算缺失值。异常值由传感器异常、操作失误或环境突变导致,会对分析结果产生极大偏差,可通过箱线图、 3σ 原则等方法识别,对异常值进行修正或删除处理,保证数据符合正常分布特征,提升数据质量。

2.3 数据预处理

数据预处理通过标准化、归一化、特征提取与选择等操作,提高数据可用性与分析效率。标准化处理将数据按比例缩放,消除不同指标间量纲和取值范围差异,常见方法如Z-score标准化,使数据服从均值为0、标准差为1的标准正态分布,便于不同类型数据对比分析。归一化处理将数据映射到特定区间,如 $[0,1]$,简化计算过程,增强模型收敛速度与稳定性。特征提取从原始数据中提炼更具代表性的特征,如通过主成分分析(PCA)将高维数据降维,提取少数几个综合特征,保留数据主要信息的同时减少数据维度,降低计算复杂度^[2]。特征选择则从众多特征中挑选出对分析目标最有价值的特征,去除冗余和无关特征,可采用相关系数法、递归特征消除法等,优化数据特征集,提升数据分析效率与模型预测性能,为特种设备事故风险预测提供高质量数据支持。

3 基于大数据分析的特种设备事故风险预测模型构建

3.1 模型选择

特种设备事故风险预测模型的构建需要选择适合的算法,以准确识别潜在风险并实现有效预警。决策树模型因其结构清晰、易于解释的特点,适用于初步分析设备故障与事故之间的关联性。该模型通过树状结构划分

数据,能够直观展示不同因素对事故风险的影响程度。随机森林在决策树基础上进行集成学习,通过多棵决策树的投票机制降低过拟合风险,提升预测稳定性。该模型在处理高维数据时表现优异,能够有效整合设备运行参数、环境因素和维护记录等多源信息。神经网络模型具备强大的非线性拟合能力,适用于复杂工况下的风险预测。深度学习中的长短期记忆网络能够捕捉时间序列数据的长期依赖关系,对设备性能退化趋势进行动态建模。卷积神经网络则适用于图像或信号数据的特征提取,可用于分析设备关键部件的磨损或裂纹等视觉检测数据。模型选择需结合特种设备的具体特点,考虑数据规模、特征维度及预测实时性要求,确保算法在准确性与计算效率之间取得平衡。

3.2 模型构建流程

特种设备事故风险预测模型的构建遵循系统化流程,从数据准备到模型优化各环节紧密衔接。数据采集阶段整合设备传感器监测数据、历史维护记录和事故报告,形成结构化数据集。数据预处理环节对原始数据进行清洗,处理缺失值和异常值,通过标准化或归一化消除量纲差异。特征工程阶段筛选关键变量,采用主成分分析或相关性分析降低数据维度,保留对风险预测贡献度高的特征。模型架构设计根据预测目标确定输入输出层结构,中间层节点数和激活函数的选择影响模型表达能力。参数初始化采用随机分布或预训练策略,避免梯度消失或爆炸问题。模型训练采用交叉验证方法划分数据集,通过迭代优化损失函数调整权重参数。训练过程中监控准确率和召回率等指标,防止过拟合或欠拟合现象。模型优化环节采用网格搜索或贝叶斯优化调整超参数,提升预测性能。模型部署阶段将训练好的算法嵌入实时监测系统,实现风险动态评估与预警。整个流程以大数据为驱动,强调数据质量与算法协同,确保模型在实际应用中的可靠性。

3.3 关键因素考量

特种设备事故风险预测模型的准确性依赖于对关键影响因素的全面考量。设备类型是首要因素,不同类别的特种设备如压力容器、起重机械和电梯等,其事故模式和风险特征存在显著差异。使用年限直接影响设备性能状态,老化程度与故障率呈非线性增长关系,需在模型中引入时间衰减因子进行修正。运行环境因素包括温度、湿度、腐蚀性介质等外部条件,这些变量会加速材料疲劳或电气系统失效。维护记录反映设备的保养状况,定期维护可降低突发故障概率,而未按规程检修则会显著增加风险等级。操作人员行为数据也是重要输

入,操作频次、违规操作历史等变量能够量化人为失误对事故的贡献度。设备负荷特性需特别关注,超载运行或频繁启停会缩短关键部件寿命^[3]。历史事故数据提供统计规律,相同型号设备在类似工况下的事故案例具有重要参考价值。传感器数据的采样频率和精度影响实时预测效果,高频振动或温度数据能更早发现异常征兆。模型构建需综合这些因素的交互作用,通过特征交叉或分层建模捕捉复杂关联,从而提升风险预测的精准度。

4 模型的验证与评估

4.1 验证方法

交叉验证是保障模型可靠性与泛化能力的常用方法。其将数据集划分为多个互斥子集,每次选取其中一个子集作为测试集,其余子集作为训练集,多次循环训练和测试,最终综合多次结果评估模型性能。这种方式能充分利用数据,避免因训练集和测试集划分不当导致的评估偏差,全面反映模型在不同数据分布下的表现。留一法作为交叉验证的特殊形式,每次仅用一个样本作为测试集,其余样本构成训练集,数据集有多少个样本就进行多少次训练和测试。该方法最大限度利用数据,评估结果较为准确,但计算成本高,适用于样本数量较少的情况。通过留一法,可细致考察模型对每个样本的预测能力,发现模型潜在的局部预测缺陷。无论是交叉验证还是留一法,都通过多轮训练与测试,模拟模型在不同数据环境下的运行情况,有效检验模型是否存在过拟合或欠拟合问题,确保模型在实际应用中具备良好的泛化能力,对新数据也能做出可靠预测。

4.2 评估指标

准确率用于衡量模型正确预测的比例,计算方式为正确预测的样本数占总样本数的比例。在特种设备事故风险预测中,准确率高意味着模型能准确识别设备正常与异常状态的大部分样本,但当数据样本分布不均衡时,准确率可能无法真实反映模型性能。召回率关注正样本的预测情况,体现模型正确预测出的正样本数量占实际正样本数量的比例。在特种设备风险预测场景下,高召回率表示模型能尽可能多地检测出存在事故风险的设备,减少漏检情况,避免因未及时发现风险设备而引发事故。F1值是综合考虑准确率和召回率的指标,通过调和平均数将两者结合,平衡模型在正样本和负样本预

测上的表现。当模型的准确率和召回率都较高时,F1值才会高,其能更全面地评估模型在分类任务中的整体性能。均方误差适用于回归问题,计算预测值与真实值之间误差平方的平均值。在量化评估特种设备事故风险等级时,均方误差越小,表明模型预测值与实际风险值越接近,模型预测的准确性越高,可直观反映模型在连续数值预测上的误差程度。

4.3 模型优化

根据验证与评估结果,改进算法是优化模型的重要方向。若模型在复杂数据模式识别上表现不佳,可尝试更换更先进的算法。将简单决策树模型替换为随机森林或神经网络模型,利用集成学习或深度学习的优势,提升模型对复杂关系的处理能力。调整参数也是优化关键。若模型出现过拟合,可通过正则化技术对参数进行约束,减少模型复杂度;对于欠拟合情况,适当增加神经网络的层数、神经元数量,或调整随机森林中树的数量等参数,增强模型对数据特征的学习能力。优化数据处理流程同样有效。若验证发现数据中存在较多干扰信息,可重新审视数据清洗步骤,加强对异常值和缺失值的处理;当特征对模型性能影响不显著时,采用特征选择方法去除冗余特征,或通过特征工程创造新的有效特征,提升数据质量,为模型训练提供更优质的数据输入,进而提高模型预测效果。

结束语

基于大数据分析的特种设备事故风险预测模型展现了良好的应用前景。通过整合多源数据与先进算法,模型能够动态识别潜在风险,为预防性维护提供依据。未来研究可进一步优化数据采集精度,探索深度学习在复杂工况下的适应性,提升模型的泛化能力与实时性。特种设备风险预测技术的持续改进,将推动安全管理向智能化方向发展。

参考文献

- [1]李翔,王鹏.大数据技术在特种设备安全监管中的应用[J].中国特种设备安全,2021,37(02):67-70.
- [2]涂震生.大数据技术在特种设备检验中的应用研究[J].中国质量监管,2024(9):102-103.
- [3]张梦生.特种设备检验中大数据技术的应用探讨[J].中国设备工程,2024(21):200-202.