

Overview of Flexible Needle Puncture Based on Reinforcement Learning Algorithms

Jintao Yu Shuo Wang Ting Yang*

School of Medical Information and Engineering, Xuzhou Medical University, Xuzhou, Jiangsu, 221004, China

Abstract

With the emphasis of puncture therapy in the field of medical therapy and sports science, how to implement puncture therapy efficiently and safely has become an important research topic. Reinforcement learning algorithms provide effective ways to solve this challenge. This paper reviews the research progress of QL reinforcement learning algorithms such as DQN, PG and AC suitable for flexible needle puncture path planning, analyzes their respective advantages and disadvantages, and points out that flexible needle puncture through reinforcement learning model can effectively improve the efficiency and safety of puncture therapy.

Keywords

flexible needle; reinforcement learning; deep reinforcement learning

基于强化学习算法的柔性针穿刺综述

余锦涛 王硕 杨婷*

徐州医科大学医学信息与工程学院, 中国·江苏 徐州 221004

摘 要

随着医学治疗和运动科学领域对穿刺治疗的重视, 如何高效且安全地实施穿刺治疗已成为一个重要的研究课题。强化学习算法为解决这一挑战提供了有效的方法。论文回顾了适用于柔性针穿刺路径规划的QL、DQN、PG、AC等强化学习算法的研究进展, 分析了它们各自的优缺点, 并指出通过强化学习模型进行路径规划的柔性针穿刺, 能够有效提升穿刺治疗的效率和安全性。

关键词

柔性针; 强化学习; 深度强化学习

1 引言

柔性针穿刺是一种在微创介入手术中广泛应用的技术, 相比传统刚性针, 柔性针具有更好的灵活性和适应性, 能够在软组织中以曲线路径前进, 从而避开重要血管、骨骼和器官, 减少对患者的损伤。然而, 柔性针穿刺也面临控制精度、路径矫正能力和实时信息感知等挑战, 这增加了临床手术的难度, 通常需要医生投入大量时间训练, 并依赖医疗影像设备进行引导。而应用强化学习算法的机器人操作柔性针穿刺能够避免缺点。适用于柔性针穿刺路径规划的强化学习算法主要包括: 适合离线学习和小状态空间的 QL、能处理高维输入和动态环境的 DQN、可以直接优化策略, 适应性强的 PG 和结合了价值评估与策略优化的 AC。

【作者简介】余锦涛(2005-), 男, 中国江苏徐州人, 本科, 从事医学信息工程研究。

【通讯作者】杨婷(1986-), 女, 博士, 从事路径规划算法、平台设计与搭建研究。

2 柔性针

2.1 柔性针介绍

柔性针在微创介入手术中广泛应用, 相比传统刚性针, 其具备显著优势。由于柔韧性, 柔性针能够避开障碍物并纠正路径偏离, 实现深度目标插入, 特别适合在肋骨、血管和神经等重要组织间进行复杂穿刺。机器人辅助系统被广泛用于操控柔性针, 以实现精准穿刺, 这些系统结合了机器人的灵巧操作与柔性针的变形控制能力, 能够实时精确控制进针位置、姿态、穿刺轨迹和目标靶点。

2.2 柔性针的特性

①弯曲变形能力: 柔性针在穿刺过程中能弯曲, 避开血管和神经, 实现靶向穿刺, 这使其在微创手术中具有明显优势, 能够到达传统刚性针无法到达的区域。

②力学特性: 柔性针的力学建模和受力分析是研究的重要部分, 例如优化针尖角度和针体设计, 提高穿刺效率和安全性。

③操控性和路径规划: 通过控制进给和旋转运动可调整针尖位姿, 达到预测穿刺路径。

2.3 柔性针穿刺的难点

柔性针穿刺手术在临床应用中面临多个技术难点，主要包括：

①路径规划的复杂性：需要避开重要血管和神经^[1]，一般直线或曲线路径无法满足要求，需采用复杂的曲线路径，同时考虑针体姿态的优化和多种路径组合。②针—组织相互作用的建模^[2]：柔性针与软组织的相互作用复杂，难以建立精确模型，涉及针体形态和针尖位姿的实时追踪，以及针与组织之间的力和摩擦建模。③实时反馈和控制：仅依靠路径规划无法实现柔性针的实时状态反馈，需要结合复合操控技术以实现更精确控制。

3 强化学习算法

3.1 强化学习算法简介

强化学习^[3]是一种通过与环境交互学习最优策略的方法。其核心思想是智能体通过试错和探索与利用的方式，最大化累积奖励。基本框架包括状态(S)、动作(A)、奖励(R)和策略(P)。智能体选择动作影响环境状态，并根据反馈调整策略，以实现最优决策。

适用于柔性针穿刺的常见强化学习算法包括QL、DQN、PG和AC。强化学习能够在复杂环境中学习有效的决策策略，并且随着深度学习技术的发展，其应用前景愈发广阔。

3.2 强化学习算法原理

强化学习通过与环境交互、试错学习和探索与利用策略，逐步优化智能体行为以最大化累积奖励，其核心要素包括以下几个方面：

①马尔可夫决策过程(MDP)：描述智能体在不同状态下采取动作并获得奖励的过程，包含状态空间、动作空间、转移概率、奖励函数和折扣因子，目标是最大化长期累积奖励。②状态、动作、奖励和策略：智能体通过与环境交互尝试不同动作，改变环境状态并获得奖励。策略是智能体在给定状态下选择动作的规则。③价值函数：用于评估状态或状态-动作对的好坏，包括状态值函数和动作值函数。④策略优化：通过不断更新策略来提升智能体表现，常见方法包括基于值函数的QL和SARSA等。⑤深度强化学习：结合深度神经网络和强化学习，处理更复杂的状态和动作空间，常见算法有DQN、DDPG和PPO。

3.3 适用于柔性针穿刺路径规划的强化学习算法实例

3.3.1 Q-learning (QL) 算法

①基本概念：QL是一种无模型的强化学习算法，旨在通过学习行动-价值函数(Q函数)优化智能体策略。它通过与环境交互获取经验，并用这些经验更新Q值，从而使智能体在给定状态下选择最佳行动。Q值表示在某一状态下采取某一行动的预期收益。

②工作原理：QL的核心是Q值的更新。智能体在每个时间步中根据当前状态选择一个行动，并从环境中获得奖励和下一个状态。然后，使用以下Q值更新公式：

其中：

$$Q(s,a) \leftarrow Q(s,a) + \alpha(r + \gamma \max_{a'} Q(s',a') - Q(s,a))$$

$Q(s,a)$ 是当前状态s下行动a的Q值。

式中： α ——学习率。

r ——当前奖励。

γ ——折扣因子。

s' ——执行行动后的新状态。

通过不断更新Q值，智能体能够逐渐学习到最优策略。

③算法步骤：

初始化：创建Q值表，通常初始化为零或随机值。

选择动作：在每个时间步，使用策略选择行动，即以的概率随机选择行动，以的概率选择当前Q值最高的行动。

执行行动：将选择的行动应用于环境，获取奖励和下一个状态。

更新Q值：根据奖励和下一个状态更新Q值。

重复：重复步骤2到4，直到满足终止条件。

④优点：

简单易懂：QL的理论基础简单，易于实现。

无模型特性：不需要环境的模型信息，适用于各种动态环境。

⑤缺点：

收敛速度慢：在复杂环境中，收敛所需的样本数量可能非常大。

高维状态空间问题：状态或动作空间较大时，Q表会变得庞大，导致计算和存储开销增大。

3.3.2 Deep Q-Network (DQN) 算法

①基本概念：

DQN是将深度学习与QL结合的强化学习算法。它通过深度神经网络(DNN)来近似Q值函数，允许智能体在高维状态空间中进行学习。

②工作原理：

DQN的核心思想是使用一个深度神经网络来表示Q值函数。具体来说，网络输入是当前状态，输出是各个可能行动的Q值。DQN通过环境的交互，使用经验回放(ER)和目标网络(TN)来稳定训练过程。

③算法步骤：

初始化：创建主Q网络和目标Q网络，初始化参数，并建立经验回放池。

选择动作：在每个时间步，使用策略选择行动，随机选择或选择当前Q值最高的行动。

执行行动：在环境中执行所选行动，获取奖励和下一个状态，并将经验存储到经验池中。

抽样训练：先从经验池随机抽取一批样本再，使用目标Q网络计算目标Q值，随后更新主Q网络的参数，最小化Q值的均方误差，最后更新目标网络：按照预定时间间隔，将主Q网络的参数复制到目标Q网络。

重复：重复步骤2到5，直到满足终止条件。

④优点:

处理高维状态空间: DQN 能有效处理复杂输入, 扩展了强化学习应用场景。

经验回放: 随机抽样训练数据, 减少相关性, 提高学习效率 and 稳定性。

⑤缺点:

超参数调整复杂: DQN 涉及多个超参数, 调整需要经验。

非最优收敛: 在复杂策略空间中, DQN 可能收敛到次优解。

3.3.3 Policy Gradient (PG) 算法

①基本概念。

策略梯度是一类强化学习算法, 直接优化策略函数。策略梯度不需要构建价值函数, 而是通过梯度上升优化策略参数, 提高行动的期望回报。策略可以是确定性的或随机的, 通常使用神经网络表示。

②工作原理。

策略梯度的核心思想是计算策略相对于预期回报的梯度, 以优化策略。算法通过采样智能体的行为并计算其回报, 然后利用这些回报更新策略。策略梯度定理提供了计算梯度的基础, 公式如下:

$$\nabla J(\theta) = E_{\tau \sim \pi_{\theta}}[t=0 \sum \nabla \log \pi_{\theta}(a_t | s_t) R(\tau)]$$

式中: $J(\theta)$ ——期望回报的目标函数。

$\pi_{\theta}(a_t | s_t)$ ——在状态下采取行动的策略。

$R(\tau)$ ——完整轨迹的回报。

③算法步骤。

初始化: 初始化策略参数 θ 和学习率 α 。

采样: 从当前策略中生成一系列的状态 - 动作轨迹。

计算回报: 对于每个轨迹, 计算回报。

估计梯度: 使用上述策略梯度公式估计梯度。

更新参数: 通过梯度上升更新策略参数:

$$\theta \leftarrow \theta + \alpha \nabla J(\theta)$$

重复: 重复步骤 2 到 5, 直到满足终止条件。

④优点:

处理高维和连续动作空间: 策略梯度方法可直接处理高维和连续动作空间, 适用性广。

收敛性: 策略梯度方法能够收敛到局部最优解。

⑤缺点:

方差大: 策略梯度的估计通常具有较高方差, 可能导致学习过程不稳定。

样本效率低: 策略梯度通常需要比值函数方法更多样本来实现相同学习效果。

局部最优: 基于局部信息优化可能导致陷入局部最优解。

3.3.4 Actor-Critic (AC) 算法

①基本概念:

AC 结合了策略梯度 Actor (A) 和价值函数 Critic (C)。

A 负责生成和更新策略, C 评估策略表现并提供反馈, 从而提高学习的稳定性和效率。

②工作原理:

AC 的核心思想是利用 C 评估 A 所选择的动作, 并通过反馈更新策略。A 根据当前策略选择行动, C 计算状态或状态 - 动作对的价值, 并提供基于优势信号给 A, 以优化策略。具体来说, C 通过估计价值函数或 Q 值函数来评估价值, A 则使用策略梯度方法调整策略参数, 以最大化预期回报。

③算法步骤:

初始化: 初始化 A 和 C 的参数。

采样: 在当前策略下与环境交互, 生成一系列状态 - 动作对和奖励。

计算优势: 使用 C 计算当前状态的价值或状态 - 动作的 Q 值, 并根据实际获得的奖励计算优势:

$$A(s, a) = r + \gamma V(s') - V(s)$$

更新 C: 通过最小化 TD 误差更新 C 的参数:

$$L(\theta_c) = (r + \gamma V(s') - V(s))^2$$

更新 A: 使用优势信号更新 A 的参数, 通常通过策略梯度方法:

$$\nabla J(\theta_a) \approx A(s, a) \nabla \log \pi_{\theta}(a | s)$$

重复: 重复步骤 2 到 5, 直到满足终止条件。

④优点:

稳定性: AC 方法通过将策略评估与更新分开, 比纯策略梯度方法更稳定。

灵活性: 同时学习策略和价值函数, 适用于多种环境和问题设置。

⑤缺点:

算法复杂性: 实现和调试相对复杂, 因为需要训练和更新两个网络。

需要调参: 不同环境和问题可能需要不同超参数设置, 增加调参难度。

4 结语

柔性针穿刺是一种在微创介入手术中广泛应用的技术, 其相对于传统刚性针有很大的优势, 包括弯曲变形能力、力学特性、操控性和路径规划等方面。但同时也面临着许多难点, 如路径规划复杂性、针 - 组织相互作用建模等问题。适用于柔性针穿刺路径规划的强化学习算法包括 QL、DQN、PG 和 AC 等, 结合这些算法可以提高动态环境适应能力、高维空间处理能力、目标导向性和多目标优化能力等。

参考文献

- [1] 高德东, 赵广伟, 王珊, 等. 针穿刺软组织变形预测模型研究[J]. 生物医学工程学杂志, 2016, 33(3): 442-447.
- [2] 高德东. 柔性针穿刺软组织变形机理及动态轨迹规划方法研究[D]. 杭州: 浙江大学, 2017.
- [3] 褚建华. Q-learning 强化学习算法改进及其应用研究[D]. 北京: 北京化工大学, 2009.