# 人工智能导论第一次报告

——在gym中实现基于规则的CartPole

何文宇 2017202012

1. 实验目的

在OpenAiGym中，选取游戏环境，并在其中实现一个基于规则的计算模型，从而在游戏中获得更好的成绩。

在此，选择CartPole车杆平衡游戏，并对其实现了随机策略、简单的选择策略与爬山策略。

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

在实验中评测算法优劣时，用当前策略运行1000次小车模型，所得到的成绩取平均值。

在实验中，可以通过改变action的决定策略来控制小车。

1. 随机策略

令action在当前环境的所有可能的动作的集合中随机选择，即：



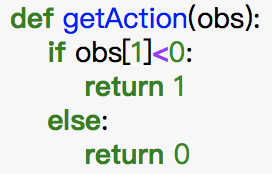
该环境中，小车可以选择左右移动，action取值为0或1。

1. 简单的选择策略

在了解了环境与所能获得的信息后，可以制定一个策略，使杆向左倾斜时，小车选择向左走；杆向右倾斜时，小车选择向右走。使用observation来记录当前状态观察到的四个属性：



observation是一个列表，其中第二个值即为所需的杆倾斜角度。写出根据杆倾斜方向选择小车走向的函数：



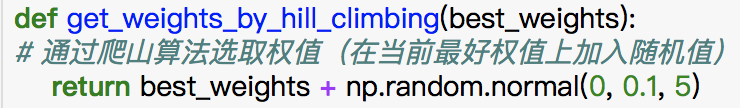
1. 爬山策略

由于可以观测到四个属性，设置五个权值，使其和这四个属性值组成一个一次多项式。要找出合适的权值使该策略能在各种情况下做出合适的action选择，从而取得更好的成绩。采取爬山策略，首先随机选取5个权值：



然后开始对权值进行微调，在现有权值的基础上加上一个小的向量，用新的微调后的权值重新做评测，将所得的成绩均值与原先的成绩均值比较，若更好则更新权值为微调后权值，反之不变。将上述步骤重复一定次数（此处重复10000次），当已达到最佳成绩阈值后停止微调。

微调：



1. 实验结果

使用随机策略时，成绩均值大约在11左右。









使用简单的选择策略时，成绩均值大约在19左右。

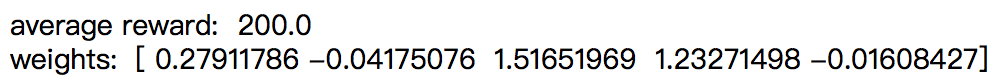


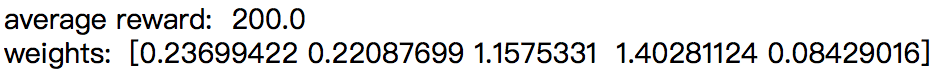


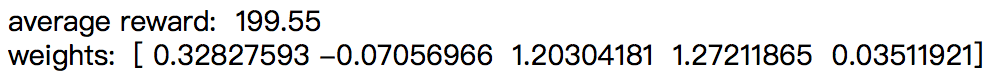


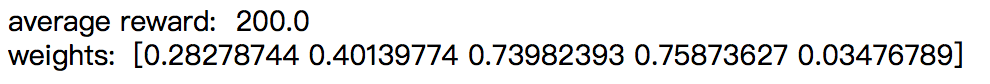


使用爬山策略时，成绩均值可达到满分200。









可以看出，由于爬山策略是局部性优化算法，故其所得权值结果依赖于随机得到的初始权值，结果较为多样。