人工智能与信息社会

基于神经网络的智能系统II：状态、动作、回报

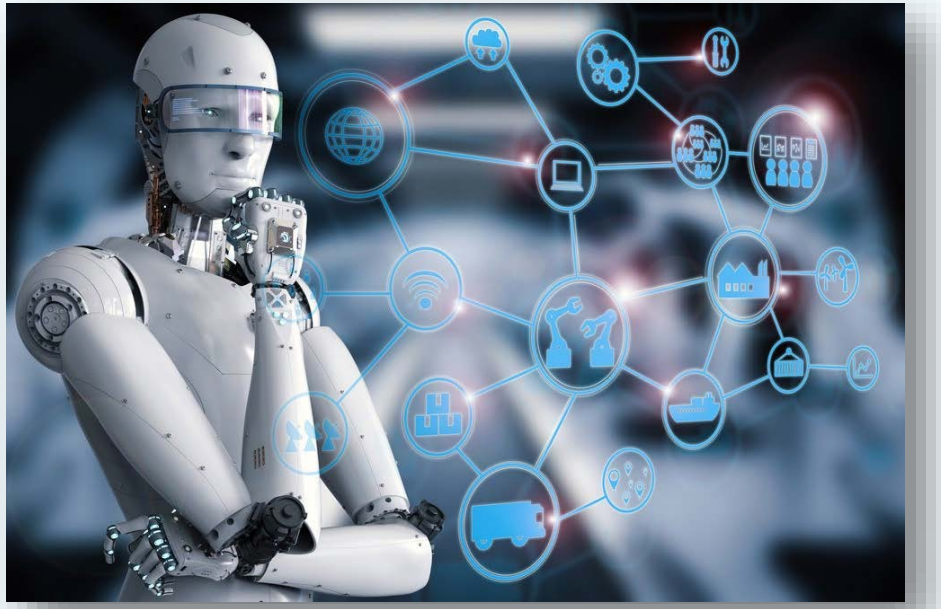
陈斌 北京大学 gischen@pku.edu.cn

强化学习的要素

主体（agent）

负责做出决策的实体。

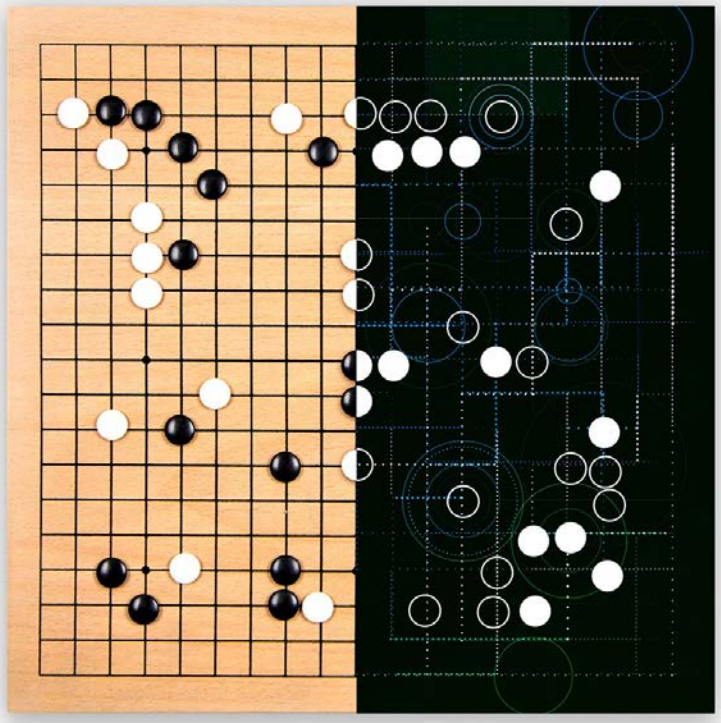
比如Alpha Go、人、玩flappy bird的AI。



强化学习的要素

环境（environment）

主体存在于环境之中，主体的行为作用于环境，并接受环境的反馈。比如一个完整的游戏程序。



强化学习的要素

- **状态（state）**

  环境的状态不断发生变化。不同时刻的棋盘状况、游戏画面各不相同。

- **动作（action）**

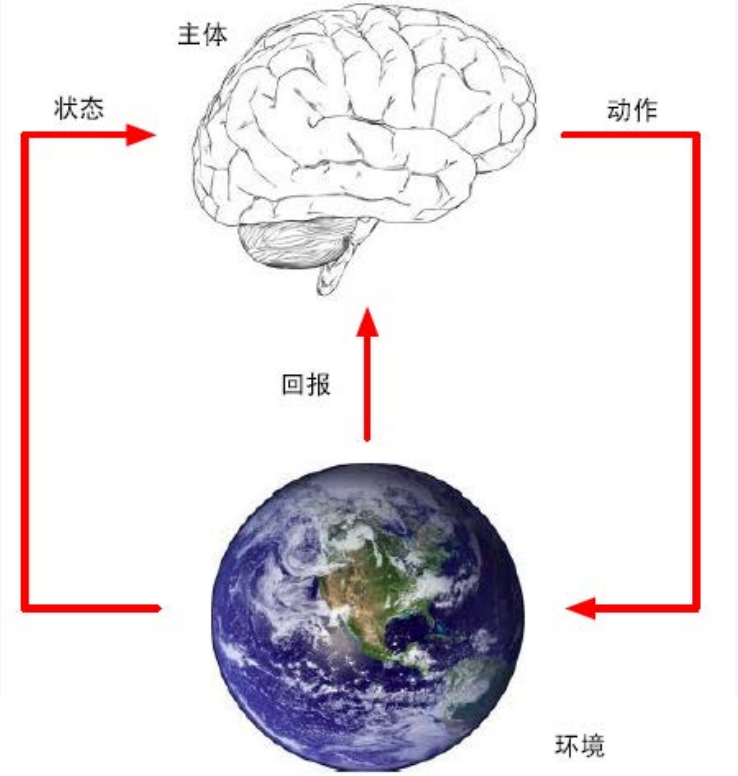
  主体通过执行动作来改变环境的状态。

- **回报（reward）**

  环境状态改变之后会返回主体一个回报，主体可以根据回报来判断动作的好坏。

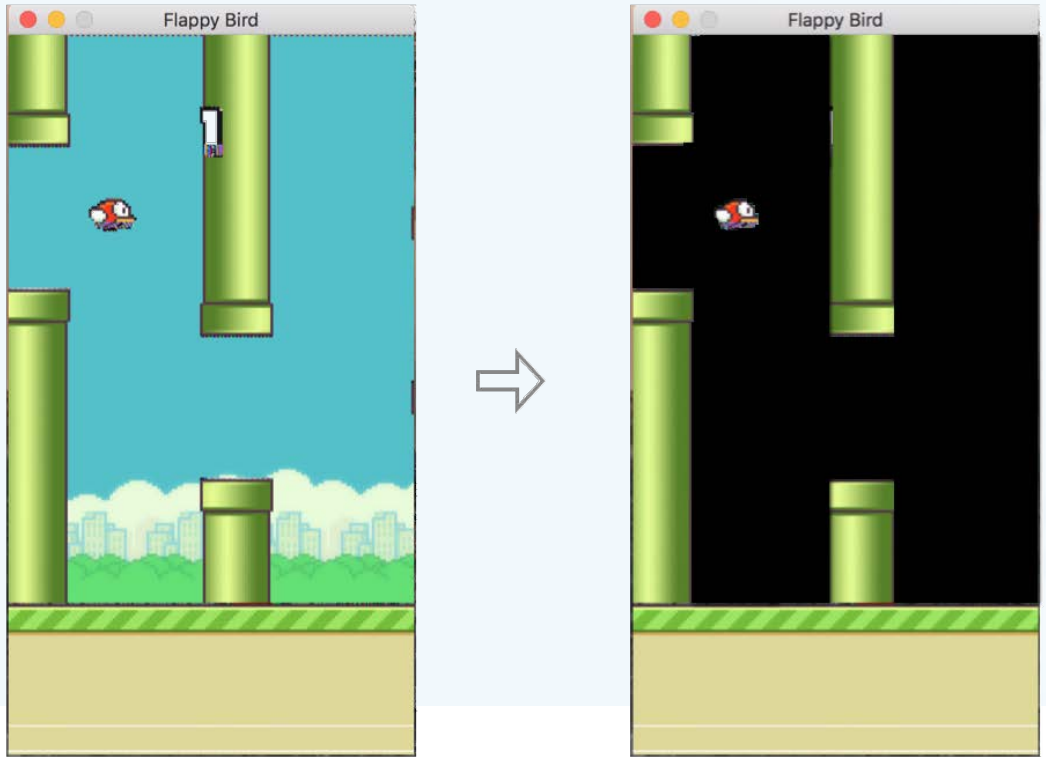
强化学习的主要流程

- 主体与环境不断地进行交互，产生多次尝试的经验，再利用这些经验去修改自身策略。经过大量迭代学习，最终获得最佳策略。



Flappy Bird 状态

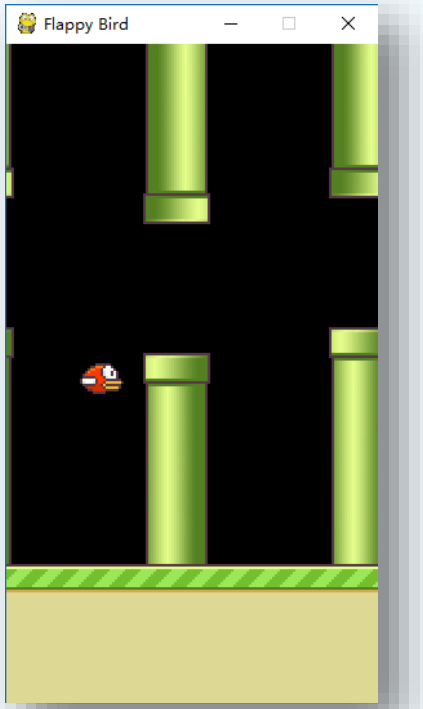
- 每一帧的画面都是一个状态
- 对画面简化，保留AI用于学习的关键信息

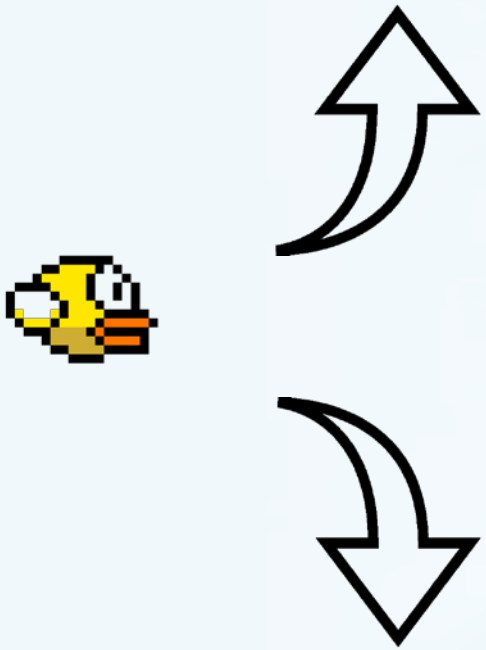


Flappy Bird 动作

每个状态下都有两个可选择的动作

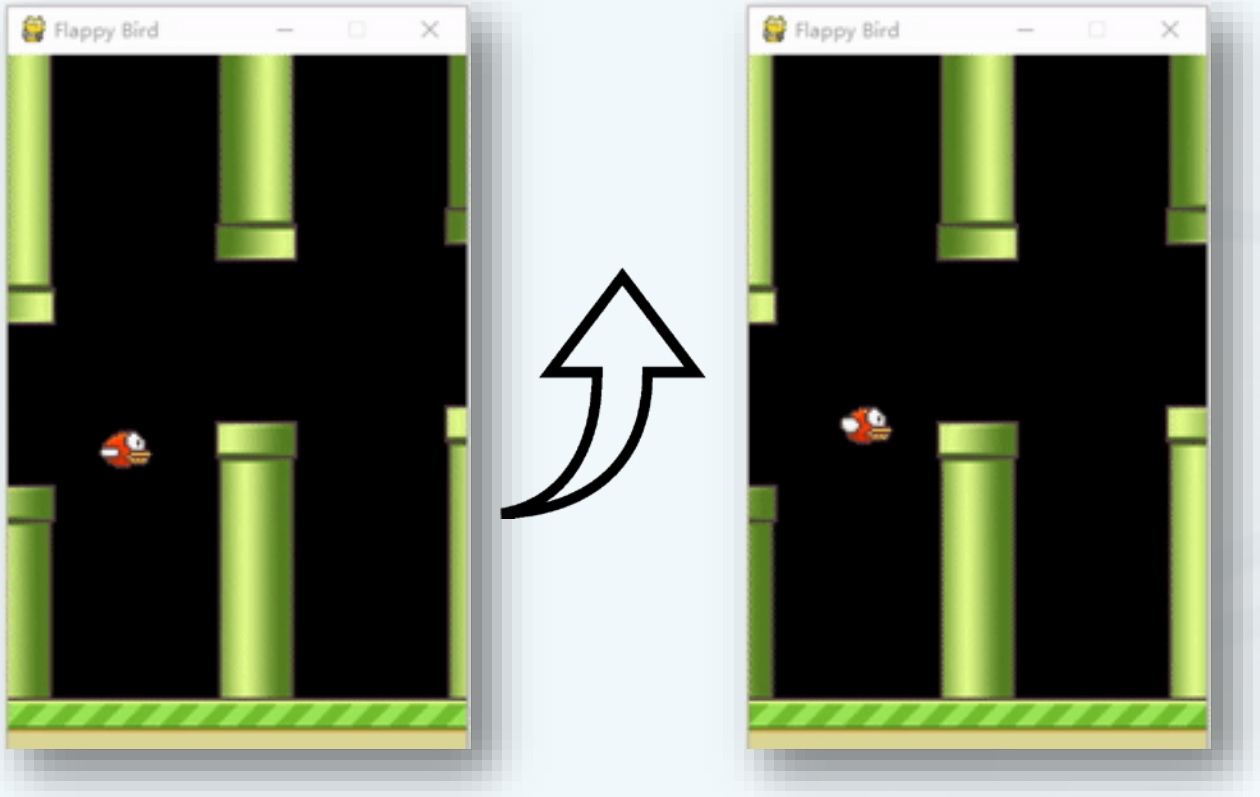
让鸟往上跳 or 什么都不做





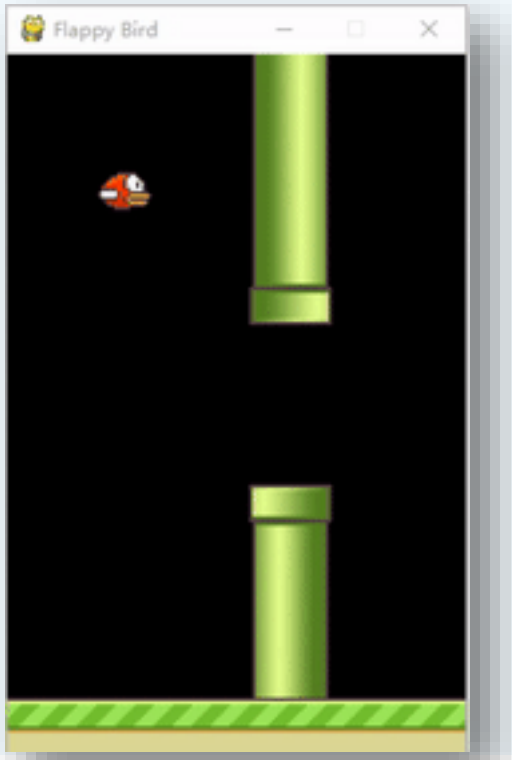
Flappy Bird 动作

- 不同的动作会产生不同的新状态

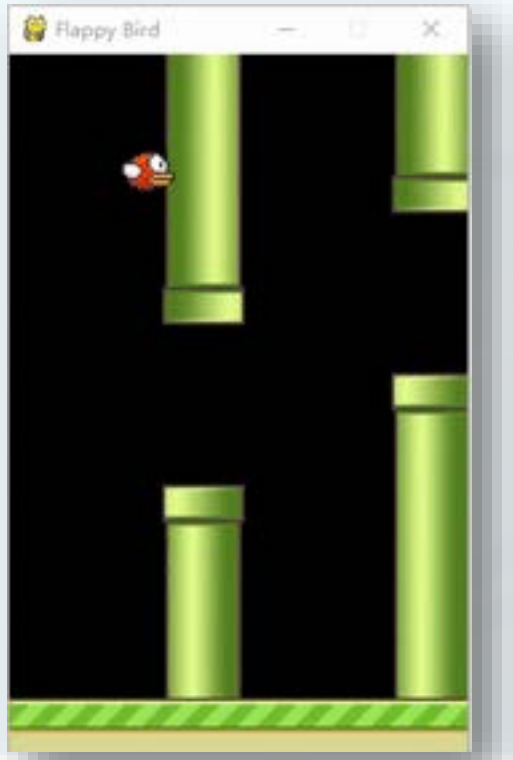


Flappy Bird 回报

- 主体在得到环境给的新状态时也会得到一个回报。简单情况下活着是1，死了是0



活着：1



死了：0