基于强化学习的动态环境中移动机器人路径实时规划策略

荆 麟[1] 张晓勇[1] 王文康[1] 薛 刚[1] 纪浩文[1] 韩孝军[2]
1. 国能宁夏大坝四期发电有限公司 宁夏 吴忠 751607
2. 北京鼎誉通科技发展有限公司 北京 100041

摘 要：随着科技飞速发展，动态环境下的路径规划愈发关键。本文聚焦于基于强化学习的动态环境中移动机器人路径实时规划策略。首先阐述强化学习在路径规划中的应用原理，涵盖基本概念与建模方式；接着介绍关键技术，如深度强化学习、多智能体协同强化学习等；然后分析该策略面临的挑战，包括状态空间爆炸、样本效率低等；最后提出针对性解决方案，如状态空间压缩、多智能体协同等。旨在为动态环境下移动机器人路径实时规划提供理论支持与实践参考，推动强化学习在该领域的有效应用与发展。

关键词：强化学习；动态环境；移动机器人；路径实时规划

引言：在科技飞速发展的当下，移动机器人在诸多领域的应用愈发广泛，如物流配送、智能巡检等。动态环境中的路径实时规划是移动机器人高效完成任务的关键。传统路径规划方法在应对动态变化时存在局限性，难以快速适应复杂多变的场景。强化学习作为一种通过智能体与环境交互来学习最优策略的机器学习方法，为动态环境下的路径规划提供了新思路。它能够使机器人在不断探索与学习中，实时调整路径以适应环境变化。本文将深入探讨基于强化学习的动态环境中移动机器人路径实时规划策略，分析其原理、技术、挑战及解决方案。

1 强化学习在路径规划中的应用原理

1.1 强化学习基本概念

强化学习主要由智能体、环境、状态、动作和奖励五要素构成。智能体是具备学习与决策能力的主体，在路径规划中就是移动机器人；环境是智能体所处的外部世界，包含各种影响机器人行动的因素；状态是对环境在某一时刻的描述，如机器人的位置、周围障碍物分布等；动作是智能体可采取的行为，如前进、转向等；奖励是环境对智能体动作的反馈，引导智能体学习最优策略。智能体通过不断感知状态、采取动作、获得奖励，逐步优化自身策略，以在长期内获得最大累计奖励。

1.2 强化学习应用于路径规划的建模

将强化学习应用于路径规划时，需构建合理的模型。把移动机器人所处的动态环境抽象为马尔可夫决策过程，其中状态空间涵盖机器人位置、速度、周围障碍物信息等；动作空间包含机器人可执行的各种移动动作。定义奖励函数，例如到达目标点给予高额正向奖励，碰撞障碍物给予负向奖励，接近目标点给予小额正向奖励等。智能体根据当前状态选择动作，环境根据动作转移到新状态并反馈奖励。通过不断迭代学习，智能体学会在不同状态下选择最优动作，从而规划出从起点到目标点的最优路径[1]。

2 基于强化学习的动态环境中移动机器人路径实时规划关键技术

2.1 深度强化学习

深度强化学习将深度学习的强大感知能力与强化学习的决策能力深度融合，为动态环境中移动机器人路径实时规划提供了有力工具。在传统强化学习里，当状态空间维度较高时，难以有效处理和表示状态信息。而深度强化学习借助深度神经网络，如卷积神经网络（CNN）、循环神经网络（RNN）及其变体，能够自动从高维的原始数据（如机器人传感器采集的图像、距离信息等）中提取有效特征，实现对复杂状态的高效表示。以深度Q网络（DQN）为例，它使用神经网络来近似Q函数，避免了传统Q学习中构建庞大Q表的不便，能更好地处理连续状态空间。在路径规划中，机器人通过深度神经网络对当前环境状态进行评估，输出不同动作的Q值，进而选择最优动作。同时，经验回放机制和目标网络的使用，提高了学习的稳定性和效率。

2.2 多智能体协同强化学习

在动态环境的移动机器人路径实时规划场景中，单个机器人受限于自身感知范围与处理能力，面对复杂任务时往往力不从心。多智能体协同强化学习则能有效解决这一问题，它让多个机器人作为一个整体系统，通过相互协作来完成路径规划任务。每个机器人作为独立的智能体，具备自身的感知、决策和行动能力。在协同强化学习框架下，智能体之间通过信息交互共享各自对环境的观察和决策意图。这种信息交互可以是直接的通信，也可

以是基于环境的间接感知。通过共享信息，智能体能够更全面地了解整体环境状况，避免重复探索和冲突。同时，多智能体协同强化学习采用联合奖励机制，将整个团队的任务完成情况作为奖励依据，促使智能体为了共同目标而协作。

2.3 分层强化学习

动态环境中的移动机器人路径实时规划面临状态空间复杂、任务多样等挑战，分层强化学习为应对这些难题提供了有效思路。它将复杂的路径规划任务分解为多个层次，每个层次负责不同抽象级别的决策。高层策略处于宏观层面，负责制定整体目标和长期规划，例如确定机器人大致的行进方向和关键目标点。它不关注具体的动作细节，而是从全局视角把握任务走向，为低层策略提供指导框架。低层策略则处于微观层面，依据高层策略的指示，处理具体的动作选择，如机器人的转向角度、移动速度等。它专注于在当前状态下实现精细的动作控制，确保机器人能够准确执行高层规划。通过这种分层结构，分层强化学习降低了每个层次的学习难度和状态空间复杂度。高层策略简化了对整体任务的理解，低层策略聚焦于具体动作优化。

2.4 模仿学习与强化学习结合

在动态环境中为移动机器人规划实时路径时，单纯强化学习需大量试错来学习最优策略，学习效率低且初期探索可能因不当动作导致危险。模仿学习与强化学习的结合则能有效改善这一状况。模仿学习通过让机器人观察专家（如人类操作或预先设定的优质路径示范）的示范行为，快速学习到有效的策略模式。它能够直接从示范数据中提取关键信息，为强化学习提供良好的初始策略，避免强化学习从零开始探索的盲目性，大大缩短学习时间。强化学习在此基础上进一步优化策略。它利用与环境交互获得的实时反馈，对模仿学习得到的初始策略进行微调和完善。在动态环境变化时，强化学习能根据新的状态和奖励信号，不断调整机器人的行为，使其更好地适应环境动态性[2]。

3 基于强化学习的动态环境路中移动机器人径实时规划面临的挑战

3.1 状态空间爆炸

在动态环境下的移动机器人路径实时规划中，状态空间爆炸是一大严峻挑战。动态环境包含众多不断变化的因素，如障碍物的位置、移动速度、形状，以及目标点的动态调整等。这些因素组合形成了庞大且复杂的状态空间。随着环境复杂度的增加，状态数量呈指数级增长。强化学习算法需要处理如此海量的状态，对计算资源提出了极高要求。传统方法难以有效表示和存储这些状态，并且在搜索最优策略时，要在如此巨大的状态空间中遍历，导致计算时间大幅增加，使得实时规划变得极为困难，严重影响机器人的响应速度和决策效率。

3.2 样本效率低

强化学习依赖大量的样本数据来学习最优策略，在动态环境移动机器人路径规划中，样本效率低的问题尤为突出。动态环境的不确定性和复杂性使得机器人需要不断探索以获取有效样本，但每次探索都可能面临风险，且不一定能获得对学习有价值的反馈。而且，不同样本之间的关联性和信息重复度较高，真正能显著提升策略性能的样本占比低。这就导致机器人需要耗费大量时间和资源去收集样本，学习过程缓慢，难以在短时间内收敛到最优策略，无法及时适应动态环境的变化，降低了路径规划的时效性和准确性。

3.3 实时性要求高

动态环境中的移动机器人路径实时规划对实时性有着极高的要求。环境处于不断变化之中，障碍物可能随时出现、移动或消失，目标点也可能动态改变。机器人必须在极短的时间内对环境变化做出反应，重新规划出可行路径。然而，强化学习算法在进行决策时，需要经过复杂的计算过程，包括状态评估、动作选择、策略更新等。若算法计算复杂度高，处理速度慢，就无法在规定时间内完成路径规划，导致机器人不能及时避开障碍物或到达目标点，影响任务的顺利执行，甚至可能引发安全问题，无法满足动态环境下对实时性的苛刻需求。

3.4 安全性保障

在动态环境中为移动机器人进行路径实时规划时，安全性保障是关键挑战。动态环境充满不确定性，机器人可能遭遇各种突发状况，如突然出现的障碍物、其他移动物体的碰撞等。强化学习算法在探索过程中，为了寻找最优策略，可能会尝试一些危险动作，如高速冲向障碍物、进入狭窄危险区域等，这极易导致机器人损坏或引发安全事故。而且，在复杂动态环境下，准确预测所有可能的安全风险难度极大，现有的安全约束机制可能无法全面覆盖各种情况，使得机器人在路径规划过程中面临较高的安全风险，难以确保其安全稳定运行[3]。

4 基于强化学习的动态环境中移动机器人路径实时规划解决方案

4.1 状态空间压缩

在动态环境移动机器人路径实时规划里，状态空间爆炸问题严重制约强化学习效率。状态空间压缩是有效应对手段。一方面，可对状态特征进行筛选，去除冗余

信息。通过分析环境要素与路径规划的相关性，仅保留对决策有关键影响的状态特征，如障碍物的关键位置、机器人的核心运动参数等，减少状态维度。另一方面，采用聚类或降维技术，将相似状态归为一类，用聚类中心代表该类状态，降低状态数量。此外，还可以构建状态抽象层次，将具体状态抽象为更高级的概念，如将多个相邻位置抽象为一个区域，使智能体在更高层次进行决策，简化状态空间，提升强化学习在动态环境路径规划中的处理能力和效率。

4.2 多智能体协同

动态环境复杂多变，单智能体能力有限，多智能体协同可提升路径实时规划效果。多个机器人作为智能体，通过信息交互共享环境感知数据和决策意图。它们能更全面地了解环境动态，避免重复探索。在协同过程中，可制定联合奖励机制，以团队整体任务完成情况为奖励依据，促使智能体为共同目标协作。比如部分智能体负责探索未知区域，为其他智能体提供环境信息；部分智能体专注于引导团队避开危险。同时，采用分布式决策架构，每个智能体根据局部信息和团队协调信号做出决策，提高决策速度。多智能体协同还能增强系统的容错性，当部分智能体出现故障时，其他智能体可继续完成任务，保障动态环境下路径规划的稳定性和可靠性。

4.3 深度强化学习结合

深度强化学习结合为动态环境移动机器人路径实时规划提供强大支持。深度学习具有强大的特征提取能力，能处理高维的传感器数据，如图像、激光雷达点云等。将其与强化学习结合，可自动从原始数据中学习有效特征，无需手动设计特征工程。例如使用卷积神经网络（CNN）处理图像数据，提取环境中的障碍物、目标点等信息，为强化学习决策提供丰富依据。同时，深度神经网络可近似复杂的价值函数或策略函数，提高决策的准确性。在动态环境中，深度强化学习能快速适应环境变化，通过不断与环境交互更新网络参数，优化路径规划策略。

4.4 离线强化学习

离线强化学习为动态环境移动机器人路径实时规划提供了新途径。在动态环境中收集大量高质量在线交互数据成本高且风险大，而离线强化学习可利用预先收集的静态数据集进行学习。这些数据集包含丰富的环境状态、动作和奖励信息，无需机器人实时与环境交互。离线强化学习算法通过对数据集的分析和学习，构建价值函数或策略模型，为机器人提供路径规划决策依据。它避免了在线学习中的探索风险，尤其适用于对安全性要求高的场景。同时，离线学习可充分利用历史数据，挖掘潜在模式和规律，提升学习效率。在动态环境变化时，可结合少量在线数据对离线学习得到的模型进行微调，使机器人快速适应新环境，实现实时路径规划，降低学习成本和风险[4]。

结束语

在动态环境这一充满挑战的舞台上，基于强化学习的移动机器人路径实时规划策略展现出巨大潜力与独特优势。它凭借对环境动态变化的敏锐感知和智能决策能力，为机器人赋予了灵活应对复杂状况的本领。尽管面临状态空间复杂、样本效率低等重重困难，但通过状态空间压缩、多智能体协同等创新策略不断突破。未来，随着强化学习理论的持续完善与技术的深度融合，该规划策略将进一步优化，大幅提升机器人在动态环境中的路径规划能力，推动其在物流、救援、探索等众多领域的广泛应用，开启智能移动新时代。

参考文献

[1]王晓峰,李鹏,张志强.基于强化学习的移动机器人路径规划研究[J].机器人技术与应用,2022,38(5):145-151.

[2]李军,杨晓静,赵敏.强化学习在动态环境中的应用及挑战[J].控制与决策,2023,38(2):115-120.

[3]周艳华,刘晓兵,周威.深度强化学习在动态路径规划中的应用[J].智能控制与自动化,2021,43(7):223-228.

[4]邓修朋,崔建明,李敏,等.深度强化学习在机器人路径规划中的应用[J].电子测量技术,2023,46(06):121-128.