智能体三要素，分类，规划与学习

有模型与免模型

智能体三要素：模型、价值函数、策略

策略：智能体会用策略来选取下一步的动作

确定性策略：左0.7,右0.3，选左

随机性策略：左0.7,右0.3，建立分布采样

价值函数：评估智能体进入某个状态后，可以对后面的奖励带来多大的影响。

$$V_\pi(s) \doteq \mathbb{E}_\pi\left[G_t \mid s_t = s\right] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s\right]$$

状态s使用策略 π 的时候，到底预期可以得到多少奖励

$$Q_\pi(s,a) \doteq \mathbb{E}_\pi\left[G_t \mid s_t = s, a_t = a\right] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a\right]$$

状态s使用动作a 的时候，到底预期可以得到多少奖励

智能体三要素：模型、价值函数、策略

模型：模型包括状态转移概率和奖励函数两个部分

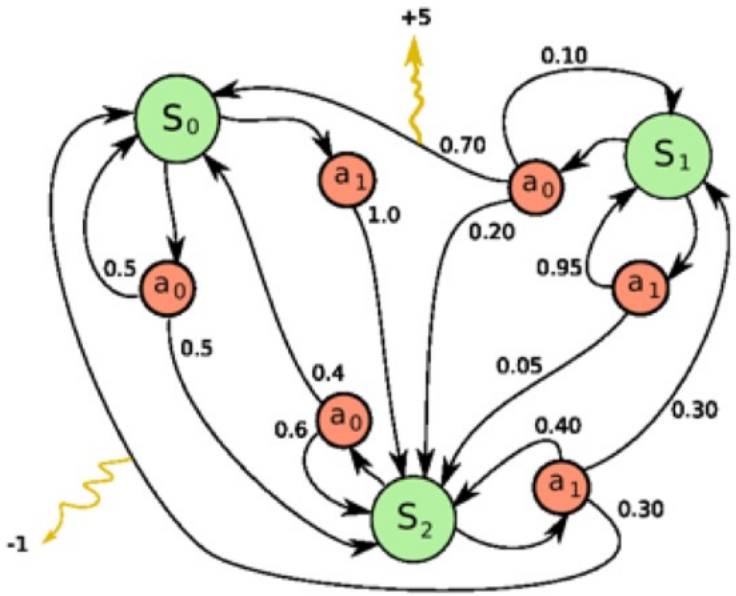
动作。它由状态转移概率和奖励函数两个部分组成。状态转移概率即

$$p_{ss'}^{a} = p\left(s_{t+1} = s' \mid s_t = s, a_t = a\right)$$

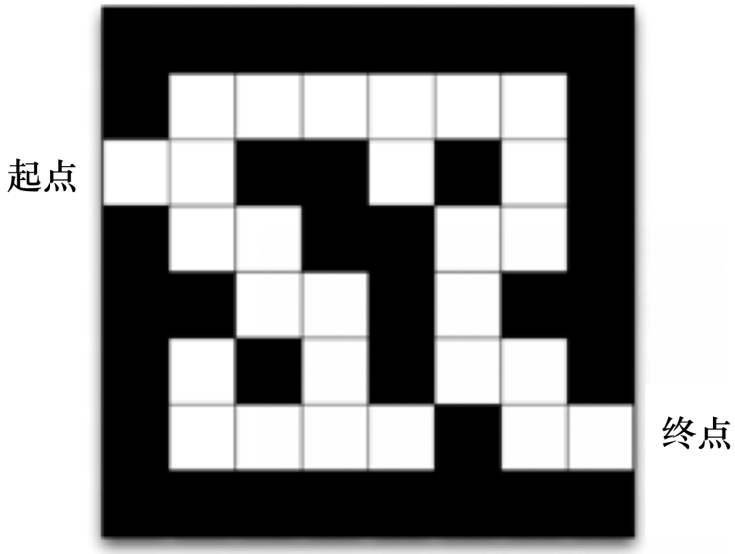
奖励函数是指我们在当前状态采取了某个动作，可以得到多大的奖励，即

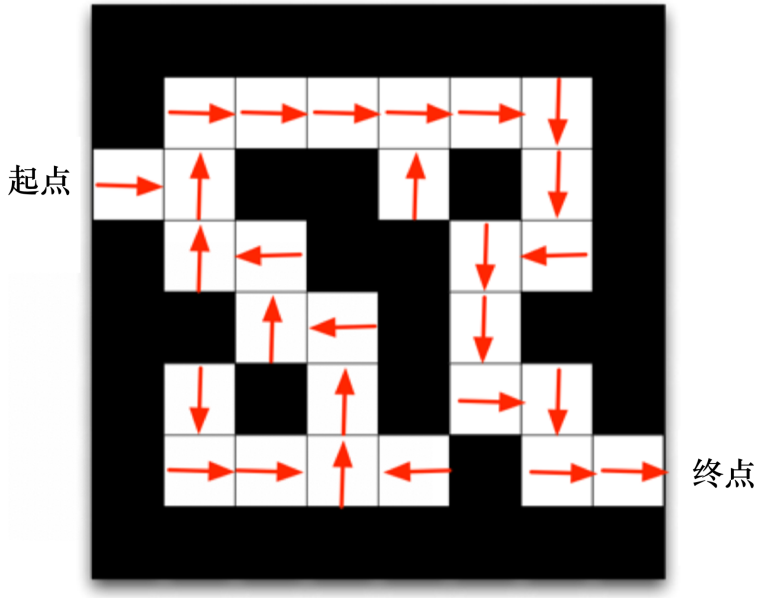
$$R(s,a) = \mathbb{E}\left[r_{t+1} \mid s_t = s, a_t = a\right]$$

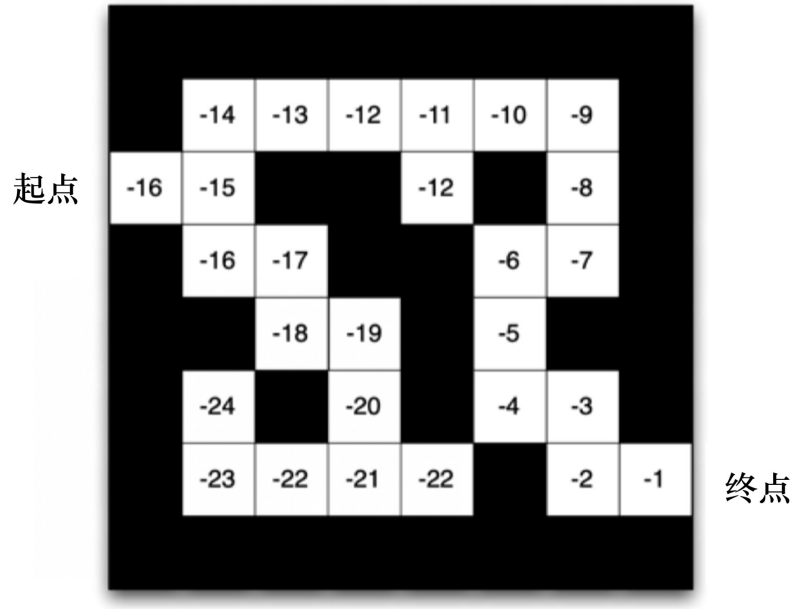
马尔可夫决策过程=模型+价值函数+策略



智能体的分类：基于策略的，基于价值的，演员-评论员





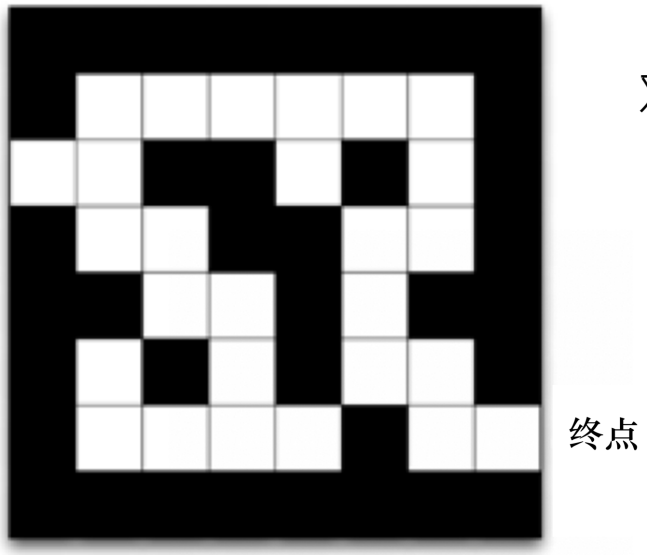


每走一步获得-1奖励

哪个是基于策略的？哪个是基于价值的？

规划（预测）与学习（控制）；有模型、免模型智能体

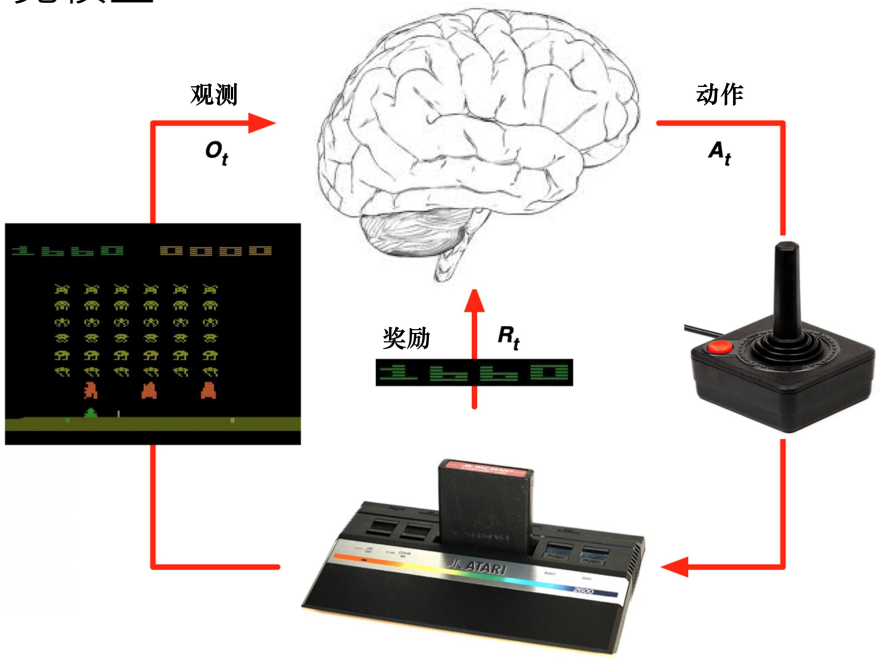
有模型



对数据要求较低

在某个状态进行上下左右运动的奖励与概率
非常方便知道
建模一个迷宫世界，用贝尔曼方程求解最佳策略

免模型



数据驱动
泛化性能好

在某个状态进行上下左右运动的奖励与概率
不清楚
让智能体通过与环境交互提升策略表现

本章小结

- 强化学习智能体的三要素是？
- 确定性策略/随机性策略；Q价值/V价值；模型的两个组成成分
- 基于策略、基于价值、演员评论员的区别
- 规划和学习有什么区别？
- 有模型与免模型有什么区别？

下一章：OpenAI Gym使用方法

Credit goes to: EasyRL