A Survey of Multi-Agent Deep Reinforcement Learning with Communication*

JAAMAS Track

Changxi Zhu
Department of Information and Computing Sciences
Utrecht, Utrecht University
c.zhu@uu.nl

Mehdi Dastani
Department of Information and Computing Sciences
Utrecht, Utrecht University
m.m.dastani@uu.nl

Shihan Wang
Department of Information and Computing Sciences
Utrecht, Utrecht University
s.wang2@uu.nl

ABSTRACT

Communication is an effective mechanism for coordinating the behaviors of multiple agents, broadening their views of the environment, and to support their collaborations. In the field of multi-agent deep reinforcement learning (MADRL), agents can improve the overall learning performance and achieve their objectives through communication. Agents can communicate various types of messages, either to all agents or to specific agent groups, or conditioned on specific constraints. With the growing body of research work in MADRL with communication (Comm-MADRL), there is a lack of a systematic and structural approach to distinguish and classify existing Comm-MADRL approaches. In this paper, we survey recent works in the Comm-MADRL field and consider various aspects of communication that can play a role in designing and developing multi-agent reinforcement learning systems. With these aspects in mind, we propose 9 dimensions along which Comm-MADRL approaches can be analyzed, developed, and compared. By projecting existing works into the multi-dimensional space, we discover interesting trends. We also propose some novel directions for designing future Comm-MADRL systems through exploring possible combinations of the dimensions.

KEYWORDS

Multi-Agent Reinforcement Learning; Deep Reinforcement Learning; Communication; Survey

ACM Reference Format:
Changxi Zhu, Mehdi Dastani, and Shihan Wang. 2024. A Survey of Multi-Agent Deep Reinforcement Learning with Communication*: JAAMAS Track. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

Many real-world scenarios, such as autonomous driving [9], sensor networks [11], robotics [3] and game-playing [1, 10], can be modeled as multi-agent systems. Such multi-agent systems can be designed and developed using multi-agent reinforcement learning (MARL) techniques to learn the behavior of individual agents, which can be cooperative, competitive, or a mixture of them. As agents are often distributed in the environment where they only have access to their local observations rather than the complete state of the environment, partial observability becomes an essential assumption in MARL [2, 5, 7]. Moreover, MARL suffers from the non-stationary issue [8], since each agent faces a dynamic environment that can be influenced by the changing and adapting policies of other agents. Communication has been viewed as a vital means to tackle the problems of partial observability and non-stationary in MARL. Agents can communicate individual information, e.g., observations, intentions, experiences, or derived features, to have a broader view of the environment, which in turn allows them to make well-informed decisions [8, 12].

Due to the recent success of deep learning [4] and its application to reinforcement learning [6], multi-agent deep reinforcement learning (MADRL) has witnessed great achievements in recent years, where agents can process high-dimensional data and have generalization ability in large state and action spaces [2, 5]. We notice that a large number of research works focus on *learning tasks with communication*, which aim at learning to solve domain-specific tasks, such as navigation, traffic, and video games, through communicating and information sharing. To the best of our knowledge, there is a lack of survey literature that can cover recent works on learning tasks with communication in multi-agent deep reinforcement learning (Comm-MADRL). Most Comm-MADRL surveys cover only a small number of research works without proposing a fine-grained classification system to compare and analyze them.

In our survey paper, we review the Comm-MADRL literature by focusing on how communication can be utilized to improve the performance of MADRL techniques. Specifically, we identify 9 dimensions that correspond to unique aspects of Comm-MADRL systems and call them: Controlled Goals, Communication Constraints, Communicatee Type, Communication Policy, Communicated Messages, Message Combination, Inner Integration, Learning Methods, and Training Schemes. These dimensions, which form the skeleton of a Comm-MADRL system, can be used to analyze and gain insights into designed Comm-MADRL approaches thoroughly. By mapping recent Comm-MADRL approaches into this multi-dimensional structure, we not only provide insight into the current state of the art in this field but also determine some important directions for designing future Comm-MADRL systems*.

*This paper is an extended abstract of an article published in Autonomous Agents and Multi-Agent System [13].

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

*The relations of dimensions and research directions are discussed in the journal

Table 1: Proposed dimensions and research questions.

Key Components	*Target Questions*	*Dimensions*	*Index*
Problem Settings	What kind of behaviors are desired to emerge with communication?	Controlled Goals	①
	How to fulfill realistic requirements?	Communication Constraints	②
	Which type of agents to communicate with?	Communicatee Type	③
Communication Processes	When and how to build communication links among agents?	Communication Policy	④
	Which piece of information to share?	Communicated Messages	⑤
	How to combine received messages?	Message Combination	⑥
	How to integrate combined messages into learning models?	Inner Integration	⑦
Training Processes	How to train and improve communication?	Learning Methods	⑧
	How to utilize collected experience from agents?	Training Schemes	⑨

2 LEARNING TASKS WITH COMMUNICATION IN MADRL

Learning tasks with communication in multi-agent deep reinforcement learning is a challenging problem. Numerous studies have emerged, developing effective and efficient Comm-MADRL systems, with overlapping characteristics. To better distinguish among these models, we propose classifying them based on several dimensions in Comm-MADRL system design. We start by focusing on three key components of Comm-MADRL systems: problem settings, communication processes, and training processes. Problem settings concern the settings of Comm-MADRL systems developed for learning, encompassing both communication-specific settings (e.g., communication constraints) and non-communication-specified settings (e.g., reward configurations). Communication processes concern the decision as to whether to communicate or not, and what message to communicate. Training processes concern the learning of both agents and communication within MADRL. Based on the three key components, we identify and summarize 9 research questions that commonly arise in Comm-MADRL system design, corresponding to 9 dimensions as detailed in Table 1. We further outline a systematic procedure for providing a guideline to effectively navigate through these dimensions when developing Comm-MADRL systems. The procedure allows us to organize the dimensions, demonstrate their relevance in system design, and guide the creation of customized Comm-MADRL systems in a step-by-step manner.

As outlined in Procedure 1, reinforcement learning agents employ communication throughout their learning and decision-making. Initially, the learning objective for the agents is set, defining rewards that induce cooperative, competitive, or mixed behaviors, as captured by dimension 1. We then consider potential communication-specified settings like limited resources, addressing the need for realistic scenarios as described in dimension 2. Dimension 3 identifies potential communicatees, determining the agents for messages to be received, which varies across domains. At each time step, agents decide when and with whom to communicate, as highlighted in dimension 4. The patterns of communication occurrences are structured like a graph, where links, either undirected or directed, aid information exchange. Subsequently, messages that encapsulate agents' understanding of the environment are generated and shared, relating to dimension 5. Given that agents often receive multiple messages, they must decide on how to combine these messages effectively. This process, crucial for integrating messages into their policies or value functions, is captured in dimensions 6 and 7. In cases of Comm-MADRL studies focusing on emergent language (i.e., learning tasks with emergent language), where messages are modeled as communicative acts emitted alongside domain-level actions, a specific rearrangement of the procedure is required. Here, messages are not observed by other agents until the next time step. Therefore, the processes outlined in dimensions 6 and 7 (lines 8 and 9) are moved to the front of those in dimension 4 (line 6). This rearrangement allows agents to combine and integrate messages from the previous time step before initiating new communication. As a result, agents make decisions and perform actions in the environment based not only on their environmental observations but also on information obtained from other agents (lines 10 and 11). During the training phase, experiences from both environmental interactions and inter-agent communication are utilized to train how agents will behave and communicate, i.e., agents' policies, value functions, and communication processes, as characterized in dimensions 8 and 9 (line 14).

Procedure 1 A guideline of Comm-MADRL systems

```
Require: N reinforcement learning agents
1: Set goals for N reinforcement learning agents                  ▷ Dim. ①
2: Set possible communication constraints                         ▷ Dim. ②
3: Set the type of communicatees                                  ▷ Dim. ③
4: for episode = 1, 2, ... do
5:     for every environment step do
6:         Decide with whom and whether to communicate            ▷ Dim. ④
7:         Decide which piece of information to share             ▷ Dim. ⑤
8:         Combine received information shared from others        ▷ Dim. ⑥
9:         Integrate messages into agents' internal models        ▷ Dim. ⑦
10:        Select actions based on communication
11:        Perform in the environment (and store experiences)
12:    end for
13:    if training is enablled then
14:        Update agents' policies, value function, and communication processes ▷ Dim. ⑧ & ⑨
15:    end if
16: end for
```

Procedure 1: MADRL systems integrate with communication across different dimensions.

3 CONCLUSIONS

Our survey proposes to classify the literature based on 9 dimensions. These dimensions constitute the basis of designing Comm-MADRL systems. We further categorize existing works under each dimension, where readers can easily compare research works from a unique perspective. Based on those dimensions, we also observe findings through the trend of the literature and identify new research directions by filling the gap among recent works. Our survey concludes that while the number of works in Comm-MADRL is notable and represents significant achievements, communication can be more fruitful and versatile to incorporate non-cooperative settings, heterogeneous players, and large-scale multi-agent systems. Agents can communicate information not only from raw image inputs or handcrafted features but also from diverse data sources such as voice and text. Furthermore, we can explore novel metrics to better understand the contribution of communication to the overall learning.

REFERENCES

[1] Noam Brown and Tuomas Sandholm. 2019. Superhuman AI for multiplayer poker. *Science* 365, 6456 (2019), 885–890.

[2] Jakob N. Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. 2018. Counterfactual Multi-Agent Policy Gradients. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, Sheila A. McIlraith and Kilian Q. Weinberger (Eds.). 2974–2982.

[3] Jens Kober, J. Andrew Bagnell, and Jan Peters. 2013. Reinforcement learning in robotics: A survey. *Int. J. Robotics Res.* 32, 11 (2013), 1238–1274.

[4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.

[5] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. 2017. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 6379–6390.

[6] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin A. Riedmiller, Andreas Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (2015), 529–533.

[7] Frans A. Oliehoek and Christopher Amato. 2016. *A Concise Introduction to Decentralized POMDPs.*

[8] Georgios Papoudakis, Filippos Christianos, Arrasy Rahman, and Stefano V. Albrecht. 2019. Dealing with Non-Stationarity in Multi-Agent Deep Reinforcement Learning. *CoRR* abs/1906.04737 (2019). arXiv:1906.04737 http://arxiv.org/abs/1906.04737

[9] Shai Shalev-Shwartz, Shaked Shammah, and Amnon Shashua. 2016. Safe, Multi-Agent, Reinforcement Learning for Autonomous Driving. *CoRR* abs/1610.03295 (2016). arXiv:1610.03295 http://arxiv.org/abs/1610.03295

[10] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy P. Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. 2017. Mastering the game of Go without human knowledge. *Nature* 550, 7676 (2017), 354–359.

[11] Meritxell Vinyals, Juan A. Rodríguez-Aguilar, and Jesús Cerquides. 2011. A Survey on Sensor Networks from a Multiagent Perspective. *Comput. J.* 54, 3 (2011), 455–470.

[12] Mohamed Salah Zaïem and Etienne Bennequin. 2019. Learning to Communicate in Multi-Agent Reinforcement Learning : A Review. *CoRR* abs/1911.05438 (2019). arXiv:1911.05438 http://arxiv.org/abs/1911.05438

[13] Changxi Zhu, Mehdi Dastani, and Shihan Wang. 2024. A survey of multi-agent deep reinforcement learning with communication. *Auton. Agents Multi Agent Syst.* 38, 1 (2024), 4. https://doi.org/10.1007/S10458-023-09633-6