TIM-MARL: Information Sharing for Multi-Agent Reinforcement Learning in Smart Environments

Ayesha Siddiqua[1], Siming Liu[1], Razib Iqbal[1], Fahim Ahmed Irfan[1], Logan Ross[2], and Brian Zweerink[3]
[1]Missouri State University, Springfield, Missouri, USA
[2]Joplin High School, Joplin, Missouri, USA
[3]Marshfield High School, Marshfield, Missouri, USA

2024 IEEE 21st Consumer Communications & Networking Conference (CCNC) | 979-8-3503-0457-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/CCNC51664.2024.10454813

***Abstract*—Information sharing among agents to jointly solve problems is challenging for multi-agent reinforcement learning algorithms (MARL) in smart environments. In this paper, we present a novel information sharing approach for MARL, which introduces a Team Information Matrix (TIM) that integrates scenario-independent spatial and environmental information combined with the agent's local observations, augmenting both individual agent's performance and global awareness during the MARL learning. To evaluate this approach, we conducted experiments on three multi-agent scenarios of varying difficulty levels implemented in *Unity ML-Agents Toolkit*. Experimental results show that the agents utilizing our TIM-Shared variation outperformed those using decentralized MARL and achieved comparable performance to agents employing centralized MARL.**

***Index Terms*—Deep reinforcement learning, multi-agent system, hierarchical information sharing, Unity ML-Agent Toolkit**

I. Introduction

Modern smart environments have evolved into information-intensive cyber-physical multi-agent systems (MAS) driven by the convergence of technologies like advanced robotics, Internet of Things (IoT), and artificial intelligence (AI). Also, the integration of machine learning (ML) and reinforcement learning (RL) with autonomous robots can greatly enhance industrial processes in smart environments. Despite the focus on single-agent automation in RL, there exists an opportunity for enhancement within the realm of multi-agent smart environments. Multi-agent reinforcement learning (MARL) extends single-agent RL and provides learning techniques to a group of evolving agents in both cooperative and competitive tasks by maximizing rewards through agents' interaction with the environment and among themselves [1]. Several challenges arise while formulating problems within the domain of MARL, with one of them being effectively enabling internal communication among agents. Training MARL agents solely on the basis of individual agents' limited local observation could hardly yield global optimal solutions in complex smart environments. In order to achieve team objectives, a comprehensive grasp of both the team's dynamics and the surrounding environment around an agent becomes imperative. Many research initiatives, such as [2], have already tackled this issue by leveraging comprehensive information to eliminate the necessity of modeling communication among agents. Several researchers have worked on modeling information sharing and communication among agents in smart environments for enhancing team performance in collaborative tasks, such as [3]. To reduce the dependency on complete information and to enable team information sharing, we present a novel multi-layer Team Information Matrix (TIM) as a shared knowledge base to enhance agents' decision-making when engaging in collaborative tasks in MAS. Moreover, we present a variation of TIM named TIM-Shared that integrates the local observations of individual agents to emulate a global state in real-world scenarios where complete information is not readily accessible. To evaluate the effectiveness of TIM, we modeled several smart environment scenarios in the *Unity ML-Agents Toolkit*, involving interactions among multiple robots with the primary goal of optimizing their team performance and enhancing inter-robot collaboration.

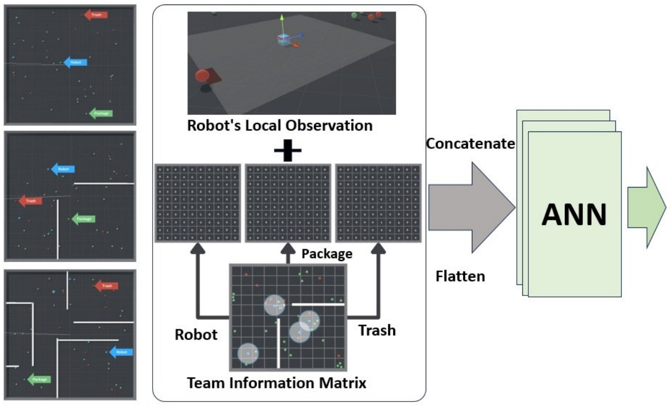


Fig. 1. Architecture of the TIM-MARL Robot Control

II. Methodology

We model the experimental scenarios using Markov games, an expansion of Markov Decision Processes (MDP) to multiple agents, as demonstrated in [4], [5]. We constructed a series of multi-agent scenarios shown in Fig. 1 with varying degrees of complexity for training a group of autonomous robots in pursuit of collaborative learning goals. In our scenarios, we focused on training four autonomous robots to collect packages and clean the trashes cooperatively. The four learning robots are able to select a package or trash and navigate autonomously in a distributed way with the shared objective of collecting as many packages and cleaning trashes as they can in a given amount of time. Each robot has a sensing area of 20×20, and observes packages, trashes, robots, and walls within their sensing limit. To effectively simulate the cleaning task and distinguish it from the package collection task, we equipped each robot with a laser capable of targeting and "cleaning" trashes, which simulate a cleaning action of the task in Unity. We trained our robots using PPO and conducted a comprehensive comparison between our TIM-MARL and approaches utilizing global and local observations. As shown in Fig. 1, a three-dimensional TIM carrying robots, packages, and trash information is used as an input to the MARL model in our TIM-MARL architecture. Based on the availability of global information during execution, we formulated two variations: TIM-Global and TIM-Shared utilizing Equation 1. In TIM-Global, we assume that the system has access to complete information about the environment, such as exact location of every object in the production area captured via cameras. In TIM-Shared, however, the system does not have access to complete environmental information, such as office building scenarios where a centralized camera is hard to deploy. Equation 1 is illustrated below:

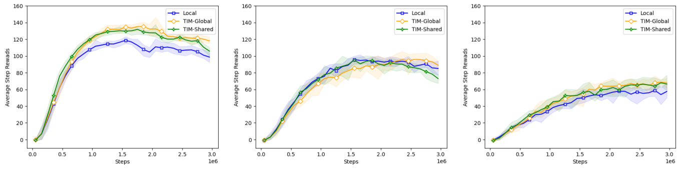
$$Pos\_TIM[i] = \frac{Scene\_Pos\_Obj_k + \frac{1}{2} \times Scene\_Len}{Dimen\_TIM[i]} \tag{1}$$
$$i \in \{0, 1, 2\}, k \in \{robot, package, trash\}$$

where $Scene\_Pos\_Obj_k$ represents the coordinate of an object in a given scenario, $Scene\_Len$ denotes the scene dimension, and $Dimen\_TIM[i]$ represents the target dimension of each matrix layer of TIM.

For scenarios with complete information available, we compute TIM based on the complete global state as referred to *TIM-Global*. Here, an intermediary controller hub receives the real-time state information of each object and updates TIM regularly. Each robot fetches and utilizes TIM-Global with its own local information via TIM-MARL architecture to make optimal actions. Considering that complete global information of a deployment might not be available in real-world scenarios, we designed the *TIM-Shared* variation to approximate the complete information by creating a collective knowledge base for the team. In this variation, robots share and upload their individual observations to a hub controller where TIM-Shared is consistently stored and updated, and subsequently shared with all robots in real-time to aid them in their task assignments and navigation processes. As illustrated in Fig. 1, for TIM-MARL, we employ multiple robots with a shared single neural network under a multi-task learning scheme, hence saving computation at training and inference time as only a single network would need to be evaluated. For each robot, its local observation combined with the TIM constitutes the input space to individual robots for collaborative task selection and navigation in the MARL learning process.

III. Results and Discussion

We first examined the collaborative performance of robots controlled by Local-MARL, TIM-Global MARL, and TIM-Shared MARL in team-based tasks in the warehouse scenario with open space. As shown in Fig. 2a, robots controlled by Local-MARL exhibited continual learning and progressive team performance enhancement throughout the 3 million training steps. Remarkably, the highest average reward achieved by Local-MARL is 123 across 7 separate runs with different random seeds. The results indicate that the robots can achieve local optimal with high performance without information sharing among agents. Fig. 2a also shows that the robots utilizing TIM-Global and TIM-Shared outperformed the Local-MARL by 19% and 11% respectively. The results illustrated in Fig. 2b and Fig. 2c are collected from the simulated warehouse scenario with simple walls and complex office building. We can observe that, although robots successfully learned to work collaboratively with each other, the introduction of walls and obstacles resulted in diminished performance. In the warehouse scenario with simple walls, there was minimal variation between the robot teams utilizing the Local-MARL, TIM-Global, and TIM-Shared approaches. However, the robots with TIM-Shared MARL showed significant improvement in the office building scenario with complex walls by outperforming Local-MARL by 8.5% and achieving a very close performance with the TIM-Global approach.



(a) Avg Reward on Open Warehouse (b) Avg Reward on Simple Walls Warehouse (c) Avg Reward on Complex Office

Fig. 2. Results of MARL Learning across Scenarios

IV. Conclusion

In this paper, we introduced a novel Team Information Matrix (TIM) to enable inter-agent communication and integrated it with local observations to effectively navigate robots for learning collaborative tasks in complex smart environments. Experimental results demonstrated that our robots successfully collaborated in both simple and complex scenarios and adeptly learned to maintain appropriate distances while carrying out their respective tasks, thereby maximizing space coverage and mitigating potential resource competition among robots. The TIM-enhanced MARL outperforms local models and achieves comparable performance similar to global models.

V. Acknowledgement

This material is based upon work supported by the National Science Foundation under Award No. 2302060.

References

[1] L. Buşoniu, R. Babuška, and B. De Schutter, *Multi-agent Reinforcement Learning: An Overview*. Springer Berlin Heidelberg, 2010, pp. 183–221.

[2] L. Panait and S. Luke, "Cooperative multi-agent learning: The state of the art," *Autonomous Agents and Multi-Agent Systems*, vol. 11, no. 3, pp. 387–434, Nov 2005.

[3] S. Sukhbaatar and R. Fergus, "Learning multiagent communication with backpropagation," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[4] A. Harris and S. Liu, "MAIDRL: Semi-centralized MARL using agent influence," *IEEE Conference on Games*, pp. 01–08, 2021.

[5] P. Brackett, S. Liu, and Y. Liu, "Sc-mairl: Semi-centralized MA imitation reinforcement learning," *IEEE Access*, vol. 11, pp. 57 965–57 976, 2023.