李昕旸(2017202105)实验一报告

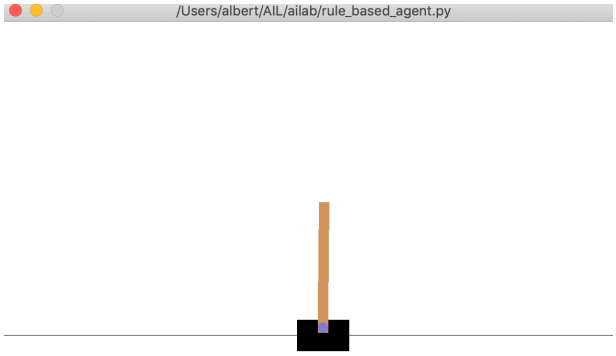
选题： open ai gym

实验环境： Mac OS X Python 3.7

实验概述：

如图所示，本实验装置由一个可以移动的黑色小车和一个细长的木棍组成，我们的主要任务是左右移动小车使得木棍在长时间内不倾倒且小车不撞上左右边缘。本实验在每个状态给我们提供一个长度为4的向量，分别代表：小车位置、小车速度、木棒角度正弦值、木棒角度变化率，上面四个值的负数

表示向左、正数表示向右；而我们需要返回一个决策给系统，返回0代表往左走，返回1代表往右走。



实验过程：

经过一些比较差的策略设计后，我试验出了一些规律：

1. 应该尽量保持小车和木棒稳定
2. 不应碰壁
3. 在木棒角度比较大(比较不稳定)的时候，以角度作为主要参数定决策
4. 在木棒角度比较小(比较稳定)的时候，以角度变化率作为主要参数定决策

由此我制定了以下决策：

1. 每次计算先把角度正弦值换算成角度
2. 当小车距离墙壁很近的时候，让小车向相反方向移动，返回决策
3. 当木棒角度距离垂直较大大于2°时候，做出纠正角度的决策
4. 否则以木棒角度变化率为准，做出纠正角度变化率的决策

主要代码如下：

```python
def get_choice(observation) -> int:
    car_x = observation[0]
    car_v = observation[1]
    pole_alpha_sin = observation[2]
    pole_top_v = observation[3]
    arcsin_alpha = arcsin(pole_alpha_sin) * 180 / 3.14
    # log_info(car_x, car_v, pole_alpha_sin, pole_top_v, arcsin_alpha)
    if car_x - Cart_Position_MIN < 0.1:
        return 1
    elif Cart_Position_MAX - car_x < 0.1:
        return 0
    if arcsin_alpha < - 2:
        return 0
    elif arcsin_alpha > 2:
        return 1
    if pole_top_v < 0.0:
        return 0
    else:
        return 1
```

实验结果：

较为理想，在全部二十次测试中均达到了200次策略的通过标准

```
Episode finished after 200 timesteps   0
Episode finished after 200 timesteps   1
Episode finished after 200 timesteps   2
Episode finished after 200 timesteps   3
Episode finished after 200 timesteps   4
Episode finished after 200 timesteps   5
Episode finished after 200 timesteps   6
Episode finished after 200 timesteps   7
Episode finished after 200 timesteps   8
Episode finished after 200 timesteps   9
Episode finished after 200 timesteps   10
Episode finished after 200 timesteps   11
Episode finished after 200 timesteps   12
Episode finished after 200 timesteps   13
Episode finished after 200 timesteps   14
Episode finished after 200 timesteps   15
Episode finished after 200 timesteps   16
Episode finished after 200 timesteps   17
Episode finished after 200 timesteps   18
Episode finished after 200 timesteps   19
end

Process finished with exit code 0
```