

시즌 RL - Lecture 07

노트북: 모두를 위한 머신러닝
만든 날짜: 2019-01-21 오후 2:13
작성자: rr
태그: #모두를 위한, .Lecture

수정한 날짜: 2019-01-21 오후 3:41

Lecture 07

= correlations between samples
샘플들 사이의 연관성

적은 양의 샘플들로 학습하면 제대로 학습되지 않음

= non-stationary targets
target이 움직임

$$\min_{\theta} \sum_{t=0}^T [\hat{Q}(s_t, a_t | \theta) - (r_t + \gamma \max_{a'} \hat{Q}(s_{t+1}, a' | \theta))]^2$$

(y-hat) Q pred target

같은 네트워크를 사용하기 때문에, 네트워크를 업데이트 시키면 target도 바뀌게 됨

= DQN's Solution

- go deep: 깊게 만듦
- experience replay

- correlations between samples 문제 해결

- action을 취하여 받아온 상태를 학습 시키지 말고, 받아오는 상태를 버퍼에 저장을 함
일정한 시간이 지나면 버퍼에서 랜덤하게 몇 개를 가져와서 학습을 시킴
가까이 있는 게 아니라 랜덤하게 된 점으로 학습함

- separate target network, copy network

- non-stationary targets 문제 해결

- 새로운 네트워크를 만듦

$$\min_{\theta} \sum_{t=0}^T [\hat{Q}(s_t, a_t | \theta) - (r_t + \gamma \max_{a'} \hat{Q}(s_{t+1}, a' | \bar{\theta}))]^2$$

θ 만 업데이트하고 어느 정도 시간이 지나고 θ -bar를 업데이트