Overview

- 总价值不易计算时，但环境状态有显式的分布时
  - 如何使用迭代法计算总价值
  - 如何使用迭代法反复改进总策略
  - 策略迭代法的收敛

----------THIS CHAPTER FOCUSED ON ----------

- 总价值不易计算时，环境状态没有显式的分布时，从连续的样本和经验中学习

    * 蒙特卡洛方法 (Monte Carlo - MC Method)
    * 计算总价值
    * 更新总策略

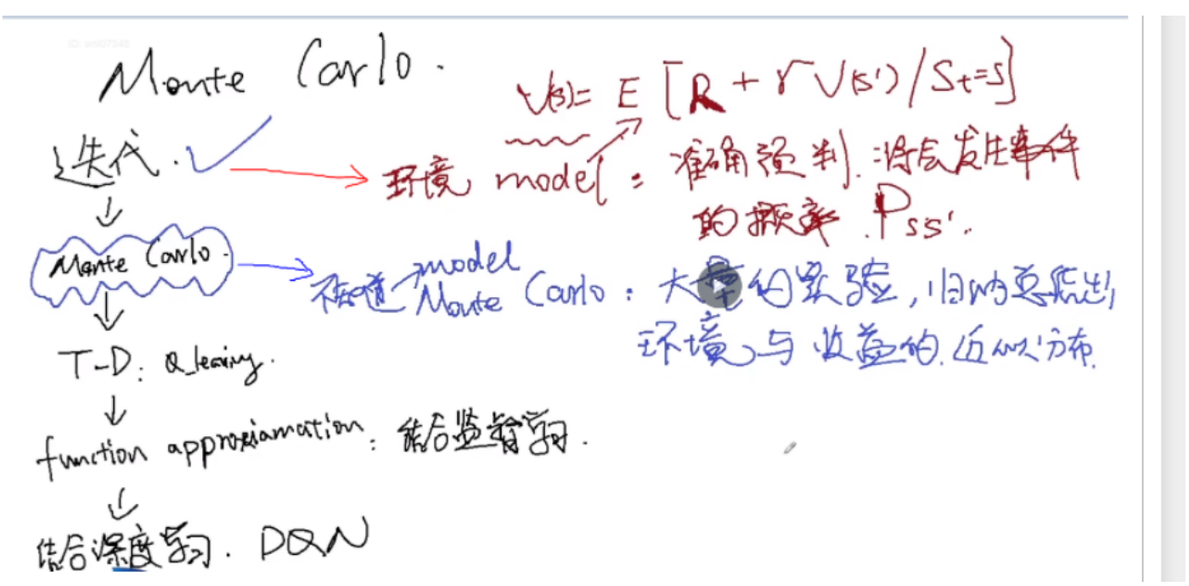
----------END----------

- 总价值不易计算时，环境状态没有显式的分布时，从每一次与环境状态的交互中学习
  - Temporal Differences
  - Temporal Differences与蒙特卡罗方法的对比
  - SARSA
  - Q-learning
- 当环境状态过多，如何将有限样本中的策略推广到更大的状态空间，作为更大状态空间的近似解？
  - 结合监督学习, function approximation
  - 线性方法等
- Q-learing+Deep-Learning
  - DQN
  - DQN的优势与特点

参考视频

参考书籍

参考中文知乎



- Iterative vs. MC:

- MC: 记住：对每一个possibility （path), 都一路走到黑 for each state，all the way until t=end;
  - t=end: example: if it is flappy bird, t=end means bird hit the pillar and game over

A Typical MC example:

In [ ]: