

基于深度学习的工业巡检机器人算法研究

马洪武¹ 王利利² (通讯作者) 张滨杰²

1 内蒙古天润化肥股份有限公司, 内蒙古鄂尔多斯市, 010321;

2 内蒙古工业大学, 内蒙古呼和浩特市, 010051;

摘要: 机器人系统应用于工业巡检的某些任务时, 不可能覆盖各种环境和传感器属性, 导致机器人在未知环境下需自动探索、自适应环境。为了解决上述问题, 提出了一种基于深度强化学习的决策算法, 通过将机器人的探索过程分解为决策、规划和地图绘制模块, 构建一个通用的探索框架, 弥补了由于学习效率低及难以从模拟到现实的可移植性而受到的损害, 提高了机器人系统的模块化程度, 为相关领域的发展提供了可行思路与实践范式。

关键词: 深度强化学习; 巡检机器人; 自主探索; TD3 算法

DOI: 10. 64216/3080-1508. 25. 09. 038

引言

巡检机器人广泛应用于电力、工业和管道等领域, 但面对未知或部分未知环境时, 传统的全局路径规划方法无法应用^[1-2]。因此, 机器人需要实时获取环境信息以自主规划路径。现有的传统算法, 包括动态窗口方法和人工势场法, 它们在复杂环境中存在局限性, 容易陷入局部最优^[3-4]。模拟退火方法虽然能解决局部最优问题, 但收敛速度慢^[5]。基于深度学习的控制方法虽然适用于复杂环境, 但学习效率低且可移植性差。为了解决这些问题, 文章提出了一个通用的探索框架, 将探索过程分解为决策、规划和映射三个模块。在此基础上, 开发了一种基于 TD3 的决策算法, 该算法利用深度神经网络从部分地图中学习探索策略。

1 研究方法

机器人需要通过传感器数据来构建周围环境的地图, 并进行自我定位。随着深度强化学习 (DRL) 的发展, 出现了多种端到端的控制方法, 这些方法直接将传感器数据 and 环境信息作为输入, 输出机器人的控制指令。端到端方法的优点在于其简洁性, 但将这些方法应用到真实机器人上存在挑战, 主要存在以下问题。

1.1 真实机器人的结构和传感器误差

在模拟环境中学习到的策略可能无法直接应用于真实机器人, 因为真实机器人的结构和传感器可能存在误差。

1.2 环境差异

真实的环境与模拟环境不完全相同, 导致深度神经网络难以泛化到真实环境中。为了解决这些问题, 提出了将整个系统分为三个独立模块的解决方案。建图定位模块: 负责构建环境地图并进行机器人的自我定位; 决策模块: 负责根据环境地图和机器人的定位信息做出导航决策; 轨迹规划模块: 负责根据决策模块的输出规划机器人的移动轨迹。

通过将系统分解为这三个模块, 每个模块可以独立运行, 减少了它们之间的相互影响, 从而提高了整体系统的稳定性和可靠性。这种方法有助于提高机器人在真实环境中的导航性能, 并且可以更好地处理真实世界的不确定性和复杂性。图1所示为各个模块之间的关系。

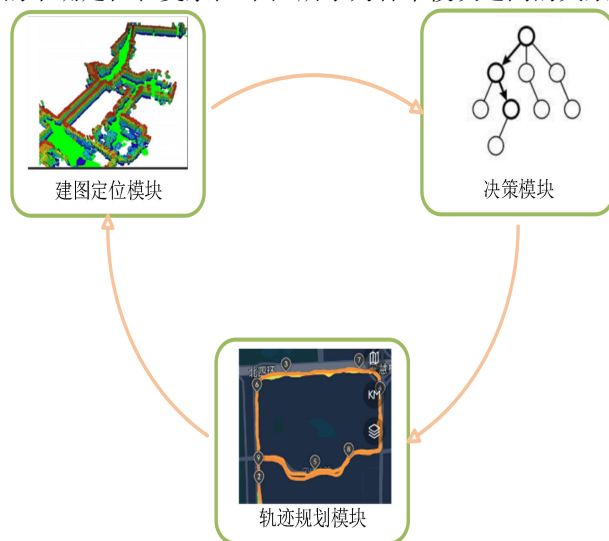


图1 模块间的关系图

①建图定位模块

采用多传感器融合的 SLAM, 数学模型可以定义为

$$M_t = f_{\text{mapping}}(I_t, L_t, C_t) \quad (1)$$

其中 M_t 表示 t 时刻的地图, I_t, L_t, C_t 分别表示 t 时刻的 IMU 信息、Lidar 信息、Camera 信息。

②决策模块

该模块的作用是选择 t 时刻的目标点。模块的输入为建图定位模块的地图和当前机器人的位姿, 决策模块的建模为:

$$G_t = f_{\text{decision}}(M_t, P_t) \quad (2)$$

其中 G_t 是目标点, P_t 是当前机器人的位姿。在本文中, 我们提出了一种基于 TD3 的决策方法。与基于规则的方法相比, 该方法能够自动学习激光特征和视觉特征, 以利于探索。我们的方法不是将规则堆叠起来进行探索, 而是通过反复试错来学习可行的策略, 并覆盖更复杂的场景, 并且提高了该决策模型在现实环境下泛化能力。

③轨迹规划模块

该模块旨在确定从当前位置到目标点的可行轨迹。规划模块建模为:

$$T_t = f_{\text{planing}}(P_t, G_t, M_t) \quad (3)$$

其中 T_t 是机器人控制的规划轨迹, 现有的规划方法包括全局规划方法和局部规划方法。前者的目的是确定地图上从当前位置到目标位置的最短路径, 后者的目的是获取周围环境信息进行实时的轨迹规划。

本文的研究非常重视构造 DRL 算法的决策模块, 以实现有效的探索。与传统的决策方法相比, 基于 DRL 的决策算法可以消除复杂规则中的琐碎细节。与端到端的方法不同, 基于该框架的方法具有更高的模块化程度。

2 基于 DRL 的决策模块

采用基于双延迟深度确定性策略梯度 (TD3) 的神经网络架构进行运动策略训练。TD3 是一种演员-评论家模型, 适用于连续动作空间的动作执行。机器人通过激光雷达获取局部环境信息, 将其与相对路径点的极坐标信息融合, 形成输入状态 s , 供 TD3 演员网络使用。该演员网络构建在两个全连接 (FC) 层上, 每层后接一个激活函数层, 增强网络的非线性表达能力。最后, 通过将最终全连接层与代表机器人线速度 v 和角速度 a 的动作参数联接, 形成了网络的输出层。使用 $\tanh$ 激活函数来确保输出值的范围在 $(-1, 1)$ 之间, 满足实际动作执行需求。在实施动作前, 需要对最大线速度和最大角速度进行缩放, 公式如下:

$$a = \left[v_{\max} \left(\frac{a_1 + 1}{2} \right), \omega_{\max} a_2 \right] \quad (4)$$

由于激光读取仅限于机器人前方的数据, 因此, 我们只考虑正向运动, 且将线速度调整为正值。Q 值, 即状态-动作对 $Q(s, a)$ 的评价指标, 由两个批评者网络进行评估。这两个批评者网络在结构上是一致的, 但其参数更新是延迟进行, 从而允许参数值之间存在差异。批评者网络将状态 s 和动作 a 一同作为输入。输入的状态 s 首先经由全连接层处理, 随后进入带有输出 L_s 的 ReLU 激活函数。该层的输出以及动作 a 都会被分别送入两个尺寸相同的全连接层, 进行独立变换。而后, 这些层按照特定的方式进行组合。

$$L_c = L_s W_{\tau_1} + a W_{\tau_2} + b_{\tau_2} \quad (5)$$

L_c 为组合全连接层 (CFC), W_{τ_1} 、 W_{τ_2} 分别是 τ_1 、 τ_2 权重, b_{τ_2} 为 τ_2 层偏差。应用 ReLU 激活函数后, 连接输出层, 输出层参数代表 Q 值。选取两个批评网络中最小的 Q 值作为最终输出, 避免状态-动作对值过高估计。网络架构见图 2。奖励函数如下:

$$r(s_t, a_t) = \begin{cases} r_g = 100 & \text{if } D_t < \eta_D \\ r_c = -100 & \text{if 碰撞} \\ v - |\omega| = 80 & \text{其他,} \end{cases} \quad (6)$$

其中时间步 t 的状态-动作对 (s_t, a_t) 的奖励 r 取决于三个条件。如果在当前时间步 D_t 到目标的距离小于阈值 η_D , 则应用正目标奖励 r_g 。如果检测到碰撞, 则应用负碰撞奖励 r_c 。如果这两个条件都不存在, 则基于当前线性速度 v 和角速度 ω 应用即时奖励。为了引导导航策略朝着给定目标前进, 在计算以下采用了延迟属性奖励方法:

$$r_{t-i} = r(s_{t-i}, a_{t-i}) + \frac{r_g}{i}, \quad \forall i = \{1, 2, 3, \dots, n\} \quad (7)$$

其中 n 是更新奖励的先前步骤的数量。这意味着正目标奖励不仅归因于达到目标的状态-动作对, 而且还会在它之前的最后 n 步上递减。该网络学习了一个本地导航策略, 该策略能够到达本地目标, 同时直接从激光输入避开障碍物。

3 实验结果

本文设计实验在 Ubuntu20.04 环境下进行, Ros 版本选择 Ros Noetic, 用 Gazebo 建立机器人仿真环境, 机器人的初始位置和目标点随机设定。如图 2 所示, 经模型训练表明: 机器人经过 200epochs 的训练之后平均奖励值稳定在正区间, 机器人可以快速到达随机目标点, 完全满足预期结果。

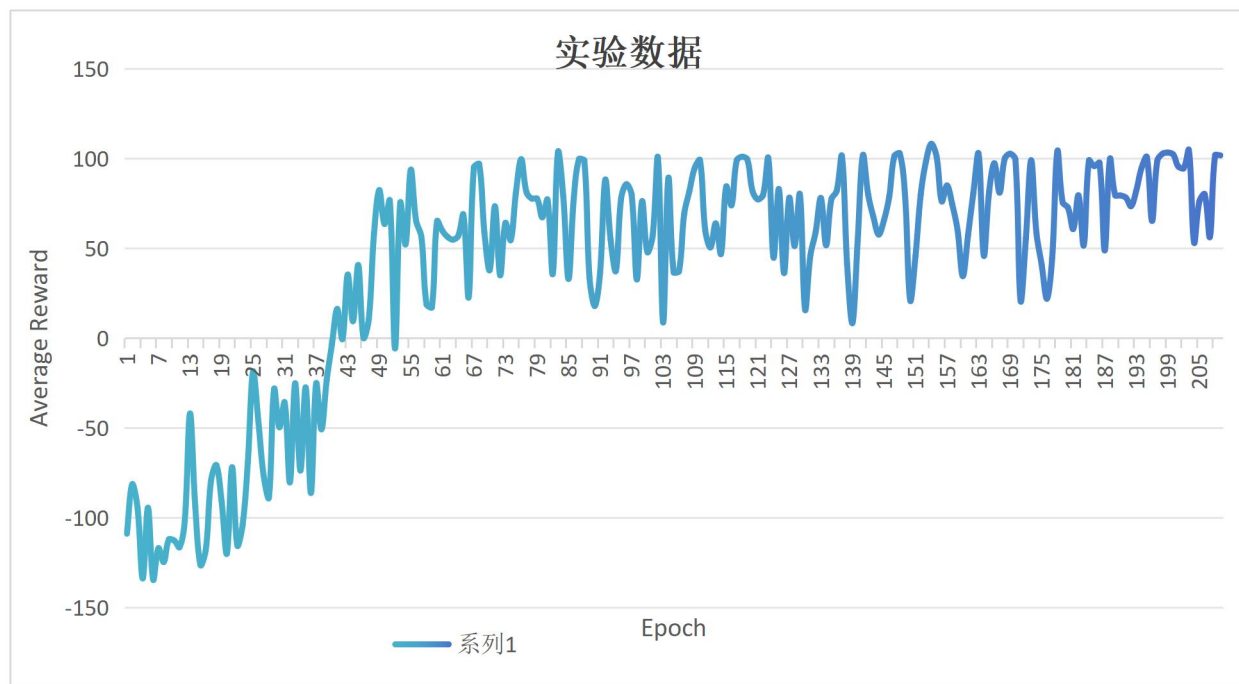


图 2 实验结果

4 结语

综上所述, 本文将机器人的探索过程进行分解, 利用决策、规划和地图绘制模块, 构建了一个通用的探索框架, 通过神经网络从局部地图中学习探索策略, 用 Gazebo 建了立机器人仿真环境进行测试, 结果表明, 该算法具有较好的学习效率和未知环境的适应性, 能够应对未知环境的挑战。同时, 在物理机器人上也进行了实验, 实验结果验证了学习的策略可以很好地从模拟转移到真实的机器人, 为相关领域的发展提供了借鉴。

参考文献

- [1] 李业谦, 陈春苗. 基于 ROS 和激光雷达的移动机器人自动驾驶系统设计[J]. 现代电子技术, 2020, 43(10): 176-178, 183.
- [2] 金书奎, 寇子明, 吴娟. 煤矿水泵房巡检机器人路径规划与跟踪算法的研究[J]. 煤炭科学技术, 2022, 50(5): 253-262.
- [3] 王清珍, 杨丽. 远距离激光机器人嵌入式视觉传感器灵敏度优化方法[J]. 激光杂志, 2022, 43(7): 184-188.

[4] 黄凌杰. 服务机器人室内投递的全局路径规划仿真[J]. 福建电 脑, 2023, 39(3): 104-106.

[5] 李春晖, 周松涛, 张谢许, 等. AGV 自动驾驶智能化计量设备搬运技术研究及应用[J]. 河南科技, 2021, 40(32): 47-50.

作者简介: 马洪武 (1976-), 男, 汉族, 籍贯: 内蒙古赤峰市, 本科, 工程师, 主要研究方向: 化工机械。王利利 (1974-), 女, 汉族, 籍贯: 内蒙古赤峰市, 硕士研究生, 正高级实验师, 主要研究方向: 机器人技术。

张滨杰 (2004-), 男, 汉族, 籍贯: 吉林省吉林市, 大学本科在读, 研究方向: 机械设计。

基金项目: 内蒙古自治区教育科学研究“十四五”规划课题“新质生产力背景下工程创新人才培养模式研究”(课题批准号: NGJGH2024019); 内蒙古工业大学教育教学改革项目“强本固基, 课赛融合, 地方高校高水平实习实训基地建设与实践”(项目批准号: 2024108)。