A Tighter Convergence Proof of Reverse Experience Replay

Nan Jiang, ***Jinzhao Li***, Yexiang Xue

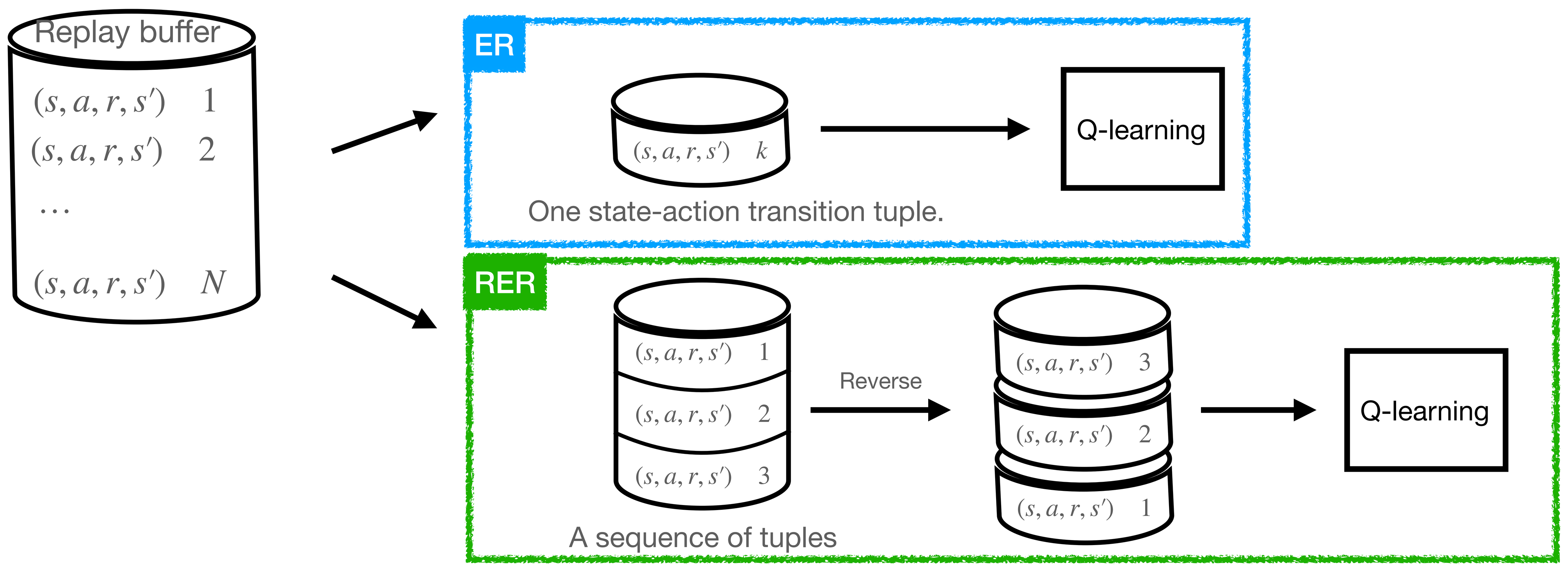
Department of Computer Science, Purdue University

Introduction

Experience Replay (ER) An agent stores past experiences and randomly samples (replays) transitions during the Q-learning process. Many variants, like Prioritized Experience Replay, and Hindsight Experience Replay

Reverse Experience Replay (RER) Inspired by sequential replay occurs in the rat hippocampus [1] — a region of the brain crucial for memory formation.

Reverse Experience Replay - based Q-learning Samples consecutive sequences of transitions (of length L) from the replay buffer. Q-learning updates are performed in the reverse order of the sampled sequences.



Background

Linear MDP Assumption: (1) Reward function: can be written as the inner product of the parameter $w \in \mathbb{R}^d$ and the feature function

$$\phi(s,a) : \mathcal{S} \times \mathcal{A} \to \mathbb{R}^d.$$

(2)Transition probability: proportional to its corresponding feature

$$P(\cdot \mid s,a) \propto \phi(s,a).$$

(3) The Q function is computed as:

$$Q(s,a;w) = \langle w, \phi(s,a) \rangle$$

the error of Q function to the error of learned parameter w by Linear MDP:

$$\varepsilon(s,a) = \hat{Q}(s,a) - Q^*(s,a) \Leftrightarrow \hat{w} - w^*$$

The error breaks into two parts (Lemma 3):

$$\hat{w} - w^* = \underbrace{\Gamma_L \left(w_1 - w^*\right)}_{\text{Bias term}} + \underbrace{\eta \sum_{l=1}^{L} \varepsilon_l \Gamma_{l-1} \phi_l}_{\text{variance term}} .$$

where $s_1 \xrightarrow{a_1,r_1} s_2 \xrightarrow{a_2,r_2} s_3 \to \ldots \to s_L$ is the sampled sequence, and we denote:

$$\Gamma_L = \left(\mathbf{I} - \eta\phi_1\phi_1^\top\right)\left(\mathbf{I} - \eta\phi_2\phi_2^\top\right)\ldots\left(\mathbf{I} - \eta\phi_L\phi_L^\top\right), \text{ with each } \phi_L = \phi(s_L, a_L)$$

The bias term can reduce to zero if $\mathbb{E}_{(s,a)\sim\mu}\left[\Gamma_L^\top \Gamma_L\right]$ is bounded (Lemma C.2), i.e.,

Need to prove: $\mathbb{E}_{(s,a)\sim\mu}\left[\Gamma_L^\top \Gamma_L\right] \preceq$ A proper bound

1. Foster, David J., and Matthew A. Wilson. "Reverse replay of behavioural sequences in hippocampal place cells during the awake state." Nature (2006)

Our Contribution

Previous Result: The expansion is:

$$\mathbb{E}_{(s,a)\sim\mu}\left[\Gamma_L^\top \Gamma_L\right] = \mathbb{E}_{(s,a)\sim\mu}\left[\left(\mathbf{I} - \eta\phi_L\phi_L^\top\right)\cdots\left(\mathbf{I} - \eta\phi_1\phi_1^\top\right)\left(\mathbf{I} - \eta\phi_1\phi_1^\top\right)\cdots\left(\mathbf{I} - \eta\phi_L\phi_L^\top\right)\right]$$

$$= \mathbf{I} - 2\eta\mathbb{E}_{(s,a)\sim\mu}\left[\sum_{l=1}^{L}\phi_l\phi_l^\top\right] + \boxed{\mathbb{E}_{(s,a)\sim\mu}\left[\sum_{k=2}^{2L}(-\eta)^k \sum_{l_1,\ldots,l_k}\phi_{l_1}\phi_{l_1}^\top\ldots\phi_{l_k}\phi_{l_k}^\top\right]} .$$

Requirement: $\eta L < \cancel{1/3}$ $\quad 0 < \eta \leq 1$ $\quad \preceq \eta \cancel{\sum_{l=1}^{L}\mathbb{E}_{(s,a)\sim\mu}\left[\phi_l\phi_l^\top\right]}$ A tighter bound

Our Result: We show it can be upper bound with a weaker assumption using the proposed combinatorial counting. The upper bound becomes:

$$\mathbb{E}_{(s,a)\sim\mu}\left[\Gamma_L^\top \Gamma_L\right] \preceq \left(1 - \frac{\eta(4-2L)L + L - (1-\eta)^{L-1}L - \eta^2 L}{\kappa}\right)\mathbf{I},$$

Our Idea: counting the big summation

To tackle: $\mathbb{E}_{(s,a)\sim\mu}\left[\sum_{k=2}^{2L}(-\eta)^k \sum_{l_1,\ldots,l_k}\phi_{l_1}\phi_{l_1}^\top\ldots\phi_{l_k}\phi_{l_k}^\top\right]$ 1

- Lemma 1: for non-zero vector $\mathbf{x}$:

$$|\mathbf{x}^\top\phi_{l_1}\phi_{l_1}^\top\ldots\phi_{l_k}\phi_{l_k}^\top\mathbf{x}| \leq \frac{1}{2}\mathbf{x}^\top\left(\phi_{l_1}\phi_{l_1}^\top + \phi_{l_k}\phi_{l_k}^\top\right)\mathbf{x}$$

  - It is a relaxation: only depend on l_1 and l_k.
- The summation containing a combinatorial number of elements becomes:

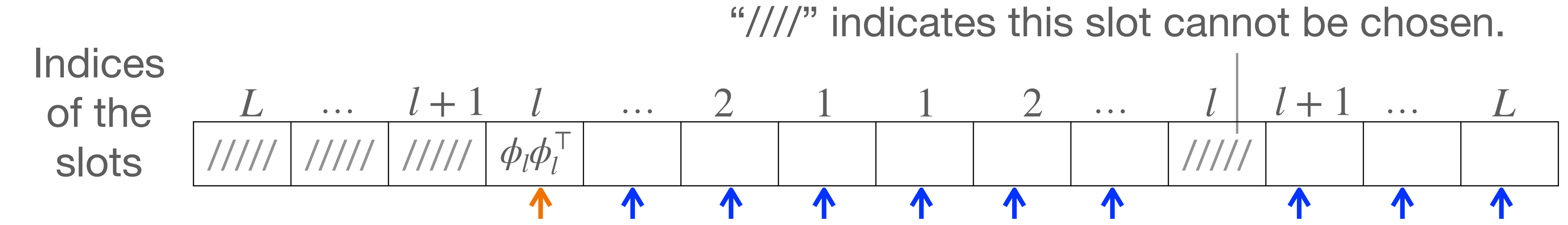
$$\sum_{l_1,\ldots,l_k}\phi_{l_1}\phi_{l_1}^\top\ldots\phi_{l_k}\phi_{l_k}^\top \preceq \sum_{(l_1,l_k)}\frac{1}{2}\left(\phi_{l_1}\phi_{l_1}^\top + \phi_{l_k}\phi_{l_k}^\top\right) = \sum_{l=1}^{L}\boxed{C_l}\cdot\phi_l\phi_l^\top$$

Count the number of combinations $\phi_{l_1}\phi_{l_1}^\top\ldots\phi_{l_k}\phi_{l_k}^\top$ that start/end with ϕ_l

Counting case 1

Count how many cases of picking valid l_1 and l_k at each possible position in the consecutive sequence of state-action-reward tuples.

"////" indicates this slot cannot be chosen.



fix $l_1 = l$ $\quad l_2,\ldots,l_k$ can be placed to $L + l - 2$ many positions.

In this case 1, the count is: $\sum_{l=1}^{L}\binom{L+l-2}{k-1}\phi_l\phi_l^\top$. Rest cases are shown in paper.

This research was funded by NSF grant CCF-1918327 and DOE – Fusion Energy Science grant: DE-SC0024583

Contacts: jiang631, li4255, yexiang@purdue.edu