

Prioritized Experience Replay

Breaking the temporal correlation

Experience Replay

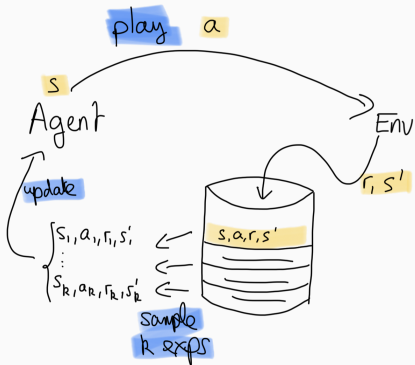
Experience Replay

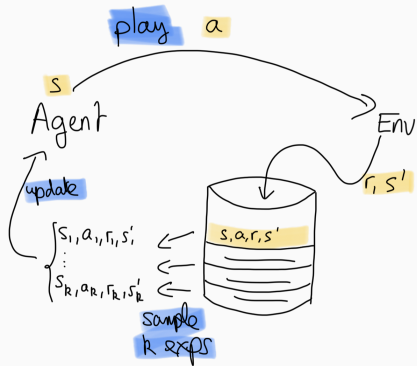
Running new samples each time is:

- waste of data
- biased towards recent experiences and the current policy

Running new samples each time is:

- waste of data
- biased towards recent experiences and the current policy
- **Idea:** a **replay memory (buffer)** stores the N last recent experiences (s, a, r, s') . To update the Q-function, sample a minibatch from the buffer and update the Q-function using the sampled experience.





Observation.

- it breaks the temporal correlation between samples
- it removes the strategy bias

Playing Atari with Deep Reinforcement Learning (2013)

Let's look at the paper.

Prioritized Experience Replay

Not all experiences are equally important!

Prioritized Experience Replay

Not all experiences are equally important!

- **Idea:** a **replay memory (buffer)** stores the N last recent experiences (s, a, r, s') together with a bias B for how surprising the experience was.

$$B = |r + \gamma \max_{a'} Q(s', a') - Q(s, a)|$$

Prioritized Experience Replay, Schaul et al., 2015

Let's look at the paper.

Algorithm - Experience Replay

- Sample a minibatch from the buffer with probability

$$\frac{B^\alpha}{\sum_{i=1}^N B_i^\alpha}$$

and update the Q-function using the sampled experience.

- The hyperparameter α controls the amount of prioritization.
- $\alpha = 0$ corresponds to uniform sampling.
- α high corresponds to high prioritization.

Recap

- **Experience Replay** is a technique that stores a buffer of past experiences and samples from it to update the Q-function. It breaks the temporal correlation between samples.
- **Prioritized Experience Replay** is a technique that stores a buffer of past experiences and samples from it with a bias for how surprising the experience was.

Question:

What if the state space is too large?

- Backgammon: 10^{20} states
- Go: 10^{170} states
- Robotics: continuous state space, real-life