Value-based Reinforcement Learning

Some Discussions

Kan Ren

Apex Data and Knowledge Management Lab
Shanghai Jiao Tong University

Aug. 3 2017

Outline

1. Revision of Value-based RL
    - Sarsa & Q-learning
    - Q-learning
2. Issues in Q-learning
    - Overestimation
    - Double Q-learning
    - Averaged Q-learning

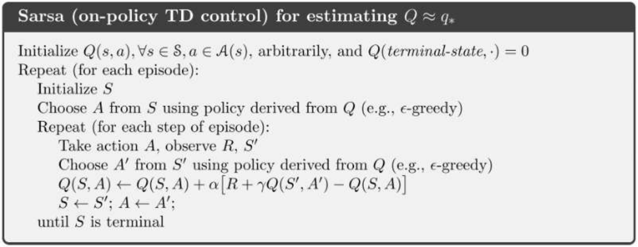
Outline

1. Revision of Value-based RL
   - Sarsa & Q-learning
   - Q-learning

2. Issues in Q-learning
   - Overestimation
   - Double Q-learning
   - Averaged Q-learning

Sarsa

Algorithm



Sarsa (on-policy TD control) for estimating $Q \approx q_*$

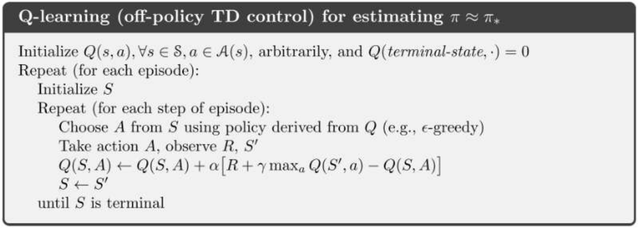
```
Initialize $Q(s,a), \forall s \in \mathcal{S}, a \in \mathcal{A}(s)$, arbitrarily, and $Q(\textit{terminal-state}, \cdot) = 0$
Repeat (for each episode):
    Initialize $S$
    Choose $A$ from $S$ using policy derived from $Q$ (e.g., $\epsilon$-greedy)
    Repeat (for each step of episode):
        Take action $A$, observe $R$, $S'$
        Choose $A'$ from $S'$ using policy derived from $Q$ (e.g., $\epsilon$-greedy)
        $Q(S,A) \leftarrow Q(S,A) + \alpha\big[R + \gamma Q(S',A') - Q(S,A)\big]$
        $S \leftarrow S'$; $A \leftarrow A'$;
    until $S$ is terminal
```

Outline

1. Revision of Value-based RL
   - Sarsa & Q-learning
   - Q-learning
2. Issues in Q-learning
   - Overestimation
   - Double Q-learning
   - Averaged Q-learning

Q-Learning

Algorithm



Q-learning (off-policy TD control) for estimating $\pi \approx \pi_*$

```
Initialize $Q(s, a), \forall s \in \mathcal{S}, a \in \mathcal{A}(s)$, arbitrarily, and $Q(\textit{terminal-state}, \cdot) = 0$
Repeat (for each episode):
    Initialize $S$
    Repeat (for each step of episode):
        Choose $A$ from $S$ using policy derived from $Q$ (e.g., $\epsilon$-greedy)
        Take action $A$, observe $R$, $S'$
        $Q(S, A) \leftarrow Q(S, A) + \alpha\big[R + \gamma \max_a Q(S', a) - Q(S, A)\big]$
        $S \leftarrow S'$
    until $S$ is terminal
```

Difference

Update Rule

- Sarsa

Outline

1. Revision of Value-based RL
   - Sarsa & Q-learning
   - Q-learning
2. Issues in Q-learning
   - Overestimation
   - Double Q-learning
   - Averaged Q-learning

Outline

1. Revision of Value-based RL
   - Sarsa & Q-learning
   - Q-learning
2. Issues in Q-learning
   - Overestimation
   - Double Q-learning
   - Averaged Q-learning

Outline

1. Revision of Value-based RL
   - Sarsa & Q-learning
   - Q-learning
2. Issues in Q-learning
   - Overestimation
   - Double Q-learning
   - Averaged Q-learning