**Model-Free Control**

# 介绍

上一节讲的是无模型的预测，本节讲无模型的控制（value function的迭代优化）。

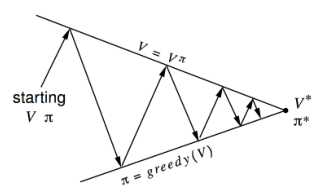
无模型的控制技术可以解决一下两大问题：

1. MDP模型未知，但是有经验数据。
2. MDP模型已知，但是样本空间太大。

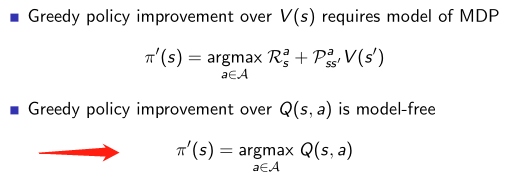
On-policy：相对于off-policy

Off-policy：用别的agent的数据来训练本模型。

一般的策略迭代过程：

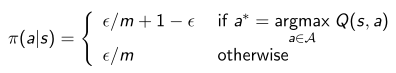


无模型的策略迭代：



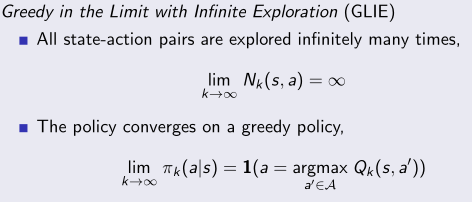
**贪心搜索：**

想法：安得世间双全法，不负如来不负卿？即保持对当前决策的最佳选择，又分一部分的概率到其他的选择上。

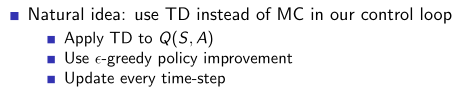


m 是action空间的大小，对最优选择，给予最大的支持。

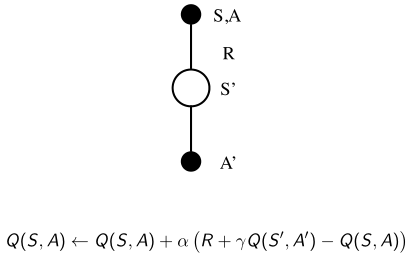
同时，需要指出的是，对非最优action的选择，可能会造成低效，所以一个这种的办法就是逐渐降低otherwise的概率，即所谓的GLIE方法：

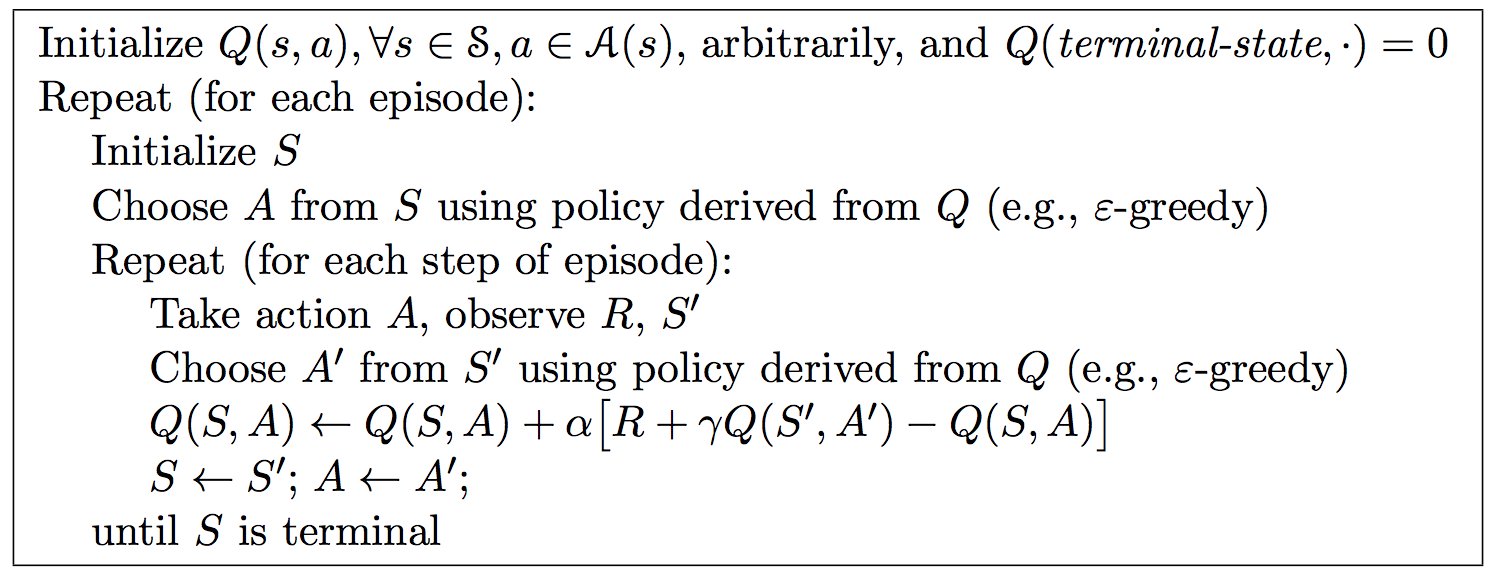


鉴于TD的优势（单步更新机制），接下来介绍TD-control方法

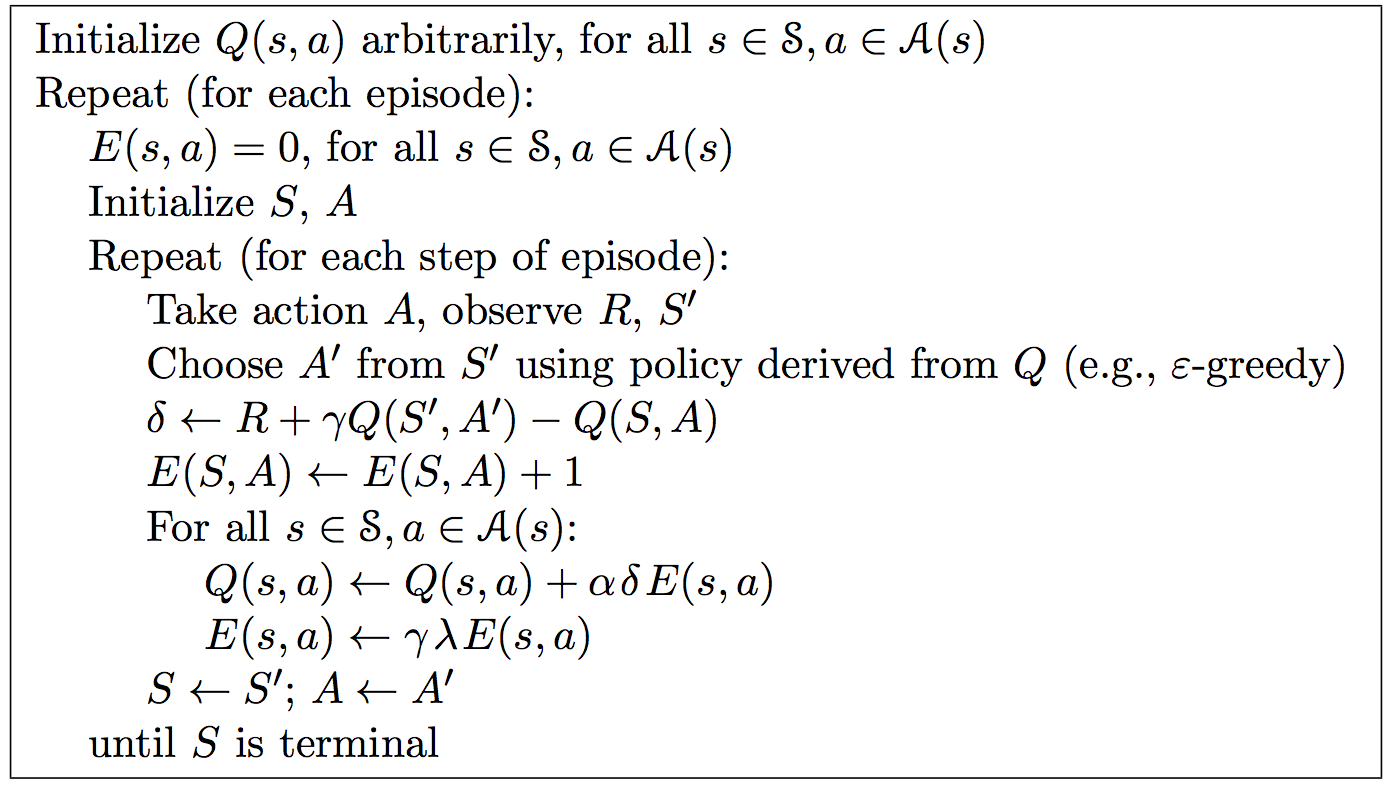


# Sarsr





Sarsr(λ)



<https://www.jianshu.com/p/91fbc682fb3e>

关于内层循环，如果一个s,a没有发生，那么不会进行更新，所以可以理解为所有的历史轨迹。相对于Sarsr而言，这一块增加了历史记忆。

离线TD-Learning—Q learning

