基于强化学习的管廊巡检机器人路径规划算法优化

陈　阳
四川君逸数码科技股份有限公司　四川成都　610000

摘　要：本文旨在研究基于强化学习的管廊巡检机器人路径规划算法优化问题。首先，介绍了强化学习的基本理论及其在路径规划中的应用，然后分析了现有强化学习算法在管廊巡检路径规划中的局限性，接着提出了一种基于深度Q网络（DQN）的优化算法，并结合环境建模和多机器人协同策略，进一步提升了路径规划的效果。最后，通过仿真和实验验证了所提出算法的有效性。本文的研究成果为管廊巡检机器人路径规划提供了一种新型的强化学习方法，具有重要的理论意义和应用价值。

关键词：强化学习；管廊巡检机器人；路径规划算法；优化策略

引言

城市地下管廊是城市基础设施中非常重要的一部分，它承载了电力、通信、燃气、供水等多种功能，它的安全运营直接影响了城市安全以及居民日常生活。为提高地下管廊巡检效率及安全性，智能巡检机器人被提出。路径规划作为管廊巡检机器人最核心的工作之一，通过合理的路径规划可以使得机器人在管廊复杂环境下高效、安全地执行巡检任务。强化学习是机器学习方法之一，它能够通过与周围环境的相互作用学习出最优策略，对于解决路径规划问题有着巨大潜力。但现有基于强化学习的管廊巡检机器人路径规划算法还存在算法收敛速度较慢、缺乏复杂环境适应性等问题。因此，开展基于强化学习的管廊巡检机器人路径规划算法的优化研究具有一定的理论意义与现实意义。

一、强化学习与管廊巡检机器人路径规划基础

（一）强化学习基础

强化学习是一种通过智能体（agent）与环境进行交互，以最大化累积奖励为目标来学习最优策略的机器学习方法。在强化学习中，智能体在每个时间步根据当前状态选择一个动作，环境根据智能体的动作给出一个奖励和下一个状态。智能体的目标是学习一个策略，使得在长期内获得的累积奖励最大。强化学习的基本要素包括状态（state）、动作（action）、奖励（reward）和策略（policy）。状态表示智能体所处的环境信息，动作表示智能体可以采取的行为，奖励是环境对智能体动作的反馈，策略是智能体根据当前状态选择动作的规则。

（二）管廊巡检机器人路径规划基础

管廊巡检机器人的路径规划问题可以描述为：在给定管廊环境地图和巡检任务的情况下，寻找一条从起始点到目标点的最优路径，使得机器人能够在最短的时间内完成巡检任务，同时避免与障碍物发生碰撞。

管廊环境通常具有复杂性和动态性，如存在各种管道、设备和障碍物，并且可能会有人员和其他移动目标。因此，管廊巡检机器人的路径规划需要考虑环境的不确定性和实时变化。

二、基于强化学习的管廊巡检机器人路径规划算法分析

（一）现有强化学习算法在路径规划中的应用

管廊巡检机器人路径规划中常采用的强化学习算法有Q-Learning、Sarsa和深度强化学习算法。Q-Learning作为基于值迭代的强化学习算法，通过构造Q值表存储各状态—行动对的期望奖励，智能体依据Q值表筛选最优行动。在路径规划的过程中，Q-Learning能够学习到从起始点到结束点的最佳路径，但在处理大规模状态空间时，Q值表的保存和更新可能会遭遇维度灾难的问题。Sarsa算法是一种基于值迭代的方法，与Q-Learning有所不同，它采用在线策略学习，即智能体根据当前的策略选择动作并更新Q值，这使其在应对动态环境方面有一定优势，但同样面临状态空间维度灾难的问题。

深度强化学习算法集深度学习与强化学习优点于一身，可以对高维感知数据例如图像、激光雷达数据进行处理。深度Q网络（DQN）被认为是深度强化学习中的经典方法之一，它采用深度神经网络来模拟Q值函数，

并结合经验回放策略和目标网络来增强学习的稳定性和高效性。管廊巡检机器人进行路径规划时，DQN能够对管廊环境信息进行复杂处理，从而实现更智能、更有效的路径规划。另外，深度强化学习算法如策略梯度算法和Actor–Critic算法也有应用于路径规划的潜力。

（二）基于深度Q网络（DQN）的路径规划算法

深度Q网络（DQN）被视为深度强化学习的关键技术，它结合了深度神经网络和Q学习策略，为复杂场景中的路径设计提供了一个高效的策略。核心是用神经网络逼近Q值函数，实现从高维状态空间到动作价值评估的映射，打破了传统Q学习应对大范围状态的储存和计算瓶颈。当执行路径规划任务时，智能体与周围环境相互作用产生状态、行动、奖励和下一个状态四元组数据，并将其保存到经验回放缓冲区，训练中的随机采样破坏了数据的相关性，有利于样本利用率的提高和训练进程的稳定。以目标网络为独立副本周期性地对主网络参数进行同步，以计算目标Q值，有效地缓解了由于主网络频繁更新而引起的目标值起伏，提高了学习的稳定性。DQN使用 ε –贪心策略来均衡探索与利用，在初期以更大的几率随机选取动作探索环境，并随训练的进行而逐渐减小探索率，转而更多地使用已知的最优动作。奖励函数设计一般与路径规划目标相结合，例如在靠近目标处给予正向奖励，在出现碰撞或背离路径处施加负向惩罚等，从而指导智能体学会安全有效地移动策略。经过不断迭代优化后，DQN可以让智能体自主地规划复杂动态环境下从始至终的最优路径，并已在机器人导航、无人机集群协同以及自动驾驶方面表现出显著优势，已成为智能决策方面的一个重要技术支柱。

（三）算法存在的问题与挑战

深度Q网络（DQN）在路径规划中虽成效显著，却也面临诸多问题与挑战。其训练过程对超参数极为敏感，学习率、折扣因子等参数的细微调整都可能引发训练不稳定或收敛困难，导致智能体难以习得可靠策略。经验回放机制虽能提升样本利用率，但随着环境复杂度增加，所需存储的数据量呈指数级增长，占用大量内存资源，对硬件配置提出更高要求。目标网络虽可缓解目标值波动，但参数同步存在延迟，在快速变化的环境中，目标网络提供的Q值可能与当前真实状态存在偏差，影响决策准确性。DQN采用离散动作空间设计，面对连续动作场景时需进行离散化处理，这一过程会丢失部分动作细节，限制智能体在精细操作任务中的表现。此外，DQN的探索机制较为单一，ε –贪心策略难以适应复杂环境中的多样化探索需求，易陷入局部最优解。在部分场景中，奖励函数设计也面临挑战，若奖励信号稀疏或反馈延迟，智能体可能因长期得不到有效指导而学习效率低下，甚至无法完成路径规划任务。

三、管廊巡检机器人路径规划算法优化策略

（一）环境建模优化

环境建模是路径规划中最基本的一环，环境建模的优化对于增强智能体的决策能力具有重要意义。传统的栅格化建模虽然简单直观，将环境分割成规则网格并标示出可通行区域等特点，但很难准确地描绘复杂场景下曲面斜坡等坡度变化的微妙特性，狭窄通道在宽度上存在起伏，而这些细节上的缺失会使智能体在规划路径过程中忽略真实的物理约束，而发生碰撞或者卡顿等问题。为了解决这个问题，可以采用三维体素建模方法，将空间细分为多个立体单元，并详细记录每个单元的高度、材料等特性，从而创建一个更接近实际环境的三维模型，使得智能体能够在垂直方向上感知空间信息，并规划更加合乎物理规律的道路。对于动态环境而言，建模需要纳入实时更新机制，并使用传感器数据来不断监测周围环境的变化，例如移动障碍物位置和临时封闭区域生成等，通过对模型参数或模型的局部重建进行动态调整，保证智能体总是根据最新的环境信息进行决策。另外，融合多源传感器的数据能够增强建模的鲁棒性，激光雷达为建模提供了高精度的距离信息，摄像头采集视觉特征并与惯性测量单元进行运动状态的互补，多模态数据经过融合处理，能够更加全面地反映环境特性，并降低单一传感器错误对建模结果造成的影响，从而为智能体进行安全高效的路径规划提供了可靠依据。

（二）强化学习算法优化

尽管强化学习算法在路径规划这一领域取得了显著的成效，但仍然存在大量的优化潜力，以突破当前的限制。传统的DQN算法主要依赖于单一网络来估计Q值，容易因为网络参数的更新而造成目标值的波动，从而影响训练的稳定性。针对这一问题，可以引入双DQN架构，从价值评估中分离出动作选择：由主网络承担根据当前状态筛选出最优动作的任务，而目标网络只对动作相应目标Q值进行评价，有效地减弱了高估偏差的影响，使得价值估计更准确，增强了智能体决策的可靠性。传统的算法探索策略在面对复杂环境时通常效率较低，容易陷入局部最优的问题，可以将噪声网络和好奇心驱动

机制相结合：噪声网络将可控的随机噪声注入到动作输出，加强了探索的随机性；好奇心驱动机制通过内部奖励函数激励智能体对未知状态进行探索。两者协同工作，指导智能体在已知高效路径和未知潜在路径之间实现探索与利用的动态平衡，扩大了搜索空间，并促进了寻找全局最优路径可能性的提高。另外，传统的强化学习算法大多是以离散的动作空间设计为基础，需要对连续动作场景进行离散化，丢失了动作的细节，因此可以引入深度确定性策略梯度算法进行研究，对连续动作空间内的确定性策略进行直接学习，以策略网络的方式输出准确的动作，并与演员–评论家架构相结合，使用评论家网络引导策略网络的优化，实现了对路径规划更加精细的控制，以满足复杂动态环境的需要。

（三）多机器人协同路径规划优化

多机器人协同路径规划是分布式系统有效运行的核心技术，优化质量的好坏直接决定着物流仓储、灾害救援和智能制造等场景下协同效率和安全性。传统的集中式规划框架存在着计算复杂度随机器人个数成倍增加、对动态环境的适应性较差、通信负载过重等深层次挑战：虽然全局优化算法能够产生理论上的最优路径，却在实时性要求高的场景中因求解耗时过长而失效；分布式规划虽然可以减轻计算压力，但由于缺少全局协调，容易造成路径冲突和资源死锁；在动态障碍物存在的情况下，局部避障策略往往会由于信息滞后而使机器人经常停滞或绕道而行，明显降低了总体任务完成率。为解决这些冲突，优化需要在全局协调架构、冲突消解机制和动态适应能力三个维度上进行突破。

在全局协调层面上采用分层混合式结构，将规划任务拆分为全局拓扑规划和局部动态调整两层：高层根据环境先验信息建立拓扑地图，采用图论算法产生与任务节点相连的无冲突路径骨架，作为底层的全局约束；下层通过分布式协商机制，使每个机器人根据骨架路径并结合实时传感器数据进行局部轨迹调整，不仅保证了全局最优性，而且减轻了集中式计算的压力。在冲突消解中引入一种基于拍卖机制的动态资源分配策略，在发现路径交叉风险后，该系统依据机器人残余任务量和能量储备这一综合指标对路径优先权进行动态分配，优先权较大的机器人维持原有路径，而另一方则通过调节速度或选择备选路径进行躲避，并结合速度障碍法进行最小安全距离的实时计算，以保证避障动作平滑且能耗最优。

结论

本论文针对管廊巡检机器人路径规划算法在强化学习基础上展开深入研究，提出相关优化策略。通过优化环境建模、强化学习算法以及多机器人协同路径规划等方法，可以有效提高管廊巡检机器人路径规划的效率与精度，提升机器人在复杂环境下的适应性与鲁棒性。但基于强化学习的管廊巡检机器人路径规划仍面临着如何应对更复杂动态环境以及如何进一步增强算法泛化能力的挑战。在未来，有机会深入研究新型的强化学习方法和优化策略，并与其他人工智能技术，例如计算机视觉和自然语言处理，进行整合，以实现更智能、更高效的管廊巡检机器人路径规划。与此同时，仍需加强实际应用方面的研究工作，对实际管廊环境下的算法进行检验与验证，并不断对其进行改进与完善，从而为城市地下管廊的安全、稳定运营提供更多可靠保障。

参考文献

[1]李刚，王童语，包仁人，等.吊轨式管廊巡检机器人结构设计及实验[J].制造业自动化，2023，45（6）：189-194.

[2]武铁峰，谢勇勇.特种机器人在隧道管廊内的应用研究[C]//2023中国油气人工智能科技大会.1.中海石油炼化有限责任公司2.中海石油宁波大榭石油有限公司，2023.

[3]俞高伦.某石化企业隧道智能巡检设计与研究[J].市场调查信息，2023（7）.

[4]高著海，雷铭，朱玲芬，等.巡检机器人在综合管廊智慧化上的应用[J].中国建设信息化，2024（18）：54-57.

[5]常城，舒志兵，陈守林，等.基于UWB的智能巡检机器人定位系统研究[J].机床与液压，2024，52（9）：22-29.