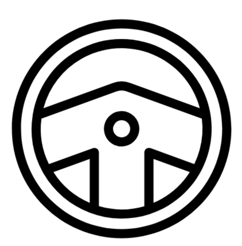
Learning to Cooperate Among Heterogeneous Agents via Intrinsic Rewards

Deeparghya Dutta Barua, Jahir Sadik Monon, Md. Mosaddek Khan
Department of Computer Science and Engineering, University of Dhaka

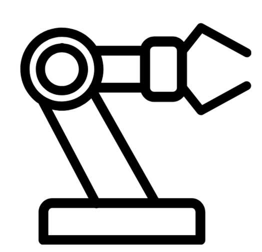
Research Summary

- **CoHet** Algorithm, designed to tackle **reward sparsity** and agent **heterogeneity** in Multi-agent Reinforcement Learning
- A novel **decentralized training** algorithm capable of training under **partial observability** that considers both challenges
- Empirical evaluation demonstrating **performance beyond the state-of-the-art** in several cooperative tasks
- Analysis of dense **intrinsic reward** calculation module and how it helps in dealing with reward sparsity

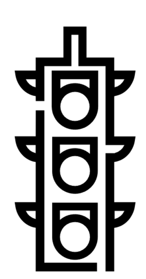
Applications



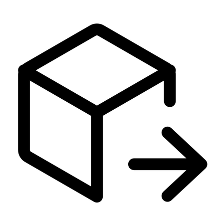
Autonomous Driving
Tesla, Waymo



Robotics
Multi-Robot System



Traffic Control
CoLight, PressLight



Package Transport
Amazon Warehouse

Problem Formulation: Multi-Agent Reinforcement Learning
Notable Aspects: Reward Sparsity, Agent Heterogeneity
Challenges: Partial Observability, Decentralized Training

Related Works

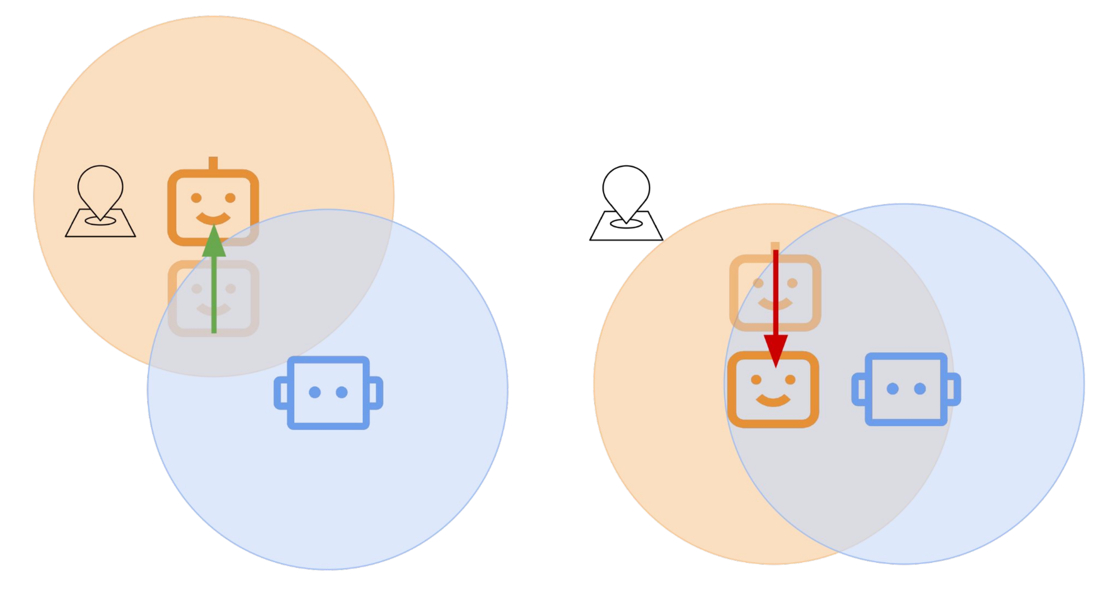
ELIGN *[Ma et al., 2022]*
- Uses Intrinsic Motivation
- Addresses reward sparsity
- Decentralized training
- Task-Agnostic

CHDRL *[Zheng et al., 2020]*
- Addresses heterogeneity
- Addresses reward sparsity
- Different definition of heterogeneity

HetGPPO *[Bettini et al., 2023]*
- Classifies heterogeneous systems
- Heterogeneous policy learning using GNN
- Decentralized

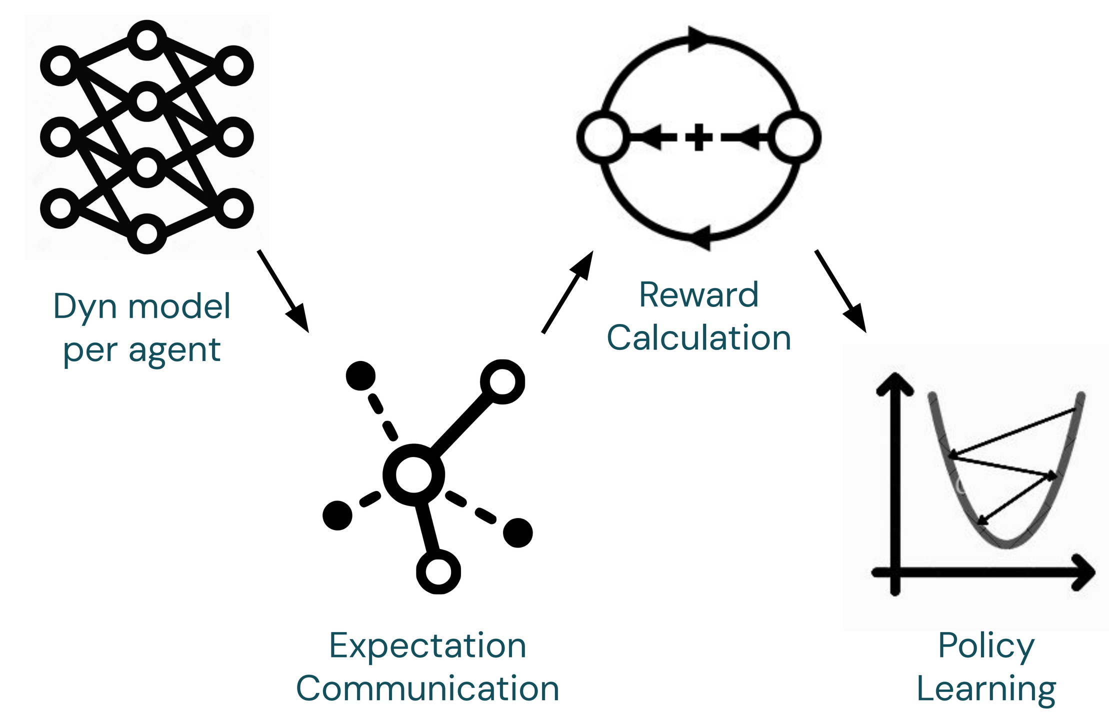
CC-based *[Andres et al., 2022]*
- Addresses heterogeneity
- Addresses reward sparsity
- Centralized training
- Parameter sharing among agents

Expectation Alignment



(a) Aligned → high r_{in} (b) Misaligned → low r_{in}

Algorithm Steps



CoHet Architecture

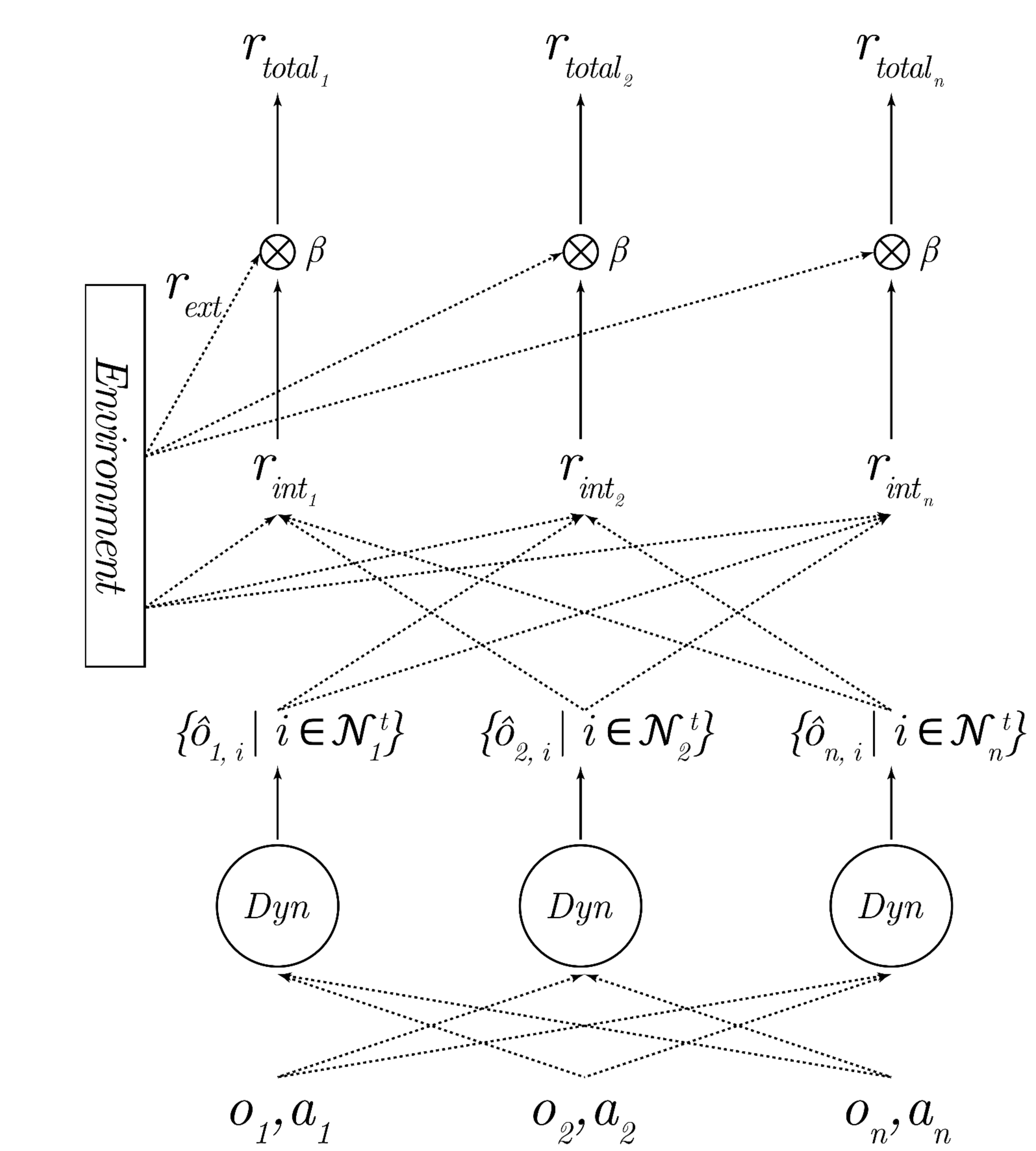


Figure 1: Reward Calculation Module

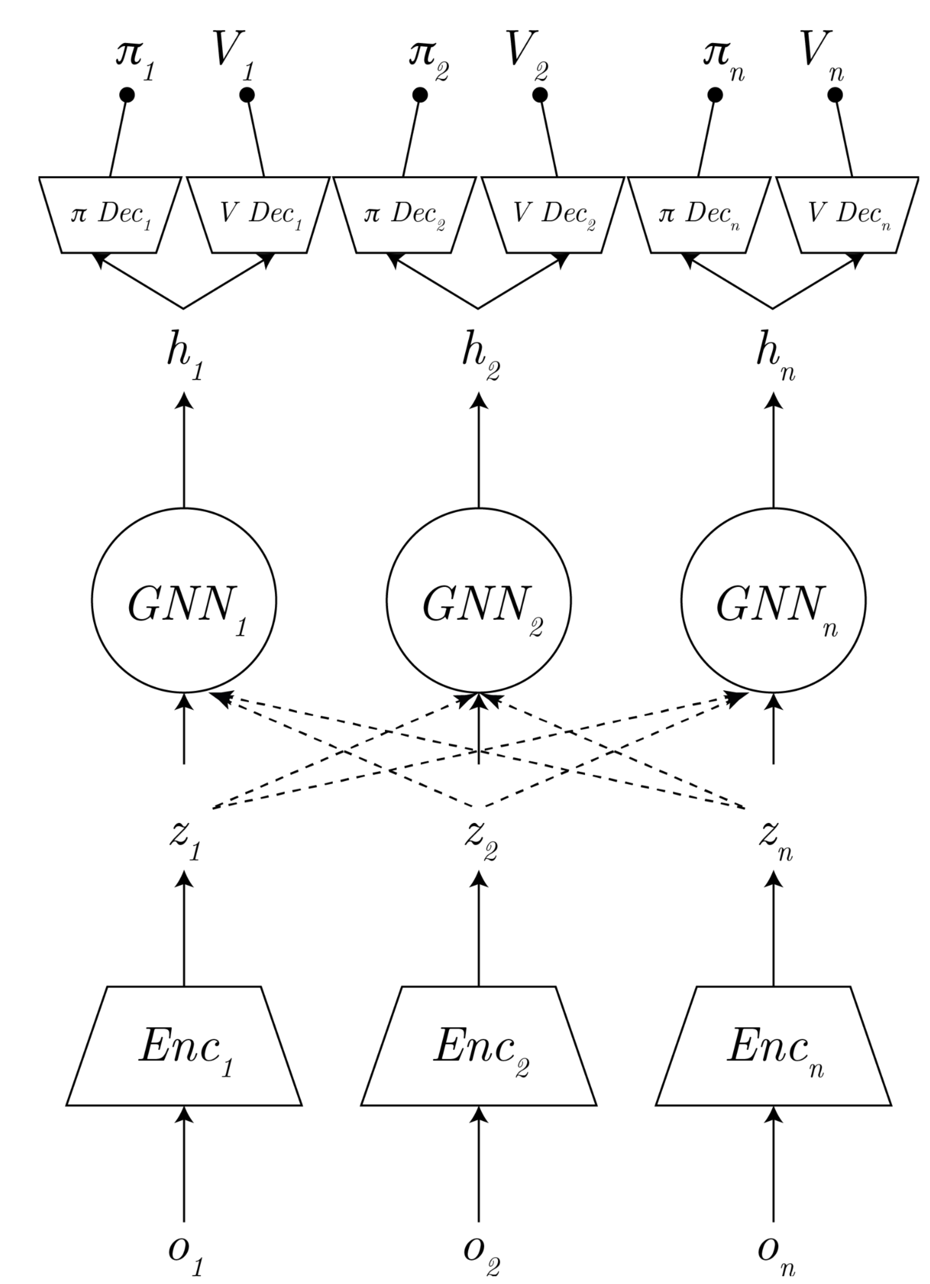


Figure 2: Policy Learning Module

Algorithm Description

1. **Random initialization of** per-agent dynamics models, the encoders, GNN multi-layer perceptrons, policy & value decoders
2. **At each time step,** the architecture of CoHet,
   a. **Calculates** dense intrinsic reward signals and **augments** those to the sparse environmental rewards, using reward calculation module in Figure 1
   b. **Optimize policies** for using policy learning module in Figure 2
   c. Train the dynamics model simultaneously

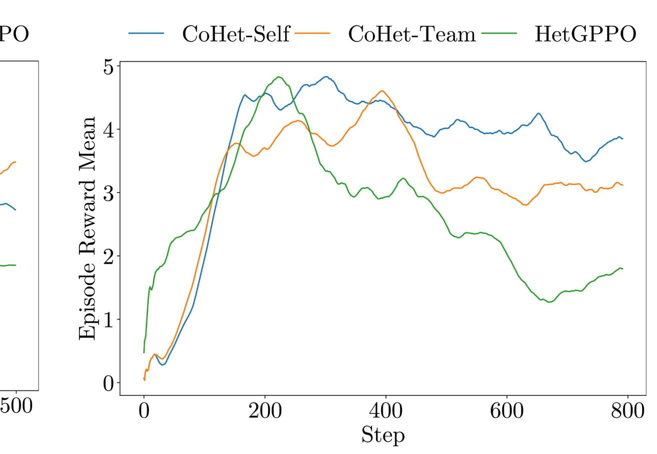
$$r^t_{int_i}(o^t_i, a^t_i) = -\sum_{j \in \mathcal{N}^t_i \cap \mathcal{N}^{t+1}_i} w_j \times \|o^{t+1}_i - \hat{o}^t_{j,i}\|$$

$$w_j = \frac{d(i,j)}{\sum_{k \in \mathcal{N}^t_i \cap \mathcal{N}^{t+1}_i} d(i,k)}$$

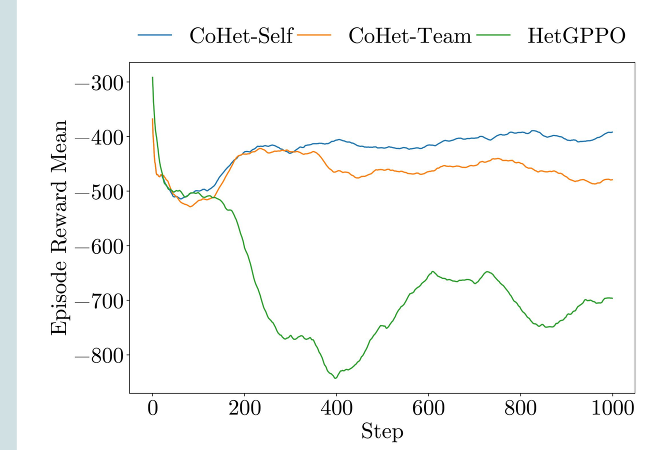
Empirical Evaluation

Performance compared to SOTA in MPE and VMAS tasks

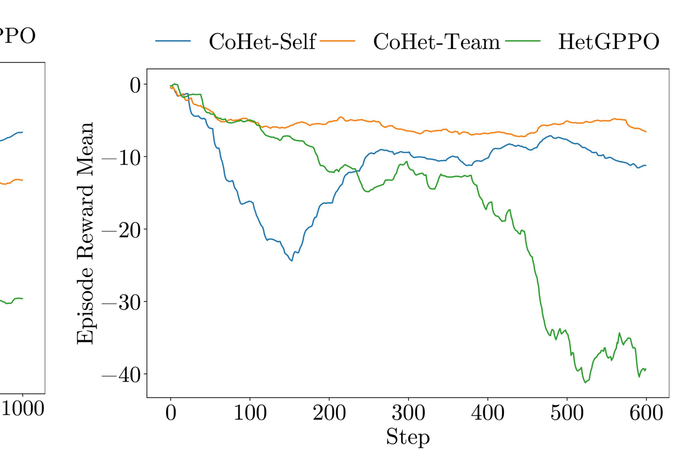
VMAS Sampling



VMAS Reverse Transport



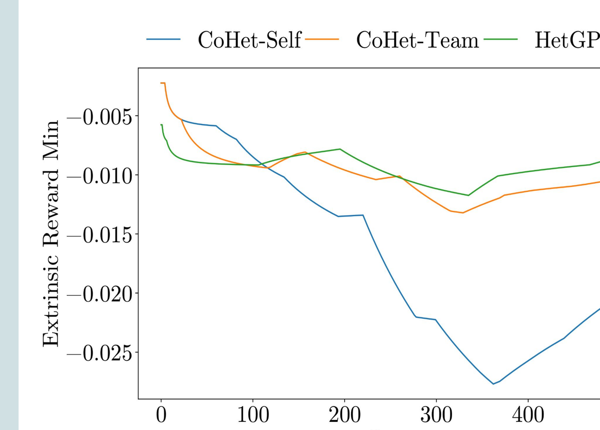
MPE Simple spread

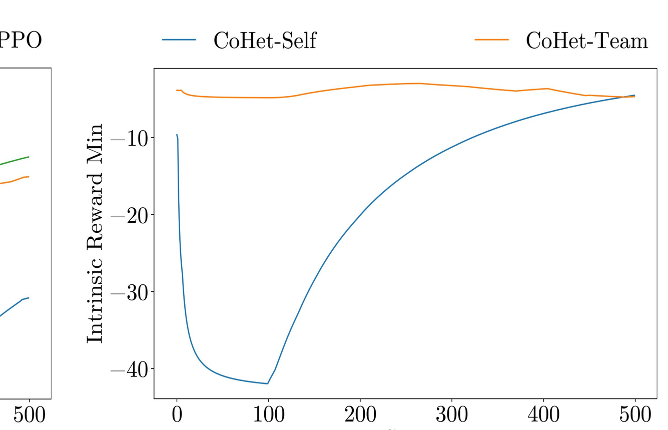


VMAS Joint Passage

Effect of intrinsic reward

Effect of intrinsic reward on environmental feedback





Contribution & Future Work

- An algorithm applicable to **real-world tasks** requiring agent **heterogeneity** where agents have **noisy sensor readings**
- Applicability to problem settings where a **central coordinator** is **infeasible** and **reward signals** are **sparse**
- A promising **direction for future work** includes – exploring the alternative types of intrinsic motivation and finding the correct balance between intrinsic and extrinsic rewards

Contact

jahirsadikmonon@gmail.com
deeparghya.csedu@gmail.com
mosaddek@du.ac.bd