Reinforcement Learning

Zhang Fengchen

Abstract

Leave it blank for now.

1. Background

(1)Probability Definition

$$p\left(s', r \mid s, a\right) = \Pr\left\{S_t = s', R_t = r \mid S_{t-1} = s, A_{t-1} = a\right\}$$

$$\sum_{s' \in S} \sum_{r \in R} p\left(s', r \mid s, a\right) = 1 \quad \forall s \in S, a \in A$$

$$p\left(s' \mid s, a\right) = \sum_{r \in R} p\left(s', r \mid s, a\right)$$

$$r(s, a) = \sum_{s' \in S} \sum_{r \in R} \left(r * p\left(s', r \mid s, a\right)\right)$$

$$r\left(s, a, s'\right) = \frac{\sum_{r \in R}\left(r * p\left(s', r \mid s, a\right)\right.}{p\left(s' \mid s, a\right)}$$

(2)Bellman Equations

v(s):

$$v_\pi(s) = E_\pi\left[G_t \mid S_t = s\right] \quad \forall s \in S$$

$$v_\pi(s) = E_\pi\left[R_{t+1} + \gamma G_{t+1} \mid S_t = s\right] \quad \forall s \in S$$

$$v_\pi(s) = \sum_{a \in A}\left(\pi(a \mid s) * q_\pi(s, a)\right) \quad \forall s \in S$$

$$v_\pi(s) = \sum_{a \in A} \pi(a \mid s) \sum_{s', r} p\left(s', r \mid s, a\right) * [r + \gamma E_\pi\left[G_{t+1} \mid S_{t+1} = s'\right]] \quad \forall s \in S$$

$$v_\pi(s) = \sum_{a \in A} \pi(a \mid s) \sum_{s', r} p\left(s', r \mid s, a\right) * [r + \gamma v_\pi\left(s'\right)] \quad \forall s \in S$$

q(s, a):

$$q_\pi(s, a) = E_\pi\left[G_t \mid S_t = s, A_t = a\right] \quad \forall s \in S, a \in A$$

$$q_\pi(s, a) = E_\pi\left[R_{t+1} + \gamma G_{t+1} \mid S_t = s, A_t = a\right] \quad \forall s \in S, a \in A$$

$$q_\pi(s, a) = \sum_{s', r} p\left(s', r \mid s, a\right) * [r + \gamma E_\pi\left[G_{t+1} \mid S_{t+1} = s'\right]] \quad \forall s \in S, a \in A$$

$$q_\pi(s, a) = \sum_{s', r} p\left(s', r \mid s, a\right) * [r + \gamma v_\pi\left(s'\right)] \quad \forall s \in S, a \in A$$

$$q_\pi(s, a) = \sum_{s', r} p\left(s', r \mid s, a\right) * \left[r + \gamma \sum_{a' \in A}\left(\pi\left(a' \mid s'\right) * q_\pi\left(s', a'\right)\right)\right] \quad \forall s \in$$

Optimal Equations:

$$v_*(s) = \max_{a \in A} q_{\pi *}(s, a) \quad \forall s \in S$$

(3)Dynamic Programming

Policy Improvement Theorem:

$$E_{\pi'}\left[q_\pi\left(s, \pi'(s)\right)\right] \geq v_\pi(s) \quad \forall s \in S$$

$$\pi' \geq \pi \quad \longleftrightarrow \quad v_{\pi'}(s) \geq v_\pi(s) \quad \forall s \in S$$

Policy Evaluation:

$$v_{k+1}(s) = E_\pi\left[R_{t+1} + \gamma v_k\left(S_{t+1}\right) \mid S_t = s\right]$$

$$= \sum_a \pi(a \mid s) \sum_{s', r} p\left(s', r \mid s, a\right)\left[r + \gamma v_k\left(s'\right)\right]$$

Policy Improvement:

$$\pi'(s) = \arg\max_a q_\pi(s, a)$$

Value Iteration:

$$v_{k+1}(s) = \max_a \mathbb{E}\left[R_{t+1} + \gamma v_k\left(S_{t+1}\right) \mid S_t = s, A_t = a\right]$$

2. Proposed Solution

(1)Frozen Lake

3. Numerical Results

Your experiment results should be here. You should add the figure/table if necessary.

. Correspondence to: Zhang Fengchen <fczhang0606@gmail.com>.

Preprint