

基于强化学习的装备智能维修决策技术研究综述

万心仪, 李传扬, 胡昌华, 张泽明, 冷明哲

(中国人民解放军火箭军工程大学智控实验室, 陕西 西安 710025)

摘要: 随着工业 4.0 背景下智能装备复杂性和运维需求的日益提升, 传统维修决策方法在动态环境下的适应性不足问题逐渐凸显。基于强化学习的维修决策技术通过与环境交互实现策略自主优化, 为装备智能维护提供了一种范式。本文聚焦 1954 至 2024 年间发表的 76 篇文献, 系统探讨了强化学习理论与维修决策的融合路径。本文深入解析了强化学习的 SARSA、Q-Learning、Actor-Critic 等核心算法; 分析了装备智能维修决策技术现状; 从工业制造、能源、航空航天和交通运输四个方面剖析了强化学习在装备维修决策中的典型应用场景; 揭示了当前技术在算法收敛速度和计算效率、模型可解释性、以及数据获取及隐私问题等面临的核心挑战。本研究为智能运维领域的算法创新与工程落地提供了理论参考, 有助于推动强化学习在装备维修决策中的深度应用。

关键词: 故障诊断; 强化学习; 智能维修决策; 人工智能

中图分类号: TH165⁺.3; TP181 **文献标志码:** A **文章编号:** 1004-4523(2025)06-1154-13

DOI: 10.16385/j.cnki.issn.1004-4523.2025.06.004

A review of reinforcement learning-based intelligent maintenance decision-making for equipment

WAN Xinyi, LI Chuanyang, HU Changhua, ZHANG Zeming, LENG Mingzhe

(Laboratory of Intelligent Control, PLA Rocket Force University of Engineering, Xi'an 710025, China)

Abstract: The increasing complexity of intelligent equipment and evolving operation and maintenance demands within Industry 4.0 highlight the inadequate adaptability of traditional maintenance decision-making methods in dynamic environments. Reinforcement learning (RL)-based maintenance decision-making technology offers a paradigm for intelligent equipment maintenance by enabling autonomous strategy optimization through environmental interaction. This paper systematically explores the integration of RL theory and maintenance decision-making, focuses on 76 peer-reviewed articles published between 1954 and 2024. Core RL algorithms, including SARSA, Q-Learning, and Actor-Critic, are thoroughly examined and analyzed. The current state of intelligent equipment maintenance decision-making technology is also analyzed in depth. Typical application scenarios for RL in equipment maintenance decision-making are comprehensively dissected across four key areas: industrial manufacturing, energy, aerospace, and transportation. The study also identifies and discusses the core challenges facing current technology, such as algorithm convergence speed, computational efficiency, model interpretability, and issues related to data acquisition and privacy. This research provides a theoretical reference for algorithm innovation and engineering implementation in the field of intelligent operation and maintenance, fostering the deeper application of RL in maintenance decision-making.

Keywords: fault diagnosis; reinforcement learning; intelligent maintenance decision; artificial intelligence

随着工业智能化的加速推进, 设备健康管理及智能维护已成为现代制造业提质增效的关键环节, 在工业制造、能源装备、航空航天等重要领域发挥着不可替代的作用^[1-3]。传统设备管理模式多基于固定阈值规则或专家经验模型, 通过集中规划的方式来优化维修资源, 虽然具有逻辑清晰、实施简便的特点, 但普遍存在动态适应性不足、多目标优化能力薄弱、大规模系统决策效率低下等显著技术瓶颈, 对于设备状态的动态变化缺乏应变能力^[4]。当

对复杂工况下设备退化状态的动态演变、多故障耦合的关联影响以及维修资源的多维度约束时, 传统方法往往难以实现精准的维修策略优化与实时决策响应。传统维修策略, 诸如定期维修与故障后维修, 虽在一定时期内发挥了作用, 但随着全球化市场竞争的加剧以及技术的飞速迭代, 其局限性愈发凸显。定期维修易引发过度维修, 徒增成本; 故障后维修则可能导致意外停机, 甚至灾难性后果。在此背景下, 人工智能领域中的强化学习 (reinforcement

收稿日期: 2025-05-15; 修订日期: 2025-05-26

基金项目: 国家自然科学基金资助项目 (62227814, 52205040); 陕西省三秦英才计划支持

learning, RL) 技术为解决这一难题提供了突破口^[5]。RL 作为机器学习当中的一个重要分支, 如图 1 所示。通过构建智能体与环境的持续交互机制, 自主学习设备状态演化规律, 并在动态不确定条件下实现维修策略的自适应优化, 展现出强大的在线学习与决策能力^[5-6]。

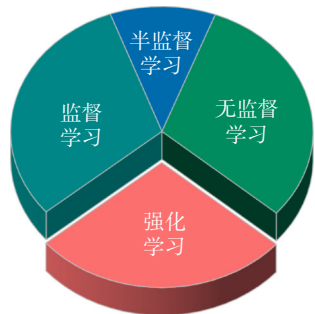


图 1 机器学习分支

Fig. 1 Branches of machine learning

RL 凭借智能体与环境动态交互、试错学习的特性, 能够依据设备实时状态与复杂环境信息, 动态优化维修策略, 有望彻底变革传统维修决策模式。

近年来, RL 在工业制造、能源、航空航天和交通运输等领域的维修决策应用研究不断涌现。学者们纷纷投身于开发 RL 模型, 以适应不同领域维修场景的需求。然而, RL 技术在迈向成熟应用的进程中仍面临诸多挑战。算法收敛速度慢、模型可解释性差、数据隐私与安全问题突出, 以及对复杂环境的适应性不足等问题, 成为 RL 装备智能维修决策技术发展道路上的一道道难题。

本文旨在全面剖析基于 RL 的装备智能维修决策技术的现状, 深入探讨其面临的挑战与问题, 以期未来研究指明方向。首先对 RL 在各主要领域的维修决策应用场景进行系统梳理, 进而聚焦于现存的关键技术瓶颈, 并对其展开深入剖析与探讨。通过在 Web of Science、IEEE Xplore 和 Scopus 这三大主要学术数据库中以“RL”和“智能维修决策”为主题词进行检索, 可以得到 2009 年至 2024 年关于 RL 在装备智能维修决策的年度科学产出, 以论文数量衡量研究进展, 如图 2 所示。该图表明, 近年来 RL 在智能维修决策领域应用的出版物数量呈上升趋势, 这凸显了该研究领域日益增长的重要性及未来潜力。因此, 本文旨在提供一份全面且详细的综述, 文章整体框架如图 3 所示, 具体目标如下:

- (1) 分析 RL 的研究进展, 涵盖其概念、历史发展以及应用 RL 所需的关键技术;
- (2) 分析装备智能维修决策技术的研究进展;
- (3) 专注于 RL 在装备智能维修决策中的应用分析与讨论;
- (4) 总结 RL 在装备智能维修决策应用中存在的

问题和挑战, 并提出建议和未来展望。

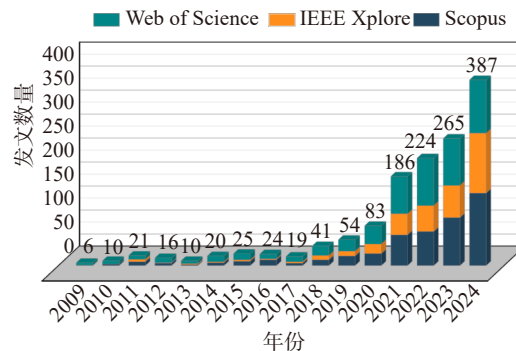


图 2 2009 年到 2024 年的发文数量

Fig. 2 The number of papers published from 2009 to 2024

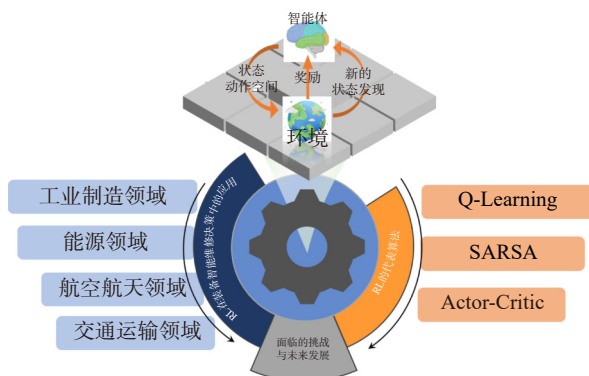


图 3 综述内容的研究框架

Fig. 3 Research framework of the review content

1 RL 问题研究现状

1.1 RL 的起源与发展历程

RL 作为通用人工智能 (artificial general intelligence, AGI) 领域的重要研究方向, 其起源及发展历程如图 4 所示。

1954 年, FARLEY 等^[7]提出了自组织系统的概念和定义, 并通过数字计算机模拟展示了学习和适应机制的实际应用, 率先提出模拟神经网络的自适应控制算法, 如图 5 所示。这一创新性研究成果为后来 RL 算法的发展及其在复杂系统中的应用奠定了坚实的基础, 也对后续 RL 和 AGI 产生了深远的影响。

1988 年, SUTTON 等^[8]系统化提出时序差分学习 (temporal-difference learning, TD) 的数学框架, 包括 TD(0) 和 TD(λ) 算法。证明 TD 方法在预测问题中比传统蒙特卡罗 (Monte Carlo, MC) 方法和动态规划更高效。通过理论分析和试验 (如随机游走任务) 验证了 TD 的收敛性。

如图 6 所示, 以一个展示监督学习方法低效率性的游戏对局示例。图中每个圆圈代表双人棋盘游戏中的某个局面或同类局面的集合。根据长期经验可

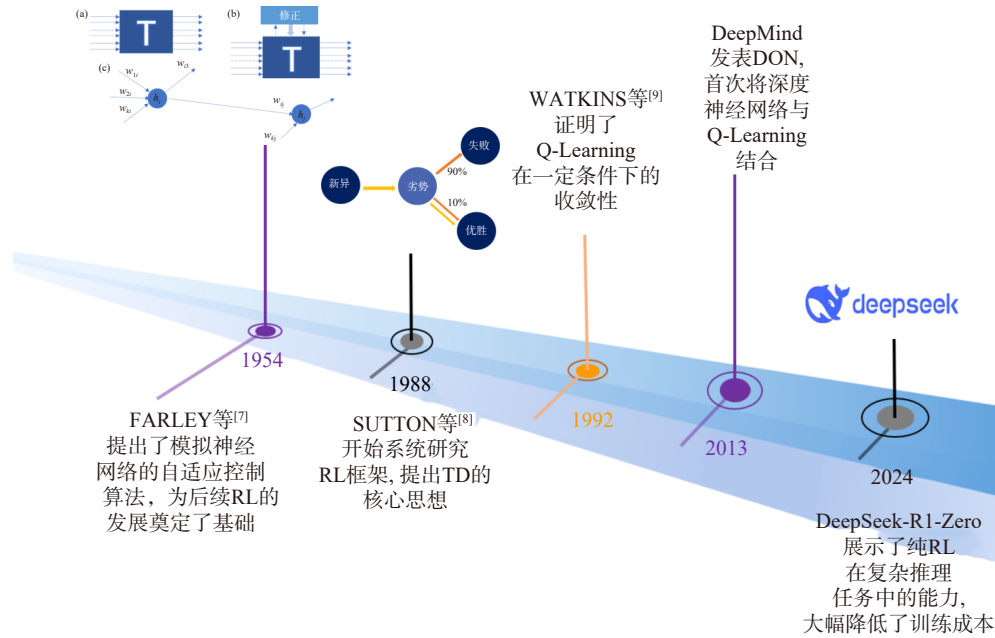


图 4 RL 的起源与发展历程

Fig. 4 The origin and development history of RL

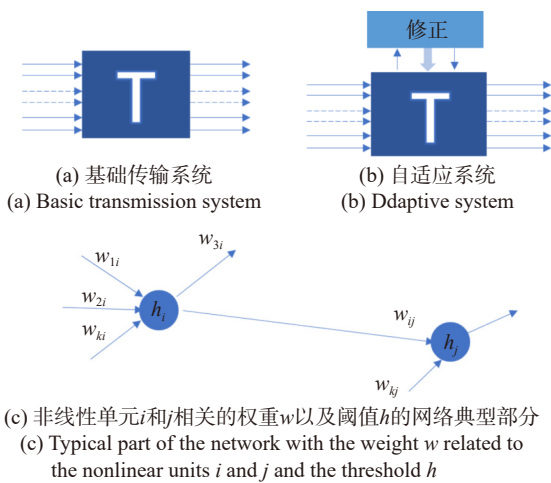


图 5 模拟神经网络的自适应控制算法机理

Fig. 5 mechanism of adaptive control algorithm for simulated neural networks

知,“劣势”局面有 90% 概率败北,仅有 10% 概率获胜。当“新异”局面首次出现时,其发展路径如黄色箭头所示。基于这次对局经验,该新异局面应获得怎样的评估:TD 方法能正确判定其应归类为劣势局

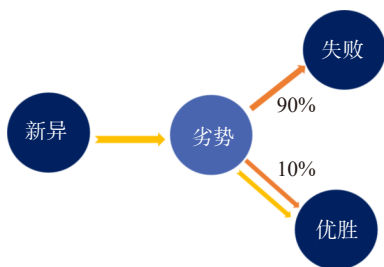


图 6 监督学习方法低效性的游戏对局

Fig. 6 Game matches showing the inefficiency of supervised learning methods

面,而监督学习方法却会因其唯一后续结果与胜利关联,从而将其完全归类为优势局面。

1992 年, WATKINS 等^[9] 发表于《Machine Learning》期刊的论文《Q-Learning》证明了 Q-Learning 在一定条件下的收敛性,从理论层面为 RL 算法的可靠性提供了坚实保障。该论文详细描述了 Q-Learning 的算法流程和更新规则,证明了在一定条件下, Q-Learning 能够收敛到最优策略。文章成果得到了学术界和工业界的广泛认可和应用,成为 RL 领域的经典文献。

2013 年, DeepMind 发表深度强化学习 (deep reinforcement learning, DRL), 创新性地 将深度神经网络 (deep neural networks, DNN) 与 Q-Learning 相结合, 为 RL 注入了新活力, 拓展了其在复杂场景下的应用潜力^[10]。DeepMind 提出了一种卷积神经网络 (convolutional neural networks, CNN), 该网络通过 Q-Learning 的变体进行训练, 输入是原始像素, 输出是估计未来奖励的值函数。该研究展示了 DNN 在复杂环境中学习控制策略的能力, 激发了后续在深度 Q 网络 (deep Q networks, DQN) 领域的大量研究和系统性改进, 使智能体在多个 Atari 游戏中超越人类水平。2015 年, MNIH 等^[11] 在《Nature》上发表的《Human-level control through deep reinforcement learning》详细分析了经验回放和目标网络的作用, 成为 DQN 领域的里程碑式工作。其中的方法不仅在 Atari 游戏上取得了优异的成果, 还为后续在机器人控制^[12-14]、自动驾驶^[15-18] 等领域应用 DQN 提供了思路。

同时, 中国杭州的深度求索人工智能基础技术研究有限公司 (DeepSeek) 于 2024 年开发了一款基于 RL 的智能体 DeepSeek-R1-Zero, 展示出 RL 在复杂

推理任务中的卓越能力,对比 ChatGPT 大幅降低训练成本,为 RL 在实际复杂任务中的应用开辟新路径,同时预示着 RL 在解决高难度问题上的巨大前景^[19]。DeepSeek R1-Zero 作为第一个完全通过 RL 训练的模型,没有依赖任何监督微调数据。它通过群体相对策略优化算法,避免了传统 RL 中复杂价值模型的构建需求,显著降低了计算成本。DeepSeek R1 的推出标志着在不依赖有监督微调的情况下,仅通过 RL 提升模型推理能力的可行性。这种创新的训练方法为 AGI 的发展提供了一种思路和方向,尤其是在模型的自主进化和推理能力方面。此外,DeepSeek R1 还探索了将推理能力蒸馏到更小模型中的潜力,为资源受限的环境提供了高效的解决方案。

1.2 RL 的关键技术

RL 是一种机器学习方法,其关键组成部分包括智能体、环境、状态、行为奖励和策略。它训练智能体通过与环境的持续互动来学习决策策略(“策略”)。这种交互是一个由人类学习启发的基于“状态-动作-奖励”循环的试错过程,在这个过程中,智能体以成功行动的奖励或失败行动的惩罚的形式接收反馈。随着时间的推移,这种反馈允许智能体改进其策略并最大化预期累积奖励。一个直观的类比是学习滑板。最初,学习者可能会在平衡方面遇到困难,当失去平衡时,会受到负强化(如摔倒);当达到平衡时,正强化(如保持直立)会鼓励重复那些成功的动作。类似地,RL 智能体根据收到的反馈改进其行为,逐渐制定一个提高决策的政策。数学上,RL 可以使用马尔可夫决策过程(Markov decision process, MDP)来描述,MDP 提供了一种结构化的方法来对决策问题建模,在这些问题中,智能体跨离散时间步与环境交互。MDP 为各种决策问题建模提供了灵活且广泛适用的框架,允许代理在复杂和不确定的环境中学习有效的决策策略,是 RL 中许多算法和技术的基础。

如图 7 所示,大脑指代智能体,地球指代环境,从当前的状态出发,在做出一个行为之后,对环境产生了一些影响,环境首先给智能体反馈了一个奖励信号,接下来智能体可以观察到一些信息,进而进入一个新的状态,在做出新的行为形成循环。

1.2.1 基于模型的 RL

在基于模型的学习中,智能体学习环境的动态模型包括转移概率和预期奖励^[20]。该模型允许智能体模拟可能的未来状态和结果,促进有效的规划和决策。通过使用模型来模拟轨迹,智能体可以在不直接与环境交互的情况下预测其行为的后果。例如:机器人迷宫的案例中,通过基于模型的 RL,机器

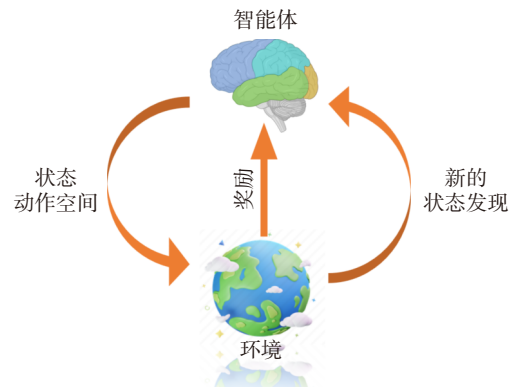


图 7 RL 的工作机制

Fig. 7 The working mechanism of RL

人将构建迷宫布局 and 动态的模型。这个模型可能包括迷宫结构的信息,如墙壁和走廊,以及机器人采取的行动的结果(例如:向前移动、向左或向右转弯)。通过使用该模型模拟可能的轨迹,机器人可以预测其行动的后果,并相应地规划其穿过迷宫的路径。此外,DeepMind 的 AlphaZero^[21] 是使用蒙特卡罗树搜索范式实现学习与优化策略的基于模型的 RL 算法的一个例子。

1.2.2 无模型的 RL

机器人在迷宫中导航的另一种方法是直接从经验中学习,即根据观察到的奖励和过渡更新策略,因而不需要明确的模拟环境。该无模型 RL 的案例^[22] 可以通过 3 种方式(基于价值、基于策略或混合方式)来实现状态或状态-动作对值的学习。

(1) 基于价值的 RL

基于价值的 RL 侧重于评估处于特定状态或采取特定行动的价值,然后使用这些价值评估来做出最大化累积奖励的决策^[23]。基于价值的 RL 的核心是价值函数的概念,它是衡量一种状态下的“好”程度,它代表了从给定状态或状态-行动对获得的预期累积奖励。在基于价值的 RL 中,智能体可以有两种不同的学习策略^[24]: 同策略方法和异策略方法^[25]。在智能体与环境交互时更新智能体的策略,意味着用于学习的数据来自正在更新的同一策略,即智能体从自己的经验中学习,并根据这些经验更新策略。其中, SARSA (state-action-reward-state-action) 算法^[26] 便是采用策略方法的经典范式。

SARSA 算法是一种基于 TD 的 RL 算法,属于使用策略学习的核心范式^[27],其核心思想是通过迭代更新状态-动作对的 Q 值函数,逼近最优策略对应的贝尔曼方程。在 SARSA 中,智能体观察当前状态,根据其当前策略采取行动,获得奖励,观察下一个状态,更新策略,然后采取相应的另一个行动。 Q 值据此进行序列更新,确保学习和行动策略是一致的。具体而言, SARSA 在每一步交互中遵循当前策略选

择动作, 并利用当前策略生成的五元组(当前状态 S_t 、动作 A_t 、即时奖励 R_{t+1} 、下一状态 S_{t+1} 、下一动作 A_{t+1})进行 Q 值更新, 其更新规则为:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha [R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t)] \quad (1)$$

式中, α 为学习率; γ 为折扣因子。

相较于禁用策略的 Q-Learning, SARSA 直接依赖当前策略生成的动作序列, 从而更注重策略的在线交互安全性, 适用于需平衡探索与利用的动态场景如机器人避障、实时决策控制^[28]。

SARSA 的理论收敛性在有限状态-动作空间及适当学习率条件下已得到严格证明, 但其性能受限于策略的探索效率与环境随机性。近年来, 研究者通过 DNN 函数逼近^[29]及自适应探索策略(如 Boltzmann 探索^[30])等机制, 显著提升了 SARSA 在高维连续状态空间与部分可观测环境中的实用性。

相反, 异策略方法^[31]将学习策略和行为策略分开, 允许智能体从不同策略下收集的经验中学习, 而不是从正在改进的策略中学习。换句话说, 智能体从一个策略生成的数据中学习, 同时尝试优化另一个策略。异策略 RL 的一个经典例子是 Q-Learning。

Q-Learning 可以通过不断探索和更新 Q 值函数来学习最优策略^[32]。 Q 值函数 $Q(s, a)$ 表示在状态 s 下执行动作 a 所能获得的最大期望奖励。Q-Learning 的更新公式:

$$Q(s, a) \leftarrow Q(s, a) + \alpha [R(s, a) + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (2)$$

式中, α 表示学习率, 控制 Q 值更新的步长; γ 为折扣因子, 表示对未来奖励的重视程度; $\max_{a'} Q(s', a')$ 表示在下一个状态 s' 下所有可能动作的最大 Q 值。

(2) 基于策略的 RL

另一种方法是基于策略的学习。基于策略的 RL 直接学习策略, 是从状态到动作的映射, 而不需要显式地估计状态或状态-动作对的值策略指定给状态的行动的概率分布^[33]。目标是找到使预期累积奖励最大化的策略。例如, 在 REINFORCE 算法中, 策略参数是基于相对于策略参数的预期收益的梯度来更新的^[31]。基于策略的方法在处理随机策略和探索行动空间方面具有优势。然而, 与基于价值的方法相比, 它们可能需要更多的数据和计算资源来收敛。

(3) 混合方式的 RL

第三类方法被称为混合方法, 在同一框架中结合了基于价值和基于策略的方法, 例如: 演员-评论家(Actor-Critic)方法。它是一种基于策略梯度与值函数估计相结合的 RL 框架, 旨在通过解耦策略优化与值函数评估来提升学习效率与稳定性。其核心架构由两个并行的神经网络模块构成: 演员负责参数

化策略, 通过梯度上升直接优化策略以最大化累积回报; 评论家则通过评估状态或状态-动作对的期望值函数(如 Q 值或状态值 V 值), 为演员提供低方差的方向性反馈。两者通过 TD 实现协同优化, 从而在策略更新的偏差与方差之间达成有效权衡。

相较于传统策略梯度方法(如 REINFORCE), 演员-评论家算法通过评论家的引导显著降低了策略更新的方差, 同时避免了纯价值函数方法(如 Q-Learning)因策略隐式依赖而导致的收敛性问题。其改进变体进一步引入优势演员-评论家^[33]等机制, 分别从分布式异步更新、策略更新约束和最大熵优化等角度提升算法鲁棒性与样本效率。因此, RL 为智能体提供了多种策略, 以在复杂环境中学习最优策略。当前, 该框架已广泛应用于机器人决策^[34]及多智能体协同^[35-36]等复杂场景, 成为 DRL 理论与应用研究的核心范式之一。

其中, RL 的技术架构总结如图 8 所示。

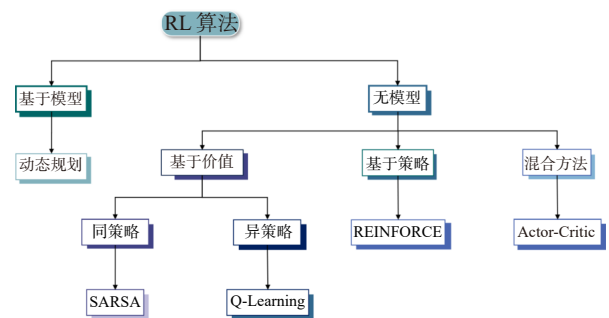


图 8 RL 技术架构

Fig. 8 The RL technical architecture

2 装备智能维修决策技术现状

2.1 数据驱动方法在故障诊断与预测中的主导地位

数据驱动方法通过深度挖掘设备运行数据中的隐含规律, 已成为装备智能维修决策的核心技术范式, 并在故障诊断与设备维护领域展现出显著价值。在故障诊断方向, 基于深度学习的端到端特征提取技术显著提升了复杂工业信号(如旋转机械振动信号)的分类精度与噪声鲁棒性。

以 CNN 为例, 其局部特征提取能力与多尺度建模特性为振动信号分析提供了创新解决方案; ZHANG 等^[37]设计的多尺度 CNN 架构通过并行卷积核高效捕获振动信号中的多频段共振特征, 在轴承故障诊断中实现 97.8% 的准确率, 较传统频谱分析方法提升 23%; XU 等^[38]进一步提出基于注意力机制的多尺度去噪残差 CNN(AM-DRCN), 通过融合残差学习与自适应噪声抑制模块, 解决了工业场景中噪声干扰导致的特征模糊问题, 经多案例验证其在高噪声环

境下仍能稳定监测机器健康状态;HUANG等^[39]引入通道注意力机制与一维卷积并行融合策略,在增强轴承故障特征表达能力的同时降低网络复杂度,为实际部署提供了轻量化解决方案。与此同时,XIE等^[40]构建的多尺度DNN通过层级化特征融合与抗噪训练策略,进一步提升了诊断模型的泛化能力与工业场景适应性。在设备维护决策层面,研究通过跨领域技术集成实现了数据驱动的优化:MA等^[41]将可靠性为中心的维护(reliability centered maintenance, RCM)与建筑信息模型(building information modeling, BIM)、地理信息系统(geographic information system, GIS)深度融合,提出动态RCM流程框架,借助蒙特卡罗仿真量化设备失效风险,并通过虚拟校园原型系统验证了该方法在降低人工成本方面有显著效果。BIM和GIS集成在一起,以支持RCM流程所需数据的采集和更新,为工业设备全生命周期管理提供了可扩展的理论与实践范式。

2.2 混合建模与物理知识融合的增强策略

为突破纯数据驱动方法在物理可解释性方面的瓶颈,融合领域知识的混合建模技术正在重塑装备智能维修决策的理论框架。通过将物理仿真模型与数据驱动算法深度耦合,研究者构建了兼具高预测精度与机理透明度的新型模型,显著提升了工业场景中的决策可信度与工程适用性。

针对传统纯物理模型(计算成本高、泛化性弱)与纯数据驱动模型(黑箱特性、机理缺失)的局限性,基于物理的混合数据驱动模型通过三类创新范式实现优势互补:(1)基于物理的机器学习(嵌入物理约束优化特征学习);(2)机器学习辅助仿真(利用数据驱动代理模型加速物理计算);(3)可解释AGI(通过因果推理增强模型透明度)。WANG等^[42]系统总结了上述模型的原理与特性,并验证其在智能制造三大核心环节——产品设计(如参数优化)、运维(如故障模式识别)及智能决策(如维护策略生成)中的应用潜力。以设备健康管理为例,GONG等^[43]提出了一种集成数字孪生与AGI的创新框架:通过构建滚动轴承的高保真孪生模型动态映射物理退化过程,结合贝叶斯深度学习量化剩余使用寿命(remaining useful life, RUL)预测中的不确定性,其平均绝对误差(MAE=0.071)与均方根误差(RMSE=0.084)较传统方法(MAE=0.100, RMSE=0.118)降低约30%,凸显了混合模型在精度与可靠性方面的优势。然而,由于物理-数据模型协同优化的复杂性(如多物理场耦合建模、异构数据融合),此类方法在流程工业中仍面临严峻挑战,尤其是可解释性与预测精度间的权衡问题。

未来研究需聚焦多模态数据标准化、超参数机理化设计及轻量化部署架构,以推动混合建模技术在工业设备全生命周期管理中的规模化落地。

2.3 数字孪生与边缘计算的实时决策支持

数字孪生与边缘计算的深度融合正在重构工业设备维护的实时决策范式。在新一代智能制造与工业物联网背景下,集成化、大规模化与复杂化的制造过程对设备状态监测与故障检测提出了更高要求^[44],尤其是多生产子过程协同耦合导致的跨层级故障传播问题日益凸显。例如,在化工或汽车制造场景中,单设备异常可能通过物料流、能量流或信息流迅速扩散至全系统,引发连锁性质量缺陷或停机风险。

针对此类挑战,ZHANG等^[45]构建了基于云边协同的质量相关分层故障检测框架:云端通过图神经网络建模全局生产流程拓扑关系,边缘端则利用轻量化卷积网络实时分析本地传感器数据,结合动态权重分配机制实现子流程间故障特征的多粒度融合。试验表明,该框架在半导体晶圆制造产线中将故障误报率降低了18.3%,检测响应时间缩短至毫秒级。

在RUL预测领域,AIVALIOTIS等^[46]提出了一种基于物理仿真与数字孪生的非侵入式预测方法,其核心在于构建高精度机电耦合模型,通过融合有限元分析与实时振动频谱数据,动态修正轴承磨损系数与润滑退化方程。在某汽车装配线工业机器人案例中,该方法成功预测了谐波减速器在连续负载波动下的RUL,预测误差较传统振动分析法降低42%,且无需安装额外传感器。然而,传统边缘客户端受限于计算能力与数据孤岛问题,难以支撑高维物理模型的实时运算。为此,GUO等^[47]设计了基于联邦学习的分布式预测框架:采用分块卷积自动编码器在边缘端提取局部退化特征,云服务器通过自适应聚合算法整合全局模型参数,并引入差分隐私机制保护客户端数据所有权。在风电齿轮箱多机组协同预测场景中,该框架仅需20%的边缘数据参与即可达到集中式训练95%的预测精度,同时减少68%的通信成本。

在控制系统验证优化方面,MYKONIATIS等^[48]开发的模块化生产系统数字孪生仿真器,创新性地将离散事件仿真(用于模拟物料流转逻辑)与基于智能体的建模(刻画设备自主决策行为)相结合,支持PLC代码的虚拟调试与动态参数优化。实际应用表明,该仿真器可将汽车电子控制单元产线的调试周期从14天压缩至8.5天,并通过数字线程技术使得工艺参数迭代效率提升55%。尽管如此,当前技术仍面临多重挑战:首先,数字孪生模型保真度受限于

多物理场耦合建模的精度损失(如热-力-电耦合效应简化假设导致 RUL 预测偏差);其次,边缘端轻量化模型与高保真仿真的矛盾突出,例如联邦学习中本地模型压缩可能丢失关键退化模式特征;此外,云边协同架构的实时性瓶颈在微秒级控制场景中尤为显著。未来需突破多尺度联合建模方法(从微观材料损伤到宏观系统行为)、边缘自适应量化感知训练技术,以及基于 5G-TSN 的确定性网络传输机制,以实现智能维护系统在复杂工业环境中的可靠落地。

3 基于 RL 的装备智能维修决策技术研究现状

近年来,随着 RL 技术重要性的提升,其在装备智能维修决策中的应用得到逐步拓展,吸引了越来越多的研究人员寻求提升装备维修效率和准确性的智能方法。使用 AGI 制定维护计划对许多行业都极具吸引力,它拥有智能的决策工具,可以提供最佳的维护策略,最大限度地降低预期的维护成本^[49]。调查显示,工业制造、能源、航空航天^[50]和交通运输是 RL 技术在维修决策中应用较多的行业。考虑到这些行业对设备可靠性、安全性和维护效率的高要求,下面将详细探讨基于 RL 的装备智能维修决策在这些领域中的应用,以便更加深入理解这一技术的实践意义。

3.1 工业制造

在工业制造领域,设备的高效运行和可靠性是保障生产效率与质量的关键。基于 RL 的装备智能维修决策技术为工业制造中的设备维护提供了动态优化的解决方案。通过将设备的运行状态、故障模式和维修资源等因素纳入考虑,RL 算法能够动态生成最优的维修排故策略,从而在保证设备可靠性的前提下,降低维修成本和停机时间。

工业制造环境中的设备通常面临复杂的运行条件和多变的故障模式。传统的维修策略(如定期维修和故障后维修)往往难以适应这种动态变化,导致维修成本过高或设备可靠性不足。近年来,RL 算法通过与环境的交互学习,能够根据设备的实时状态动态调整维修策略。MISHRA 等^[51]应用了一种名为 Jaya 算法的新元启发式算法,以及模拟退火和粒子群优化这两种流行算法来优化目标函数,通过先进的推迟平衡方法实现预防性维护间隔的调整,与传统的集成调度和维护优化方法相比,该方法具有更好的经济性能。

预测性维护是通过监测设备状态,预测设备故

障时间,并提前安排维修以避免意外停机。RL 算法在预测性维护中的应用主要体现在利用设备的历史数据和实时状态信息,动态调整维护策略。ONG 等^[52]提出了一种基于 DRL 的预测性维护模型,用于有效管理工业物联网中的资源。这种模型能够根据设备的实时状态动态调整维护策略,适应复杂的工业环境。此外,通过优化资源分配,RL 算法还能够提高设备的利用率和生产效率。

针对行星齿轮箱在多工况下故障表征具有差异性,现有方法中存在特征提取不足,且泛化性和诊断准确率低的问题,王辉等^[53]提出一种基于多尺度 DQN 的 DRL 行星齿轮箱多工况智能诊断方法。通过行星齿轮箱的多工况试验和实际案例的测试与分析,表明所提方法具有更高的诊断准确率和较强的工况适应性。

在复杂的工业制造系统中,多个设备或系统之间的协同维护是一个重要问题。多智能体 RL 通过多个智能体之间的协作,能够更有效地管理复杂的维护任务。NAJAFI 等^[54]为半马尔可夫决策过程提出了一种改进的 DRL 算法,将所提算法应用于具有损伤自愈特性的多机组水力发电系统,以证明 DRL 算法在降低成本方面优于替代策略,并解释了提高系统可靠性如何通过特定技术的学习过程降低成本。

3.2 能源

在能源领域,设备的可靠性和高效运行对于保障能源供应的稳定性和经济性至关重要。基于 RL 的智能维修决策技术在能源系统中的应用主要集中在发电设备、输电网络和分布式能源系统等方面。通过动态优化维修策略,RL 算法能够降低能源系统的运行成本,提高设备的可靠性和可用性。

发电设备(如风力发电机、燃气轮机和太阳能电池板)通常面临复杂的运行环境和多变的故障模式。RL 算法通过监测设备的实时状态,动态调整维修策略,以降低维修成本和停机时间。海上风能是可持续发电的基石,然而随着风电场为追求成本而扩大规模,其运营挑战日益加剧,包括需要应对尾流效应和天气波动导致的电力波动。LEE 等^[55]引入了一个域知情的深度 Q 网络框架,旨在优化维护资源的分配和维护任务的战略选择,并通过卷积层的集成,加快了算法的收敛,强调了该模型在提高大型海上风电场运营生产力方面的巨大潜力。雨水侵蚀损害被认为是导致风力涡轮机叶片前沿侵蚀的最关键因素之一,FANG 等^[56]提出了一种基于 DRL 的优化方法,用于寻找不同降雨强度和风速下的最佳转子转速,该方法与传统的梯度和进化设计优化方法相

比,计算时间显著缩短。董文康等^[57]为提高海上风电机组设计寿命周期内的运维经济性,构建了状态维护与备件库存联合优化策略,验证了所提联合优化方法的优越性,并讨论了RL的探索率、风电场的可达率对运维成本的影响。燃气轮机的维护成本明显高于其原始购买成本。WAN等^[58]提出了一种利用多环境时间相似性理论的新型条件预后维持方法。该方法将同类型的燃气轮机视为参考系统,并选择运行时间和启动次数作为主要劣化指标,同时提出了实时RL、历史RL及其计算算法,以获取更准确的历史退化数据,用于决策。YANG等^[59]提出了一种融合光伏、热电及热能利用的混合太阳能系统,通过RL动态优化多能协同控制策略,实时调节能量分配以匹配环境波动(如光照、温度),最终实现比传统方法更高的综合能源转换效率。

RL算法可以通过监测输电线路的实时状态,动态调整维修策略,以降低故障风险和维修成本。DU等^[60]提出了一个两阶段学习框架,以确定最佳恢复策略,并与具有代表性的基于模型的方法和标准的深度确定性策略梯度方法进行了比较,证明求解的准确性和计算效率的提高。微电网的运行面临诸多挑战,包括间歇性能源供需的波动性、传统电网意外故障的随机发生以及受运行条件影响较大的能源存储系统的退化。为应对这些挑战,PINCIROLI等^[61]提出了一种基于DRL的微电网ESS运维联合管理方法。该工作着眼于整个微电网生命周期内ESS的退化及传统电网的意外故障问题,实现了在微电网整个生命周期内有效最大化利润和最小化维护干预的规范性维护。

3.3 航空航天

在航空航天领域,装备的可靠性和安全性是至关重要的^[62]。基于RL的智能维修决策技术在航空航天中的应用主要集中在飞机发动机、航空电子设备和飞行控制系统等方面^[63]。通过动态优化维修策略,RL算法能够降低维修成本,提高设备的可靠性和安全性。

飞机发动机的可靠性和安全性是飞行安全的关键。RL算法通过监测发动机的实时状态,动态调整维修策略,以降低维修成本和停机时间。何超等^[64]针对航空发动机内部损伤实时检测问题,提出了基于YOLOv4框架下的目标识别网络模型,基于训练好的模型,对图片和视频流上的损伤进行实时检测,具有很强的工程实用价值。

此外,多目标优化方法也被用于飞机发动机的维修策略优化,通过建模多目标奖励函数,平衡维修成本和可靠性,提高了复杂装备的维修效率和环境

适应性。

航空电子设备的可靠性和安全性对于飞行安全同样重要。RL算法通过实时监测设备状态,动态调整维修策略,降低了故障风险和维修成本。有效优化维护策略对于降低复杂设备的成本至关重要。目前该领域的大多数研究主要集中在单目标优化上,通常以基于成本的方法为中心。然而,这种方法可能不适用于航空发动机等复杂设备,这些设备在复杂条件下运行并表现出非线性退化,从而带来严峻的可靠性挑战。为了解决这个问题,WEI等^[65]集成了分解策略、基于邻域的参数转移策略和DRL,并将此框架应用于考虑经济依赖性的航空发动机维护策略的优化,能够在保证可靠性的同时降低维护成本,并且能够适应环境的动态变化。同时,LEE等^[66]提出了从传感器测量到数据驱动的概率性RUL预测,再到维修规划的预测性维修路线图,可用于航空电子设备的智能维修,以降低故障风险并优化维修资源分配。

SILVA等^[67]提出基于RL的航空维修调度法,融合静态与自适应算法,为航空维修效率提升提供新途径。此外,LI等^[68]尝试通过加强符号学习方法有效预测涡轮机叶片寿命,该方法有望应用于涡轮机叶片以外的各种寿命有限的部件。

未来,随着技术的发展,基于RL的智能维修决策技术有望在航空航天领域发挥更大的作用,进一步提高设备的可靠性和安全性,降低维修成本。

3.4 交通运输

在交通运输领域,基于RL的智能维修决策已成为应对日益复杂的基础设施和载具健康管理需求的变革性范式。

桥梁的维护规划通常面临多重挑战,主要与复杂性和规模有关。这些挑战源于每座桥梁中大量的结构构件,以及围绕其健康状况的不确定性,这些状况是通过构件级别的目视检查进行监控的。最近的发展依赖于通过DRL技术来解决维护规划问题,目的是最大限度地降低长期成本^[69]。桥梁数量庞大,桥梁资产管理问题的复杂性和维度要求更智能的顺序决策支持工具,以最佳方式分配有限的资源进行主动维护规划,同时最大限度地提高长期社会经济回报,提高国家交通网络的弹性和可持续性,从而提高经济性。DU等^[70]提出了一个参数化的DRL框架,通过案例研究公路桥梁组合以评估和证明所提出的方法的优点。从结果中发现,参数化DRL策略提供了卓越的顺序维护决策,可以更好地帮助桥梁结构适应长期恶化,同时最大限度地降低累积的社会经济成本。

铁路轨道系统由包括轨道、轨枕、道砟和底基层等各种部件组成,桥梁、隧道和路堤等配套设施在决定运行安全性和乘坐舒适性方面也起着重要作用。LEE等^[71]提出了一种基于随机轨道劣化模型和DRL方法的数据驱动铁路轨道养护调度框架,形成了具有成本效益的最佳轨道维护计划,通过将DRL模型与LAM框架相结合,以提高维护效率,制定最佳的铁路更新和维护计划,以最大限度地降低长期成本和故障风险,但该方法仍存在内在的不确定性和约束问题,可能引发其他具有挑战性的工程问题。MOHAMMADI等^[72]开发了一种基于DRL的DDQN方法来优化维护和更新计划,该方法通过考虑成本效益和风险,优化了规划范围内的更新和维护,将DDQN方法应用于I类铁路网络,有望减少预算的同时,提高网络的可靠性和安全性。

高速公路里程和使用寿命的增加推动了对道路维护的需求。CAI等^[73]提出了一种细粒度的维护决策模型,该模型通过专家反馈RL将数据驱动的方法与专业知识相结合,在决策模型中引入的专家经验可以提高决策的工程应用能力,继而更好地适应工程场景和应用,并实现卓越的成本效益。

为路面健康制定适当的维护计划是保持良好服务水平和承载能力的关键。为应对日益增长的路面养护需求,HAN等^[74]提出了一种基于近端策略优化算法的路面养护智能决策模型,将该模型应用于江苏省公路养护决策,验证RL模型的决策总体准确率为82.2%,较人工神经网络模型提高了17.2%。

4 现存问题与挑战

尽管基于RL的装备智能维修决策技术在工业制造、能源、航空航天和交通运输等领域展现出了巨大的应用潜力,但目前仍存在着诸多亟待解决的问题。

一、从算法本身来看,RL算法普遍存在收敛速度慢和计算资源需求高的问题。这在面对大规模、复杂多变的实际维修场景时,严重制约了其决策效率和实时性,难以满足工业制造中快速响应的生产需求以及能源领域对实时维修决策的要求。

二、模型的可解释性差是一大障碍。在工业制造和航空航天等对安全性要求极高的领域,由于RL模型往往是基于复杂的函数近似(如DNN),其决策过程难以理解,工程师和技术人员难以信任模型所给出的维修策略,进而限制了该技术在工程中的推广应用。

三、数据获取与隐私问题同样不容忽视。一方面,在实际的维修场景中,获取高质量、完整且准确

的设备状态数据难度较大,如航空航天领域中飞行控制系统和航空电子设备的状态数据采集成本高且易受多种因素干扰;另一方面,工业制造和能源领域的设备状态数据常包含企业的核心技术和商业秘密,数据隐私和安全性问题突出,一旦泄露,将给企业带来巨大的经济损失,而目前数据加密、访问控制等隐私保护措施尚不完善,无法充分保障数据的安全性。

四、在多智能体系统的应用场景下,智能体之间的通信和协作效率低下。信息传递的延迟或不准确导致无法实现全局最优的维修策略,且不同智能体的目标可能存在冲突。王冲等^[75]提出了一种基于多智能体DRL的随机事件驱动故障恢复策略:将电力交通耦合网故障恢复中的随机事件驱动问题描述为半马尔可夫随机决策过程问题,综合考虑系统故障恢复优化目标^[76],利用多智能体DRL算法对所构建的随机事件驱动模型进行求解。

综合来看,基于RL的装备智能维修决策技术面临以下挑战,如图9所示。



图9 RL技术在智能维修决策应用方面面临的问题与挑战
Fig. 9 The problems and challenges faced by the RL technology in the application of intelligent maintenance decision-making

(1) 改进与创新算法

通过优化模型结构、改进学习策略等方式来加快RL算法收敛速度。增强算法的鲁棒性,设计更具鲁棒性的RL算法,如采用多种数据增强技术和鲁棒优化方法,提高模型在面对数据噪声和环境变化时的适应能力,使其在面对复杂多变的维修环境时能够保持稳定的性能。

(2) 提升模型的可解释性

聚焦RL模型可解释性工具与方法研发,设计适配RL特性的解释体系;探索模型架构构建,从底层设计实现决策逻辑的透明化与可追溯性。

(3) 优化数据质量

随着数据采集和预处理技术的提升,加强大规模、高质量设备状态数据的采集和预处理技术研究。同时,在工业制造和能源领域,开发更先进的传感器技术和数据清洗算法,提高数据的质量和可用性。

(4) 强化数据隐私保护

强化数据隐私和安全保护,采用先进的加密技

术和数据访问控制机制,确保设备状态数据在采集、传输和存储过程中的隐私和安全。

5 结论与展望

本文系统梳理了RL理论与装备维修决策交叉领域的研究进展。深入解读SARSA、Q-Learning、Actor-Critic等核心算法,分析了装备智能维修决策技术现状及工业制造、能源、航空航天和交通运输等四大典型应用场景,明确了当前研究在算法收敛、计算效率、可解释性和数据隐私等方面的瓶颈。为了推动RL技术在复杂装备维修场景的深度落地应用同时为智能运维领域提供更具工程适应性的解决方案,未来研究需聚焦下述四个方面:(1)开发更高效的RL算法以加快收敛速度、降低计算复杂度,并结合轻量级神经网络架构以适应实时性要求高的维修决策场景;(2)提升算法的鲁棒性以应对复杂多变的环境,借助可视化技术等手段开发可解释性工具和方法,构建具有可解释性的模型架构,如结合符号推理和机器学习,使模型能提供清晰的决策依据,增强在实际应用中的可信度,加强数据采集和预处理技术,提高数据质量,同时采用先进加密技术和数据访问控制机制,保障设备状态数据的隐私和安全;(3)提升RL算法对复杂环境的建模能力和动态适应性,使其能及时调整维修策略以应对实际场景中的不同情况;(4)优化多智能体系统中的通信和协作机制,设计高效的协调方法以解决智能体间的利益冲突问题。此外,应拓展RL在更多领域的应用场景,如海洋工程设备、医疗设备等,并探索其在多领域融合系统中的应用,实现整个系统的协同优化。

RL技术正处于快速发展阶段,其在装备智能维修决策领域的应用前景十分广阔。通过克服现有挑战,并持续创新和改进,RL技术将成为推动工业4.0和智能制造发展的关键技术之一。未来的研究应聚焦于提高其智能化水平、实时性能和应用广度等方面,同时确保安全性和隐私性,以实现其在更广泛领域的应用和发展。

参考文献:

[1] 周东华,魏慕恒,司小胜.工业过程异常检测、寿命预测与维修决策的研究进展[J].自动化学报,2013,39(6):711-722.
ZHOU Donghua, WEI Muheng, SI Xiaosheng. A survey on anomaly detection, life prediction and maintenance decision for industrial processes[J]. Acta Automatica Sinica, 2013, 39(6): 711-722.

[2] 韩中,程林,熊金泉,等.大数据结构化与数据驱动的复杂系统维修决策[J].自动化学报,2020,46(2):385-396.
HAN Zhong, CHENG Lin, XIONG Jinqun, et al. Complex system maintenance decisions based on big data structuration and data-driven[J]. Acta Automatica Sinica, 2020, 46(2): 385-396.

[3] 孙见忠,易杨,文洪,等.航空发动机全寿命维修概率建模与仿真[J].航空动力学报,2022,37(3):573-588.
SUN Jianzhong, YI Yang, WEN Hong, et al. Life cycle probabilistic modeling and simulation of aero-engine maintenance[J]. Journal of Aerospace Power, 2022, 37(3): 573-588.

[4] 高贵兵,王俊深,岳文辉,等.基于脆弱性的制造设备故障智能诊断与维护[J].机械工程学报,2020,56(23):141-149.
GAO Guibing, WANG Junshen, YUE Wenhui, et al. Fault diagnosis and maintain of manufacturing equipment based on vulnerability[J]. Journal of Mechanical Engineering, 2020, 56(23): 141-149.

[5] 袁焯,张永,丁汉.工业人工智能的关键技术及其在预测性维护中的应用现状[J].自动化学报,2020,46(10):2013-2030.
YUAN Ye, ZHANG Yong, DING Han. Research on key technology of industrial artificial intelligence and its application in predictive maintenance[J]. Acta Automatica Sinica, 2020, 46(10): 2013-2030.

[6] 赵俊华,文福拴,黄建伟,等.基于大语言模型的电力系统通用人工智能展望:理论与应用[J].电力系统自动化,2024,48(6):13-28.
ZHAO Junhua, WEN Fushuan, HUANG Jianwei, et al. Prospect of artificial general intelligence for power systems based on large language model: theory and applications[J]. Automation of Electric Power Systems, 2024, 48(6): 13-28.

[7] FARLEY B, CLARK W. Simulation of self-organizing systems by digital computer[J]. Transactions of the IRE Professional Group on Information Theory, 1954, 4(4):76-84.

[8] SUTTON R S. Learning to predict by the methods of temporal differences[J]. Machine Learning, 1988, 3(1): 9-44.

[9] WATKINS C J C H, DAYAN P. Q-Learning[J]. Machine Learning, 1992, 8(3): 279-292.

[10] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning[J]. arXiv preprint arXiv: 1312.5602, 2013.

[11] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[12] AL-HAMADANI M N A, FADHEL M A, ALZUBAIDI L, et al. Reinforcement learning algorithms and applications in healthcare and robotics: a comprehensive and systematic review[J]. Sensors, 2024, 24(8): 2461.

[13] CHEN Y L, CAI Y R, CHENG M Y. Vision-based robotic

- object grasping: a deep reinforcement learning approach[J]. *Machines*, 2023, 11(2): 275.
- [14] XIE Z, LIANG X Q, ROBERTO C. Learning-based robotic grasping: a review[J]. *Frontiers in Robotics and AI*, 2023, 10: 1038658.
- [15] CHEN B M, CHEN X, WU Q, et al. Adversarial evaluation of autonomous vehicles in lane-change scenarios[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(8): 10333-10342.
- [16] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[J]. arXiv preprint arXiv: 1509.02971, 2015.
- [17] LOWE R, WU Y, TAMAR A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. ACM, 2017: 6382-6393.
- [18] 王淳浩, 闭家铭, 阮利, 等. 自动驾驶安全关键场景生成技术综述[J]. *信息与控制*, 2024, 53(1): 17-32.
WANG Chunhao, BI Jiaming, RUAN Li, et al. Survey on automatic driving safety-critical scenario generation technology[J]. *Information and Control*, 2024, 53(1): 17-32.
- [19] GIBNEY E. China's cheap, open AI model DeepSeek thrills scientists[J]. *Nature*, 2025, 638(8049): 13-14.
- [20] MOERLAND T M, BROEKENS J, PLAAT A, et al. Model-based reinforcement learning: a survey[J]. *Foundations and Trends in Machine Learning*, 2023, 16(1): 1-118.
- [21] SILVER D, HUBERT T, SCHRIITWIESER J, et al. A general reinforcement learning algorithm that Masters chess, shogi, and Go through self-play[J]. *Science*, 2018, 362(6419): 1140-1144.
- [22] HUANG Q Y. Model-based or model-free, a review of approaches in reinforcement learning[C]//Proceedings of International Conference on Computing and Data Science (CDS). IEEE, 2020: 219-221.
- [23] MCKENZIE M C, MCDONNELL M D. Modern value based reinforcement learning: a chronological review[J]. *IEEE Access*, 2022, 10: 134704-134725.
- [24] POOLE D L, MACKWORTH A K. *Artificial Intelligence: Foundations of Computational Agents*[M]. Cambridge: Cambridge University Press, 2010.
- [25] Rummery G A, Niranjan M. *On-line Q-Learning using connectionist systems*[M]. Cambridge, UK: University of Cambridge, Department of Engineering, 1994.
- [26] SINGH S P, SUTTON R S. Reinforcement learning with replacing eligibility traces[J]. *Machine Learning*, 1996, 22(1): 123-158.
- [27] FIGUEIREDO PRUDENCIO R, MAXIMO M R O A, COLOMBINI E L. A survey on offline reinforcement learning: taxonomy, review, and open problems[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(8): 10237-10257.
- [28] UEHARA M, SHI C C, KALLUS N. A review of off-policy evaluation in reinforcement learning[J]. arXiv preprint arXiv: 2212.06355; 2022.
- [29] 司彦娜, 普杰信, 于晓升, 等. 基于径向神经网络的多步 Sarsa 控制算法[J]. *控制与决策*, 2023, 38(4): 944-950.
SI Yanna, PU Jiexin, YU Xiaosheng, et al. Multi-step Sarsa control algorithm based on RBF neural network[J]. *Control and Decision*, 2023, 38(4): 944-950.
- [30] CESA-BIANCHI N, GENTILE C, LUGOSI G, et al. Boltzmann exploration done right[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. ACM, 2017: 6287-6296.
- [31] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[J]. *IEEE Transactions on Neural Networks*, 1998, 9(5): 1054.
- [32] MAJID A Y, SAAYBI S, FRANCOIS-LAVET V, et al. Deep reinforcement learning versus evolution strategies: a comparative survey[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(9): 11939-11957.
- [33] PETERS J, SCHAAL S. Natural actor-critic[J]. *Neurocomputing*, 2008, 71(7-9): 1180-1190.
- [34] ZHOU Z Y, LIU G J, ZHOU M C. A robust mean-field actor-critic reinforcement learning against adversarial perturbations on agent states[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(10): 14370-14381.
- [35] LI H P, HE H B. Multiagent trust region policy optimization[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(9): 12873-12887.
- [36] SIMÕES D, LAU N, PAULO REIS L. Multi-agent actor centralized-critic with communication[J]. *Neurocomputing*, 2020, 390: 40-56.
- [37] ZHANG W, PENG G L, LI C H, et al. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals[J]. *Sensors*, 2017, 17(2): 425.
- [38] XU Y D, YAN X A, FENG K, et al. Attention-based multi-scale denoising residual convolutional neural networks for fault diagnosis of rotating machinery[J]. *Reliability Engineering & System Safety*, 2022, 226: 108714.
- [39] HUANG Y J, LIAO A H, HU D Y, et al. Multi-scale convolutional network with channel attention mechanism for rolling bearing fault diagnosis[J]. *Measurement*, 2022, 203: 111935.
- [40] XIE Y N, LIU W, LIU X, et al. Multi-scale deep neural network for fault diagnosis method of rotating machinery[J]. *Ferroelectrics*, 2023, 602(1): 215-230.
- [41] MA Z L, REN Y, XIANG X L, et al. Data-driven decision-making for equipment maintenance[J]. *Automation in Construction*, 2020, 112: 103103.
- [42] WANG J J, LI Y L, GAO R X, et al. Hybrid physics-based

- and data-driven models for smart manufacturing: modelling, simulation, and explainability[J]. *Journal of Manufacturing Systems*, 2022, 63: 381-391.
- [43] GONG F J, MA P, ZHANG H L, et al. Rolling bearings remaining useful life estimation using digital twin and physics-informed methods with uncertainty quantification[J]. *Engineering Applications of Artificial Intelligence*, 2025, 154: 111070.
- [44] HU C H, ZHANG Z M, LI C Y, et al. A state of the art in digital twin for intelligent fault diagnosis[J]. *Advanced Engineering Informatics*, 2025, 63: 102963.
- [45] ZHANG X Y, MA L, PENG K X, et al. A cloud-edge collaboration based quality-related hierarchical fault detection framework for large-scale manufacturing processes[J]. *Expert Systems with Applications*, 2024, 256: 124909.
- [46] AIVALIOTIS P, GEORGOULIAS K, CHRYSOLOURIS G. The use of Digital Twin for predictive maintenance in manufacturing[J]. *International Journal of Computer Integrated Manufacturing*, 2019, 32(11): 1067-1080.
- [47] GUO L, YU Y X, QIAN M G, et al. FedRUL: a new federated learning method for edge-cloud collaboration based remaining useful life prediction of machines[J]. *IEEE/ASME Transactions on Mechatronics*, 2023, 28(1): 350-359.
- [48] MYKONIATIS K, HARRIS G A. A digital twin emulator of a modular production system using a data-driven hybrid modeling and simulation approach[J]. *Journal of Intelligent Manufacturing*, 2021, 32(7): 1899-1911.
- [49] YOUSEFI N, TSIANIKAS S, COIT D W. Dynamic maintenance model for a repairable multi-component system using deep reinforcement learning[J]. *Quality Engineering*, 2022, 34(1): 16-35.
- [50] 刘宝鼎, 于劲松, 韩丹阳, 等. 基于贝叶斯网络强化学习的复杂装备维修排故策略生成 [J]. *北京航空航天大学学报*, 2024, 50(4): 1354-1364.
- LIU Baoding, YU Jinsong, HAN Danyang, et al. Generation of troubleshooting strategy for complex equipment maintenance based on Bayesian network reinforcement learning[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2024, 50(4): 1354-1364.
- [51] MISHRA A K, SHRIVASTAVA D, TARASIA D, et al. Joint optimization of production scheduling and group preventive maintenance planning in multi-machine systems[J]. *Annals of Operations Research*, 2022, 316(1): 401-444.
- [52] ONG K S H, WANG W B, NIYATO D, et al. Deep-reinforcement-learning-based predictive maintenance model for effective resource management in industrial IoT[J]. *IEEE Internet of Things Journal*, 2022, 9(7): 5173-5188.
- [53] 王辉, 徐佳文, 严如强. 基于多尺度注意力深度强化学习网络的行星齿轮箱智能诊断方法 [J]. *机械工程学报*, 2022, 58(11): 133-142.
- WANG Hui, XU Jiawen, YAN Ruqiang. Multi-scale attention based deep reinforcement learning for intelligent fault diagnosis of planetary gearbox[J]. *Journal of Mechanical Engineering*, 2022, 58(11): 133-142.
- [54] NAJAFI S, LEE C G. A deep reinforcement learning approach for repair-based maintenance of multi-unit systems using proportional hazards model[J]. *Reliability Engineering & System Safety*, 2023, 234: 109179.
- [55] LEE N, WOO J, KIM S. A deep reinforcement learning ensemble for maintenance scheduling in offshore wind farms[J]. *Applied Energy*, 2025, 377: 124431.
- [56] FANG J H, HU W F, LIU Z Y, et al. Wind turbine rotor speed design optimization considering rain erosion based on deep reinforcement learning[J]. *Renewable and Sustainable Energy Reviews*, 2022, 168: 112788.
- [57] 董文康, 吴雨芯, 姚琦, 等. 基于深度强化学习的海上风电机组状态维护与备件库存联合优化 [J]. *太阳能学报*, 2023, 44(12): 190-199.
- DONG Wenkang, WU Yuxin, YAO Qi, et al. Joint optimization of state maintenance and spare parts inventory of offshore wind turbines based on deep reinforcement learning[J]. *Acta Energetica Sinica*, 2023, 44(12): 190-199.
- [58] WAN A P, GU F, CHEN J H, et al. Prognostics of gas turbine: a condition-based maintenance approach based on multi-environmental time similarity[J]. *Mechanical Systems and Signal Processing*, 2018, 109: 150-165.
- [59] YANG D J, YIN H M. Energy conversion efficiency of a novel hybrid solar system for photovoltaic, thermoelectric, and heat utilization[J]. *IEEE Transactions on Energy Conversion*, 2011, 26(2): 662-670.
- [60] DU Y, WU D. Deep reinforcement learning from demonstrations to assist service restoration in islanded microgrids[J]. *IEEE Transactions on Sustainable Energy*, 2022, 13(2): 1062-1072.
- [61] PINCIROLI L, BARALDI P, COMPARE M, et al. Optimal operation and maintenance of energy storage systems in grid-connected microgrids by deep reinforcement learning[J]. *Applied Energy*, 2023, 352: 121947.
- [62] 曹明, 王鹏, 左洪福, 等. 民用航空发动机故障诊断与健康现状、挑战与机遇 II: 地面综合诊断、寿命管理和智能维护维修决策 [J]. *航空学报*, 2022, 43(9): 625574.
- CAO Ming, WANG Peng, ZUO Hongfu, et al. Current status, challenges and opportunities of civil aero-engine diagnostics & health management II: comprehensive off-board diagnosis, life management and intelligent condition based MRO[J]. *Acta Aeronautica et Astronautica Sinica*, 2022, 43(9): 625574.
- [63] HU Y, MIAO X W, ZHANG J, et al. Reinforcement learning-driven maintenance strategy: a novel solution for long-term aircraft maintenance decision optimization[J]. *Computers & Industrial Engineering*, 2021, 153: 107056.

- [64] 何超, 陈果, 王雨薇. 基于深度学习的航空发动机内部损伤实时检测方法 [J]. 航空动力学报, 2023, 38(8): 1857-1864.
HE Chao, CHEN Guo, WANG Yuwei. Real-time detection method of aero-engine internal damage based on deep learning[J]. Journal of Aerospace Power, 2023, 38(8): 1857-1864.
- [65] WEI Z Q, ZHAO Z B, ZHOU Z, et al. A deep reinforcement learning-driven multi-objective optimization and its applications on aero-engine maintenance strategy[J]. *Journal of Manufacturing Systems*, 2024, 74: 316-328.
- [66] LEE J, MITICI M. Deep reinforcement learning for predictive aircraft maintenance using probabilistic Remaining-Useful-Life prognostics[J]. Reliability Engineering & System Safety, 2023, 230: 108908.
- [67] SILVA C, ANDRADE P, RIBEIRO B, et al. Adaptive reinforcement learning for task scheduling in aircraft maintenance[J]. *Scientific Reports*, 2023, 13: 16605.
- [68] LI P, CHOI J H, ZHANG D Y, et al. Reinforced symbolic learning with logical constraints for predicting turbine blade fatigue life[J]. *Aerospace Science and Technology*, 2025, 158: 109888.
- [69] HAMIDA Z, GOULET J A. Hierarchical reinforcement learning for transportation infrastructure maintenance planning[J]. Reliability Engineering & System Safety, 2023, 235: 109214.
- [70] DU A, GHAVIDEL A. Parameterized deep reinforcement learning-enabled maintenance decision-support and life-cycle risk assessment for highway bridge portfolios[J]. *Structural Safety*, 2022, 97: 102221.
- [71] LEE J S, YEO I H, BAE Y. A stochastic track maintenance scheduling model based on deep reinforcement learning approaches[J]. Reliability Engineering & System Safety, 2024, 241: 109709.
- [72] MOHAMMADI R, HE Q. A deep reinforcement learning approach for rail renewal and maintenance planning[J]. Reliability Engineering & System Safety, 2022, 225: 108615.
- [73] CAI W Y, DU Y C, WU D F, et al. Engineering-adaptive pavement maintenance decision-making model: a reinforcement learning approach from expert feedback[J]. IEEE Transactions on Intelligent Transportation Systems, 2025(99): 1-16.
- [74] HAN C J, MA T, CHEN S Y. Asphalt pavement maintenance plans intelligent decision model based on reinforcement learning algorithm[J]. *Construction and Building Materials*, 2021, 299: 124278.
- [75] 王冲, 石大夯, 万灿, 等. 基于多智能体深度强化学习的随机事件驱动故障恢复策略 [J]. 电力自动化设备, 2025, 45(3): 186-193.
WANG Chong, SHI Dahang, WAN Can, et al. Uncertain event-driven fault recovery strategy based on multi-agent deep reinforcement learning[J]. Electric Power Automation Equipment, 2025, 45(3): 186-193.
- [76] MIKOŁAJEWSKAE, MIKOŁAJEWSKID, MIKOŁAJCZYK T, et al. Generative AI in AI-based digital twins for fault diagnosis for predictive maintenance in industry 4.0/5.0[J]. *Applied Sciences*, 2025, 15(6): 3166.

第一作者: 万心仪 (2002—), 女, 硕士研究生。

E-mail: guqing0412@qq.com

通信作者: 李传扬 (1991—), 男, 博士, 副教授。

E-mail: li_chuan yang@yeah.net