MDP宏观思路

## 思路

马尔可夫过程

马尔可夫奖励过程

马尔可夫决策过程

## 计算方式

Bellman方程计算

数值迭代

## 概念

### 基本定义

动力

策略

### 推论定义

状态转移概率

给定“状态—动作”期望奖励

给定“状态—动作—下一状态”期望奖励

## 奖励、回报与价值函数

回报（return）

折扣（discount）

状态价值函数

动作价值函数，

## Bellman期望方程：用来进行价值评估（value evaluate）

### 基本内容

描述了状态价值与动作价值之间的关系。

用时刻的动作价值表示时刻的状态价值

用时刻的状态价值表示时刻的动作价值

### 扩展形式

## Bellman最优方程

由于最优策略有以下特点：

所以，将该特性应用于Bellman期望方程，得到