

# 基于大数据分析的泵站机组运行状态监测与故障预报技术研究

冯润文<sup>1</sup> 何首炜<sup>2</sup> 于树利<sup>3</sup>

1 甘肃省景泰川电力提灌水资源利用中心, 甘肃省白银市景泰县, 730400;

2 甘肃省水利厅石羊河流域水资源利用中心, 甘肃省兰州市, 730000;

3 唐山现代工控技术有限公司河北省灌区量水测控系统及仪表技术创新中心, 河北省唐山市, 063020;

**摘要:** 作为水利工程运行中的关键, 泵站在期间承包着水资源利用以及规划等重要内容, 这一部分的稳定运行也对水资源在利用中的科学调配环节起着积极影响。但是泵站机组在运行期间, 由于设备老化或者环境等问题, 会出现一定故障。故障的存在不仅会造成整个机组的停止运行, 也会对水利工程其他水资源运转环节造成阻碍, 对社会后续的经济的发展造成不同程度的影响。因此, 本文就从大数据技术的角度出发, 集中对泵站机组的运行状态监测和相关故障处理基础进行研究。

**关键词:** 大数据分析; 泵站机组; 运行状态监测; 故障预报技术

**DOI:** 10.69979/3060-8767.25.05.025

## 引言

以往在进行泵站机组和故障的判断过程中, 传统方式依赖的人工监测存在较大的局限性。因此, 随着现代化技术的进步, 大数据技术的应用价值逐渐显现。实际应用期间, 大数据技术也因自身效率高价值、数据量大的优势被社会各个领域应用。而这一技术也对泵站机组中的相关故障诊断提供了较为便捷的路径, 不仅为故障预报和状态监测提供了较为科学丰富的数据支持, 也为各项工作的开展提供了高质量高效率的应用工具。有关人员利用技术提供的数据资源就可以及时发现异常环节, 体现进行处理措施的制定, 提升泵站机组的精准维护和科学管理。

## 1 基于大数据分析的泵站机组运行状态监测系统的设计

### 1.1 数据采集与传输

在机组关键部位, 温度传感器被大量部署, 用于测量电机定子、电机轴承以及水泵轴承等位置的温度。这些部位的温度变化是反映机组运行状态的重要指标, 电机定子温度过高可能暗示电机存在过载或散热不良等问题, 而水泵轴承温度异常则可能意味着轴承磨损或润滑不足。压力传感器则安装在水泵进出口管道以及其他关键管段上, 用于监测水压。水泵进出口压力的变化可以直观反映水泵的工作性能, 压力异常波动可能是水泵内部部件损坏、管道堵塞或泄漏等故障的前兆。流量传感器被应用于测量水流量, 这对于评估泵站的供水能力

和机组运行效率至关重要。除了上述物理量传感器, 数据采集还涵盖了电气参数的监测, 通过统计设备的运行时间和启停次数, 可以合理安排设备的维护计划, 提前预防设备故障的发生。在数据传输方面, 石羊河流域泵站根据不同泵站的地理位置和现场条件, 采用了有线和无线相结合的传输方式, 以确保数据能够安全、可靠、快速地传输到数据存储层和数据分析层。

### 1.2 数据存储与管理

石羊河流域泵站机组运行数据的存储需求具有多维度特点。从数据量角度来看, 随着监测时间的持续和监测范围的扩大, 数据量呈现快速增长趋势。从数据的时效性来讲, 实时数据对于泵站的实时监控和运行调整至关重要, 需要能够快速存储和读取, 以满足实时性要求。综合考虑石羊河流域泵站机组运行数据的存储需求, 可以选择分布式数据库 Hadoop 分布式文件系统 (HDFS) 与关系型数据库 MySQL 相结合的方式进行数据存储与管理。HDFS 具有高容错性、高扩展性和适合处理大规模数据的特点, 能够很好地满足泵站运行数据量大、类型多样的存储需求。MySQL 作为一种成熟的关系型数据库, 在处理结构化数据方面具有优势, 适合存储泵站机组的设备基本信息、运行参数、故障记录等结构化数据。对于设备的型号、规格、生产厂家等基本信息, 以及每次运行的时间、流量、压力等参数, 都可以按照数据库表的结构进行有序存储, 方便进行数据的查询、统计和分析。

### 1.3 状态监测模型构建

首先,对经过数据清洗和特征提取后的数据进行主成分分析(PCA)。PCA是一种常用的降维技术,其目的是将高维数据转换为低维数据,同时尽可能保留原始数据的主要特征,以减少数据处理的复杂性和计算量。在石羊河流域泵站机组运行数据中,经过特征提取后可能会得到大量的特征参数,这些特征之间可能存在相关性,直接使用这些高维特征进行模型训练不仅会增加计算负担,还可能导致模型过拟合。通过PCA,将这些高维特征转换为一组新的相互正交的主成分。其次,在计算过程中,先计算数据的协方差矩阵,然后求解协方差矩阵的特征值和特征向量,根据特征值的大小对特征向量进行排序,选择前 $k$ 个特征向量组成变换矩阵。通过该变换矩阵对原始数据进行变换,得到降维后的主成分数据。在实际应用中,根据累计贡献率来确定主成分的个数 $k$ 。一般认为,当累计贡献率达到85%以上时,所选择的主成分就能够较好地代表原始数据的主要信息。假设经过计算,选择前5个主成分,其累计贡献率达到了90%,则这5个主成分就可以作为后续模型训练的输入特征,这样就将高维的原始特征数据降低到了5维,显著的减少了数据的维度。

## 2 泵站机组故障预测建模的算法筛选及其优化过程探究

### 2.1 算法选取与技术改进

在故障预测研究领域展开的是基于海量数据处理的算法选取工作。具有较高适用性的支持向量机模型被优先考虑,通过交叉验证法验证其有效性。由可见实例表明,传统BP神经网络存在收敛速度慢的缺陷,而经遗传算法优化的SVM模型展现出明显优势。关键参数的确定过程中采用的是网格搜索结合粒子群算法的混合策略。研究显示,当惩罚因子 $C$ 处于 $10^{-2}$ 量级、核函数参数 $\gamma$ 取值0.01时,分类准确率达到92.7%这一峰值。并且模型结构方面实施了三阶段优化。首先,特征选择环节引入的是基于互信息的动态权重分配机制;其次,训练阶段采用的是自适应学习率调整策略;最终模型集成环节运用的则是Bagging框架下的多模型投票方法。由此可见,经过上述技术改进后的复合模型,其预测精度较基线版本提升了15.6个百分点。并且在机器学习算法体系中,神经网络以其显著的非线性映射特性与自主学习能力而著称。通过大量数据的自动学习过程,复杂模式与规律得以被该算法所捕获,这使其在故障预测研究领域占据重要位置。特别针对泵站机组故障预测问题,历史运行数据与故障记录的深度分析被神经网络

所执行,由此构建起输入特征与输出故障类别或发生概率之间的复杂关联模型。考虑到泵站机组运行数据具有明显的时间序列特征,以及各类故障间存在的非线性关联关系,长短期记忆网络(LSTM)这一特殊结构被研究组织所采用。

### 2.2 训练过程与验证工作

为实现泵站机组故障预测模型基于大数据分析的有效性提升,石羊河流域历史运行数据被采用于模型训练与验证过程。异常数据的识别与处理通过合理阈值的设定得以完成,这是数据预处理阶段的关键步骤。训练集与测试集的划分比例为7:3,该比例经反复实验验证确认为最优选择。期间足够的学习样本由训练集提供,测试集的代表性则通过30%的数据占比才能实现保障。神经网络架构的构建过程中,支持向量机模型与多种优化算法被同步测试,参数调整方面,较多应用的是随机梯度下降法及Adam算法,后续的实践表明,Adam算法在石羊河流域数据的处理中展现出更优性能,这里要注意的是,隐藏层数量与神经元个数的确定需经过多次调试实验。三层隐藏层的网络结构最终被采用,神经元配置分别为64、32、16个。由此可见,此种结构既能维持预测精度又可防止过拟合现象的发生。

## 3 案例介绍与应用

### 3.1 研究背景与学术价值

在国内水资源调配体系中,泵站作为核心枢纽设施,其运行状态直接关乎区域供水安全与生态平衡。甘肃省水利厅石羊河流域水资源利用中心肩负着维系民勤绿洲生态、保障区域农业灌溉的重任,其管辖泵站长期处于高负荷运转状态,设备老化、故障频发等问题日益凸显。传统运维模式依赖人工巡检与经验判断,难以应对复杂工况下的潜在风险。在此背景下,基于大数据分析的泵站机组运行状态监测与故障预报技术应运而生,通过构建多源数据融合分析模型,实现设备健康状态的精准评估与可预测预警,为西北干旱地区水资源高效利用提供技术支撑。

### 3.2 技术应用案例 1: 2023 年 7 月民勤调水工程预警

2023年夏天,民勤调水工程进入用水高峰期,泵站机组持续满负荷运行。7月某日凌晨3时,大数据监测系统的振动传感器捕捉到某关键泵站机组前轴承异常波动。具体异常中发现,在3-5kHz高频振动频段,能量值较正常水平突增15dB。这一细微变化立即触发了系统的三级预警机制,该频段对应的叶轮气蚀特征识别算法迅速启动,结合近3年同工况下2000余组振动数据,

通过机器学习模型进行模式匹配与风险推演。系统诊断模型基于深度神经网络算法,综合考虑流量、压力、温度等 12 参数个关联,最终以 87.6%的置信度判定为叶轮气蚀初期。基于这一现象,技术团队立即启动应急预案,停运故障机组并组织专业检修。拆解后发现,叶轮表面已出现直径约 5mm 的气蚀坑,若继续运行,预计 48 小时内将导致叶片断裂,直接威胁整个调水线路的安全运行。经核算,此次精准预警成功避免了 200 万元的设备更换成本及因停水造成的生态补水损失,更重要的是保障了民勤绿洲生态恢复工程的连续性。

### 3.3 技术应用案例 2: 设备的冬季运行优化

在地区中,由于西北冬季极端低温的出现,对泵站设备运行构成一定挑战。通过对 2018-2022 年冬季运行数据的时序分析,研究团队发现,当环境温度低于 $-15^{\circ}\text{C}$ 时,电机启动电流超标概率骤增 4.2 倍,主要原因为润滑油粘度上升、绕组绝缘性能下降导致启动阻力增加。为破解这一难题,技术团队构建了温度-电流耦合预测模型,模拟不同预加热方案对启动性能的影响。最终在 2023 年冬季,试点泵站加装智能预加热装置,该装置可根据环境温度自动调节预热功率,在启动前将电机温度提升至 $5^{\circ}\text{C}$ 。运行数据显示,改造后机组故障率由 18%降至 6.8%,下降幅度达 62%,同时启动电流波动范围缩小 35%,延长了电机使用寿命。更值得关注的是,该优化方案使冬季泵站整体能耗降低 12%,在前期基础上进一步实现了安全性与经济性的双重提升。

### 3.4 经济效益与学术贡献

技术应用带来经济与社会效益较为显著,在期间,平均故障间隔时间(MTBF)从 2021 年的 2158 小时跃升至 2023 年的 3742 小时,设备可靠性提升 73.4%;年度维修成本从 340 万元降至 127 万元,降幅达 62.6%。此外,精准的故障预警减少非计划停机时间 43%,保障了民勤绿洲每年 1.2 亿立方米的生态补水量。在学术研究层面,团队围绕大数据驱动故障诊断算法,在某期刊发表两篇高影响力论文。其中《基于多尺度特征融合的泵站振动故障诊断方法》提出改进型卷积神经网络模型,将故障识别准确率提升至 95.2%;《极端环境下泵站设备健康状态预测模型构建》创新性融合气象数据与设备参数,为高寒地区泵站运维提供理论依据。实践成果方面,结合石羊河流域应用经验,编制完成《西北高扬程泵站大数据监测技术规范》(草案),涵盖数据采集标准、分析流程、预警阈值设定等 12 项核心内容,填补了西北干旱地区泵站智能化运维的标准空白,为行业技

术推广奠定基础。

## 4 结语

综上所述,通过本研究可以看出,借助大数据技术强大的数据采集、存储和处理能力,可以构建高质量的泵站机组运行状态监测系统,实现了对机组多维度运行数据的实时精准捕捉。并运用先进的数据分析算法,挖掘数据的在后续的发展趋势和变化特征,为故障预报提供了科学的依据。但本研究仍存在一定的局限性,还需要在更多不同工况和环境条件下进行验证。对此,在未来也将继续深化大数据技术在泵站机组运行监测与故障预报领域的应用,完善数据分析方法和故障预报模型,为泵站机组的稳定运行提供更加高效的技术支撑。

## 参考文献

- [1] 郭新帅. 泵站机组远程状态监测与故障诊断系统研发[D]. 河南工业大学, 2024.
- [2] 冯宛露. 泵站数据平台和机组振动故障分析模型研究与开发[D]. 扬州大学, 2023.
- [3] 闫静. 考虑抽水装置特性差异的多机组泵站优化调控研究[D]. 江苏大学, 2023.
- [4] 宋长松, 彭恒义, 王齐领, 等. 阜阳泵站机组状态智能在线监测系统的设计与应用[J]. 现代制造技术与装备, 2022, 58(10): 53-57.
- [5] 王志远, 翟光耀, 朱劲木, 等. 鲤鱼洲泵站水泵机组大范围调速运行及出水池出水状态分析[J]. 水利水电技术(中英文), 2022, 53(S1): 330-335.
- [6] 沈雪梅, 张磊, 朱晨亮. 泵站机组运行状态监测系统在通吕运河水利枢纽水泵机组中的应用[J]. 治淮, 2021(01): 35-38.
- [7] 张印, 陶东, 邹红美. 泉眼山泵站机组在线监测系统研究与应用[J]. 中国水能及电气化, 2020, (12): 30-33.
- [8] 李涛, 汪尚红, 刘藐. 金口泵站水泵机组运行状态在线监测分析系统的应用[J]. 南方农机, 2019, 50(09): 92-93.
- [9] 冯雪峰, 周嘉言, 郑源. 大型泵站设备远程状态监测系统的研究与开发[J]. 水利建设与管理, 2019, 39(01): 11-16.

项目类型: 2025 年甘肃省水利科学试验研究及技术推广项目; 项目名称: 泵站机组故障预警预报系统研发项目编号: 25GSLK056