ICLR 2024

PRIORITIZED SOFT Q-DECOMPOSITION FOR LEXICOGRAPHIC REINFORCEMENT LEARNING

FINN RIETZ, ERIK SCHAFFERNICHT, STEFAN HEINRICH, JOHANNES A. STORK

MOTIVATION

- Improve RL's sample-efficiency by means of **knowledge transfer**
- We assume a fixed MDP: $\langle \mathcal{S}, \mathcal{A}, \mathbf{r}, p, \gamma \rangle$ with continuous $\mathbf{a} \in \mathbb{R}^d$
  - Multiple "subtask" reward functions $\mathbf{r} = [r_1, \ldots, r_n]$
  - Solve subtasks seperately, transfer knowledge between tasks

- Challenges:
  - Q-Decomposition: $Q^*_\Sigma \neq Q^*_1 + Q^*_2$ (Russel & Zimdars 2003)
  - Reward engineering: Incompatible subtasks?

LEXICOGRAPHIC REINFORCEMENT LEARNING

- Multi-objective RL $\langle \mathcal{S}, \mathcal{A}, \mathbf{r}, p, \gamma \rangle$ with priority ordering over subtasks
  - Notation: $r_1 \succ r_2 \succ \cdots \succ r_n$
  - Priority is set by practitioner
  - Priority resolves subtask conflicts

- Knowledge transfer?

PRIORITIZED SOFT Q-DECOMPOSITION: INTUITION

- Lexicographic RL problems with continuous actions
- Knowledge transfer from higher to lower priority subtasks
- Learn r_1 once
- Transfer π_1^* and Q_1^* to $n-1$ lower-priority subtasks
- Lower-priority subtasks are constrained to keeping higher-priority subtask near-optimal

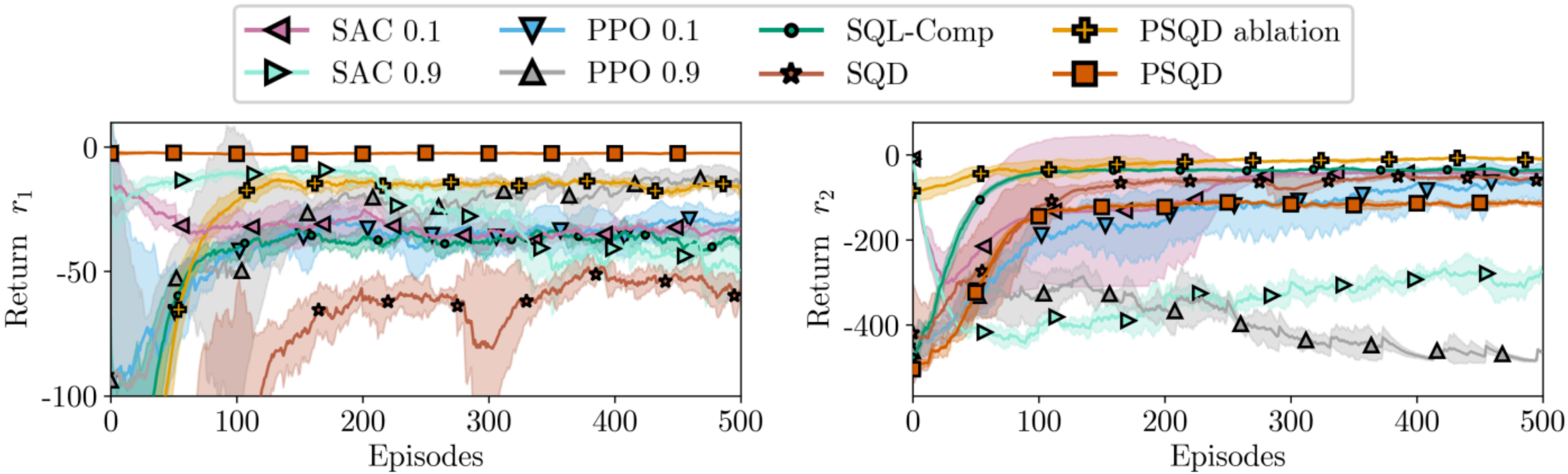
PRIORITIZED SOFT Q-DECOMPOSITION: TECHNICAL

- Assume access to already-learned Q_1^* for r_1
- Define performance threshold ε_1 for Q_1^*
  - $\max_{\pi_2} J(\pi_2) = \mathbb{E}_{(\tau \sim \pi_2)}\left[\sum_{t=0}^{\infty} \gamma^t r_2(\mathbf{s}_t, \mathbf{a}_t)\right]$
  - subject to: $\underbrace{\max_{\mathbf{a}' \in \mathcal{A}} Q_1^*(\mathbf{s}_t, \mathbf{a}')}_{\text{best Q-val in } \mathbf{s}} - Q_1^*(\mathbf{s}_t, \pi_2(\mathbf{s}_t)) \leq \varepsilon$
  - $\max_{\mathbf{a}' \in \mathcal{A}} Q_i^*(\mathbf{s}_t, \mathbf{a}') - Q_i^*(\mathbf{s}_t, \pi_n(\mathbf{s}_t)) \leq \varepsilon_i, \ \forall i \in \{1, \ldots, n-1\}$

- Knowledge transfer via constraints on already-learned Q-functions

RESULTS

- High-priority subtask r_1: Obstacle avoidance
- Low-priority subtask r_2: Goal navigation



THANK YOU

Prioritized Soft Q-Decomposition for Lexicographic Reinforcement Learning

Finn Rietz, Erik Schaffernicht, Stefan Heinrich, Johannes A. Stork

https://github.com/frietz58/psqd/