

强化学习讨论的问题是智能体怎么在复杂、不确定的环境中最大化它能获得的奖励。

这一帧：
向上移动期望奖励：100 v
向下移动期望奖励：60

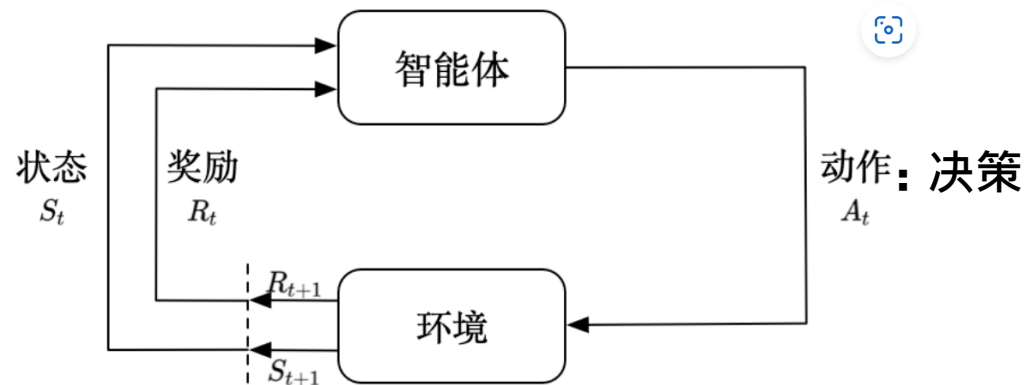


图 1.1 强化学习示意

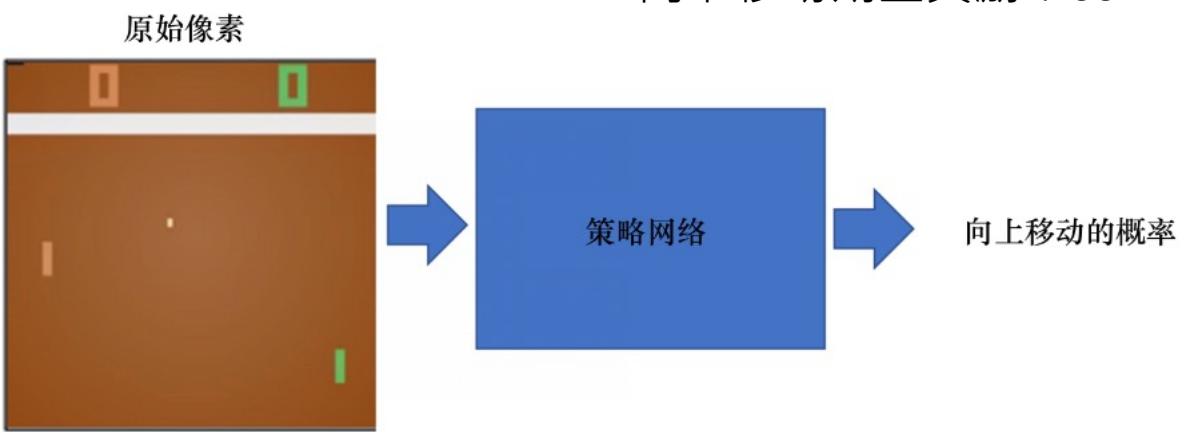


图 1.6 强化学习玩 Pong

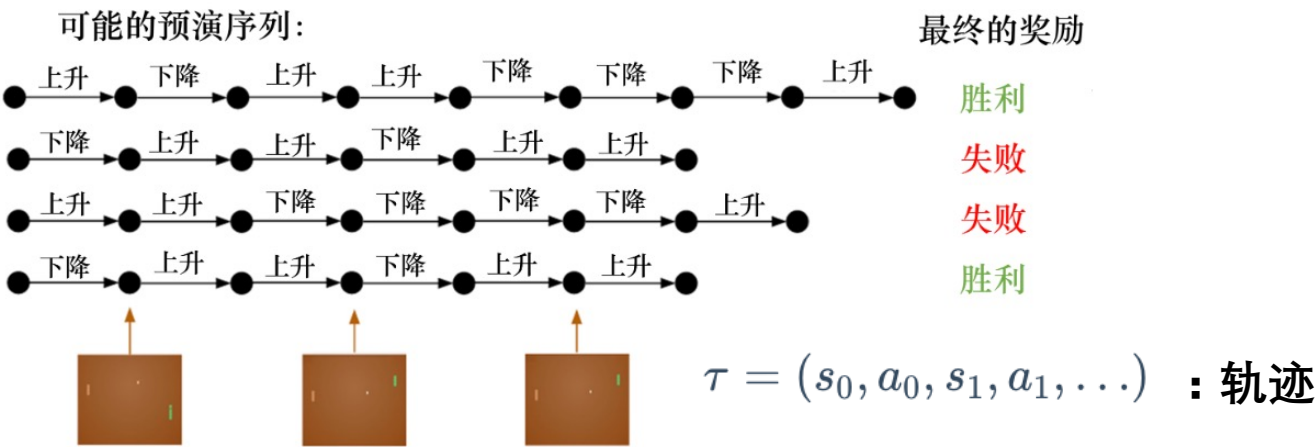


图 1.8 可能的预演序列

强化学习的目标
输入：观测/状态
输出：当前采取动作
学习方法：利用轨迹与奖励

本章小结

- 什么是强化学习？
- 智能体、环境是怎么交互的？（状态、奖励、动作）
- 什么是轨迹？

下一章：奖励、序列决策、动作空间

Credit goes to: EasyRL

