Fully Independent Communication in Multi-Agent Reinforcement Learning

Extended Abstract

Rafael Pina
Loughborough University London
London, United Kingdom
r.m.pina@lboro.ac.uk

Varuna De Silva
Loughborough University London
London, United Kingdom
v.d.de-silva@lboro.ac.uk

Corentin Artaud
Loughborough University London
London, United Kingdom
c.artaud2@lboro.ac.uk

Xiaolan Liu
Loughborough University London
London, United Kingdom
xiaolan.liu@lboro.ac.uk

ABSTRACT

Several recent works have focused on communication approaches in Multi-Agent Reinforcement Learning (MARL). However, the multiple proposed communication methods might still be too complex and not easily transferable to more practical contexts. One of the reasons is due to the use of the famous parameter sharing trick. In this paper, we investigate how independent learners in MARL that do not share parameters can communicate. We demonstrate that this setting might incur into some problems, to which we propose a new learning scheme as a solution. Our results show that, despite the challenges, independent agents can still learn communication strategies following our method. Additionally, we use this method to investigate how communication in MARL is affected by different network capacities, both for sharing and not sharing parameters.

KEYWORDS

Multi-Agent Reinforcement Learning, Communication, Independent Learning, Deep Learning

ACM Reference Format:
Rafael Pina, Varuna De Silva, Corentin Artaud, and Xiaolan Liu. 2024. Fully Independent Communication in Multi-Agent Reinforcement Learning: Extended Abstract. 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

Communication in Multi-Agent Reinforcement Learning (MARL) has been an important subject in the field of MARL [2, 4, 7]. Usually, in the standard approaches, agents learn the tasks by making decisions based on their local observations. However, if they have communication capabilities, other information can be used to make a better decision. From the perspective of practical applications, communication is also seen as a feasible way of improving learning due to progresses in diverse fields [1, 5, 10].

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.

In conventional MARL approaches, communication can often be applied by using an additional network that learns how to produce messages [3, 9, 14]. This network can be used by the agents to produce a message that represents their knowledge or experience at a certain moment, and is then sent to the others. In most approaches, the parameter sharing setting is adopted, meaning that, under this configuration, it is used only a single network (or two, if there is a mixer network or a communication network, for instance) that is shared by all the agents [6, 7, 12]. However, when we look at practical applications, sharing parameters becomes unrealistic [13].

Within the multiple proposed communication methods, when sharing parameters is not feasible a question arises: can communication still be conducted successfully when the agents do not share parameters? In this paper, we intend to study the challenges of communication in independent MARL when parameters are not shared and agents have distinct networks for their policy and for generating communication messages, which is an understudied setting in MARL that can be beneficial for practical applications [13]. Thus, we propose a way of successfully communicating under these conditions. We further show how communication affects learning when the agent networks have higher or lower capacities. For an extended version of this work with further results refer to [11].

2 COMMUNICATION IN MARL FOR FULLY INDEPENDENT LEARNERS

We start by formally describing the implications of communication with fully independent agents in MARL that do not share parameters. Sharing parameters is taken for granted in most MARL approaches, and thus it is often forgotten to consider the implications of not sharing parameters.

We consider agents that are controlled by a policy network and have a distinct network whose purpose is to generate messages for communication. Let us first consider the case of Independent Q-learning (IQL) **with** parameter sharing and communication. For simplicity of notation, in the demonstration, we use the observations o_i instead of the history τ_i. Let also f_i and g_i denote two certain functions such that $f_i \rightarrow Q$ and $g_i \rightarrow \mathcal{M}$, for a set of all Q-values Q and a set of all messages $\mathcal{M}$. We have that

$$\{Q_i\}_{i=1}^{N} = \{f_i(o_i, m_{-i}, a_i; \theta)\}_{i=1}^{N}, \tag{1}$$

where m_{-i} corresponds to the messages from all agents except i, that is produced by a neural network denoted by a function g_j with parameters μ

$$m_{-i} = \{g_j(o_j;\mu)\}_{j=1,j\neq i}^{N} \wedge m_i = g_i(o_i;\mu). \tag{2}$$

This is the standard procedure for IQL **with** parameter sharing. However, when we do not share parameters of the networks, the case can be very different. We consider now IQL with **no** parameter sharing and with communication. In this configuration, the agents are fully self-contained and do not share any parameters. However, we allow them to communicate. In this case, if we follow an equivalent communication scheme as in the previous case (i.e., learning from the incoming messages from the others), we now have that

$$\{Q_i\}_{i=1}^{N} = \{f_i(o_i, m_{-i}, a_i;\theta_i)\}_{i=1}^{N}, \tag{3}$$

where m_{-i} corresponds once again to the messages from all agents except i, that are produced by a neural network denoted by a function g_j with parameters μ_j

$$m_{-i} = \{g_j(o_j;\mu_j)\}_{j=1,j\neq i}^{N} \wedge m_i = g_i(o_i;\mu_i). \tag{4}$$

In this case, j does not share parameters with i, and thus μ_i will never be updated $\forall i \in \{1,\ldots,N\}$. As a solution, we propose instead the following learning scheme for independent communication:

$$\{Q_i\}_{i=1}^{N} = \{f_i(o_i, m_{-i}, m_i, a_i;\theta_i)\}_{i=1}^{N}, \tag{5}$$

where m_{-i} corresponds again to the messages from all agents except i, that are produced by a function g_j with parameters μ_j (Eq. (4)).

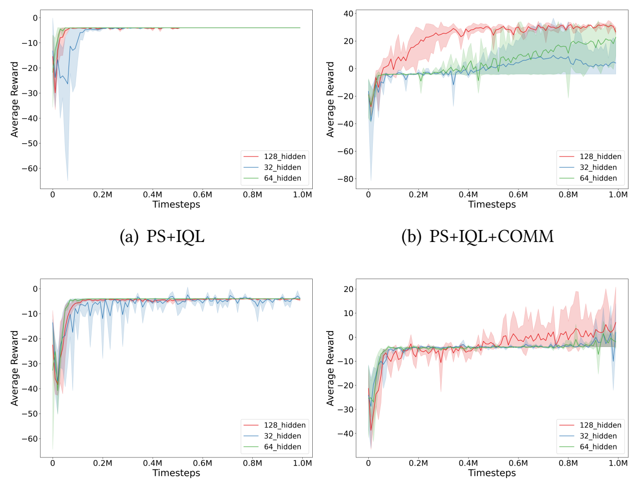
From the above, we note the existence of a second problem (that is independent of our solution to the initial problem), since μ_{-i} would be updated for N times, causing losses of gradient when propagating through the same values several times. To overcome this problem described, for each agent i, we detach m_{-i} from the computational graph, ensuring that all $\theta_i \wedge \mu_i, i \in \{1,\ldots,N\}$ are updated exactly once.

With this learning scheme, which can be summarized by Eq. (5), we solve both identified problems that occur in fully independent learning with communication and without parameter sharing. This scheme allows all the parameters to be updated, enabling learning with communication, as we show in the results ahead.

3 EXPERIMENTS AND RESULTS

We consider different settings to analyse the effects of communication with and without parameter sharing: PS+IQL (IQL and the agents share the parameters of the same network), NPS+IQL (IQL but no parameter sharing), PS+IQL+COMM (like PS+IQL, but now adding a communication network), and NPS+IQL+COMM (that corresponds to the scheme in section 2 - like NPS+IQL, but now each agent has its own communication network). Importantly, one of the key points of these experiments is to evaluate whether the proposed method in section 2 enables successful communication for independent learners who do not share parameters.

We observe that sharing parameters naturally brings advantages. In line with other works, we observe in Fig. 1 that sharing parameters (PS) works mostly as a way of speeding up learning and saving computational resources. When parameters are not shared (NPS), the problem becomes much more complex since there is no link between networks of different agents, and thus the communication network does not receive direct feedback of the messages produced by itself. However, despite this limitation inherent to the fact that parameters are not shared, we can still see the strong improvements of communication in fully independent learners in Fig. 1(d), where the agents manage to achieve positive rewards, as opposed to when they do not communicate (Fig. 1(c)). This demonstrates that our framework for communication when parameters are not shared enables learning in this challenging configuration.



(a) PS+IQL (b) PS+IQL+COMM (c) NPS+IQL (d) NPS+IQL+COMM

Figure 1: Rewards achieved in the PredatorPrey task [8] with 4 agents and 2 prey, with a penalty for non-cooperative behaviours of $-0.75 \times N$ every time an agent attempts to capture a prey alone. This punishment has shown to be important to evaluate communication approaches [7, 9, 14].

In order to study the amount of network capacity needed for learning and how communication helps with this information, we have also experimented with different sizes for the hidden layers of the agent network but fixed the communication network hidden dimensions to 64. As it was expected, when the agents have a higher network capacity, their performance is drastically improved. On the other hand, when the network capacity is not enough, it might take them longer to learn the tasks. This means that networks with a higher network capacity have a higher sample efficiency as they can learn faster with the same amount of samples.

Importantly, when we look at Fig. 1, we can see that the agents can only solve the task with communication and, while increasing the network capacity without communication does not have any impact, increasing it together with adding communication makes a big difference. In summary, note the important remark that while increasing the network capacity might be enough for some tasks, when communication is necessary simply increasing the network capacity is not enough, and both are needed.

ACKNOWLEDGMENTS

This work was funded by the Engineering and Physical Sciences Research Council, grant number EP/X028631/1: ATRACT: A Trustworthy Robotic Autonomous system to support Casualty Triage.

REFERENCES

[1] Kwang-Cheng Chen and Hsuan-Man Hung. 2019. Wireless Robotic Communication for Collaborative Multi-Agent Systems. In *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*. 1–7. https://doi.org/10.1109/ICC.2019.8761140

[2] Abhishek Das, Théophile Gervet, Joshua Romoff, Dhruv Batra, Devi Parikh, Michael Rabbat, and Joelle Pineau. 2019. TarMAC: Targeted Multi-Agent Communication. In *Proceedings of the 36th International Conference on Machine Learning*, Vol. 97. 1538–1546. arXiv: 1810.11187.

[3] Ziluo Ding, Weixin Hong, Liwen Zhu, Tiejun Huang, and Zongqing Lu. 2022. Sequential Communication in Multi-Agent Reinforcement Learning. https://openreview.net/forum?id=xzeGP-PtPMI

[4] Jakob N. Foerster, Yannis M. Assael, Nando de Freitas, and Shimon Whiteson. 2016. Learning to Communicate with Deep Multi-Agent Reinforcement Learning. In *Advances in Neural Information Processing Systems*, Vol. 29. arXiv: 1605.06676.

[5] Andrei Furda, Laurent Bouraoui, Michel Parent, and Ljubo Vlacic. 2010. Improving Safety for Driverless City Vehicles: Real-Time Communication and Decision Making. In *2010 IEEE 71st Vehicular Technology Conference*. 1–5. https://doi.org/10.1109/VETECS.2010.5494179

[6] Jayesh K. Gupta, Maxim Egorov, and Mykel Kochenderfer. 2017. Cooperative Multi-agent Control Using Deep Reinforcement Learning. In *Autonomous Agents and Multiagent Systems*, Gita Sukthankar and Juan A. Rodriguez-Aguilar (Eds.). Springer International Publishing, Cham, 66–83.

[7] Daewoo Kim, Sangwoo Moon, David Hostallero, Wan Ju Kang, Taeyoung Lee, Kyunghwan Son, and Yung Yi. 2019. Learning to Schedule Communication in Multi-agent Reinforcement Learning. arXiv:1902.01554 [cs.AI]

[8] Anurag Koul. 2019. ma-gym: Collection of multi-agent environments based on OpenAI gym. https://github.com/koulanurag/ma-gym.

[9] Zeyang Liu, Lipeng Wan, Xue sui, Kewu Sun, and Xuguang Lan. 2021. *Multi-Agent Intention Sharing via Leader-Follower Forest.* Technical Report arXiv:2112.01078. arXiv. http://arxiv.org/abs/2112.01078 arXiv:2112.01078 [cs] type: article.

[10] Hangyu Mao, Zhibo Gong, Zhengchao Zhang, Zhen Xiao, and Yan Ni. 2019. Learning Multi-agent Communication under Limited-bandwidth Restriction for Internet Packet Routing. arXiv:1903.05561 [cs.MA]

[11] Rafael Pina, Varuna De Silva, Corentin Artaud, and Xiaolan Liu. 2024. Fully Independent Communication in Multi-Agent Reinforcement Learning. arXiv:2401.15059 [cs.LG]

[12] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z. Leibo, Karl Tuyls, and Thore Graepel. 2017. Value-Decomposition Networks For Cooperative Multi-Agent Learning. arXiv:1706.05296 [cs.AI]

[13] Annie Wong, Thomas Bäck, Anna V Kononova, and Aske Plaat. 2023. Deep multiagent reinforcement learning: Challenges and directions. *Artificial Intelligence Review* 56, 6 (2023), 5023–5056.

[14] Sai Qian Zhang, Qi Zhang, and Jieyu Lin. 2019. Efficient Communication in Multi-Agent Reinforcement Learning via Variance Based Control. arXiv:1909.02682 [cs.LG]