

马尔可夫决策过程 (一)

关键词:

#马尔可夫决策

#state

state

状态 (state) 是代理对环境的观察，并不一定完全等同于环境。

比如：“抽鬼牌”游戏中，代理无法看穿对手的手牌（虽然这个信息对于代理的决策是有益的）；也就是说，环境中存在隐藏信息不会作为状态参与代理的决策。

我们希望代理接受一个状态，并映射到动作空间，实现决策过程。

马尔可夫模型的特点

马尔可夫模型的核心特点是：**下一个状态只与当前状态当前和采取的行为有关。**

即 $P(s_{t+1}|s_t, a_t) = P(s_{t+1}|s_0, a_0, r_1, s_1, a_1, r_2, \dots, s_t, a_t)$

一个典型的例子是棋局，当前状态可以定义为 棋子在棋盘上的摆放。棋子的摆放和当前落下的棋子位置完全决定了下一个状态。不过，这并不意味着 **状态不包含历史信息**（当前状态是由历史动作导致的，只是，我们完全忽略了‘路径’。）

当最终结果和路径无关时，我们往往可以将问题抽象为马尔可夫决策问题。

假设我们已知一个状态（问题满足马尔可夫模型），我们可以推算出下一个状态；根据下一个状态，又可以推算出下下个状态……最终可以推算出未来所有可能的状态（当然，也许计算量会比较大）。

基于马尔可夫模型的特点，我们如此构造代理的策略的映射：

$$f: \text{当前状态} \rightarrow \text{行为}$$

马尔可夫模型在强化学习中十分重要，虽然实际复杂问题中很多问题并不满足马尔可夫性，但是我们依然可以近似为马尔可夫模型处理。当前状态信息越是丰富、表达越是有效，问题就越近似于马尔可夫模型。所以，很多时候，我们需要构造 informative 的当前状态。