FedHQL: Federated Heterogeneous Q-Learning

The exploration-exploitation dilemma in the multi-agent setting

Flint Xiaofeng Fan[1,2], Yining Ma[1], Zhongxiang Dai[1], Cheston Tan[2], Bryan Kian Hsiang Low[1]

[1]National University of Singapore, [2]Agency for Science, Technology and Research

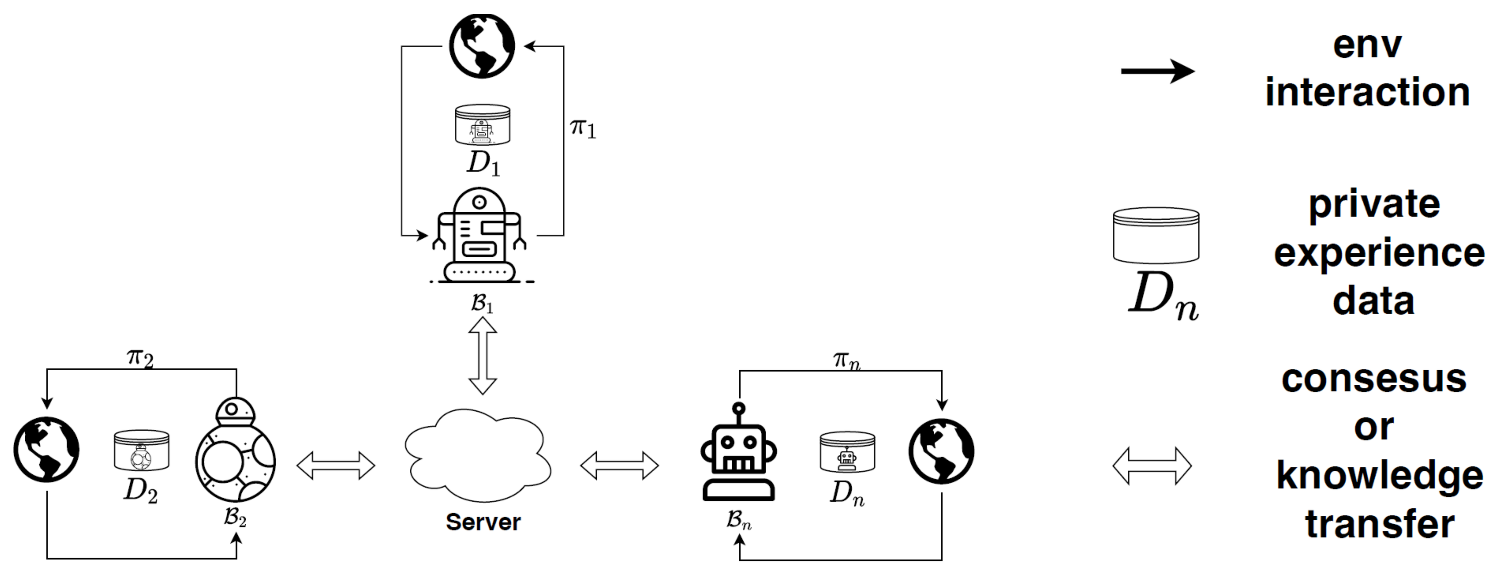
Motivation

Federated Reinforcement Learning (FedRL)

Aim. Improve the sample efficiency of RL agents through federated learning.

Challenge. Practical agents exhibit disagreement regarding their choices of policy parameters, training configurations, and exploration strategies.

Objective. Can **heterogeneous** agents learn collectively?



The Multi-agent Exploration Problem

Intra-agent. Balance between exploring the new knowledge and exploiting the current knowledge of an agent.

Inter-agent. Make decisions that are deemed promising by all agents or explore decisions for which the agents have inconsistent estimations.

Federated Upper Confidence Bound (FedUCB) Algorithm

We propose an UCB-like algorithm to upper-bound the optimal Q^* by Q^{UCB} for any (s,a) , as defined below:

$$\bar{Q}(s,a) = \frac{1}{N}\sum_{n=1}^{N} Q_n(s,a),$$

$$Q^{\text{std}}(s,a) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}[\bar{Q}(s,a) - Q_n(s,a)]^2},$$

$$Q^{\text{UCB}}(s,a) \simeq \underbrace{\bar{Q}(s,a)}_{\text{exploitation}} + \lambda \underbrace{Q^{\text{std}}(s,a)}_{\text{exploration}},$$

Then the group decision can be made by:

$$\bar{a}_t \leftarrow \arg\max_a Q^{\text{UCB}}(s_t, a).$$

FedUCB controls the degree of exploration using the inter-agent exploration coefficient λ:

- larger λ encourages more exploratory behavior

- smaller λ exploits more the current knowledge of the group

FedHQL Algorithm

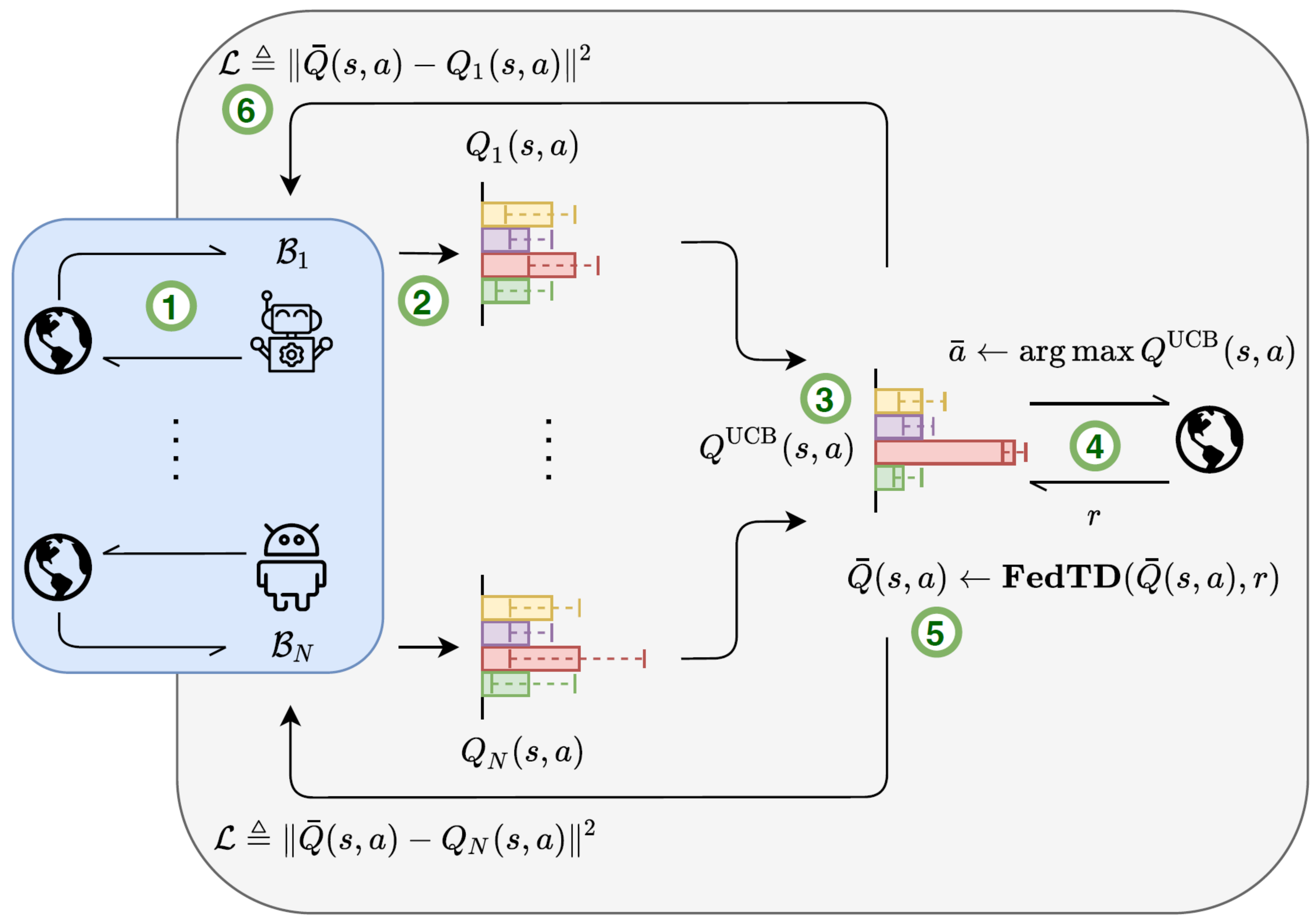
Federated Temporal Difference (FedTD)

We propose to regularize the group decision by FedTD:

$$\bar{Q}(s_t, \bar{a}_t) \leftarrow \bar{Q}(s_t, \bar{a}_t) + \alpha_s\Big(r_t + \gamma \max_b \bar{Q}(s_{t+1}, b) - \bar{Q}(s_t, \bar{a}_t)\Big)$$

Federated Heterogeneous Q-Learning (FedHQL)

We propose FedHQL algorithm to enable collective intelligence in decision-making among a group of heterogeneous agents.

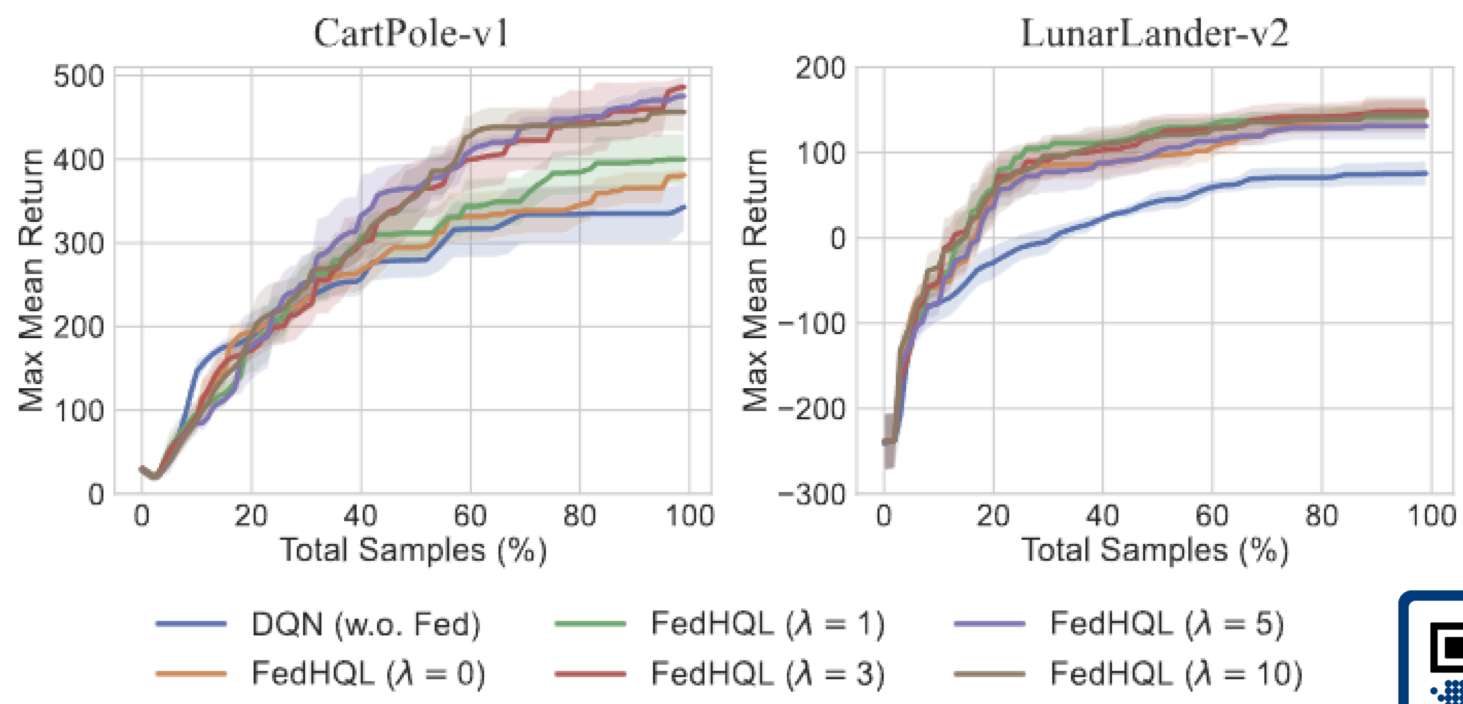


Empirical Evaluation

We conduct experiments with heterogeneous agents with different configurations:

Agent	*Network*	*Learning rates*	*Intra-exploration coefficient*
1	64x64 (Tanh)	0.005	0.01
2	128x128 (ReLU)	0.01	0.1
3	32x32 (Tanh)	0.01	0.05
4	16x16 (ReLU)	0.02	0.01
5	8x8x8 (ReLU)	0.001	0.01

We test the efficacy of FedHQL in boosting the sample efficiency of agents using OpenAI gym environments:



SCAN ME

AAMAS 2023 LONDON