



基于强化学习的智能控制系统设计 与工程实践研究

牟多铎

(马来西亚理工大学, 马来西亚 柔佛州 士古来 81310)

摘要：在智能制造、无人系统、智慧交通与能源管理等领域快速发展的背景下，控制系统所面对的对象呈现出强非线性、多变量耦合、扰动频繁与运行条件动态变化等特点。传统控制方法在工业现场仍占据主导地位，但在复杂工况下容易受到模型误差、参数漂移与环境变化的影响，导致控制性能下降，甚至出现稳定性风险。强化学习作为一种以交互反馈为基础的智能决策方法，能够在缺乏精确数学模型的情况下，通过环境试错与奖励反馈实现控制策略优化，为复杂系统控制提供了新的技术路径。本文围绕强化学习在智能控制系统中的应用，系统梳理其理论基础与算法特征，提出面向工程落地的强化学习控制系统设计框架，并从状态建模、动作约束、奖励构建、训练机制、部署架构、安全控制与工程可维护性等方面总结关键技术要点。在此基础上，结合工业过程控制、移动机器人控制、无人机姿态调节与能源管理调度等典型应用场景，对强化学习控制的工程实践过程进行分析，归纳其在真实工业系统中面临的挑战，包括样本效率不足、训练成本高、仿真与现实差距、安全探索风险与可解释性不足等问题。最后，本文提出强化学习控制系统工程化优化路径，如数字孪生驱动训练、强化学习与传统控制融合、安全约束机制设计以及在线迁移与持续学习策略。研究表明，强化学习在智能控制领域具有显著应用潜力，但其大规模工程应用仍需在安全性、稳定性与可验证性方面进一步突破。

关键词：强化学习；智能控制；深度强化学习；控制系统设计；工程实践；安全约束

收稿日期：2026 年年 1 月 20 日

中图分类号：TP29

通讯作者：牟多铎，马来西亚理工大学

Design and Engineering Practice of Intelligent Control Systems Based on Reinforcement Learning

Duoduo Mou

(Universiti Teknologi Malaysia, Skudai 81310, Johor, Malaysia)

Abstract: With the rapid development of intelligent manufacturing, unmanned systems, smart transportation, and energy management, control systems are increasingly required to handle objects characterized by strong nonlinearity, multivariable coupling, frequent disturbances, and dynamically changing operating conditions. Traditional control methods still dominate industrial applications; however, under complex working conditions, they are often affected by model errors, parameter drift, and environmental variations, which may lead to degraded control performance and even stability risks. Reinforcement learning, as an intelligent decision-making approach based on interactive feedback, can optimize control policies through trial-and-error interactions and reward signals without relying on accurate mathematical models, thereby providing a novel pathway for complex system control. This paper systematically reviews the theoretical foundations and algorithmic characteristics of reinforcement learning in



intelligent control applications. It proposes an engineering-oriented design framework for reinforcement learning-based control systems and summarizes key technical elements, including state modeling, action constraints, reward function construction, training mechanisms, deployment architectures, safety control, and engineering maintainability. Furthermore, by examining typical application scenarios such as industrial process control, mobile robot control, UAV attitude regulation, and energy management scheduling, this study analyzes engineering implementation procedures and identifies major challenges in real industrial environments, including low sample efficiency, high training cost, simulation-to-reality gaps, unsafe exploration risks, and insufficient interpretability. Finally, this paper proposes practical optimization strategies for engineering deployment, such as digital twin-driven training, integration of reinforcement learning with traditional control methods, safety constraint mechanism design, and online transfer and continual learning strategies. The results indicate that reinforcement learning has significant potential in the field of intelligent control; however, large-scale engineering applications still require further breakthroughs in safety, stability, and verifiability.

Keywords: Reinforcement learning; Intelligent control; Deep reinforcement learning; Control system design; Engineering practice; Safety constraints

0 引言

控制技术是现代工业系统与智能装备的重要支撑,广泛应用于化工生产、智能制造、机器人系统、无人机飞行控制以及能源调度等关键领域。随着工业系统向复杂化、智能化与网络化方向演进,被控对象的运行环境更加开放,系统结构更加复杂,外部扰动更加频繁。传统控制方法依赖系统模型或经验参数整定,在面对非线性耦合、参数时变、噪声干扰与多目标优化需求时往往存在局限性。尤其在新型装备系统中,控制对象可能具备高度不确定性,难以通过传统建模方法建立高精度数学模型,导致控制器设计成本高、调试周期长。

近年来,人工智能方法不断渗透到工业控制领域。强化学习作为一种能够通过交互反馈实现策略优化的学习机制,具备自主学习、在线适应与长期回报优化等特征,逐渐成为智能控制研究的重要方向^[1]。与监督学习依赖标注样本不同,强化学习可以在动态环境中不断试错,从而获得最优控制策略。这一特征使其特别适用于复杂系统控制、智能装备自主决策与高维控制问题。

强化学习在控制领域的引入,为传统控制理论体系提供了新的补充。其突出价值主要体现在以下方面:首先,强化学习具有弱模型依赖性,能够在缺乏精确系统模型的条件下进行控制策略学习,降低建模与参数辨识成本。其次,强化学习

能够通过长期收益优化实现多目标控制,适合解决工业系统中稳定性、经济性与能耗控制等多重约束问题^[2]。再次,强化学习在复杂动态环境中具备自适应能力,可在负载波动、外界扰动与系统参数漂移条件下持续优化控制策略,提高系统鲁棒性与智能化水平。

然而,强化学习控制在工程实践中也面临显著挑战,包括训练过程探索风险大、控制输出稳定性不足、对训练数据与算力要求高以及策略可解释性弱等问题。如何在保证安全性与可维护性的前提下实现强化学习控制的工程落地,是当前研究的重点与难点。

1 国内外研究现状

国外学术界较早将强化学习应用于控制系统研究,尤其在机器人控制、自动驾驶与智能调度领域形成了较为系统的研究成果。随着深度学习技术的发展,深度强化学习逐渐成为解决高维控制问题的重要手段,相关算法在连续控制任务中表现出较强的性能优势。与此同时,安全强化学习、约束强化学习与鲁棒强化学习逐渐成为研究热点,以满足工业控制对安全性与稳定性的高要求^[3]。

国内强化学习控制研究起步稍晚,但近年来发展迅速。研究重点主要集中在工业过程优化控制、移动机器人路径控制、无人机自主控制以及电力系统调度优化等领域。部分研究表明,强化学习能够在复杂的工况下实现较好的控制效果,并

在一定程度上提升系统自适应能力^[6]。但总体而言，强化学习在国内工业现场的规模化应用仍较少，主要受限于仿真环境构建困难、工业数据获取受限以及安全验证机制不足等因素。

本文围绕强化学习智能控制系统的设计与工程实践展开研究，主要包括：

第一，梳理强化学习控制理论基础与关键概念，明确强化学习在控制系统中的建模方法；

第二，提出面向工程应用的强化学习控制系统设计框架，给出状态、动作与奖励设计原则；

第三，总结强化学习控制系统训练与部署过程中的关键技术，包括仿真训练、安全保障与实时推理；

第四，结合典型应用场景分析强化学习控制的工程落地过程，提炼工程实践中存在的问题；

第五，提出强化学习控制系统工程化优化路径，为实际应用提供参考。

2 强化学习在智能控制中的理论基础

2.1 强化学习基本思想

强化学习是一种通过交互反馈实现策略优化的学习机制，其核心是智能体在环境中通过不断尝试不同动作，获得奖励反馈，并根据奖励信息

调整决策策略。强化学习强调“试错学习”与“长期收益”，智能体不仅关注即时回报，更关注未来回报累积，从而能够在复杂动态环境中形成全局优化的控制策略^[4]。

在控制系统中，强化学习智能体可以被视为控制器，环境对应被控对象。智能体接收系统状态信息，输出控制动作，并依据系统运行效果获得奖励反馈。通过持续学习，智能体逐渐形成能够稳定控制系统并优化性能的控制策略。

2.2 强化学习控制问题的建模思路

强化学习控制建模的关键在于将控制问题转化为智能体与环境交互的决策过程。一般而言，状态由传感器测量数据构成，动作对应控制输入，奖励函数反映控制目标，例如误差减小、能耗降低、系统稳定运行等。折扣机制用于反映未来收益的重要性，使智能体能够在短期目标与长期目标之间实现平衡^[5]。

控制系统建模应尽可能保证状态信息能够反映系统动态特征，并能够支撑智能体进行有效决策。对于存在延迟或滞后特性的系统，可引入历史状态信息或采用序列输入方式，以增强策略对动态过程的感知能力。

表 1 强化学习主要算法与控制任务适配性对比

算法	动作空间	样本效率	训练稳定性	实时性/部署难度	安全性风险	典型应用场景
Q-learning	离散	低	中	低	中	离散调度、简单过程控制
DQN	离散	中	中	中	中	离散控制、路径决策、资源分配
DDPG	连续	中	偏低	中	高	机器人控制、机械臂、连续执行器
TD3	连续	中	中-高	中	中-高	连续控制、无人系统、工业过程
SAC	连续	较高	高	中-高	中	高维连续控制、鲁棒控制任务
PPO	离散/连续	中	高	中	中	复杂策略学习、仿真训练到部署

2.3 价值评估与策略优化机制

强化学习算法的核心在于对决策策略的评价与优化。价值评估机制用于衡量某一状态或某一动作在长期收益上的优劣，从而为策略更新提供依据。策略优化机制则通过不断调整决策规则，使

智能体在相同环境下获得更高的累积奖励。在智能控制应用中，价值评估不仅反映控制精度，也体现系统运行的经济性与安全性。例如，在能源管理系统中，价值评估可能综合考虑供电稳定性与成本控制，在工业过程控制中则可能同时考虑



输出误差与执行机构损耗^[6]。

2.4 强化学习算法类型及控制适配性

强化学习算法可分为基于价值估计的方法与基于策略优化的方法。基于价值估计的方法适合动作空间较小的离散控制问题,但是在连续控制任务中往往难以直接应用。基于策略优化的方法能够直接输出连续控制动作,更适合工业控制与机器人控制等场景。近年来,结合神经网络的深度强化学习算法显著提升了对高维状态空间的处理能力,使强化学习能够直接处理复杂感知输入,如视觉信息、激光雷达数据等,并在连续动作控制任务中展现出较强性能^[7]。

3 强化学习智能控制系统设计框架

3.1 控制系统总体架构设计

强化学习智能控制系统通常包括数据采集模块、状态处理模块、策略决策模块、执行控制模块与反馈评估模块。数据采集模块负责从传感器或工业控制系统获取实时运行数据,状态处理模块对原始数据进行归一化、滤波与特征提取,策略决策模块基于强化学习模型输出控制动作,执行模块将动作转换为实际控制信号并驱动执行机构,反馈评估模块根据系统输出计算奖励并记录交互数据,用于后续训练与优化^[8]。

在工程系统中,强化学习控制器往往不直接替代原有控制器,而是以辅助控制或上层优化的形式嵌入控制系统,以保证系统运行的可靠性与安全性。常见架构包括“传统控制器主导、强化学习辅助优化”以及“双控制器并行冗余”的部署方式。

3.2 状态变量选取与特征构建

状态设计是强化学习控制系统成功的关键之一。状态变量通常包括系统关键输出、控制误差、执行机构反馈以及外部扰动因素。对于复杂工业系统,状态还可包括系统运行模式、生产工艺参数、设备健康状态等信息,以增强策略对系统运行全局状态的感知能力。

在状态构建过程中,应避免状态维度过高导致训练效率下降,同时也需避免状态信息不足导致策略难以学习有效控制规律。工程实践中常采用降维方法或特征提取方法,将高维传感器数据转换为可用于决策的核心特征集合^[9]。

3.3 动作输出与执行机构约束

强化学习控制动作对应系统输入信号,例如电机转矩、阀门开度、速度指令或功率调度决策。工程控制系统通常具有明确的动作约束,例如控制信号幅值上限、变化率限制与安全阈值约束。若强化学习输出超出执行机构能力范围,可能造成设备损坏或引发系统失稳。

因此,在动作设计阶段必须引入约束机制,包括动作归一化映射、动作限幅处理以及输出平滑滤波。对于高风险系统,还应设置安全控制模块,对策略输出进行实时检查与修正,确保控制信号满足工程安全规范^[10]。

3.4 奖励函数构建与多目标权衡

奖励函数决定强化学习策略优化方向,其设计直接影响控制系统最终性能。控制问题奖励函数通常由误差惩罚、能耗惩罚、稳定性惩罚与安全约束惩罚构成。误差惩罚用于提升跟踪精度,能耗惩罚用于降低控制输入成本,稳定性惩罚用于抑制振荡行为,安全惩罚用于避免输出超限。

在多目标控制场景中,奖励函数往往需要实现不同目标之间的平衡。例如,在工业过程控制中,过度追求误差最小化可能导致执行机构频繁动作,从而加速设备磨损。因此奖励函数应综合考虑控制精度与执行机构寿命,避免策略输出过于激进^[11]。

4 强化学习控制系统训练与部署关键技术

4.1 训练数据采集与样本效率提升

强化学习训练依赖大量交互样本,而工业现场采集样本成本高且存在安全风险。为解决这一问题,工程上常采用仿真训练方式,通过数字孪生环境快速生成交互数据。为了提升样本利用效率,可采用经验回放机制,将历史交互数据重复利用,并通过优先采样方式提高关键样本的训练权重,从而提升学习效率^[12]。

此外,可采用迁移学习方式,将已有控制任务训练得到的策略迁移至相似任务,减少从零开始训练的成本。对于具有相似动态特征的系统,例如不同规格的电机控制系统或不同容量的储能系统,迁移学习具有较强适用性。

4.2 仿真环境构建与数字孪生支撑

仿真环境是强化学习控制工程化的重要基础。



仿真环境需尽可能准确反映真实系统动态特性，包括系统响应特性、执行机构约束、传感器噪声与外部扰动模型。高保真仿真能够提高训练策略的泛化能力，并减少策略在真实系统部署时的性能衰减。

数字孪生技术能够通过实时数据驱动不断更新仿真模型，使仿真环境与真实系统保持一致性，从而形成“仿真训练—现实验证—仿真修正”的闭环迭代机制。这种机制能够显著提升强化学习控制策略的工程适用性^[13]。

表 2 强化学习智能控制系统工程化设计关键要素与建议

设计要素	核心作用	常见问题	工程实践建议
状态空间构建	决定策略可观测性与学习效率	状态冗余、缺失关键变量、噪声干扰	结合控制机理选取关键变量；进行归一化、滤波与特征压缩
动作空间与约束	决定控制执行可行性与安全边界	动作超限、震荡、不可执行	引入动作裁剪、平滑约束；结合执行器模型设置上下限与速率限制
奖励函数设计	决定学习目标与收敛方向	奖励稀疏、目标冲突、局部最优	采用分层奖励与稠密奖励；将能耗、误差、稳定性指标综合设计
安全机制与惩罚项	降低探索风险并提升稳定性	探索导致失稳、设备损伤	引入安全约束RL、屏障函数或安全过滤器；设置终止条件与惩罚策略
仿真环境与数字孪生	支撑训练与测试，减少真实试错成本	仿真与现实差异大、泛化不足	采用域随机化与参数扰动；逐步从仿真迁移到真实系统
离线预训练与在线微调	提升训练效率与现场适应性	在线学习导致性能波动	离线训练获得基线策略；在线微调采用小步更新与稳定性监控
控制架构融合	实现与传统控制系统兼容	难以接入现有PLC/DCS系统	采用RL+PID/MPC混合控制；以RL作为增益调节或策略补偿模块
模型部署与推理延迟	影响实时控制性能	推理速度慢、硬件不适配	模型压缩与量化；边缘计算部署；设置推理周期与控制周期匹配
监测与故障诊断	保障长期运行可靠性	异常状态不可解释、难追溯	建立日志系统与可解释性模块；实时监测奖励、误差与控制量变化
评估指标与验收	保证策略性能可度量	仅看奖励忽略工程指标	同时评估超调量、稳态误差、响应时间、能耗、安全事件次数等

4.3 控制稳定性保障与安全机制设计

强化学习策略在训练初期容易产生随机动作，可能导致控制系统振荡甚至失稳。为确保系统稳定运行，工程实践中常采用安全控制机制，包括动作平滑滤波、控制输出限幅、异常状态快速切换至传统控制器等方法。

此外，可采用约束强化学习思想，在策略学习过程中将安全约束作为重要优化目标，避免策略输出进入危险区域。对于高风险系统，如化工过程与无人机飞行控制，安全机制通常是强化学习控制能否落地的决定性因素^[14]。

4.4 在线部署与实时推理优化

强化学习控制器部署到工程系统时需满足实时性要求，尤其在高频控制任务中，策略推理延迟必须控制在毫秒级。工程实践中可采用模型轻

量化方法，如网络剪枝、模型量化与推理加速框架，提高策略推理速度。

同时，部署系统需要与工业控制平台实现数据接口对接，例如与 PLC、嵌入式控制器或工业通信协议进行兼容。为了提升可维护性，强化学习控制器通常需要提供监控接口，用于实时查看策略输出、系统状态与奖励趋势，便于工程人员进行调试与故障诊断^[15]。

5 强化学习控制的典型工程实践分析

5.1 工业过程控制中的应用实践

工业过程控制系统具有显著的非线性与多变量耦合特征，且运行过程中存在负载扰动与参数漂移现象。传统 PID 控制虽然应用广泛，但在复杂工况下需要频繁调参，且难以在多目标条件下实现全局优化。强化学习能够通过在线学习不断



适应工况变化，在一定程度上提升系统控制性能。

在工业过程控制工程实践中，强化学习通常以“上层优化器”的形式应用，例如用于自动调节 PID 参数、优化设定值或辅助进行控制策略切换。该方式能够在保持系统稳定性的同时引入强化学习的自适应优势，提高控制系统智能化水平^[16]。

5.2 移动机器人运动控制工程实践

移动机器人控制任务包括速度控制、轨迹跟踪与避障决策。传统控制方法依赖预先规划路径与精确环境建模，在复杂动态环境下容易出现路径失效或避障不及时问题。强化学习能够通过环境交互学习自主避障策略，并结合路径规划实现更高水平的自主控制。

工程实践中，移动机器人强化学习控制往往采用仿真训练方式，通过构建多场景、多障碍环境进行策略训练，再通过迁移学习部署到真实机器人平台。为了提升安全性，真实部署阶段通常会增加安全保护层，确保策略输出不会导致碰撞或失控^[17]。

5.3 无人机姿态与轨迹控制应用

无人机飞行控制对实时性与稳定性要求极高，

且系统受到风场扰动、载荷变化与传感器误差影响较大。强化学习能够在仿真环境中学习复杂控制策略，并在一定程度上提升系统对扰动的适应能力。

工程实践中，强化学习控制无人机通常采用分层控制结构：传统控制器负责基本稳定控制，强化学习策略负责上层轨迹优化与自主决策。这种融合方式能够在保证飞行安全的基础上提升无人机自主飞行能力^[18]。

5.4 能源系统调度与储能控制实践

在微电网与储能系统中，控制任务往往涉及多目标优化，如降低运行成本、提高供电稳定性、减少电池损耗等。强化学习能够在动态负荷与新能源波动条件下，通过学习历史运行规律形成最优调度策略。

工程实践中，强化学习常用于储能系统充放电策略优化，并能够根据电价变化、负荷预测与电池状态实现动态决策。相比传统规则控制，强化学习能够在长期收益优化上取得优势，但需要充分考虑电池安全约束与寿命管理问题^[19]。

6 工程落地挑战与优化路径

表 3 强化学习控制典型工程应用场景对比

应用场景	状态特征	动作形式	主要控制目标	仿真依赖度	安全风险等级	工程落地难点
工业过程控制 (化工/冶金)	多变量耦合、扰动频繁	连续调节阀门/泵速	稳定性、能耗最优、质量一致性	高	高	时延、强约束、设备损伤风险
移动机器人导航	位姿、障碍物、传感器融合	速度/转向控制	避障、路径最短、效率最优	中-高	中	泛化能力不足、动态环境适应
无人机姿态与轨迹控制	姿态角、速度、风扰动	连续推力/角速度	轨迹跟踪、抗扰、稳定飞行	高	高	实时性要求高、训练与验证成本高
智能制造柔性产线调度	订单状态、设备状态、库存信息	离散决策/任务分配	效率最大化、延误最小化	中	中	状态空间大、奖励稀疏、系统复杂
能源管理与微电网调度	负荷、价格、储能状态	充放电/功率分配	成本最小、稳定供能	中-高	中	预测误差、长期优化、约束复杂
智能交通信号控制	车流密度、排队长度、路网状态	离散信号相位控制	通行效率、拥堵缓解	中	低-中	路网耦合强、策略迁移难

6.1 训练成本与数据依赖性挑战

强化学习训练成本高是工程应用的主要瓶颈之一。控制任务往往需要大量交互数据才能收敛，而工业现场难以提供大规模试错空间。仿真训练能够部分解决该问题，但高保真仿真环境构建成本高，且需要持续维护更新。针对训练成本问题，

未来可采用基于模型的强化学习方法、离线强化学习方法以及经验迁移机制，以减少真实交互需求，提高训练效率。

6.2 仿真与现实差距导致的策略失效

强化学习策略在仿真中训练后部署到真实系统时，往往出现性能下降甚至失效，主要原因在



于仿真环境与真实系统在动态特性、噪声分布与外部扰动方面存在差异。工程实践中常采用域随机化训练方法,通过在仿真中随机扰动系统参数增强策略泛化能力。此外,通过数字孪生技术动态校准仿真模型,使仿真环境持续接近真实系统,也是解决仿真到现实差距的重要方向。

6.3 安全探索与工业规范约束问题

强化学习依赖探索机制,但工业控制系统通常不允许试错。若探索行为导致系统超限,将带来重大安全风险。因此强化学习控制在工业场景中必须引入安全约束机制,例如在控制输出前增加安全校验模块,或在策略学习过程中引入约束优化目标。未来强化学习控制研究应进一步加强安全强化学习算法研究,并结合工业控制标准构建可验证的安全策略框架,从而提高强化学习控制在高风险系统中的可用性。

6.4 策略可解释性不足与工程信任问题

强化学习控制策略多基于神经网络,缺乏明确物理意义,导致工程人员难以理解其决策逻辑。工业控制系统强调可验证性与可追溯性,缺乏解释机制会降低工程应用信任度。为解决该问题,可通过策略可视化分析、关键状态敏感性分析与控制行为归因方法提升可解释性。同时,可将强化学习策略输出转化为规则或局部线性控制器形式,提高工程可维护性。

6.5 强化学习与传统控制融合的工程化趋势

强化学习在短期内难以完全替代传统控制器,因此融合模式更符合工程需求。融合方式主要包括强化学习用于传统控制器参数整定、强化学习用于上层决策优化以及强化学习与模型预测控制结合等路径。融合策略能够充分利用传统控制的稳定性优势,同时发挥强化学习的自适应优化能力,是强化学习控制工程化应用的重要发展方向^[20]。

7 结论与展望

本文围绕强化学习在智能控制系统中的设计与工程实践展开研究,系统分析了强化学习控制的理论基础、系统架构设计方法以及训练与部署关键技术,并结合工业过程控制、移动机器人、无人机控制与能源管理等典型场景总结了强化学习控制的工程应用模式。研究表明,强化学习能够

在复杂动态环境中实现自适应控制与多目标优化,为解决传统控制方法难以应对的复杂控制问题提供了新的思路。

同时,强化学习控制在工程落地过程中仍面临训练成本高、仿真与现实差距、安全探索风险与可解释性不足等挑战。为推动强化学习控制的工业应用,未来研究需要进一步提升样本效率与训练稳定性,加强安全强化学习与约束机制研究,完善数字孪生支撑下的仿真训练体系,并探索强化学习与传统控制融合的工程化路径。随着工业智能化与边缘计算能力的提升,强化学习控制有望在更多复杂工业系统中实现可靠应用,为智能制造与智慧系统发展提供关键技术支撑。

参考文献:

- [1] 徐东伟,周磊,王达,等.基于深度强化学习的城市交通信号控制综述[J].交通运输工程与信息学报,2022,20(1):15-30.
- [2] 董豪,杨静,李少波,等.基于深度强化学习的机器人运动控制研究进展[J].控制与决策,2022,37(2):278-292.
- [3] 刘志荣,姜树海.基于强化学习的移动机器人路径规划研究综述[J].制造业自动化,2019,41(3):90-92.
- [4] 陈宇轩,王国强,罗贺,等.基于 Actor-Critic 算法的多无人机协同空战目标权重分配方法[J].无线电工程,2022,52:1266-1275.
- [5] 吴昭欣,李辉,王壮,等.基于深度强化学习的智能仿真平台设计[J].战术导弹技术,2020(4):193-200.
- [6] 侯贤乐,李辉,王壮,等.基于 DDPG 算法的导弹末端制导律设计[J].战术导弹技术,2021(4):110-116.
- [7] 陈昊皓,谭泰,蒋太敏,等.基于深度强化学习的固定翼无人机编队控制算法[J].战术导弹技术,2025(3):83-94.
- [8] 侯磊,贾贝熙,杜子亮,等.深度强化学习在无人机智能控制中的应用研究[J].战术导弹技术,2024(6):108-114.
- [9] 李天任,杨奔,汪初,等.基于 Q-Learning 算法的再入飞行器制导方法[J].战术导弹技术,2019(5):44-49.
- [10] 张秦浩,敖百强,张秦雪.Q-learning 强化学习制导律[J].系统工程与电子技术,2020,42(2):414-



419.

[11] 唐文泉, 孙莹, 杨奇, 等. 一种面向 2V2 近距空战的强化学习算法 [J]. 战术导弹技术, 2022 (1): 120-130.

[12] 郑成辰, 李辉, 陶伟, 等. 基于深度强化学习的导弹末端约束角制导律 [J]. 战术导弹技术, 2022 (6): 93-102+110.

[13] 吴婷, 叶林奇, 杨君, 等. 基于强化学习的多无人机避障编队控制 [J]. 飞控与探测, 2025, 8 (2): 9-17.

[14] 相晓嘉, 闫超, 王菡, 等. 基于深度强化学习的固定翼无人机编队协调控制方法 [J]. 航空学报, 2021, 42 (4): 420-433.

[15] 张宇宸, 段海滨, 魏晨. 基于深度强化学习的无人机集群数字孪生编队避障 [J]. 工程科学学报, 2024, 46 (7): 1187-1196.

[16] 唐恒, 孙伟, 吕磊, 等. 融合动态奖励策略的无人机编队路径规划方法 [J]. 系统工程与电子技术, 2024,

46 (10): 3506-3518.

[17] 张斌, 何明, 陈希亮, 等. 改进 DDPG 算法在自动驾驶中的应用 [J]. 计算机工程与应用, 2019, 55 (10): 264-270.

[18] 周友行, 赵哈斌, 刘汉江, 等. 采用 DDPG 的双足机器人自学习步态规划方法 [J]. 计算机工程与应用, 2021, 57 (6): 254-259.

[19] 王馨雪, 黄佳欣, 赵婷婷, 等. 基于有效动作表示的策略搜索强化学习方法 [J]. 天津科技大学学报, 2025, 40 (5): 57-65.

[20] 王金强, 苏日新, 刘莉, 等. Q-learning 强化学习协同拦截制导律 [J]. 导航定位与授时, 2022, 9 (5): 84-90.

作者简介: 牟多铎 (1991-), 男, 汉族, 四川自贡人, 马来西亚理工大学博士研究生, 研究方向为人工智能、计算机视觉、增强现实、算法治理与视觉传播等。