**人工智能导论cartpole第二次报告**

杨春陶 2017202130

1. 实验目的

在OpenAiGym中，在先前选取的CartPole游戏环境中实现一个基于机器学习的计算模型，从而在游戏中获得更好的成绩。

1. 环境介绍

CartPole是一个车杆平衡游戏，游戏规定小车只能在原点左右各2.4个单位距离之间移动，且杆摆动的幅度不能超过15°。在游戏中，可以观察到每个状态的小车位置、杆摆动幅度、小车速度以及杆角速度。游戏每持续一步，成绩加一，最高分为200，200分后游戏仍可进行，但成绩不再增长。

1. 算法思想

Policy Gradient不通过误差反向传播，它通过观测信息选出一个行为直接进行反向传播，当然出人意料的是他并没有误差，而是利用reward奖励直接对选择行为的可能性进行增强和减弱，好的行为会被增加下一次被选中的概率，不好的行为会被减弱下次被选中的概率。

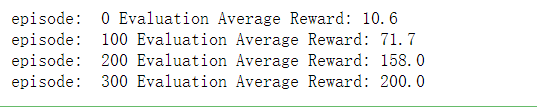
DQN是2013年DeepMind提出来的使用Q-learning与神经网络相结合的方法，其实和Q-learning的思想相同，只不过是计算的时候使用神经网络计算Q值。Q-learning简要说一下，就是使用函数逼近的方法，在选择动作时使用epsilon-greedy的方法，在更新Q函数的时候使用Qmax。

1. 实验结果

策略网络的实验结果



DQN网络中



最后，在小车游戏模型中，进行多次效果检测：



在检验中，小棒模型的每一次运行均能达到最高分200。

可以看到策略网络中，需要250左右每个iteration，每个iteration有10个episode，而DQN修改参数后能在400个episode内达到最优。