# 强化学习算法---MDP算法分析

本实验分析”马尔可夫决策过程.ipynb”算法。理解马尔可夫决策过程对于理解强化学习具有轴心作用。阅读理解随机过程、马尔可夫过程、马尔可夫奖励过程、回报、价值函数、马尔可夫决策过程、策略、状态价值函数、动作价值函数。

1. 价值函数、状态价值函数、动作价值函数有什么区别和联系？
2. 贝尔曼方程和贝尔曼期望方程有什么区别？
3. 怎样求解MDP中每个状态的价值？
4. 将ipynb粘贴形成.py，利用断点和单步分析法分析sample()子程序，理解情节(episode)和幕的编程技巧。
5. 占用度量计算公式解析：

for episode in episodes:  
 for i in range(len(episode)):  
 (s\_opt, a\_opt, r, s\_next) = episode[i]  
 total\_times[i] += 1  
 if s == s\_opt and a == a\_opt:  
 occur\_times[i] += 1  
for i in reversed(range(timestep\_max)):  
 if total\_times[i]:  
 rho += gamma\*\*i \* occur\_times[i] / total\_times[i]  
return (1 - gamma) \* rho

对应于：