

基于机器学习的核电厂设备故障预测模型研究

杜岩松 钟 骏 范志岚 童 超

浙江省中核浙能能源有限公司 浙江 象山 315700

摘 要：本文深入研究了基于机器学习的核电厂设备故障预测模型。概述了核电厂设备构成及重要性，介绍了故障预测技术的理论基础，涵盖物理模型、数据驱动和知识方法。详细阐述了数据采集、特征分析、预处理等关键环节，及故障预测模型设计流程，包括处理、特征工程、模型选择与构建、训练与优化。以主冷却剂泵为例，展示了模型的实际应用。本研究为核电厂设备故障预测提供有效方法，有助于提升核电厂安全运行水平和经济效益。

关键词：机器学习；核电厂设备；故障预测模型

1 核电厂设备故障预测相关理论与技术基础

1.1 核电厂设备概述

核电厂作为高效、低碳的能源供应基地，其稳定运行对于保障国家能源安全和促进经济发展具有重要意义。核电厂内部设备复杂多样，包括反应堆系统、蒸汽与动力系统、冷却系统、电气系统等多个关键部分，每一部分都承载着特定的功能，共同协作以维持核电厂的正常运行。反应堆系统是核电厂的心脏，负责将核能转化为热能。其中，燃料组件、控制棒、反应堆压力容器等核心部件的可靠性和安全性至关重要。蒸汽与动力系统则利用反应堆产生的热能加热蒸汽发生器，产生蒸汽推动汽轮机旋转，进而驱动发电机发电。冷却系统则负责将反应堆产生的余热导出，确保反应堆在安全温度下运行，包括主冷却系统和辅助冷却系统。电气系统则负责电能的传输、分配和控制，确保电力输出的稳定与高效。因此，对核电厂设备进行定期检测、维护和故障预测，及时发现并处理潜在问题，是确保核电厂安全运行的重要手段。

1.2 故障预测技术基础

故障预测技术是通过分析设备运行数据，利用数学模型和算法预测设备未来可能出现的故障，从而提前采取措施进行干预，避免故障发生或降低故障影响的技术。在核电厂设备故障预测中，常用的技术包括基于物理模型的预测方法、基于数据驱动的预测方法和基于知识的预测方法。基于物理模型的预测方法依赖于对设备物理特性和运行机理的深入理解，通过建立精确的数学模型来预测设备故障。这种方法在理论上具有较高的准确性，但需要对设备内部结构和运行过程有深入的了解，且模型建立过程复杂，计算量大；基于数据驱动的预测方法则主要依赖于设备运行数据的收集和分析。通过采集设备的运行参数、振动信号、温度等信息，利用

机器学习、数据挖掘等技术对数据进行处理和分析，提取故障特征，建立故障预测模型^[1]。这种方法具有适应性强、预测速度快等优点，但数据的质量和完整性对预测结果有较大影响；基于知识的预测方法则结合了专家经验和领域知识，通过建立故障知识库和推理机制，对设备运行数据进行解释和推理，从而预测设备故障。这种方法能够充分利用领域专家的经验知识，但在知识获取和更新方面存在挑战。

1.3 机器学习基础理论

机器学习是人工智能的一个重要分支，它利用统计学、概率论、优化理论等数学工具，通过让计算机从数据中学习规律，从而实现了对未知数据的预测和分类。在核电厂设备故障预测中，机器学习技术具有广泛的应用前景。机器学习算法可以分为监督学习、无监督学习和强化学习三大类。监督学习是指在训练过程中，算法通过已知的数据集（包括输入和输出）进行学习，以建立输入与输出之间的映射关系。在核电厂设备故障预测中，可以利用历史故障数据和正常运行数据，训练监督学习模型，实现对未来故障的预测；无监督学习则是在没有标签的情况下，通过发现数据中的隐藏结构和模式来进行学习。在核电厂设备故障预测中，无监督学习可以用于异常检测，即识别出与正常数据模式显著不同的数据点，这些异常点可能预示着设备故障的发生；强化学习则是一种通过与环境进行交互来学习最优策略的方法。在核电厂设备故障预测中，强化学习可以用于优化设备的维护策略，即在保证设备安全的前提下，通过调整维护计划来降低维护成本和提高设备利用率。

2 核电厂设备运行数据处理

2.1 数据采集与来源

核电厂设备运行数据的采集主要来源于两个方面：一是直接监测设备的关键参数；二是通过核电厂的信息

管理系统获取历史运行记录。直接监测设备的关键参数通常涉及传感器技术的应用。核电厂中的各类设备，如反应堆、蒸汽发生器、冷却系统等，都配备有大量的传感器，用于实时监测设备的温度、压力、流量、振动等关键参数。这些传感器数据通过数据采集系统实时上传至中央数据库，为后续的故障预测提供基础数据；另一方面，核电厂的信息管理系统也是数据的重要来源。这些系统记录了核电厂设备的运行历史、维护记录、故障报告等信息。这些信息虽然不如传感器数据实时，但提供了设备长期运行的全面视角，有助于理解设备的退化趋势和故障模式。在数据采集过程中，要注意数据的准确性和完整性。传感器可能存在误差或故障，导致数据不准确；信息管理系统中的数据也可能因为记录错误或遗漏而不完整。在数据采集阶段，需要建立严格的数据质量控制机制，确保数据的准确性和完整性^[2]。

2.2 数据特征分析

在核电厂设备运行数据中，常见的特征包括时域特征、频域特征和统计特征。时域特征描述了数据随时间变化的趋势和规律，如均值、方差、峰值等；频域特征则通过分析数据的频谱，揭示数据在不同频率上的分布特性；统计特征则利用统计学方法对数据进行分析，提取出数据的分布特性、相关性等信息；除了这些基本特征外，还可以根据设备的特性和故障模式，提取出更具针对性的特征。例如，对于振动数据，可以提取频谱特征、包络谱特征等，用于识别设备的振动故障；对于温度数据，可以提取温度变化趋势、温度波动范围等特征，用于评估设备的热状态。在特征分析过程中，需要注意特征的冗余性和相关性。过多的冗余特征会增加模型的复杂度，降低预测性能；而高度相关的特征则可能导致模型训练过程中的过拟合问题。在特征选择时，需要综合考虑特征的有效性、冗余性和相关性，筛选出对故障预测最有用的特征。

2.3 数据预处理

通过对原始数据进行清洗、转换和归约等操作，可以提高数据的质量，降低模型的复杂度，从而提高预测的准确性和效率。数据清洗是数据预处理的第一步，主要目的是去除数据中的噪声、异常值和缺失值。噪声通常是由于传感器误差或数据传输错误导致的，可以通过滤波、平滑等方法进行处理；异常值则可能是由于设备故障或数据采集错误导致的，需要通过统计方法或领域知识进行识别和剔除；缺失值则可能由于传感器故障或数据采集过程中的遗漏导致，需要通过插值、填补等方法进行补充；数据转换是将原始数据转换为更适合模型

训练的形式。例如，对于非线性数据，可以通过对数变换、幂变换等方法进行线性化处理；对于类别型数据，可以通过独热编码、标签编码等方法进行数值化处理。另外，还可以通过特征缩放、标准化等方法对数据进行归一化处理，以提高模型的收敛速度和预测性能；数据归约则是通过降维技术减少数据的维度，降低模型的复杂度。常用的降维技术包括主成分分析（PCA）、线性判别分析（LDA）等。通过降维，可以去除数据中的冗余信息，保留对故障预测最有用的特征，从而提高模型的泛化能力。

3 核电厂设备故障预测模型设计

3.1 数据采集与处理

在核电厂设备故障预测模型的设计过程中，数据采集与处理是首要步骤，数据采集主要聚焦于核电厂关键设备的实时监测数据与历史运行记录；实时监测数据通过安装在设备上的各类传感器获取，包括温度、压力、振动、流量等关键参数，这些数据能够实时反映设备的运行状态。而历史运行记录则涵盖了设备过去的维护日志、故障报告及操作记录，为理解设备的长期运行趋势和故障模式提供了宝贵信息；数据处理阶段则涉及数据清洗、整合与格式化。数据清洗旨在剔除异常值、填补缺失值，并纠正数据中的错误，以确保数据的准确性和一致性。数据整合则是将来自不同来源的数据进行合并，形成一个全面的数据集。而数据格式化则是将数据转换为适合模型训练的格式，如时间序列数据、分类标签等。

3.2 特征工程

特征工程是构建高效故障预测模型的关键步骤，在核电厂设备故障预测中，特征工程可能包括时域特征提取、频域特征分析、统计特征计算等。时域特征如均值、方差、峰值等能够描述数据随时间变化的趋势；频域特征则通过频谱分析揭示数据在不同频率上的分布特性；统计特征则利用统计学方法评估数据的分布、相关性等特性。另外，特征选择也是特征工程的重要组成部分。通过筛选最具预测价值的特征，可以降低模型的复杂度，提高预测性能。特征选择方法可能包括基于统计检验、相关性分析、机器学习算法等^[3]。

3.3 模型选择与构建

在核电厂设备故障预测中，模型的选择与构建直接决定了预测的准确性和效率。根据数据的特性和预测目标，可以选择不同类型的机器学习算法进行建模。对于分类问题，如预测设备是否发生故障，可以选择支持向量机（SVM）、决策树、随机森林等算法。这些算法能

够基于历史数据学习设备的故障模式，并用于预测未来设备的运行状态；对于回归问题，如预测设备的剩余使用寿命（RUL），则可以选择线性回归、岭回归、支持向量回归（SVR）等算法。这些算法能够基于设备的退化趋势和历史故障数据，预测设备在未来某个时间点的性能状态；在模型构建过程中，还需要考虑模型的结构和参数设置。例如，对于神经网络模型，需要确定网络的层数、神经元数量、激活函数等；对于支持向量机，则需要选择合适的核函数和正则化参数。

3.4 模型训练与优化

在训练阶段，需要使用处理好的数据集对模型进行训练，使其能够学习到数据的内在规律和故障模式。训练过程中，选择合适的损失函数和优化算法，损失函数用于衡量模型预测结果与真实结果之间的差异，如均方误差（MSE）、交叉熵损失等。优化算法则用于调整模型的参数，以最小化损失函数。常用的优化算法包括梯度下降、随机梯度下降、Adam等。模型优化则涉及多个方面，包括特征选择、参数调优、模型集成等。通过优化特征选择，可以提高模型的预测性能；通过参数调优，可以找到最佳的模型参数设置；通过模型集成，如使用袋装法、提升法等，可以进一步提高模型的准确性和鲁棒性。另外，还需要对模型进行交叉验证和性能测试，以确保其在未见过的数据上也能表现出良好的预测性能。交叉验证是一种常用的模型评估方法，它通过将数据集划分为训练集和测试集，多次训练并测试模型，以评估模型的泛化能力。性能测试则是在独立的测试集上评估模型的预测准确性、鲁棒性等指标。

4 案例分析

4.1 案例核电站设备介绍

本案例选取某核电站的主冷却剂泵作为分析对象。主冷却剂泵是核电站反应堆冷却系统中的关键设备，负责将冷却剂在反应堆堆芯与蒸汽发生器之间循环，确保反应堆的安全运行。该泵由电机驱动，具有高转速、大功率的特点，其运行状态直接影响到反应堆的冷却效率和安全性。主冷却剂泵在运行过程中可能面临多种故障模式，包括但不限于轴承磨损、密封失效、振动过大等。这些故障不仅可能导致泵的性能下降，还可能引发更严重的安全事故^[4]。对主冷却剂泵进行故障预测，及时

发现并处理潜在故障，对于保障核电站的安全运行具有重要意义。

4.2 基于机器学习的故障预测模型应用实践

针对主冷却剂泵的故障预测需求，采用基于机器学习的故障预测模型进行应用实践。在数据采集与处理阶段，收集主冷却剂泵过去一年的实时监测数据，包括振动信号、温度数据、压力数据和流量数据等。通过对数据进行清洗、整合和格式化处理，得到一个高质量的数据集，用于后续的模式训练；在特征工程阶段，从原始数据中提取多种特征，包括时域特征（如均值、方差、峰值等）、频域特征（如频谱能量、主频等）和统计特征（如相关性、分布特性等）。通过特征选择方法，筛选出对故障预测最有价值的特征子集，用于构建故障预测模型；在模型选择与构建阶段，选择支持向量机（SVM）作为分类算法，用于预测主冷却剂泵是否发生故障。同时，还采用了随机森林算法进行特征重要性评估，以确保所选特征的有效性；在模型训练与优化阶段，使用交叉验证方法对模型进行了训练和测试。通过调整模型的参数设置，如核函数类型、正则化参数等，优化模型的预测性能。最终，得到一个具有较高准确性和鲁棒性的故障预测模型。

结束语

综上所述，基于机器学习的核电站设备故障预测模型在提高设备安全性、降低维护成本方面具有重要意义。通过对核电站设备运行数据的深入分析，结合先进的机器学习算法，可以实现对设备故障的准确预测和及时干预。未来，随着机器学习技术的不断发展和数据资源的日益丰富，核电站设备故障预测模型将更加智能化和精准化。

参考文献

- [1]姚亚敏,金兰.基于机器学习的工程成本预测模型研究[J].土木工程信息技术,2020,8(2):12-16.
- [2]赵芳,司马飞.基于深度学习的工程成本预测研究[J].工程建设,2021,33(4):56-61.
- [3]范国栋,李博涵.基于机器学习的工业机械设备故障预测方法[J].自动化与信息工程,2023,44(04):13-18+50.
- [4]陈彩珠.工业机械设备的安全使用及保养措施研究[J].造纸装备及材料,2021,50(08):28-29.