

2024 STA303 人工智能课程大作业

2024年10月

作业介绍

背景介绍:

强化学习（Reinforcement Learning, RL）是一种动态、实时地与环境交互并学习的机器学习方法。RL在训练过程中持续收集经验并更新策略模型。

作业简介:

基于FrozenLake-V1环境探索RL算法的原理，可以调研实现多种强化学习算法（如Q-learning, DQN, TRPO, PPO等），鼓励对实现的方法进行深入分析，如：算法参数及不同模块对性能的影响，也鼓励根据对已有算法的认识进一步修改提出创新性的算法。

示例代码提供了经典的Q-learning和DQN算法供同学们参考。


分组要求

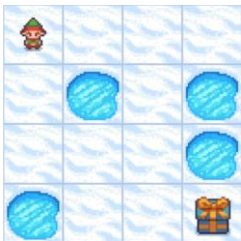
本次大作业要求分组完成，每组3人。组队表格链接在：

<https://docs.qq.com/sheet/DUUpjeGZhVUtPcHFu?tab=BB08J2>

组队截至日期为 **10月31日晚上24:00 (GMT+8)**

环境概述

FrozenLake-V1是Gymnasium库中的经典环境，智能体在冰冻湖面上导航，从起点移动到目标，同时避免掉入洞中（）。冰湖环境主要挑战在于冰面可能是滑的（is slippery=True 时玩家将以1/3的概率朝预定方向移动；以相等的1/3概率朝任一垂直方向移动）。因此，任务目标是通过训练智能体，学习如何在不稳定的冰面上进行最优的动作选择，从而找到最佳路径，在避免危险的同时并成功到达目标位置。



动作、状态空间

FrozenLake-V1环境由一个网格组成。以4x4网格为例，每个位置用一个离散状态表示，网格大小为4x4，共有16个状态，状态编号为0到15。

状态空间

- **状态0**: 起点
- **状态1 - 14**: 冰冻湖面上的位置
- **状态15**: 目标

动作空间

- **向左 (0)**: 向左移动
- **向下 (1)**: 向下移动
- **向右 (2)**: 向右移动
- **向上 (3)**: 向上移动



大作业任务介绍：超参数分析

超参数在模型更新、探索与利用平衡、以及长期奖励优化中扮演着重要角色。然而强化学习算法对多种超参数非常敏感。同学们可以从算法的超参数中选取若干关键项，通过多次实验，考察并分析其在不同设置下对模型训练表现的影响。

- **实验过程：**以DQN算法为例，你需要在frozenlake_dqn.ipynb中选择若干超参数组合，设置不同的实验条件并记录模型训练曲线，包括每个 episode 的累积奖励和损失值曲线。具体的超参数调整可以比较包括但不限于以下超参数比如回合数（train_episodes），学习率（learning_rate）、折扣因子（discount_factor）、epsilon-greedy 策略中的参数（epsilon_max, epsilon_min, epsilon_decay），以及经验回放缓冲区大小（memory_capacity）等。
- **分析结果：**记录不同超参数组合下模型的训练曲线，包括每个 episode 的累积奖励，损失值曲线。尝试使用对比图展示不同超参数的效果，总结不同的超参数调整对训练效果的影响，**给出你对所观察到现象的思考和讨论。**

大作业任务介绍：探索其他RL算法（加分项）

除了给出的基线算法外，同学们可以探索其他强化学习算法在环境中的表现（具体算法不做要求，此任务鼓励有能力同学完成）。

- **实验过程：**同学们可以探索并实现其他更先进的强化学习算法完成FrozenLake任务。例如：Double DQN，Dueling DQN，Prioritized Experience Replay (PER)，TRPO，PPO等。有能力的可以实现更高级的算法
- **分析结果：**通过实验对比改进方法和baseline的表现，例如从以下几个角度进行分析：训练稳定性，模型收敛性，最大化累积奖励等。讨论这些算法的优势和不足，并对这些算法进行超参数分析，给出你对所观察到现象的思考和讨论。

要多表达这一部分，给出自己的理解

参考材料

- 动手学强化学习教程: <https://hrl.boyuai.com/>
- Gymnasium环境介绍: https://gymnasium.farama.org/environments/toy_text/frozen_lake/
- Watkins C J C H, Dayan P. Q-learning[J]. Machine learning, 1992, 8: 279-292.
- Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. nature, 2015, 518(7540): 529-533.
- Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double q-learning[C]//Proceedings of the AAAI conference on artificial intelligence. 2016, 30(1).
- Wang Z, Schaul T, Hessel M, et al. Dueling network architectures for deep reinforcement learning[C]//International conference on machine learning. PMLR, 2016: 1995-2003.

评分标准及时间安排

Project 占总成绩的25分，其中performance 5分/report 10分/presentation 10分。

大作业提交时间安排如下：
**并不鼓励大家刷榜刷性能
后两者也很重要**

第15周

周日：Blackboard提交代码，模型评估文件Project_{group number}.zip (包括代码文件及相应的检查点)。（当实现多个算法时，需要注明用于评估performance的一个，即你认为最好的那个）
必须要运行通

第16周

周四：Blackboard提交presentation slides。

周五：答辩，**展示实现的算法**，实验结果与自己的思考/分析等

周日：提交最终report。

注：本次大作业不强调模型性能，**主要关注同学们对于项目/算法本身的思考与分析。**

作业文件以及代码提交要求

课程补充文件 project.zip 已上传至blackboard, project.zip 包括

- forzenlake_dqn.ipynb & forzenlake_q_learning.ipynb (基线代码)
- utils.py辅助函数库
- requirements.txt环境所需的库列表

准备文件：

- Notebook
- Readme file
- Report
- Slides , pre

代码提交要求

- 格式为Project_{group number}.zip
- 对于超参数分析, 提供相应的可视化代码及其对应的检查点, 对于其他RL算法实现, 提供对应的代码例如forzenlake_ppo.ipynb, 相应的可视化代码及其对应的检查点 (./output) 。
- 提醒：小组需提交一个说明文件ReadMe.md, 说明如何运行代码, 若代码无法复现则视情况扣分。

Slides提交及答辩要求

Presentation slides :

- 格式为.ppt或.pdf
- 该文件将在答辩环节被使用
- 在第一页，请注明小组编号、小组成员
- 所有团队成员都应该在场，可以自行决定是由一个成员还是多个成员完成答辩
- 更具体的答辩要求将在16周前公布

Report提交要求

Report :

- 格式为**.pdf**
- Report使用ICLR 2025 Style Files，正文部分不超过8页（包含图表），Reference部分不作限。模板文件可见<https://github.com/ICLR/Master-Template/raw/master/iclr2025.zip>
- 在第一页，请注明小组编号、小组成员（脚注应说明角色和相应的贡献百分比，无备注情况下默认所有成员均等贡献）
- 迟交按照统计系文件视情况扣分