



计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331, CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目：基于机器学习的设备预测性维护方法综述
作者：李杰其，胡良兵
网络首发日期：2020-09-30
引用格式：李杰其，胡良兵. 基于机器学习的设备预测性维护方法综述. 计算机工程与应用. <https://kns.cnki.net/kcms/detail/11.2127.TP.20200930.0858.002.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于机器学习的设备预测性维护方法综述

李杰其^{1,2}, 胡良兵¹

1. 中国科学院 合肥物质科学研究院, 等离子体物理研究所, 合肥 230031

2. 中国科学技术大学, 合肥 230026

摘要: 机器学习算法能够处理高维和多变量数据, 并在复杂和动态环境中提取数据中的隐藏关系, 在预测性维护技术中具有很好的应用前景。然而, 预测性维护系统的性能取决于机器学习算法的选择, 本文对目前应用与预测性维护中的机器学习算法进行综述, 详细比较了几种机器学习算法的优缺点, 并对机器学习在预测性维护研究中的应用进行了展望。

关键词: 预测性维护; 寿命预测; 机器学习; 人工神经网络; 支持向量机; 聚类算法; 随机森林

文献标识码:A 中图分类号: TP18 doi: 10.3778/j.issn.1002-8331.2006-0016

李杰其, 胡良兵. 基于机器学习的设备预测性维护方法综述. 计算机工程与应用

LI Jieqi, HU Liangbing. Review of machine learning for predictive maintenance. Computer Engineering and Applications

Review of machine learning for predictive maintenance

LI Jieqi^{1,2}, HU Liangbing¹

1. Institute of Plasma Physics Chinese Academy of Sciences, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China

2. University of Science and Technology of China, Hefei 230026, China

Abstract: Machine learning algorithms can process high-dimensional and multi-variable data, and extract hidden relationships in the data in complex and dynamic environments, and have good application prospects in predictive maintenance technology. However, the performance of predictive maintenance systems depends on the choice of machine learning algorithms. This paper reviews the current machine learning algorithms used in predictive maintenance system, compares the advantages and disadvantages of several machine learning algorithms characteristic in detail. The application of the machine learning in predictive maintenance was prospected in the future.

Key words: predictive maintenance; life prediction; machine learning; artificial neural networks; support vector machines; cluster analysis; random forests

基金项目: 安徽省自然科学基金面上项目(No.1808085ME124); 聚变堆主机关键系统综合研究设施(No.2018-000052-73-01-001228)。

作者简介: 李杰其(1994-), 男, 硕士研究生, 研究领域为低温系统、机器学习, E-mail: lijieqi@mail.ustc.edu.cn; 胡良兵(1978-), 男, 博士, 高级工程师, 研究领域为大型氦低温系统模拟控制及优化。

1 引言

在工业中,设备维护是很重要的环节,直接影响到设备的运行时间和设备的工作效率。然而,设备剩余寿命需要被准确预测,故障需要被及时识别和处理,才能避免在生产过程中停机。根据文献^[1]所述,可将维护策略分为如下三类:

(1)故障维护:一种简单的维护策略,设备发生故障时停止生产并更换零件。单一零件的突然失效会导致相关零件的故障率上升,增加生产成本;

(2)预防性维护:一种定期执行的维护策略,通过计划时间表或生产经验,以预测过程、设备故障。但是由于可能会采取不必要的维护措施,导致成本增加;

(3)预测性维护:使用预测工具对设备的剩余寿命进行预测,这种方法要求对机器状态或运行过程进行持续监控,需要历史数据处理(例如机器学习技术)、完整性因素分析(例如视觉方面、磨损、不同于原始颜色等)、统计推断方法和工程方法的推理,所以这种方法可以预测设备早期的故障并提供解决方案。

每种维护方式各有特色,选择故障维护策略会使得维护行为具有滞后性,设备发生风险性很高;选择预防性维护策略,这种预见性维护的干预会导致设备的使用时间下降。高效的维护策略可以改善设备的状况,降低设备故障率,降低维护成本,同时延长设备寿命。预测性维护是通过对设备状态实施周期性监视、预测剩余寿命、评价设备状况和产生相应逻辑的一套技术。传统的预测性维护方法主要是基于物理模型,基本思想是基于物理性质的固定的物理公式去拟合设备、零部件的健康状态或寿命。这种方法要求研究人员有很深的工程物理背景知识,通常采用复杂的物理公式,借助于过多的物理模型而忽略了数据本身的规律,预测准确度较低。

目前,伴随着 5G 技术的工业 4.0 时代到来,工业系统可以采集大量的数据,包括工业生产线的运行过程、发生的事件和报警信息等。预测性维护系统可对设备数据进行系统分析和处理,建立和训练预测模型,制定维护策略并及时调整,可以有效的降低维修成本、减少机器故障、减少维修停机、减少备件库存、增加备件寿命、增加产量、提高操作员的安全性、提升维修验证可靠性、提升整体利润等。人工智能中的机器学习已经成为许多应用中开发智能预测算法的有力工具。

2 机器学习简介

机器学习,通过自主学习大量数据中存在的规律,获得新经验和知识从而提高计算机智能,使得计算机拥有类似人类的决策能力。基于学习形式的不同通常可将机器学习算法分为监督学习、无监督学习以及强化学习三类:

(1)监督学习:给学习算法提供标记的数据和所需的输出,对于每一个输入,学习者都被提供了一个回应的目标。监督学习被用于解决分类和回归的问题。常见的算法有:决策树、人工神经网络算法、支持向量机、朴素贝叶斯、随机森林等。

(2)无监督学习:给学习算法提供的数据是未标记的,并且要求算法识别输入数据中的模式,主要是建立一个模型,对输入的数据进行解释,并用于下次输入。主要用于解决聚类 and 降维问题,常见的算法有:聚类算法、降维算法。

(3)强化学习:该算法与动态环境相互作用,把环境的反馈作为输入,通过学习选择能达到其目标的最优动作。强化学习这一方法背后的数学原理与监督、非监督学习略有差异。监督、非监督学习更多地应用了统计学,而强化学习更多地结合了离散数学、随机过程这些数学方法^[2]。常见的算法有:马尔可夫决策过程等。

作为人工智能的核心,机器学习的主要功能是使得计算机模拟或实现人类的学习行为,通过获取新的信息,不断对模型进行训练以提高模型的泛化能力^[3]。由于机器学习具有强大的数据处理能力,该方法广泛应用于数据挖掘、语音识别、计算机视觉、故障诊断与寿命预测等领域。基于机器学习的剩余寿命预测方法是预测性维护方法的一个重要分支,能够较为准确地对剩余寿命进行预测,成为维护策略评估的重要指标。

建立用于预测性维护应用的机器学习模型包括以下几个步骤:历史数据选择;数据预处理;模型选择、模型训练和模型验证;模型维护^[4]。历史数据选择,即收集和存储设备运行过程中产生的数据集,以便为机器学习模型设计选择有价值的数据,其目的是获取与系统状况相关的信息^[5]。通过数据预处理步骤,对数据进行处理和分析收集的数据,可以使系统更好地解释数据。数据预处理步骤包括数据转换(标准化)、数据清理(缺失数据处理和异常值去除)和数据缩减(维度缩减和数量缩减),以保证数据能够被机器学习模型有效地处理。模型的开发步骤包括:选择模型;训练模型;验证模型。模型维护可以随着时间的推移保持模型性能,因为工业生产可能会随着时间的推移而发生变化,导致模型性能下降。

机器学习算法在寿命预测中应用最多是基于神经网络、支持向量机、聚类算法和随机森林，并在这些算法的基础上灵活运用其他算法或模型进行优化改进，从而提高寿命预测的精度。

3 机器学习算法介绍
3.1 人工神经网络

人工神经网络是受生物神经元启发的智能计算技术^[6]，是一种模仿动物神经网络行为特征，进行分布式并行信息处理的算法数学模型。这种网络依靠系统的复杂程度，通过调整内部大量节点之间相互连接的关系，从而达到处理信息的目的，具有自动学习和总结的能力。其主要包含输入层、隐含层及输出层，常用于解决分类、回归等问题。人工神经网络模型以原始测量数据或基于原始测量数据所提取的特征为输入，通过一定的训练算法不断调整网络的结构和参数，利用优化后的网络在线预测设备的剩余寿命，并根据预测结果判断设备的维护策略。

Biswal 等人^[7]采用人工神经网络提出一种用于模拟风力涡轮机的模型，该模型能够模拟风力涡轮机关键部件的故障工况，也可以用缺陷的部件替换健康的部件，从而收集健康和故障工况下的涡轮机的振动数据，再通过人工神经网络模型预测，对设备健康状态进行预测识别并决定维护策略，实验结果准确性为 92.6%。Gebrael 等人^[8]基于 BP 神经网络建立单个轴承和批量轴承的寿命预测模型，利用寿命周期内的振动监测信息，以一定权重融合各轴承指数回归模型参数，同时进行参数在线更新，有效保证寿命预测的准确性。Bezazi 等人^[9]采用基于多层感知器的人工神经网络对复合材料结构监测数据进行建模，通过极大似然估计和贝叶斯推理对神经网络模型进行训练，结果表明该模型具有较好的泛化能力。Wang 等人^[10]提出一种极限学习机模型，避免隐含层神经元的数目选择问题，实现了非平稳序列的在线预测。Kolokas 等人^[11]使用运行期间的过程传感器数据，实时检测某工业阳极生产设备的实时故障，并将人工神经网络与其他机器学习进行了比较。Sheng 等人^[12]针对齿轮寿命预测问题，提出一种基于权值放大的长短时记忆神经网络，并加入一种将隐含层的输入权值和递归权值进行不同程度放大的注意机制，结果表明预测方法具有更高的精度。更多基于人工神经网络的预测性维护研究的最新论文见表 1。

表 1 人工神经网络在预测性维护中的最新应用

文献	年份	内容
[13]	2019	针对涡扇发动机的预测性维护问题，创新性的提出一种将长短时记忆 (LSTM) 和卷积神经网络 (CNN) 相结合的有向无环图 (DAG) 网络预测模型。通过退化机理和分段剩余寿命函数代替传统的线性函数，可以对发动机的运行状态进行有效分级。
[14]	2019	针对机床早期故障检测的问题，提出一种时变条件下的深度学习模型，该方法不受时变条件的影响，具有较好的早期检测能力。
[15]	2019	为研究轴承剩余寿命计算方法，创新的提出一种基于时频分析法和多尺度卷积神经网络的深度特征学习方法。时频分析法能有效地揭示轴承退化信号的非平稳特性，该方法在预测精度上有明显的提高。
[16]	2018	针对锂离子电池剩余寿命预测问题，提出一种将自编码器与深度神经网络相结合的深度学习的预测模型，该模型能够提高剩余寿命预测的精度。
[17]	2018	针对一种旋转部件的剩余寿命预测问题，利用齿轮试验台采集的数据和轴承从运行到失效的试验数据，提出一种基于深度神经网络算法的模型，可以对系统的退化状态进行有效评估。
[18]	2018	针对锂离子电池剩余寿命预测问题，创新的提出一种基于 LSTM-RNN 算法的预测模型，该方法可以独立于离线训练数据进行预测，可以比传统方法提前预测电池的剩余寿命。

基于人工神经网络的预测性维护技术的主要优势包括：训练模型基于历史数据，不需要专家知识决策；人工神经网络的稳定性较强，数据的少量波动不会影响模型。然而，人工神经网络有一些缺点：训练模型有概率获得有悖于维护策略的结论；为获得最合适的人工神经网络模型，可能会花费很多时间；人工神经网络是一种“黑箱”算法，无法对神经网络模型输出的预测结果进行解释；人工神经网络需要大量的数据集才能正确学习。

3.2 支持向量机

支持向量机是执行分类和回归任务的另一种广泛使用的机器学习方法，具有高精度^[19]。它是一类按监督学习方式对数据进行二元分类的广义线性分类器，其决策边界是对学习样本求解的最大边距超平面，具有稀疏性和稳健性。支持向量机可以通过核方法进行非线性分类，是常见的核学习方法之一。支持向量机根据用途

可分为支持向量回归机 SVR 和支持向量分类机 SVC。

Susto 等人^[20]提出基于支持向量机的多类分类器，可以识别由于累积效应和违规使用而在机器上发生的故障。Nieto 等人^[21]提出一种基于混合粒子群优化算法和支持向量机算法的模型，用于预测航天器发动机的工况，解决了支持向量机训练过程中超参数的优化问题，提高了设备工况预测精度。Maior 等人^[22]提出一种将经验模态分解与支持向量机相结合的方法，可以实现退化数据分析和剩余寿命预测，结果表明该方法能够改善电动机的剩余寿命预测性能。Mathew 等人^[23]使用改进回归核的支持回归向量的支持向量机来预测剩余寿命，通过一组模拟时间序列对算法进行测试，其结果表明所提出的支持向量回归模型优于标准支持向量回归模型。Song 等人^[24]采用支持向量机对电池容量进行预测，利用卡尔曼滤波对输出信号进行优化，该方法减少了系统噪声和测量不确定性因素的影响，实现了对电池剩余寿命的长期预测。Lasisi 等人^[25]比较了支持向量机、随机森林和三层贝叶斯概率模型来检测铁路轨道的缺陷。戴邵武等人^[26]提出基于时域特征和支持向量机的滚动轴承退化趋势预测方法，准确地预测了滚动轴承全寿命故障周期。王瀛洲等人^[27]提出一种基于蚁狮优化和支持向量回归的方法，利用蚁狮优化算法优化核参数，有效提高锂离子电池剩余使用寿命预测的准确性和鲁棒性。Liu 等人^[28]利用支持向量机建立了橡胶材料疲劳寿命预测模型，提出一种重力搜索算法来优化支持向量机的参数，提升了预测精度和预测效率。更多基于支持向量机的预测性维护研究的最新论文见表 2。

表 2 支持向量机在预测性维护中的最新应用

文献	年份	内容
[29]	2020	针对航空电池的剩余寿命预测问题，创新的提出一种将遗传算法应用于 LS-SVM 的参数优化和 BP 神经网络中初始权值选择的预测模型。改进后的 LS-SVM 算法实现简单、预测精度高、运算速度快。
[30]	2019	针对某型聚苯乙烯丁二烯橡胶绝缘电缆的剩余寿命预测问题，创新的提出一种基于支持向量回归 (SVR) 算法，通过油雾温度和油雾浓度为特征特征，降低噪声对预测的干扰，提升剩余寿命预测准确率。
[31]	2019	针对轴承的剩余寿命问题，提出一种基于相关特征和支持向量机的预测模型。利用主成分分析 (PCA) 得到了反映轴承运行状态和退化趋势的相对特征，消除特征之间的冗余性和相关性，再将相关特征输入到支持向量机中得到预测模型。

[32]	2018	针对水利发电厂的预测性维护问题，提出采用分类学习器应用程序来训练模型，该应用程序支持 22 种分类器类型，可分为决策树、判别分析、支持向量机、逻辑回归、k 近邻和集成分类等六大分类算法，并筛选出最佳控制模型。为最优模型选择和获得全局最优解的问题提供了解决思路。
[33]	2018	针对应用金属磁记忆对螺栓残余疲劳寿命进行预测的问题，提出一种通过提取 48 个特征参数作为输入的支持向量机模型。模型采用网格搜索、粒子群和遗传算法三种方法对参数进行优化。为参数优化问题提出一种有效的解决思路。

支持向量机算法的优势在于：学习过程简单、训练速度快；在预测性维护中可以解决设备工况识别问题，简单分类问题处理较为精确；在参数优化方面可以结合先进的算法进行优化，提高分类准确率。

但支持向量机也存在一些问题：选择合适的核函数较为困难；在进行大规模数据训练时间时间较长；最终模型不容易理解和解释；将实际应用与仿真模型相适应比较困难^[34]；无法解决多分类问题；理论上只能提供欠优解。

3.3 聚类算法

聚类是一种无监督的学习算法，可以对大量未知标注的数据集，按数据的内在相似性将数据集划分为多个类别，主要目的是找到数据集的聚类，以便彼此接近的样本与同一个聚类相关联，而彼此远离的样本与不同的聚类相关联^[29]。传统的聚类算法可以分为五类：基于划分的聚类，基于网格的聚类，基于层次的聚类，基于密度的聚类，基于模型的聚类。近年来产生了多种新的聚类算法：基于粒度的聚类算法，基于熵的聚类算法，不确定聚类算法，谱聚类算法，核聚类算法等^[35]。

Durbhaka 等人^[36]通过对振动信号使用聚类来分析风力涡轮机的工况，预测涡轮机的工况并对故障类型进行分类，通过对零部件的替换和校正，进而避免突然故障，通过加入 CRA 方法，使得聚类算法预测准确性达到 93%。Eke 等人^[37]提出聚类算法可以用于自动提取变压器绝缘油中溶解气体数据中的簇，识别导致故障的每个集群的特征，使用聚类算法识别出四个聚类：高能电弧放电、油温变化、生成气体速率、油液循环周期，结果表明该模型可以较好预测设备工况。Mathew 等人^[38]比较多种机器学习(聚类算法，支持向量机等)来预测涡轮增压风扇发动机的故障，并给出各种方法的优缺点。Uhlmann 等人^[39]提出一种聚类算法来识别机床数据(平台温度、处理室中氧气百分比和处理室压力)的簇，识

别数据中的四个聚类：操作条件、气体保护系统的故障条件、压力系统的故障条件以及使机床保持待机状态的故障条件。Amruthnath 等人^[40]比较大量聚类算法，用于从排气风扇的振动数据中进行故障预测，获得较为准确的结果。张旺等人^[41]提出了首先对双通道信号分别进行 Hilbert 包络解调去除噪声，对处理后的信号进行全矢融合提取主振矢，再采用聚类方法对主振矢信号进行分析，解决了轴承突发故障特征并不明显的问题，其预测精度达到 90.64%。李媛媛等人^[42]提出采用模糊 C 均值算法，并以待测状态隶属于正常状态的程度作为性能评判指标的方法，解决了转盘轴承剩余寿命预测过程中的提取能够反映转盘轴承寿命状态的特征向量和建立合理的预测模型问题，为转盘轴承的故障诊断和剩余寿命预测的研究提供了一种新思路。更多基于聚类算法的预测性维护研究的最新论文见表 3。

表 3 聚类算法在预测性维护中的最新应用

文献	年份	内容
[43]	2019	针对一种半导体设备预测性维护问题，提出一种进行单因素优化设计 DBSCAN 聚类算法模型，通过应用加速度传感器数据和虚拟误差数据，跟踪每个聚类中心点的移动。通过跟踪中心点，可以
		确定设备维护时间。
[44]	2019	针对晶圆运输机器人的预测性维护问题，提出一种通过跟踪频率数据簇的中值来预测机器人故障的技术，减少噪音数据对聚类模型的影响，提升剩余寿命预测的准确性。
[45]	2018	针对设备剩余寿命预测问题，提出一种基于退化曲线相似性匹配的方法。使用线性回归和多项式拟合方法将多维传感器参数融合为一维 HI 曲线，判别曲线与测试单元 HI 轨迹相关性，完成设备剩余寿命估计，引入 K-means++ 聚类应用，极大提
		高了剩余寿命预测的稳定性。
[46]	2018	为研究无监督机器学习算法在预测性维护中的早期故障检测问题，采集一台排气扇的振动数据，分别采用层次聚类、k 均值聚类、模糊 c 均值聚类和基于模型聚类等不同无监督学习算法来构建预测模型，通过实验测试不同聚类方法的准确性、
		性能和鲁棒性。

聚类算法可以通过新样本再训练来改变簇的中心，在处理大数据集上表现出良好的性能；聚类问题的研究趋势将会向更快、更有理论保证的方向发展，随着新的分析技术的发展，新的理论结果依然在不断出现，有良

好的发展前景^[47]。

但其也存在一些问题：训练数据中的噪声会对最终结果造成重大影响；无法自动确定簇的数量；数据输入顺序对最终结果有影响；无法处理可分离非线性数据集；初始点选取会极大的影响聚类的结果。

3.4 随机森林

随机森林法最初是由 Leo^[48]提出的。随机森林在以决策树为基学习器构建 bagging 集成的基础上，进一步在决策树的训练过程中引入了随机属性选择进行预测。它用随机的方式建立一个森林，森林由许多决策树组成，每一棵决策树之间没有关联，利用多棵树对样本进行训练并预测。当变量的数量大于样本的数量时，随机森林表现出良好的性能^[49]。决策树算法有 ID3、C4.5、CRAT、SLIQ 等。

随机森林被用于生成动态预测模型。Kusia 等人^[50]使用状态数据和运行数据来设计随机森林模型，为风力涡轮机工况监测提出了改进方案。Santos 等人^[51]提出随机森林来预测鼠笼式感应电动机的定子绕组短路故障。Su 等人^[52]的研究提出一种硬盘故障检测系统，用于实时预测故障硬盘和驱动器故障，其系统过程包括两个阶段：批量训练，其中使用历史数据生成和训练随机森林模型；实时预测，它使用从终端用户设备收集的数据来进行预测。实时预测的准确性达到 85%。Kulkarni 等人^[53]应用随机森林模型来预测制冷和冷藏系统故障，其方法正确性能够达到 89%。Paolanti 等人^[54]提出一个利用来自各种传感器、可编程逻辑控制器和通信协议的数据来预测设备工况的随机森林模型，虽然决策树可能会过度拟合，但在大多数情况下可以避免，因为决策树可以处理随机特征子集并从这些子集构建更小的树^{[48][49][55]}。更多基于随机森林的预测性维护研究的最新论文见表 4。

表 4 随机森林在预测性维护中的最新应用

文献	年份	内容
[56]	2020	针对现有人工智能方法在滚动轴承剩余寿命预测中存在精度差、运算效率低的问题，提出一种新的特征 DIF，解决了因特征趋势较差导致的预测精度低的问题；引入 CNN 网络改进多粒度扫描结构，同时构建 CasCatBoost 改进级联森林，将计算时间缩短了 25%~50%；使用深度策略构建 CasCatBoost，加强了模型的泛化能力。
		针对机械系统机油状况分析问题，探讨随机森林、前馈神经网络及使用油液分析资料训练的逻辑回归模型来分类机器工况。随机森林模型在所有机器条件下都优于其他模型，为选择随机森林进行预测提供了支持。
[57]	2020	

[58]	2020	为研究小型蒸汽灭菌器的预测性维护问题,提出一种基于随机森林算法的预测维护模型,该模型将小型蒸汽灭菌器中的真空泵和蒸汽发生器两个关键部件的健康状况进行分类。随机森林算法在真空泵和蒸汽发生器数据集的准确性分别为 83.5 和 82.0%。
[59]	2018	针对硬盘故障预测问题,创新的提出随机森林和长短期记忆(LSTM)递归神经网络两种不同类型的机器学习方法。随机森林方法更适合于短期预测,而 LSTM 更适合于长时间的模型。

随机森林的优势在于:随机森林是一类判别模型,支持分类问题、回归问题和多分类问题;随机森林结合其他算法,得到更优的预测结果,可以将新的理论引入随机森林,结合不同的算法提升模型预测的性能^[60];决策树越多,可以提供越多的观察结果作为预测的依据^[38];在某些情况下,决策树可以减少波动,增强随机森林模型泛化能力^[61]。

随机森林方法也有缺点:少量的决策树的分类能力有限,只有产生大量的决策树才有可能得到有效的分类效果^[60],与其他机器学习相比需要更多的计算时间;不能很好地处理非平衡数据,随机选取的训练集加剧了数据的非平衡性;对连续性变量的处理需要进行离散化,分析计算节点分裂标准花费大量时间;通过剪枝的方法避免决策树的过拟合问题会提高算法的复杂性,性能提升有局限性^[62]。

4 未来展望

4.1 系统架构

预测性维护是工业大数据和人工智能方向的一个重要的应用场景,针对设备的故障和失效问题,从被动的故障维护到主动的预测和综合规划管理,研究人员不断提供新思路和新方法。

应用预测性维护的设备或系统,其产生的故障是由于渐进式的功能衰退超过临界值导致的。对设备进行性能退化评估和剩余寿命预测具有很好的前景,其预测模型具有较高的准确率,是预测性维护策略决策的可靠依据。本文提出一种面向本地设备的预测性维护系统架构(如图 1 所示),为后续研究提供思路。

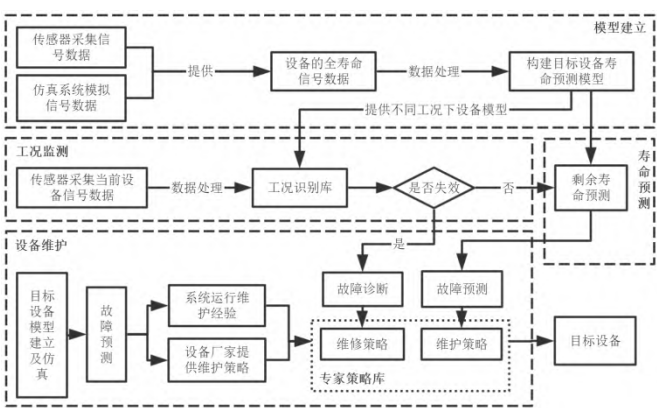


图 1 面向本地设备的预测性维护系统框架

本系统架构主要是基于对设备运行时产生的大量的历史、实时数据进行收集、处理、分析、建模来实现的,对模型预测的设备剩余寿命进行分级,通过模拟仿真构造专家策略库,根据分级对应的策略对设备进行维护,它所遵循的预测性运维流程主要包括以下步骤:

(1)数据获取:通过模拟仿真和传感器测量获得目标设备或系统的全寿命数据。

(2)数据处理:包括数据预处理和特征提取,对数据进行过滤和整理,识别数据中工况信息,剔除非重要变量,通过特征提取的方法得到衰退特征,供模型训练使用。

(3)模型训练:选择适当机器学习模型,利用经处理后的全寿命数据进行训练,获得在不同工况下可以对设备或系统剩余寿命进行准确预测的模型。

(4)工况检测:通过传感器收集当前设备或系统数据,并通过模型提供的工况识别库判断设备或系统是否失效,并作出反馈。

(5)寿命预测:在设备或系统未失效情况下,用预测模型对其剩余寿命进行预测,并作出反馈。

(6)策略制定:根据系统故障预测的仿真,可以验证维护和维修策略的可行性,并将论证结果导入专家策略库中作为方案。

(7)策略执行:当系统失效时,根据工况识别库的反馈信息进行故障诊断,决定设备或系统的维修策略;当系统未失效时,根据剩余寿命预测结果进行故障预测,决定设备或系统的维护和保养策略。

此外,根据实际情况还可以加入剩余寿命可视化、数据存储再训练、数据检查等步骤。

4.2 预测模型

由于预测性维护面向不同的目标设备时,选取的预测模型不唯一,故对模型选择提出了比较高的要求。机器学习,包括人工神经网络、支持向量机、聚类算法和

随机森林等，已经成功地应用于预测性维护应用中。随着工业设备维护要求的不断提高，预测模型也需要不断发展出新的理论和方法来适应日趋复杂的生产环境。

(1)将新的理论引入机器学习算法。以随机森林为例，由于随机森林算法本身就是 Bagging 算法与 Random Subspace 算法结合而得到的新的算法，因此在随机森林的基础上再增加新的算法一直是许多专家学者研究的重点^[60]。新算法与预测性维护技术进行正确结合，可以提升预测准确性。

(2)将新的数据预处理方法与机器学习算法结合。训练数据集的质量决定了模型训练的准确度，因此将数据进行有效数据处理之后进行模型训练，将提高模型的质量。同时，机器学习模型训练需要大量的训练数据，而实际中获得的小样本数据的情况居多，因此如何利用机器学习算法对小样本数据的设备进行寿命预测与故障诊断也是需要攻坚的一个难点^[2]。

4.3 物联网技术

由于 5G 技术和工业 4.0 时代的到来，面对大规模生产的工业设备，通过工业物联网将设备的工况数据上传至服务器，生产厂家可以将获得的大量设备数据进行分析处理、训练建模、制定维护策略、实时更新、专家会诊，进而优化服务。基于物联网技术的预测性维护策略相比于面向本地设备的预测性维护策略不同，其效率更高、适配性更强、准确率更高。本文提出一种基于物联网技术的预测性维护系统架构(如图 2 所示)，为后续研究提供思路。

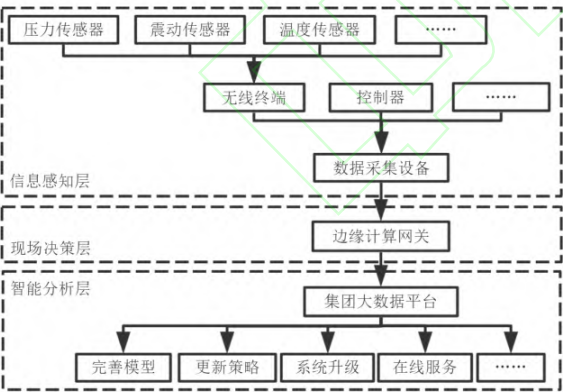


图 2 基于物联网技术的预测性维护系统框架

本系统架构主要是基于物联网技术，通过对设备维护策略分别提供线上线下服务，现场收集设备数据和控制数据，经过现场决策系统判断和反馈，并上传数据至集团大数据平台进行智能分析和数据挖掘，最新成果下载至现场决策系统并应用于之后的预测性维护策略。它所遵循的预测性运维流程主要包括以下步骤：

(1)信息感知：通过各类传感器收集现场设备数据，由无线终端上传至数据采集系统，同时上传控制器等设备数据，经数据整理后进入现场决策系统。通过为设备开发传感技术，提高数据数量和质量，当有更多高质量训练数据可用时，预测性维护策略将更具可靠性^[63]。

(2)现场决策：边缘计算网关主要完成感知信息处理、在线诊断分析和实时反馈控制三个方面的工作^[64]。感知信息处理和前文中数据处理相同，为设备状态数据分析和挖掘提供数据基础；在线诊断分析通过不同工况下的预测模型，对设备运行数据的实时分析和状态的在线识别；实时反馈控制，根据状态判断结果，通过本地专家策略库，自动生成相应控制逻辑。

(3)智能分析：现场设备数据、决策数据和维护结果等通过网络上传至设备生产集团的大数据平台进行智能分析和数据挖掘，产生新的预测模型和维护策略，通过网络下载至现场决策系统进行更新。智能分析系统还可根据数据反馈调整设备运行参数，以达到增加设备寿命、提升生产效率、提升生产安全性等目的。集团大数据平台利用智能分析解决的问题和解决方案如表 5 所示。

表 5 问题与解决方案

问题	解决方案
优化预测模型	为提升设备寿命预测模型的准确率，可以利用生产现场上传至集团大数据平台的数据，通过研究机器学习新算法，训练更精确的模型，经集团实验车间和仿真平台对模型进行多轮验证之后，下载至边缘计算网关，即可提升现场决策效率。
更新维护策略	针对设备的故障维护和维修策略，集团专家可以对所有设备故障策略和维护结果进行研究，提出更高效的策略，经集团实验车间和仿真平台验证之后，下载至边缘计算网关。
数据价值密度低	由于生产现场情况复杂，传感器采集数据存在多源异构、噪声污染等导致数据价值密度低的问题 ^[64] ，需要研究高效的设备工况模型，从现场数据中挖掘出状态特征信息，提高现场数据采集效率。经验证之后，下载至边缘计算网关。
故障样本缺失	针对一类零部件和系统存在大量运转重复数据，而应用于新系统时产生故障的问题，可研究迁移学习等技术手段实现单一子系统到复杂系统的泛化学习 ^[64] ，利用集团实验车间和仿真平台，对故障预测、维护策略和应急策略进行验证之后，下载至边缘计算网关和设备控制器。
复杂耦合	利用集团实验车间设备和设备仿真平台，研究数字孪生等技术手段，加强设备系统运行状态特性与故障机理的研究，结合数据驱动方法，建立机理和数据混合驱动设备状态诊断模型 ^[64] 。经验证后下载至现场设备控制器，优化设备控制方法，提高生产效率。
特殊故障	针对设备突发的疑难杂症，通过大数据平台开展专家会诊，利用线上平台对现场指导并制定方案，缩短故障排查时间，减少维修成本。

物联网技术为预测性维护策略提供了高效的解决方案,同时也提出了更高的要求,需要相关技术的不断革新。目标系统越复杂,现场采集的信息越多,传输效率越低,而复杂的数据也考验现场设备的计算力。5G技术的出现,为数据传输问题提供了解决方案,同时也为现场决策系统提升计算能力提供了更多选择,但由于5G技术与工业生产结合经验较少,还需专家学者进一步研究。

5 总结

预测性维护技术作为工业生产领域的一个重要研究方向,近年来受到了越来越多的关注。鉴于机器学习领域不断进行创新,本文系统地介绍了基于机器学习的预测性维护技术相关研究现状,从机器学习算法入手,重点介绍了四种常见的算法,分析了面向特定设备的剩余寿命预测方法和特性,列举了近几年来最新的研究成果和优势,并从系统架构、预测模型和物联网技术的角度对未来研究进行展望。随着工业4.0时代的到来,预测性维护技术将会为工业生产带来革新,极大提高设备的安全性、提升保养效率、降低生产成本、提升生产效率。预测性维护技术在5G技术和人工智能技术的背景下,具有广阔的应用前景和研究价值!

参考文献:

- [1] Susto, G. A., Schirru, A., Pampuri, S., McLoone, S. & Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics*, 2015, 11: 812-820.
- [2] 郭一帆,唐家银.基于机器学习算法的寿命预测与故障诊断技术的发展综述[J].*计算机测量与控制*, 2019,27(3).
- [3] MITCHELL T M, CARBONELL J G, MICHALSKI R S. machine learning[M]. Beijing: China Machine Press, 2003.
- [4] Soares, S. G., & Araújo, R. An on-line weighted ensemble of regressor models to handle concept drifts. *Engineering Applications of Artificial Intelligence*, 2015, 37: 392-406.
- [5] Jardine, A. K., Lin, D., & Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 2006, 20: 1483-1510.
- [6] SIMON H. Neural network: A comprehensive foundation[M]. New Jersey: Prentice Hall PTR, 1994.
- [7] Biswal, S., & Sabareesh, G. R. Design and development of a wind turbine test rig for condition monitoring studies. 2015 International conference on industrial instrumentation and control, ICIC 2015: 891-896.
- [8] GEBRAEEL N, LAWLEY M, LIU R, et al. Residual life predictions from vibration-based degradation signals: A neural network approach[J]. *IEEE Transactions on Industrial Electronics*, 2004, 51(3): 694-700.
- [9] BEZAZI A, PIERCE S G, WORDEN K, et al. Fatigue life prediction of sandwich composite materials under flexural tests using a Bayesian trained artificial neural network[J]. *International Journal of Fatigue*, 2007, 29(4): 738-747.
- [10] WANG X Y, HAN M. Online sequential extreme learning machine with kernels for nonstationary time series prediction[J]. *Neurocomputing*, 2014, 145(145): 90-97.
- [11] Kolokas, N., Vafeiadis, T., Ioannidis, D., & Tzovaras, D. Forecasting faults of industrial equipment using machine learning classifiers. *International symposium on innovations in intelligent systems and applications (INISTA)* 2018: 1-6.
- [12] Sheng Xiang, Yi Qin, Caichao Zhu, Yangyang Wang, Haizhou Chen. Long short-term memory neural network with weight amplification and its application into gear remaining useful life prediction. *Engineering Applications of Artificial Intelligence*, 2020, 91: 1-11.
- [13] Jialin Li, Xueyi Li, David He. A Directed Acyclic Graph Network Combined With CNN and LSTM for Remaining Useful Life Prediction[J]. *IEEE Access*, 2019, 7.
- [14] Bo Luo, Haoting Wang, Hongqi Liu, et al. Early Fault Detection of Machine Tools Based on Deep Learning and Dynamic Identification[J]. *IEEE Transactions on Industrial Electronics*, 2019, 66(1).
- [15] Zhu Jun, Chen Nan, Peng Weiwen. Estimation of bearing remaining useful life based on multiscale convolutional neural network[J]. *IEEE Transactions on Industrial Electronics*, 2019, 66(4).
- [16] Zhang Yongzhi, Xiong Rui, He Hongwen, et al. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries[J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(7).
- [17] Deutsch J, He D. Using deep learning-based approach to predict remaining useful life of rotating components[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, 48(1).
- [18] Ren Lei, Zhao Li, Hong Sheng, et al. Remaining useful life prediction for lithium-ion battery: a deep learning approach[J]. *IEEE Access*, 2018, 6: 50587-50598.
- [19] Sexton T, Brundage M P, Hoffman M, et al. Hybrid datafication of maintenance logs from AI-assisted human tags[C]//*IEEE International Conference on Big Data*, 2017: 1769-1777.

- [20] Susto G A, McLoone S, Pagano D, et al. Prediction of integral type failures in semiconductor manufacturing through classification methods[C]//IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 2013: 1-4.
- [21] Nieto P J G, García-Gonzalo E, Lasheras F S, et al. Hybrid PSO-SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability[J]. Reliability Engineering & System Safety, 2015, 138: 219-231.
- [22] Maior C B S, Moura M D C, Lins I D, et al. Remaining useful life estimation by empirical mode decomposition and support vector machine[J]. IEEE Latin America Transactions, 2016, 14 (11) : 4603-4610.
- [23] Mathew J, Luo M, Pang C K. Regression kernel for prognostics with support vector machines[C]//IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 2017: 1-5.
- [24] Song Y C, Liu D T, Hou Y D, et al. Satellite lithium-ion battery remaining useful life estimation with an iterative updated RVM fused with the KF algorithm[J]. Chinese Journal of Aeronautics, 2018 (1) : 31-40.
- [25] Lasisi A, Attoh-Okine N. Principal components analysis and track quality index: a machine learning approach[J]. Transportation Research Part C: Emerging Technologies, 2018, 91: 230-248.
- [26] 戴邵武, 陈强强, 丁宇. 基于时域特征的滚动轴承寿命预测[J]. 计算机测量与控制, 2019, 27 (10) : 60-63.
- [27] 王瀛洲, 倪裕隆, 郑宇清, 等. 基于 ALO-SVR 的锂离子电池剩余使用寿命预测[J/OL]. 中国电机工程学报, 2019: 1- 12. [2020-05-18]. <http://kns.cnki.net/kcms/detail/11.2107.TM.20200427.1423.009.html>.
- [28] Liu Qiaobin, Shi Wenku, Chen Zhiyong. Fatigue life prediction for vibration isolation rubber based on parameter optimized support vector machine model[J]. Fatigue & Fracture of Engineering Materials & Structures, 2019, 42 (3) : 710-718.
- [29] Zou Tao, Yang Dongjin, Yang Haitao. Residual life prediction of Aviation Battery[C]//2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2020.
- [30] Guo Bingxiu, Wang Xiaohui, Wang Yanyan, et al. Application of support vector regression to predict the remaining useful life of polymerized styrene butadiene rubber of cable insulation[C]//2019 Prognostics and System Health Management Conference (PHM), Qingdao, 2019.
- [31] Ma Hailong, Li Zhen. Remaining life predictions of bearing based on relative features and support vector machine[C]//2019 Prognostics and System Health Management Conference (PHM), Qingdao, 2019.
- [32] Xayyasith S, Promwungkwa A, Ngamsanroaj K. Application of machine learning for predictive maintenance cooling system in nam ngum-1 hydropower plant[C]//2018 16th International Conference on ICT and Knowledge Engineering (ICT&KE), 2018.
- [33] Xu Changhang, Cui Jinxiu. Prediction of residual fatigue life of bolts based on metal magnetic memory[C]//2018 International Applied Computational Electromagnetics Society Symposium-China (ACES), 2018.
- [34] Cawley G C, Talbot N L. On over-fitting in model selection and subsequent selection bias in performance evaluation[J]. Journal of Machine Learning Research, 2010, 11: 2079-2107.
- [35] 衣俊艳, 吴博雅, 雍巧玲. 具有加权特性的弹性网络聚类算法研究[J/OL]. 计算机工程与应用, 2020: 1-13. [2020-06-04]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20200529.1507.012.html>
- [36] Durbhaka G K, Selvaraj B. Predictive maintenance for wind turbine diagnostics using vibration signal analysis based on collaborative recommendation approach[C]//International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2016: 1839-1842.
- [37] Eke S, Aka-ngnui T, Clerc G, et al. Characterization of the operating periods of a power transformer by clustering the dissolved gas data[C]//IEEE 11th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED), 2017: 298-303.
- [38] Mathew V, Toby T, Singh V, et al. Prediction of remaining useful lifetime (RUL) of turbofan engine using machine learning[C]//IEEE International Conference on Circuits and Systems (ICCS), 2017: 306-311.
- [39] Uhlmann E, Pontes R P, Geisert C, et al. Cluster identification of sensor data for predictive maintenance in a selective laser melting machine tool[J]. Procedia Manufacturing, 2018, 24: 60-65.
- [40] Amruthnath N, Gupta T A. A research study on unsupervised machine learning algorithms for fault detection in predictive maintenance[C]//5th International Conference on Industrial Engineering and Applications (ICIEA), 2018: 355-361.
- [41] 张旺, 陈磊, 陈超宇, 等. Hilbert-全矢 HMM 轴承剩余寿命预测[J]. 机械设计与制造, 2020 (3) : 95-98.
- [42] 李媛媛, 陈捷, 洪荣晶, 等. 基于模糊 C 均值的转盘轴承剩余寿命预测[J]. 轴承, 2017 (3) : 50-55.
- [43] Yoon H S, Han S S. Clustering parameter optimization of predictive maintenance algorithm for semiconductor equipment using

one-way factorial design[C]//2019 19th International Conference on Control, Automation and Systems (ICCAS), 2019.

[44] Kim H G, Yoon H S, Yoo J H, et al. Development of predictive maintenance technology for wafer transfer robot using clustering algorithm[C]//2019 International Conference on Electronics, Information, and Communication (ICEIC), 2019.

[45] 李劲松, 张惠娟, 杨忠, 等. 基于退化曲线相似性的剩余使用寿命估计方法[J]. 应用科技, 2018, 45 (5) : 82-86.

[46] Amruthnath N, Gupta T. A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance[C]//2018 5th International Conference on Industrial Engineering and Applications (ICIEA), 2018.

[47] 任远航. 面向大数据的 K-means 算法综述[J/OL]. 计算机应用研究 2020: 1-7. [2020-06-05]. <https://doi.org/10.19734/j.issn.1001-3695.2019.10.0581>.

[48] Leo B. Random forests[J]. Machine Learning, 2001, 45: 5-32.

[49] Biau G, Scornet E. A random forest guided tour[J]. TEST: An Official Journal of the Spanish Society of Statistics and Operations Research, 2016, 25 (2) : 197-227.

[50] Kusiak A, Verma A. Prediction of status patterns of wind turbines: A data mining approach[J]. Journal of Solar Energy Engineering, 2011, 133: 1-10.

[51] Santos T, Ferreira F J, Pires J M, et al. Stator winding shortcircuit fault diagnosis in induction motors using random forest[C]//IEEE International Electric Machines & Drives Conference (IEMDC), 2017: 1-8.

[52] Su C J, Huang S F. Real-time big data analytics for hard disk drive predictive maintenance[J]. Computers and Electrical Engineering, 2018, 71: 93-101.

[53] Kulkarni K, Devi U, Sirighee A, et al. Predictive maintenance for supermarket refrigeration systems using only case temperature data[C]//The American Control Conference (ACC), 2018: 4640-4645.

[54] Paolanti M, Romeo L, Felicetti A, et al. Machine learning approach for predictive maintenance in industry 4.0[C]//4th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA), 2018: 1-6.

[55] Prytz R, Nowaczyk S, Rögnvaldsson T, et al. Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data[J]. Engineering Applications of Artificial Intelligence, 2015, 41: 139-150.

[56] 王玉静, 王诗达, 康守强, 等. 基于改进深度森林的滚动轴承剩余寿命预测方法[J]. 中国电机工程学报, 2020, 40 (15) : 5032-5043.

[57] Kearthland S, van Zyl T L. Automating predictive maintenance using oil analysis and machine learning[C]//2020 International SAUPEC/RobMech/PRASA Conference, 2020.

[58] Musabayli M, Osman M H, Dirix M. Classification model for predictive maintenance of small steam sterilisers[J]. IET Collaborative Intelligent Manufacturing, 2020, 2 (1) : 3.

[59] Anantharaman P, Qiao Mu, Jadav D. Large scale predictive analytics for hard disk remaining useful life estimation[C]//2018 IEEE International Congress on Big Data (BigData Congress), 2018.

[60] 孙明喆, 毕瑶家, 孙驰. 改进随机森林算法综述[J]. 现代信息科技, 2019, 20 (3) : 28-30.

[61] Amihai I, Gitzel R, Kotriwala A M, et al. An industrial case study using vibration data and machine learning to predict asset health[C]//20th IEEE International Conference on Business Informatics (CBI), 2018: 178-185.

[62] 吕红燕, 冯倩. 随机森林算法研究综述[J]. 河北省科学院学报, 2019, 36 (3) : 37-41.

[63] Hashemian H M, Bean W C. State-of-the-art predictive maintenance techniques[J]. IEEE Transactions on Instrumentation and Measurement, 2011, 60: 3480-3492.

[64] 丁恩杰, 俞啸, 廖玉波, 等. 基于物联网的矿山机械设备状态智能感知与诊断[J]. 煤炭学报 2020 (6) : 2308-2319.