

# FedHPD: Heterogeneous Federated Reinforcement Learning via Policy Distillation --Introduce Knowledge Distillation to Black-Box Federated Reinforcement Learning

Wenzheng Jiang<sup>1</sup>, Ji Wang<sup>1</sup>, Xiongtao Zhang<sup>1</sup>, Weidong Bao<sup>1</sup>, Cheston Tan<sup>2</sup>, Flint Xiaofeng Fan<sup>3</sup>



1 Laboratory for Big Data and Decision, National University of Defense Technology  
2 Institute of High Performance Computing, A\*STAR  
3 Department of Computer Science, National University of Singapore

## Background

### Federated Reinforcement Learning

- Improves sample efficiency
- Preserving data privacy

### Agent Homogeneity

- Common assumption of existing FedRL work
- Identical policy networks/training configurations
- Limiting real-world applicability

### Heterogeneous agents in black box setting

- Divergent architectures/configurations
- Traditional aggregation methods ineffective

## Theoretical Analysis

### Convergence

**THEOREM 1 (CONVERGENCE OF REINFORCE WITH KNOWLEDGE DISTILLATION).** Under Assumptions 1, 2, 3, if we choose  $L = L_J - \lambda L_{KL} > 0$  and learning rates  $\alpha^i$  satisfy Robbins-Monro conditions, the objective function  $J'(\theta_k)$  can converge to a stationary point as the number of iterations increases.

### Fast Convergence

**COROLLARY 1 (FAST CONVERGENCE OF FEDHDP).** In the setting of Theorem 1, if we choose  $\lambda = 1$ , FedHDP can achieve fast convergence by reducing variance, leading to:

$$N_{\text{REINFORCE with KD}} \leq N_{\text{FedHDP}} \leq N_{\text{REINFORCE}}, \quad (10)$$

where  $N$  represents the sample size, and REINFORCE with KD refers to FedHDP without distillation intervals.

## Discussion

### Diversity of Public State Set

### Knowledge Aggregation Method

### Local RL Algorithms .....

(see our full paper on arXiv)

## If you are interested in

1. Federated Reinforcement Learning
2. RL in Large Language Models
3. Knowledge Distillation

Feel free to contact with me!

Email: jiangwenzheng@nudt.edu.cn

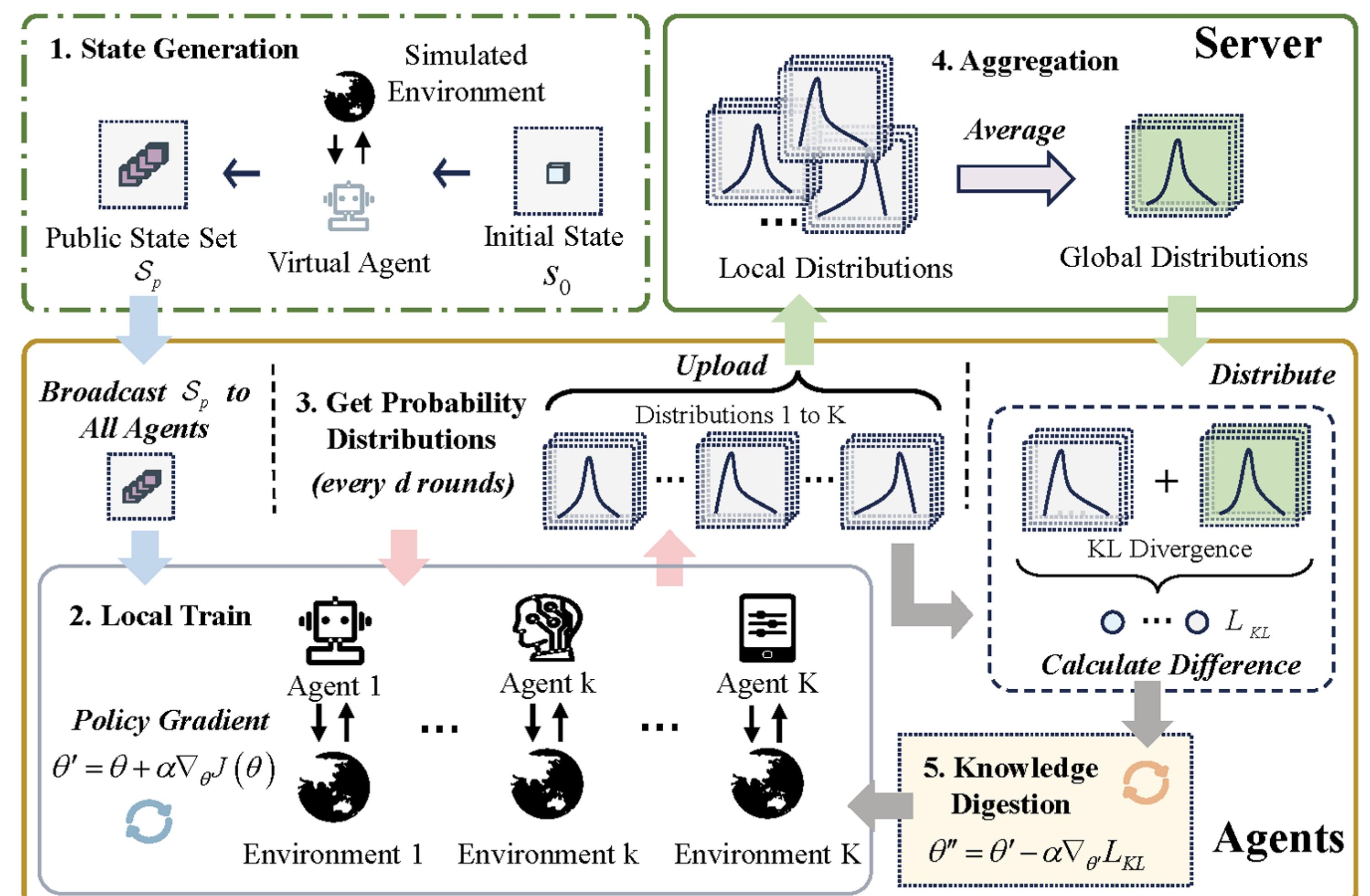


Thanks for the Era of Experience!!!

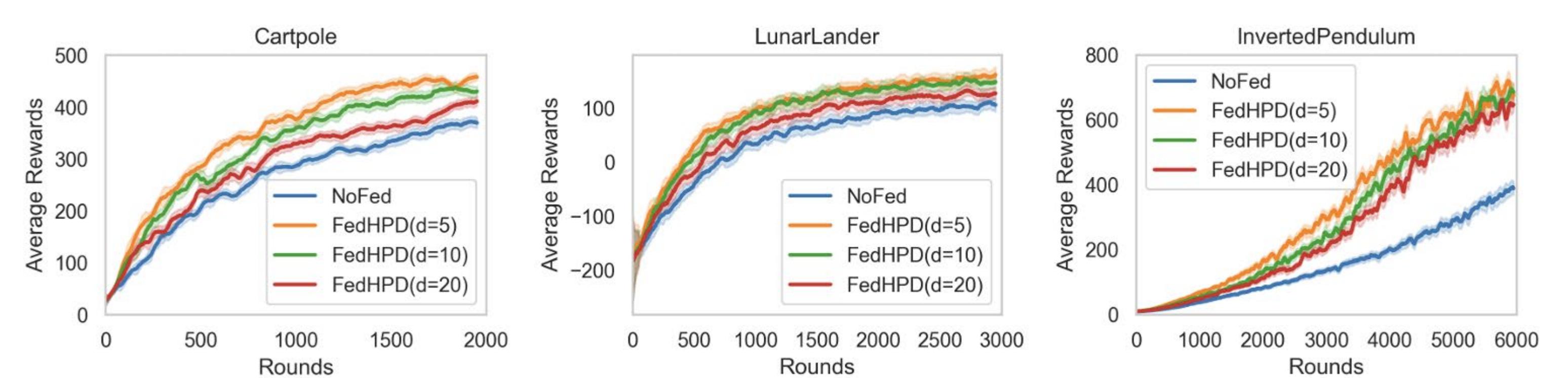
## Method: FedHPD

### Knowledge Distillation

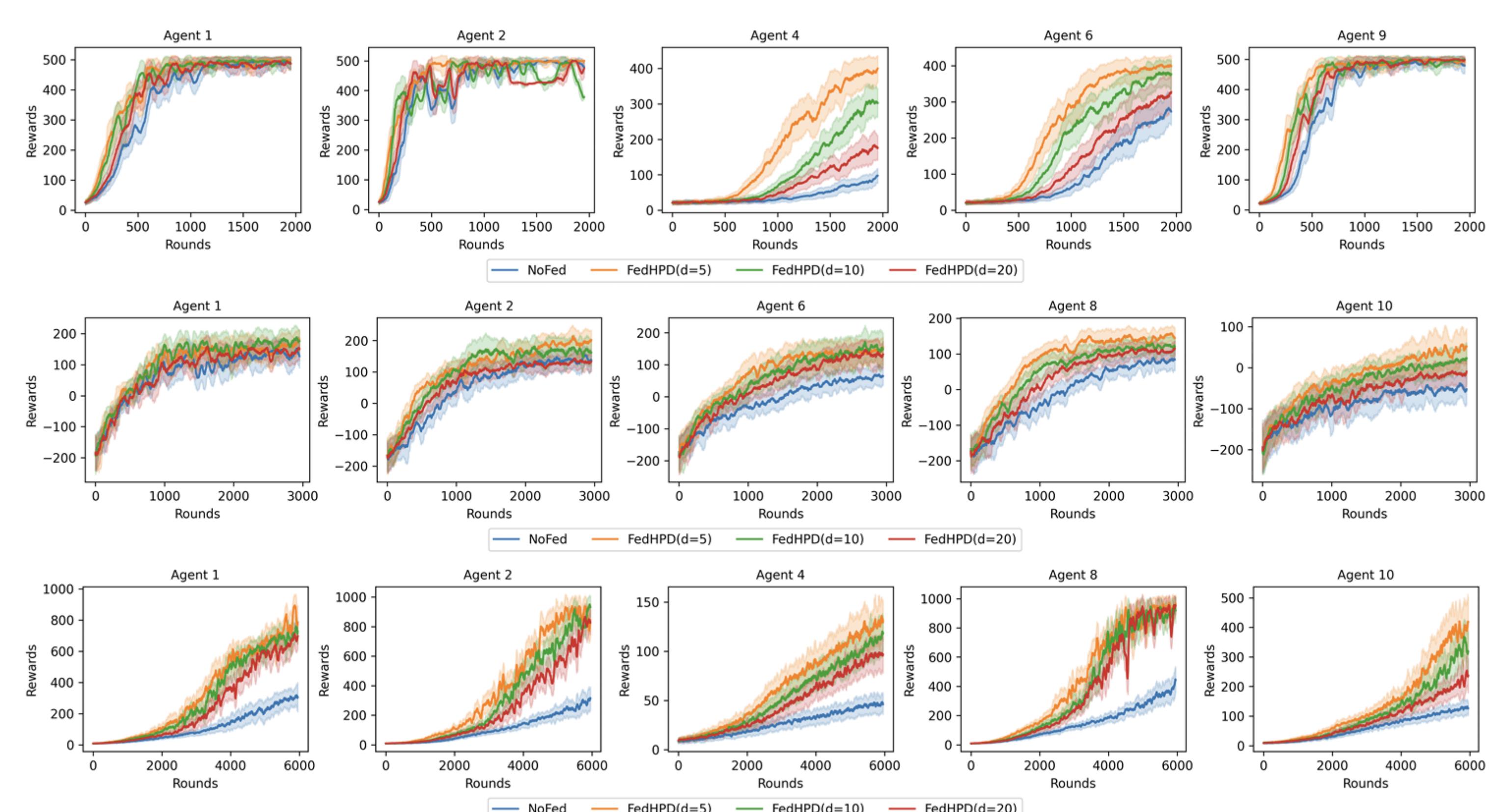
1. **Local Training:** Agents independently update policies using REINFORCE.
2. **Collaborative Training (every d rounds):**
  - **Knowledge Extraction:** Upload action distributions on a shared public state set.
  - **Knowledge Aggregation:** Server computes global consensus (average distribution).
  - **Knowledge Digestion:** Agents minimize KL divergence between local and global distributions.



## Experimental Results



### System Welfare



### Individual Improvement