Markov Decision Processes (MDP)

esgl Hu

November 10, 2017

1 Background

aaaaaaaaa

2 Notation & Definition

A reinforcement learning task taht satisfies the Markov property is called a *Markov decision process*, or *MDP*. If the state and action spaces are finite, then it is called a *finite Markov decision process (finite MDP)*

Given any state and actin s and a, the probability of each possible pair of next state and reward, $s^{'}$, r, is denoted

$$p(s^{'}, r|s, a) \doteq Pr(\{S_{t+1} = s^{'}, R_{t+1} = r|S_t = s, A_t = a\}) \tag{1}$$

Given the dynamics as specified by (1), one can compute anything else one might want to know about the environment, such as the expected rewards for state-action pairs.

$$r(s, a) \doteq \mathbb{E}[R_{t+1}|S_t = s, A_t = a] = \sum_{r\in\mathcal{R}} r \sum_{s^{'}\in\mathcal{S}} p(s^{'}, r|s, a), \tag{2}$$

the *state-transition probabilities*,

$$p(s^{'}|s, a) \doteq Pr\{S_{t+1} = s^{'}|S_t = s, A_t = a\} = \sum_{r\in\mathcal{R}} p(s^{'}, r|s, a) \tag{3}$$

and the expected rewards for state-action-next-state triples,

$$r(s, a, s^{'}) \doteq \mathbb{E}[R_{t+1}|S_t = s, A_t = a, S_{t+1} = s^{'}] = \frac{\sum_{r\in\mathcal{R}} rp(s^{'}, r|s, a)}{p(s^{'}|s, a)} \tag{4}$$