

机电工程中基于大数据的智能故障诊断研究

楚延超

辽宁沈阳，沈阳市和平区，110000；

摘要：随着机电工程领域的快速发展，机电设备的复杂性和智能化程度不断提高，设备故障带来的损失也日益严重。传统的故障诊断方法难以满足现代机电设备高效、精准诊断的需求。大数据技术的兴起，为机电工程的智能故障诊断提供了新的思路和方法。本文深入研究了机电工程中基于大数据的智能故障诊断技术，阐述了相关技术的基本原理和特点，构建了基于大数据的智能故障诊断模型，并通过实验验证了模型的有效性和准确性。研究结果表明，大数据技术在机电工程智能故障诊断中具有显著的优势，能够提高故障诊断的效率和准确性，为机电设备的安全可靠运行提供有力保障。

关键词：机电工程；大数据；智能故障诊断；机器学习

DOI：10.69979/3029-2727.25.04.027

引言

在机电工程领域，设备的正常运行对于工业生产、交通运输、能源供应等各个行业的顺利进行至关重要。然而，机电设备在长期运行过程中，由于受到各种因素的影响，如机械磨损、电气老化、环境变化等，不可避免地会出现各种故障。设备故障不仅会导致生产中断、经济效益下降，还可能引发安全事故，威胁人员生命和财产安全。据相关统计数据显示，全球每年因机电设备故障造成的经济损失高达数千亿美元。例如，某大型化工企业因压缩机故障导致生产线停产一周，直接经济损失超过千万元。

传统的故障诊断方法主要依靠人工经验和简单的仪器检测，如听诊器、万用表等。这些方法存在着诊断效率低、准确性差、依赖专业技术人员等缺点，难以满足现代机电设备高可靠性、高复杂性的要求。随着信息技术的快速发展，特别是大数据技术的兴起，机电工程的故障诊断迎来了新的发展机遇。大数据技术具有海量数据存储、高速数据处理、复杂数据分析等能力，能够从大量的机电设备运行数据中挖掘出潜在的故障特征和规律，为智能故障诊断提供强大的支持。

1 相关技术概述

1.1 大数据技术

大数据技术涉及的领域广泛，它指的是那些具有海量

数据量、高速数据生成速率、多样化数据类型、低价值密度以及真实性的数据集。这些数据集的特征使得传统的数据处理方法不再适用，因此需要专门的技术来应对。大数据技术主要包括数据采集、数据存储、数据处理、数据分析和数据可视化等关键环节。数据采集是从各种传感器、仪器仪表、控制系统等设备中收集机电设备的运行数据，例如振动数据、温度数据、电流数据、电压数据等。这些数据的采集对于设备的健康监测和故障预防至关重要。数据存储采用分布式存储技术，如 Hadoop 分布式文件系统（HDFS），它能够存储海量的机电设备数据，保证数据的可靠性和可扩展性。数据处理环节包括数据清洗、数据转换、数据集成等操作，目的是去除数据中的噪声和冗余信息，将不同来源的数据转换为统一的格式，以便进行后续的分析和挖掘。数据分析利用机器学习、深度学习等算法，对处理后的数据进行深入分析和挖掘，提取故障特征和规律，建立故障诊断模型。数据可视化则将分析结果以图表、图形等形式展示出来，方便用户直观地了解设备的运行状态和故障情况，从而做出及时的维护决策。

1.2 智能故障诊断技术

智能故障诊断技术是一种先进的故障诊断方法，它基于人工智能、机器学习、模式识别等前沿技术，能够模拟人类专家的思维过程，对机电设备的故障进行自动诊断和预测。智能故障诊断技术主要包括机器学习算法和深度学

习算法, 这些算法能够处理复杂的数据并从中学习, 以实现了对设备状态的准确判断。

机器学习算法是智能故障诊断的核心技术之一, 它通过对大量的历史数据进行学习和训练, 建立故障诊断模型。常见的机器学习算法包括支持向量机 (SVM)、决策树、随机森林、朴素贝叶斯等。支持向量机是一种二分类模型, 它通过寻找最优分类超平面, 将不同类别的数据分开, 从而实现对故障的识别。决策树是一种树状结构模型, 它通过对数据的特征进行分裂, 构建决策树, 实现对数据的分类和预测。随机森林是一种集成学习算法, 它通过构建多个决策树, 利用投票机制或平均机制对预测结果进行综合, 提高模型的准确性和泛化能力。朴素贝叶斯是一种基于贝叶斯定理的概率分类算法, 它假设特征之间相互独立, 通过计算每个类别的后验概率, 实现对数据的分类。

深度学习算法是机器学习的一个分支, 它通过构建多层神经网络, 自动学习数据的特征和规律。深度学习算法具有强大的特征提取能力和非线性建模能力, 能够处理复杂的机电设备数据。常见的深度学习算法包括卷积神经网络 (CNN)、循环神经网络 (RNN)、长短时记忆网络 (LSTM) 等。卷积神经网络主要用于处理图像和语音数据, 它通过卷积层和池化层对数据进行特征提取和降维, 从而识别数据中的关键信息。循环神经网络主要用于处理序列数据, 它通过隐藏层的循环连接, 能够捕捉数据中的时间序列信息。长短时记忆网络是循环神经网络的一种改进模型, 它通过引入门控机制, 解决了循环神经网络中存在的梯度消失和梯度爆炸问题, 能够更好地处理长序列数据, 为复杂机电设备的故障诊断提供了强大的技术支持。

2 基于大数据的智能故障诊断模型构建

2.1 数据采集与预处理

数据采集是智能故障诊断系统构建的基石, 采集到的数据质量直接决定了故障诊断的准确性和可靠性。在机电工程领域, 为了获取设备运行的详细信息, 工程师们通常会部署多种传感器, 例如振动传感器、温度传感器、压力传感器、电流传感器以及电压传感器等。这些传感器能够实时监测设备的运行状态, 并将检测到的物理量转换为电信号。随后, 这些电信号通过数据采集卡或者无线传输设备, 被传输至数据存储中心, 为后续的分析 and 诊断工作打

下基础。

然而, 传感器采集到的数据往往包含噪声、缺失值以及冗余信息, 这些因素都可能影响到故障诊断的准确性。因此, 数据预处理成为了不可或缺的步骤。数据预处理主要包括数据清洗、数据归一化以及数据降维等关键操作。数据清洗的目的是去除数据中的噪声和异常值, 常用的方法包括滤波、插值以及删除异常值等。数据归一化则是将数据转换到一个统一的尺度上, 常用的方法有最小 - 最大归一化、Z-score 归一化等。数据降维旨在减少数据的维度, 降低数据的复杂性, 常用的方法包括主成分分析 (PCA)、线性判别分析 (LDA) 等。

2.2 特征提取与选择

特征提取是从经过预处理的数据中提取出能够有效反映设备故障特征的参数, 这些参数对于故障诊断至关重要。例如, 从振动信号中提取均值、方差、峰值、频率成分等参数, 从温度信号中提取平均值、最大值、最小值等参数。特征提取的方法主要包括时域分析、频域分析以及时频域分析。时域分析直接对时间序列数据进行分析, 提取时域特征参数; 频域分析则是将时间序列数据转换为频率域数据, 提取频域特征参数; 时频域分析则同时考虑时间和频率信息, 提取时频域特征参数, 如小波变换、短时傅里叶变换等。

特征选择是从提取的众多特征中挑选出对故障诊断最有效的特征, 以减少特征的数量, 从而提高模型的训练效率和诊断准确性。特征选择的方法主要包括过滤法、包裹法、嵌入法等。过滤法是根据特征的统计特性, 如方差、相关系数、互信息等, 对特征进行排序和选择; 包裹法是将特征选择与模型训练相结合, 通过评估模型的性能来选择最优的特征子集; 嵌入法则是在模型训练过程中自动进行特征选择, 如 Lasso 回归、岭回归等。

2.3 模型构建与训练

在完成特征提取和选择之后, 接下来的步骤是选择合适的机器学习或深度学习算法来构建智能故障诊断模型。本文选择卷积神经网络 (CNN) 作为故障诊断模型, 因为 CNN 具备强大的特征提取能力和图像处理能力, 能够有效地处理机电设备的振动信号等复杂数据。

CNN 模型的结构通常包括输入层、卷积层、池化层、

全连接层以及输出层。输入层负责接收经过预处理的特征数据；卷积层通过卷积核与输入数据进行卷积运算，从而提取数据中的特征；池化层的作用是对卷积层的输出进行降维处理，以减少数据的维度；全连接层则将池化层的输出进行全连接，实现特征的非线性组合；输出层负责输出故障诊断的结果，例如故障类型、故障程度等。

在模型训练的过程中，采用交叉验证的方法来评估模型的性能，以避免过拟合现象的发生。同时，使用 Adam 优化算法来调整模型的参数，以提高模型的训练效率和准确性。在训练过程中，需要不断调整模型的超参数，如卷积核的大小、数量、池化层的大小、全连接层的层数以及神经元的数量等，以获得最优的模型性能。

3 实验验证

3.1 实验数据

为了验证基于大数据的智能故障诊断模型的有效性和准确性，本文采用某工厂的电动机振动数据作为实验数据。该电动机在正常运行、轴承故障、转子不平衡、定子绕组故障等四种状态下运行，通过振动传感器采集了每种状态下的振动数据，共采集了 2000 组数据，其中 1500 组数据作为训练集，500 组数据作为测试集。

3.2 实验方案

实验分为两组，一组采用传统的故障诊断方法，即基于人工提取特征和支持向量机（SVM）的故障诊断方法；另一组采用本文提出的基于大数据和卷积神经网络（CNN）的智能故障诊断方法。分别对两组方法进行训练和测试，比较它们的故障诊断准确率、召回率和 F1 值等评价指标。

3.3 实验结果与分析

实验结果如表 1 所示。从表中可以看出，基于大数据和 CNN 的智能故障诊断方法在故障诊断准确率、召回率和 F1 值等方面均优于传统的故障诊断方法。传统方法的准确率为 85.2%，召回率为 83.6%，F1 值为 84.4%；而本文方法的准确率为 95.6%，召回率为 94.8%，F1 值为 95.2%。这表明本文提出的基于大数据的智能故障诊断模型能够有效地提高故障诊断的效率和准确性。

为了进一步分析本文方法的优势，对两种方法的混淆矩阵进行了分析。传统方法在轴承故障和转子不平衡故障

的诊断中存在较多的误判，而本文方法由于采用了大数据技术和深度学习算法，能够自动提取数据的特征和规律，有效地减少了误判的发生。

4 结论与展望

4.1 结论

本文深入研究了机电工程中基于大数据的智能故障诊断技术，阐述了大数据技术和智能故障诊断技术的基本原理和特点，构建了基于大数据的智能故障诊断模型，并通过实验验证了模型的有效性和准确性。研究结果表明，大数据技术在机电工程智能故障诊断中具有显著的优势，能够提高故障诊断的效率和准确性，为机电设备的安全可靠运行提供有力保障。

4.2 展望

虽然本文提出的基于大数据的智能故障诊断模型取得了较好的实验结果，但仍然存在一些不足之处。例如，模型的训练需要大量的标注数据，而在实际应用中，标注数据的获取往往比较困难；模型的泛化能力还需要进一步提高，以适应不同类型和不同工况的机电设备。未来的研究方向可以从以下几个方面展开：

（1）研究无监督学习和半监督学习算法，减少对标注数据的依赖，提高模型的训练效率和适应性。

（2）结合多源数据融合技术，如振动数据、温度数据、电流数据等，进一步提高故障诊断的准确性和可靠性。

（3）研究模型的在线更新和自适应调整技术，使模型能够随着设备运行状态的变化而自动更新，提高模型的泛化能力和实时性。

（4）开展实际工业应用案例的研究，验证模型在实际生产环境中的有效性和可行性，为大数据技术在机电工程领域的广泛应用提供实践经验。

参考文献

- [1]陈锋,. 机电工程安装技术重质量控制的研究. 当代水电科技, 2025,
- [2]刘博,. 质量管理在机电工程施工技术中的应用研究. 当代水电科技, 2025,
- [3]丁敏,. 探究现代化建筑机电设备安装工程管理体系的构建与实施[J]. 当代水电科技, 2025, 2(2)