强化学习：作业四

杨思航 191180166

December 17, 2021

1 作业内容

在 gridworld 环境下实现 Model-based Q-learing 算法。

2 实验环境

- NAME = Ubuntu
- VERSION = 20.04.2 LTS(Focal Fossa)
- Tensorflow = 2.7.0

3 实验探究

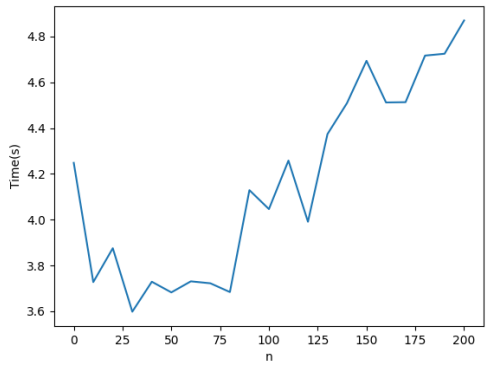
1. Dyna-Q 算法

   (1) 核心代码

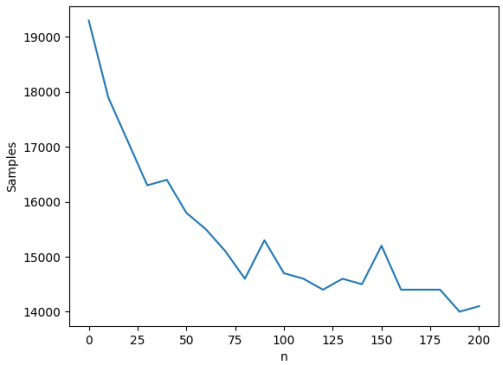
   - policy 的学习部分基于 HW2, 但是将 observation 从三维映射到一维以便于后续 model 的学习
   - DynaModel 的实现借鉴了框架代码中的 NetworkModel
   - 具体实现流程请见 algo.py -> DynaModel

(2) 参数 n 对算法收敛性的影响

- 收敛
  i. 由于框架代码并未显示定义收敛, 因此我自行定义了收敛
  ii. converge_threshold = 87, 87 为最优解于最坏初始情况 (距离钥匙和门的距离之和最大) 下的奖赏
  iii. 若连续 10 次测试均满足 min(reward_episode_set) $\geq$ converge_threshold, 则认为算法收敛
- 实验结果
  i. 收敛时间
  (曲线呈现增长趋势的原因: 学习模型需要额外消耗时间)



  ii. 消耗的样本量 (Total Steps)
  (对于每一个 n 重复执行 10 次取平均值, 由于学习存在随机性, 因此曲线存在小幅度震荡)



- 实验结论: 当 $n = 120$ 时, 收敛时所消耗的样本数量基本趋于稳定, 不再呈明显下降趋势

2. NetworkModel

(1)

4 实验效果

5 复现实验

6 小结