**Assignment 5：A3C & DDPG**

叶增渝 123033910090

由于Pendulum的v0版本已经不适用，所以这里采用v1版本，使用最新的gym0.26来完成任务。

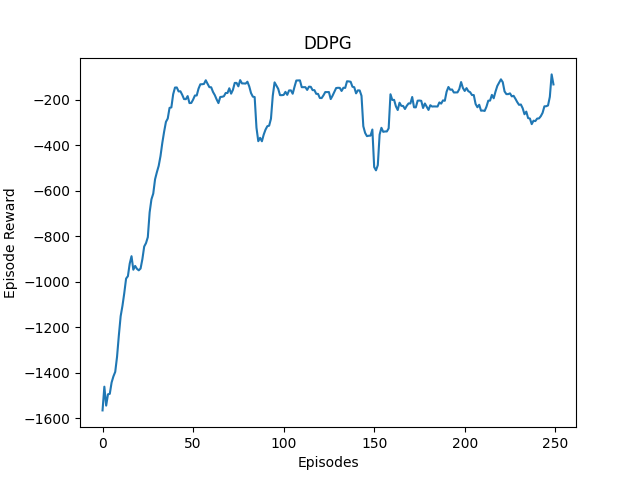
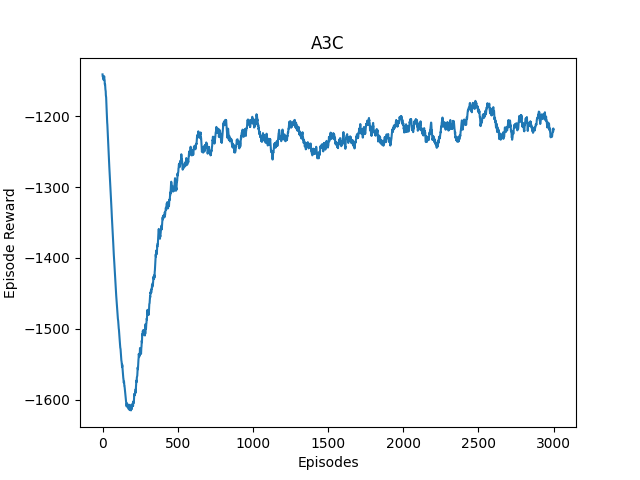
由于v1有完整的关于速度与角度的Reward函数计算公式，因此直接使用该Reward作为buffer Reward即可。

1）A3C为异步AC算法。在传统AC算法中，Actor进行动作选择，Critic进行该动作的价值评估，两者互相联动，实现任务的学习。A3C在AC算法的基础上引入了多个local worker同时进行AC算法更新参数，然后定期地提交参数到global net上进行参数更新。这里我们直接使用torch.multiprocessing进行多线程实现

2）DDPG是一种可用于连续动作空间的强化学习算法，它除了基本的AC算法结构以外，使用了确定性策略、target net与replay buffer，是一个在连续空间上的优秀算法

以下为两个网络实验所得的Episode Reward随Episode的变化曲线图：

我们不难发现，A3C网络虽然在逐步收敛，但是速度缓慢，在近3000的episodes下依然离收敛有很长的路要走；而DDPG网络则是很快能够收敛，在后续训练中虽然有波动但是基本维持在收敛状态。



Tips：源代码在本文件对应的文件夹下，A3C.py与DDPG.py直接执行便可进行训练并保存reward-episode图像。