

2024 Fall - Machine Learning Assignment

介绍

在本次作业中，你将动手实现单智能体和多智能体强化学习的两个典型算法PPO和MADDPG，并在简单环境上测试算法的性能。代码中已经提供了整体的框架，数据处理都已写好，你需要做的是在TODO标出的地方实现核心功能。代码中的所有参数经过了简单的调试，如果你正确实现了所有缺失的核心代码，应该可以获得理想的训练效果。提交内容包括可以运行的完整代码、训练曲线图、实验报告，不需要提交训练出来的模型。

环境配置

```
conda create -n RLIntro python=3.8
conda activate RLIntro
pip install gym[classic_control] pytorch numpy matplotlib pettingzoo
```

Task 1

实现PPO算法，并在CartPole环境上测试。环境介绍可参考

https://gymnasium.farama.org/environments/classic_control/cart_pole/，我们这里使用的是CartPole-v0。

你要完成的任务如下：

- 实现`rl_utils.py`中`ReplayBuffer`的采数据并且处理的机制。
- 实现`ppo.py`中actor和critic的初始化。
- 实现`ppo.py`中actor输出动作的函数。
- 实现`ppo.py`中PPO actor loss的计算。

Task 2

实现MADDPG算法，并在Speaker Listener环境上测试。环境介绍可见

https://pettingzoo.farama.org/environments/mpe/simple_speaker_listener/。

你要在`maddpg.py`完成的任务如下：

- 实现单个actor和critic的初始化
- 实现actor输出动作的函数
- 实现target network的soft update
- 实现MADDPG的actor loss和critic loss