

无人机路径规划中的深度强化学习模型训练与优化

朱晓¹ 朱朝阳^{2*}

1. 广西工业职业技术学院, 中国·广西 南宁 530000
2. 广西大学, 中国·广西 南宁 530004

摘要: 随着无人机技术的迅猛进步, 实现高效且安全的路径规划已成为提升无人机任务执行效能的核心要素。鉴于传统路径规划方法受限于对环境模型的精确依赖及手动策略设计的局限性, 难以灵活应对复杂多变的任务场景, 深度强化学习技术应运而生。该技术融合了深度学习的强大感知能力与强化学习的策略决策优势, 为无人机在复杂环境中自主规划最优路径开辟了新路径。论文深入剖析了深度强化学习模型在无人机路径规划应用中的训练与优化流程, 涵盖状态空间构建、行动策略界定、奖励机制设计、模型筛选与训练以及后续优化等关键环节。同时, 论文也探讨了模型训练与优化过程中面临的挑战, 并提出了针对性的解决策略, 旨在为无人机路径规划的学术研究与实践应用提供有价值的参考框架。

关键词: 无人机路径规划; 深度强化学习; 模型训练; 优化

Training and Optimization of Deep Reinforcement Learning Models in UAV Path Planning

Xiao Zhu¹ Chaoyang Zhu^{2*}

1. Guangxi Vocational & Technical Institute of Industry, Nanning, Guangxi, 530000, China
2. Guangxi University, Nanning, Guangxi, 530004, China

Abstract: With the rapid advancement of drone technology, achieving efficient and safe path planning has become a core element in improving the efficiency of drone mission execution. Given that traditional path planning methods are limited by their precise dependence on environmental models and the limitations of manual strategy design, making it difficult to flexibly respond to complex and changing task scenarios, deep reinforcement learning techniques have emerged. This technology combines the powerful perceptual ability of deep learning with the strategic decision-making advantage of reinforcement learning, opening up a new path for unmanned aerial vehicles to autonomously plan the optimal path in complex environments. This paper deeply analyzes the training and optimization process of deep reinforcement learning models in unmanned aerial vehicle path planning applications, covering key links such as state space construction, action strategy definition, reward mechanism design, model screening and training, and subsequent optimization. At the same time, the article also explores the challenges faced in the process of model training and optimization, and proposes targeted solutions, aiming to provide a valuable reference framework for academic research and practical applications of drone path planning.

Keywords: drone path planning; deep reinforcement learning; model training; optimization

0 前言

在无人机技术迅猛发展的当下, 其应用范畴已横跨军事侦察、民用摄影及物流运输等多个关键领域, 而路径规划能力成为制约无人机任务执行成效的核心要素。传统路径规划方法受限于对精确环境模型的依赖及人工策略设计的局限性, 面对复杂多变的实际场景, 如紧急灾害响应或密集城市导航, 常显得应对不足。深度强化学习技术的引入, 为破解这一难题提供了创新途径。该技术凭借其在交互试错中自主学习最优策略的能力, 赋能无人机在复杂动态环境中实现高效、安全的自主路径规划, 极大提升了任务执行的灵活性与可靠性, 标志着无人机智能化进程的一大飞跃。

1 深度强化学习在无人机路径规划中的应用

1.1 状态表征

在深度强化学习的框架下, 无人机路径规划的状态表征是对无人机所处环境状况的一种精确而全面的映射, 构成了模型决策与学习的基石。状态表征不仅要涵盖无人机自身的关键参数, 如位置坐标、飞行速度及姿态角度, 还需将外部环境因素纳入考量, 诸如障碍物分布、地形特征乃至实时天气状况, 这些均可通过机载传感器与外部信息系统融合获取。合理构建状态表征体系, 旨在向模型提供全面且精炼的信息基石, 避免冗余数据干扰, 从而加速学习进程并提升决策精度。通过精细设计状态表征, 我们确保模型能够精准捕捉环境的动态变化, 为无人机在复杂环境中的高效路径规划

奠定坚实基础。

1.2 行动空间定义

在构建深度强化学习模型以优化无人机路径规划的过程中，行动空间的精确界定是一个至关重要的环节。这一环节要求明确无人机可执行的动作集合，这些动作将直接构成模型决策的输出范围。对于无人机而言，行动空间可能涵盖了基础的方向性移动，如上升、下降、左移、右移、前进及后退，同时也可能包含更高级的操作指令，如悬停、加速或减速等。合理设定行动空间对模型训练成效具有决定性影响，因为它直接关联到模型所能探索的策略多样性和实用性。一个过于宽泛的行动空间可能会使学习过程变得复杂且难以收敛，而行动空间过于狭窄则可能束缚模型的学习潜力，限制其在复杂任务中的表现。因此，在定义行动空间时，必须综合考虑无人机的物理属性、任务特性和实际运行环境，旨在使模型能够掌握既实用又高效的路径规划策略。此外，随着任务要求的演变和环境条件的变化，行动空间也需具备相应的灵活性，以便及时调整，满足新的任务需求。

1.3 奖励函数设计

奖励函数在深度强化学习模型中扮演着举足轻重的角色，特别是在无人机路径规划这一应用场景中，它如同指南针一般，引导着模型不断逼近最优路径策略。此函数的核心任务在于，对无人机的每一步行动进行精准而全面的评估，并通过明确的奖惩机制，激励模型在追求任务目标的同时，也能确保飞行的安全稳定。在设计奖励函数时，我们必须深思熟虑，力求在任务执行的高效性与路径的安全性之间找到完美的平衡点。这就意味着，无人机在成功抵达目标点时应获得丰厚的正向奖励，以此鼓励其积极追求任务目标；而一旦与障碍物发生碰撞或路径出现偏离，则应给予相应的负向惩罚，以确保其始终保持警惕，维护飞行的安全。奖励函数的设计并非易事，它要求我们深入剖析任务需求与环境特性，确保函数既不过于简单以至于产生次优策略，也不过于复杂而增加学习难度。因此，在设计过程中，我们需要进行大量的实验与精细的调整，以期打造出一个能够精准引领模型探索并习得最佳路径规划方案的奖励函数。

奖励函数的一般形式可以表示为：

$$R(s,a)=f(s,a)$$

其中， s 代表无人机的当前状态， a 代表其即将采取的动作， f 代表奖励函数的核心，它根据当前状态和动作，计算并返回相应的奖励值。这个奖励值，正是驱动模型不断学习、不断优化的关键所在。

1.4 模型甄选与训练

在选择深度强化学习模型时，需根据无人机路径规划的具体需求和环境特性进行考虑。常见的模型包括深度 Q 网络 (DQN)、深度确定性策略梯度 (DDPG) 等。模型选定后，需利用仿真或真实场景数据进行训练。模型训练涉及大量数据与环境交互，目标是最大化累积奖励。以下是一些

关键公式和算法。

1.4.1 Q-learning 与 DQN

Q-值函数： $Q(s,a)=R(s,a)+\gamma \max_{a'} Q(s',a')$ ，其中 γ 为折扣因子。

DQN 的损失函数： $L(\theta)=E[(y-Q_{\theta}(s,a))^2]$ ，其中 $y=R(s,a)+\gamma \max_{a'} Q_{\theta}(s',a')$ 为目标 Q 值。

其中， θ 表示神经网络的参数， θ' 是目标网络的参数， r 是当前状态下选择动作所获得的即时奖励， γ 是折扣因子， $\max_{a'} Q(s',a';\theta')$ 是下一个状态 s' 下所有可能的动作 a' 的最大 Q 值。

1.4.2 策略梯度算法

策略： $\pi_{\theta}(a|s)$ ，表示在状态 s 下采取动作 a 的概率。

策略梯度： $\nabla_{\theta} J(\theta)=E[(\nabla_{\theta} \log \pi_{\theta}(a|s))A(s,a)]$ ，其中 $A(s,a)$ 为优势函数，用于评估动作 a 相对于平均表现的好坏。

1.5 模型优化

在深度强化学习应用于无人机路径规划的领域，模型优化构成了提升算法效能的核心环节。为达成这一目标，需精细调整超参数配置，这一过程需借助系统的实验验证与迭代调优，力求发掘出能最大化模型训练成效与实战表现的参数集。同时，对奖励函数的精心调校同样关键，它要求我们通过权重的微妙平衡与函数形式的创新设计，精准导向模型学习，促使其掌握更加贴合复杂任务需求的路径规划智慧。此外，扩充训练数据集亦不失为一条增强模型泛化能力的有效途径，丰富的数据滋养能助力模型在未知环境中展现更稳健的表现。在此基础上，我们还应积极探索如分布式训练架构的融入，以及经验回放、优先级经验回放等高级策略的应用，以期在模型的稳定性、学习效率上实现新的跃升。通过这一系列综合而深入的优化举措，深度强化学习模型方能在复杂多变的无人机路径规划挑战中游刃有余，持续输出高精度、高效率的解决方案。

2 深度强化学习模型训练与优化的挑战及应对策略

2.1 挑战

2.1.1 数据获取与处理

在深度强化学习模型训练过程中，数据获取与处理构成了首要挑战。对于无人机路径规划这一应用场景而言，真实世界数据的稀缺性和高成本成为显著障碍。实际环境的多变性，诸如天气条件的波动、障碍物的动态变化等不确定因素，进一步加剧了数据收集的复杂性。此外，即便成功获取数据，如何高效管理和运用这些数据亦成为关键问题。数据的品质、规模及其能否全面反映场景特性，均对模型的训练成效具有决定性影响。因此，在数据资源受限的情况下，如何确保模型能够精准掌握路径规划策略，成为亟待攻克的技术难题。

2.1.2 模型泛化能力的强化需求

深度强化学习模型在特定训练环境中常展现出优异性

能,但置身于新环境时,其表现却可能显著下滑。这一现象的根源在于模型易过度拟合训练数据,从而削弱了对新环境的适应力。在无人机路径规划的复杂多变场景中,模型需具备出色的泛化能力,以灵活应对不同任务挑战。然而,当前深度强化学习模型在维持高性能与增强泛化能力之间常陷入两难境地。因此,探索如何在不牺牲性能的前提下,有效提升模型的泛化能力,成为该领域亟待解决的关键问题。

2.1.3 训练耗时与稳定性难题

深度强化学习模型在训练时面临效率低下与稳定性不足的双重考验,成为制约其广泛应用的关键瓶颈。在无人机路径规划等复杂场景下,模型需通过大量与环境的交互来逐步优化策略,这一过程极为耗时且资源密集。同时,训练过程中的种种不确定性,如策略更迭的频繁、奖励信号的稀疏,都会导致模型收敛路径的不平稳,进而影响其最终表现。因此,提升训练效率并确保稳定性,成为深度强化学习在无人机路径规划领域亟待攻克的核心挑战。

2.1.4 安全性与可靠性保障

在无人机路径规划的复杂场景中,确保安全飞行与可靠执行是核心要务。然而,深度强化学习模型在自主探索与决策过程中,存在潜在的安全风险。模型或许会倾向于选择高风险策略以获取更高奖励,例如贴近障碍物飞行或穿越危险区域,此类行为无疑增加了无人机事故的风险,并可能给周围环境及人员带来严重后果。同时,模型的可靠性也是实际应用中不可忽视的一环。面对多变的环境条件,如极端天气或设备突发故障,模型需展现出高度的稳定性和适应性。因此,如何在深度强化学习的框架下,构建既安全又可靠的无人机路径规划模型,成为亟待攻克的技术难题。

2.2 应对策略

2.2.1 利用仿真环境和迁移学习解决数据获取与处理问题

在深度强化学习应用于无人机路径规划的实践中,面对数据获取与处理的难题,可以采纳一种创新策略:结合高精度仿真环境与迁移学习技术。通过构建精细化的仿真环境,我们能够全方位模拟复杂多变的飞行场景,为模型提供充足且多样的训练数据,这不仅有效缓解了实际数据采集的局限性,还增强了模型适应不同情境的能力。同时,迁移学习机制的引入,使得模型能够在某一任务中积累的知识和技能高效地转移到新任务上,显著降低了新任务学习过程中的数据依赖,加速了训练进程并提升了整体性能。这一系列举措,不仅展现了深度强化学习在无人机路径规划中的强大潜力,也为我们克服数据挑战提供了切实可行的路径。

2.2.2 采用域随机化和多样性训练提高模型泛化能力

为了提升深度强化学习模型在无人机路径规划中的泛化能力,可应用域随机化和多样性训练策略。域随机化策略通过在训练时不断随机调整环境参数,如障碍物的分布、形状及数量等,为模型提供了丰富的环境变化,促使其学习到

更为鲁棒和适应性强的路径规划策略。此外,多样性训练策略则通过引入多样化的训练场景和任务,使模型能够面对并处理更多不同类型的挑战,从而积累更全面的知识与策略储备。这种综合训练方法不仅增强了模型在复杂多变环境中的表现能力,还显著提升了其应对未知情况时的灵活性和准确性,为无人机路径规划的实际应用奠定了坚实基础。

2.2.3 利用分布式训练和经验回放提高训练效率与稳定性

为了切实提升深度强化学习模型在无人机路径规划中的训练效率和稳定性,可以采纳并实施分布式训练架构以及经验回放机制。具体而言,分布式训练策略通过高效利用计算资源,实现训练任务的并行处理与数据共享,从而大幅度缩减模型训练所需时间,提高训练过程的整体效率。与此同时,经验回放技术及其进阶版——优先级经验回放,为我们提供了更为稳健的学习路径。前者通过构建经验数据库,存储并循环利用历史数据,有效减轻了模型对即时数据的过度依赖,进而降低了训练过程中的不确定性。后者则在此基础上,通过智能评估每条经验的价值,并据此进行有偏采样,确保模型能够迅速捕捉到关键信息,实现更为精准的学习与决策。这些策略的综合运用,不仅提升了深度强化学习模型的训练效率与稳定性,更为其在无人机路径规划等复杂任务中的广泛应用提供了强有力的技术支撑。

2.2.4 加入安全性约束和冗余设计确保安全性与可靠性

在深度强化学习应用于无人机路径规划的实践中,确保模型的安全性和可靠性是至关重要的。为此,我们需将严格的安全性约束融入训练流程,有效限制模型的行动边界,从而规避潜在风险行为。例如,明确设定无人机的禁飞区及最小安全间距等关键参数,以保障其飞行操作始终遵循安全规范。同时,必须执行严苛的测试与验证程序,对模型在实际应用场景中的可靠性进行全面评估。此外,我们还需采纳冗余设计策略,通过部署多架无人机或增设传感器,提升系统的整体容错性能;并结合高效的故障检测与恢复机制,确保在无人机遭遇故障时能迅速识别并妥善处理,有效防止任何意外事故的发生。这一系列综合措施,旨在深度强化学习模型辅助下的无人机路径规划中,构筑起坚不可摧的安全与可靠性防线,为无人机的广泛且安全运用奠定坚实基础。

3 结语

深度强化学习为无人机路径规划带来了前所未有的机遇,通过不断训练与优化,模型能够在复杂环境中实现高效、安全的自主决策。面对数据、泛化、效率和安全等挑战,论文提出的解决方案为无人机路径规划的进一步发展提供了有力支持,展现了深度强化学习在实际应用中的巨大潜力。

参考文献:

- [1] 杨思明,单征,曹江,等.基于模型的强化学习在无人机路径规划中的应用[J].计算机工程,2022,48(12):255-260.
- [2] 陈昱宏,高飞飞.基于深度强化学习的多目标无人机路径规划[J].

无线电通信技术,2022,48(6):957-970.

- [3] 赵恬恬.基于深度强化学习的无人机路径规划研究[D].成都:中国民用航空飞行学院,2024.
- [4] 司鹏搏,吴兵,杨睿哲,等.基于多智能体深度强化学习的无人机路径规划[J].北京工业大学学报,2023,49(4):449-458.
- [5] 周从航,李建兴,石宇静,等.深度强化学习在无人机编队路径规划中的应用[J].电光与控制,2024(10).

作者简介:朱晓(1982-),女,中国广西南宁人,硕士,

副高,从事无人机通信研究。

通信作者:朱朝阳(1981-),男,中国广西贺州人,硕士,副高,从事无人机通信数据安全与隐私保护研究。

课题项目:2023年度广西高校中青年教师科研基础能力提升项目,课题名称:基于深度强化学习 Q-Learning 的无人机自适应路径规划算法,项目编号:2023KY1318;广西工业职业技术学院 2022 年度第一批校级科研项目,课题名称:基于深度强化学习 Q-Learning 的无人机自适应路径规划算法,项目编号:GYKY2022007B。