

深度强化学习在无人驾驶路径规划中的应用

陈江涛

天津滨海汽车工程职业学院, 天津 300352

[摘要] 随着人工智能技术的快速发展, 深度强化学习 (Deep Reinforcement Learning, DRL) 在无人驾驶技术中的应用日益广泛。路径规划作为无人驾驶系统的核心技术之一, 直接影响车辆行驶的安全性及效率。文中系统综述了深度强化学习在无人驾驶路径规划中的关键技术、典型算法及其应用现状, 分析当前存在的挑战与不足, 并探讨未来的发展方向。通过对比传统路径规划方法与基于 DRL 的方法, 验证其在复杂动态环境下的优越性, 旨在为无人驾驶技术的进一步发展提供理论依据与技术支持。

[关键词] 深度强化学习; 无人驾驶; 路径规划; 智能决策; 自动驾驶算法

DOI: 10.33142/sca.v8i4.15943

中图分类号: TP242.6

文献标识码: A

Application of Deep Reinforcement Learning in Path Planning for Autonomous Driving

CHEN Jiangtao

Tianjin Binhai Vocational Institute of Automotive Engineering, Tianjin, 300352, China

Abstract: With the rapid development of artificial intelligence technology, the application of Deep Reinforcement Learning (DRL) in autonomous driving technology is becoming increasingly widespread. Path planning, as one of the core technologies of autonomous driving systems, directly affects the safety and efficiency of vehicle operation. The article systematically summarizes the key technologies, typical algorithms, and application status of deep reinforcement learning in unmanned driving path planning, analyzes the current challenges and shortcomings, and explores future development directions. By comparing traditional path planning methods with DRL based methods, we aim to verify their superiority in complex dynamic environments and provide theoretical basis and technical support for the further development of autonomous driving technology.

Keywords: deep reinforcement learning; autonomous driving; path planning; intelligent decision-making; autonomous driving algorithm

引言

无人驾驶汽车作为未来智能交通的重要组成部分, 路径规划技术直接影响行车安全与效率。传统路径规划方法在面对复杂动态环境时存在局限, 而深度强化学习凭借其强大的环境适应能力和自学习能力, 成为解决路径规划问题的重要手段。本文将深入探讨深度强化学习在无人驾驶路径规划中的理论基础、关键算法、应用实践与未来发展趋势。

1 深度强化学习基础理论

1.1 深度学习概述

深度学习 (Deep Learning) 是人工智能领域中的一种重要方法, 属于机器学习的分支。其核心是通过构建深层神经网络模型来自动学习和提取数据中的高层次特征。常见的深度神经网络结构包括前馈神经网络 (FNN)、卷积神经网络 (CNN) 和循环神经网络 (RNN)。在无人驾驶领域, 卷积神经网络被广泛应用于图像识别与目标检测任务, 如识别道路、车辆、行人及交通标志。CNN 通过局部感受野、权值共享和池化等机制, 有效地降低了计算复杂度并提升模型的泛化能力。而循环神经网络则适合处理时间序列数据, 能够实现了对车辆状态信息和传感器数据的动态建模。这些深度学习模型为无人驾驶系统的感知与决策提供了坚实的基础。

1.2 强化学习基本概念

强化学习 (Reinforcement Learning, RL) 是一种基于

试错机制和奖励反馈的学习方法。其核心思想是通过智能体 (Agent) 与环境 (Environment) 的交互行为, 学习最优策略以实现长期累积回报最大化。强化学习问题通常使用马尔可夫决策过程 (Markov Decision Process, MDP) 进行建模, MDP 包含状态 (State)、动作 (Action)、状态转移概率和奖励函数 (Reward Function) 等基本要素。在无人驾驶路径规划任务中, 智能体根据当前环境状态决定采取何种动作, 如加速、减速或转向等。当智能体做出正确决策并到达目标位置时, 将获得正向奖励; 反之, 碰撞障碍物或偏离车道则会受到负向反馈。通过不断试验和调整, 强化学习模型能够逐步优化路径规划策略, 提高系统的自主决策能力。

1.3 深度强化学习模型框架

深度强化学习 (Deep Reinforcement Learning, DRL) 融合了深度学习和强化学习的优势, 具备强大的自主学习与决策能力。通过引入深度神经网络, DRL 能够有效处理高维复杂状态空间, 并实现对动态环境的精准建模。具体而言, 深度神经网络被用于近似强化学习中的值函数或策略函数, 从而提升模型的泛化能力与预测精度。典型的 DRL 系统框架通常由状态感知模块、决策模块和执行模块构成。状态感知模块通过多传感器融合技术, 感知和提取环境特征信息; 决策模块根据环境状态输出相应动作策略; 执行模块则完成车辆控制指令的下发与路径执行。深度强

化学习方法主要分为两大类：基于值函数的算法和基于策略梯度的算法。前者以深度 Q 网络(Deep Q-Network, DQN)为代表，通过学习状态-动作值函数指导决策；后者包括 PPO (Proximal Policy Optimization)、TRPO (Trust Region Policy Optimization) 等算法，直接对策略函数进行优化。Actor-Critic 方法结合了值函数评估和策略优化的优点，进一步提升了学习效率。如图 1 所示。

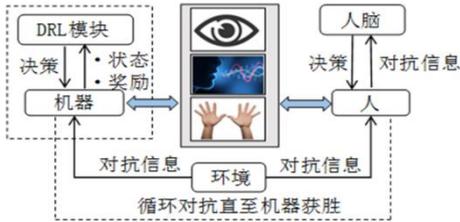


图 1 DRL 系统框架

2 无人驾驶路径规划问题描述

2.1 路径规划的基本需求

路径规划是无人驾驶系统实现自动驾驶功能的关键环节，其主要目标是在复杂的道路环境中，为车辆规划出一条安全、平滑、高效的行驶路径。首先，路径规划必须确保行车安全，即避免与其他车辆、行人或障碍物发生碰撞，遵守交通规则和道路标志。路径需要具备良好的舒适性，尽量减少急加速、急减速和频繁转向，保障乘客体验。路径规划算法应追求时间与能耗的最优化。在保证安全与舒适的前提下，合理选择行驶路线，缩短行程时间，降低能源消耗，提高无人驾驶系统的整体效率。最后，路径规划必须满足实时性需求，能够在动态变化的环境中迅速做出决策与调整，对突发情况进行及时响应。这要求系统具有较强的鲁棒性和应急处理能力。

2.2 无人驾驶环境模型

无人驾驶车辆在执行路径规划时，依赖于对环境的全面感知与建模。首先，环境模型需准确描述道路信息，包括车道线、交叉路口、交通标志及信号灯等。其次，需要动态感知和预测其他道路参与者(如车辆、行人、自行车)的行为轨迹。静态环境建模方面，依靠高精地图提供详细的地理信息和道路结构数据，确保车辆在全局范围内具备精准定位能力。动态环境感知依赖多种传感器(如激光雷达、摄像头、毫米波雷达等)对周围障碍物进行实时检测与跟踪，结合数据融合技术提高环境感知的准确性与鲁棒性。环境模型还需动态更新，及时反映突发情况，如道路封闭、交通事故等，确保路径规划依据的环境信息始终有效。

2.3 路径规划问题的数学建模

无人驾驶路径规划问题通常被建模为一个最优化问题或基于强化学习的决策过程。在经典路径规划方法中，首先定义状态空间(State Space)，表示车辆当前位置、速度、朝向等信息；然后定义动作空间(Action Space)，包括车辆的加速、制动和转向等操作。状态和动作的组合

决定了车辆的运动轨迹。路径规划的目标是设计一个代价函数或奖励函数，用于评估路径优劣。代价函数通常考虑路径长度、行驶时间、能源消耗、安全距离等因素，通过求解代价函数最小化问题获得最优路径。而在强化学习框架中，智能体通过与环境交互学习最优策略，奖励函数则综合考虑安全性、效率与舒适性等指标，激励智能体学习最优行为策略。路径规划还需满足车辆动力学约束和环境约束。例如，车辆转向角速度、加速度等物理限制，以及避免碰撞和遵守交通规则的要求，这些都对路径规划算法的设计提出了更高的数学建模要求。

目标函数：

目标是找到最优路径，使成本函数 J 最小化：

$$y = \omega_1 J_{length} + \omega_2 J_{time} + \omega_3 J_{energy} + \omega_4 J_{safety}$$

其中： $J_{length} = \int_0^T \sqrt{\dot{x}^2 + \dot{y}^2} dt$;

$J_{time} = T$ (行驶时间)；

$J_{energy} = \int_0^T a^2 dt$ (能耗)；

J_{safety} 取决于障碍物的距离，常用于： $J_{safety} = \sum_{i=1}^{Nobs} \frac{1}{d_i + \epsilon}$ 。

其中： d_i 是车辆到第 i 个障碍物的最小距离， ϵ 是一个小正数防止分母为零。权重 ω_i 决定不同因素的重要性。

3 深度强化学习在路径规划中的算法与实现

3.1 基于值函数的方法

基于值函数的深度强化学习方法主要通过估计状态或状态-动作对的值函数(Value Function)，指导智能体选择最优动作。在无人驾驶路径规划中，最具代表性的是深度 Q 网络(Deep Q-Network, DQN)算法。DQN 使用卷积神经网络来逼近 Q 值函数，将复杂的高维状态空间映射为对应的动作价值，智能体依据 Q 值选择具有最大长期奖励的动作。为了提升算法的稳定性和收敛性，DQN 引入了经验回放(Experience Replay)机制，通过打乱训练样本的时间相关性，提高样本利用率。此外，目标网络(Target Network)在 DQN 中用于固定一定时间内的 Q 值更新目标，有效缓解了训练过程中的不稳定问题。在无人驾驶路径规划任务中，DQN 能够帮助自动驾驶车辆学会在复杂动态环境下避障和选择最优行驶路径。然而，标准 DQN 存在 Q 值高估的问题，因此后续发展了 Double DQN 和 Dueling DQN 等改进方法。Double DQN 通过分离动作选择与动作评估，有效降低了 Q 值的高估偏差；Dueling DQN 则将状态价值和优势函数分开估计，提高了模型对状态价值的评估精度，进一步增强路径规划的表现。

3.2 基于策略梯度的方法

基于策略梯度的深度强化学习方法直接对策略函数进行参数优化，避免了值函数方法中的高估和不稳定问题。该类方法通过最大化期望回报来优化策略，使得无人驾驶

车辆能够根据当前状态灵活选择动作。典型的策略梯度方法包括 PPO (Proximal Policy Optimization) 和 TRPO (Trust Region Policy Optimization)。PPO 通过引入裁剪概率比值的方式,限制策略更新幅度,在保证策略改进的同时提高训练效率与稳定性。TRPO 则通过优化问题的约束条件,确保策略更新步长在一个信赖域(Trust Region)内,从而保证每次更新都能带来性能的提升。在无人驾驶路径规划中,基于策略梯度的方法能够高效处理高维、连续的动作空间问题。例如,车辆在变道、超车或避障时,PPO 算法可实现动态平滑路径调整,保证车辆决策的灵活性与连贯性。相比基于值函数的方法,策略梯度算法更适合需要高精度控制与动作连续性的复杂驾驶任务。

3.3 混合方法及优化

混合方法结合了值函数和策略梯度两大类算法的优势,在路径规划任务中表现出更强的学习能力和决策效率。最常见的混合方法是 Actor-Critic 架构,该方法由 Actor (策略网络)负责决策动作,Critic (值函数网络)负责评估策略性能。Soft Actor-Critic (SAC) 是当前应用较为广泛的混合算法之一。SAC 在传统 Actor-Critic 结构的基础上引入熵正则化机制,提升智能体探索能力,增强鲁棒性与收敛性。在无人驾驶路径规划任务中,SAC 能够帮助车辆在复杂环境中灵活选择最优路径,并兼顾安全性与效率。此外,SAC 支持连续动作空间,适用于高速公路驾驶、城市道路复杂交互等多场景路径规划任务。随着无人驾驶系统复杂度的增加,多智能体深度强化学习 (Multi-Agent DRL) 也逐渐成为研究热点。该方法支持多个智能体协同决策,提升车辆在车队行驶、交叉路口等场景下的协调性与整体优化效果。未来,通过优化模型结构和引入强化学习与传统算法的融合,有望进一步提高无人驾驶路径规划的智能化水平和应用可靠性。

4 深度强化学习路径规划的应用案例与挑战

4.1 典型应用案例分析

深度强化学习 (DRL) 已在多个无人驾驶路径规划应用中取得了显著成效。Waymo 是全球领先的自动驾驶公司之一,其自动驾驶系统在路径规划中应用了深度强化学习 (DRL) 技术,以提升车辆在复杂城市环境中的决策能力。Waymo 采用基于深度 Q 网络(DQN)和近端策略优化(PPO)的强化学习算法,使车辆能够在动态交通环境中进行实时决策。例如,在遇到交叉路口、行人穿行或其他车辆突然变道时,Waymo 车辆能够通过 DRL 预测未来可能的交通演化趋势,并选择最优路径以确保安全和高效通行。此外,该系统结合了模仿学习 (Imitation Learning) 和规则约束强化学习 (Constrained RL),使得训练的智能体不仅能遵守交通规则,还能适应未见过的复杂场景。

4.2 当前技术挑战

尽管 DRL 在无人驾驶路径规划中表现出广泛应用前景,但仍存在诸多技术挑战。首先,样本效率低和训练时

间长是当前主要瓶颈。由于路径规划涉及高维状态和动作空间,训练 DRL 模型通常需要大量交互数据和反复迭代,导致开发周期较长。泛化能力不足使得训练好的模型在不同环境下的迁移性能较差。许多算法只能在特定场景中表现良好,难以适应天气变化、道路状况突变或突发交通事件,严重限制了无人驾驶系统的实用性。深度强化学习算法的安全性鲁棒性仍需进一步加强。路径规划决策错误可能导致交通事故,而 DRL 模型在面对未知环境或极端情况时,可能出现不可预测的行为,带来潜在安全风险。因此,如何提升模型在复杂动态环境中的决策稳定性与可靠性,是当前技术攻关的重要方向。

4.3 未来研究方向

深度强化学习在无人驾驶路径规划中的研究将重点围绕以下几个方向展开。提升样本效率和训练算法的稳定性,将是核心目标之一。通过引入迁移学习、元学习等方法,使模型能够快速适应新环境,减少对大量训练样本的依赖。多模态感知与数据融合技术的引入,将提升环境建模的准确性。通过融合激光雷达、摄像头、毫米波雷达等多传感器信息,增强路径规划系统对环境变化的响应能力。联邦学习与隐私保护也是未来的重要研究方向。无人驾驶车辆作为数据收集和处理终端,采用联邦学习框架能够在保护用户隐私的前提下,实现分布式模型训练和共享,促进路径规划算法的协同优化。强化学习与传统控制方法的融合将为路径规划算法提供更多可靠性保障。将强化学习的灵活性与传统基于规则或优化的控制算法相结合,有望解决现有 DRL 模型在安全性和可解释性方面的不足,为自动驾驶系统提供更安全、更可靠的路径规划解决方案。

5 结束语

深度强化学习为无人驾驶路径规划带来了新的技术突破,其强大的自主学习和决策能力使自动驾驶系统在复杂环境下更加智能和高效。然而,现有算法仍面临样本效率低、泛化能力差等挑战。未来,随着感知、控制与智能算法的进一步发展,深度强化学习将在无人驾驶路径规划中实现更加广泛和成熟的应用,推动智能交通系统的创新与落地。

[参考文献]

- [1]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,2024,31(10):27-33.
- [2]杨思明.深度强化学习在无人机升空平台路径规划中的应用[D].郑州:战略支援部队信息工程大学,2021.
- [3]李书杰.基于深度强化学习的室外无人车辆路径跟踪控制系统研究[D].哈尔滨:哈尔滨工业大学,2023.
- [4]黄国明.自动驾驶汽车三维路径规划与路径跟踪控制方法研究[D].长沙:湖南大学,2022.

作者简介:陈江涛(1992.4—),女,天津市,汉族,研究生,就职于天津滨海汽车工程职业学院,从事智能网联专业相关教学工作。