HexaJungle: a MARL Simulator to Study the Emergence of Language

Kiran Ikram, Esther Mondragón, Eduardo Alonso, Michael Garcia Ortiz
Artificial Intelligence Research Centre (CitAI)
City, University of London
London EC1V 0HB, UK
{ kiran.ikram, e.mondragon, e.alonso, michael.garcia-ortiz } @city.ac.uk

Abstract

Multi-agent reinforcement learning in mixed-motive settings allows for the study of complex dynamics of agent interactions. Embodied agents in partially observable environments with the ability to communicate can share information, agree on strategies, or even lie to each other.In order to study this, we propose a simple environment where we can impose varying levels of cooperation, communication and competition as prerequisites to reach an optimal outcome. Welcome to the jungle.

1. Introduction and Motivation

Training agents to act while accounting for the (potentially adversarial) motives of other agents is a fundamental challenge in AI research. This ties to Theory of Mind: the intent (communicated or inferred) of one agent's influence an others behavior [6]. Learning to communicate is imperative to allow agents to share intent as well as information, leading to potential agreement on collaboration. Most of the prior work at the intersection of MARL[5] and Emergent Communication assumes full cooperation, rarely considering the inclusion selfish motives. In such settings emergence of communication is usually only assessed in relation to an increase in shared global reward value [8],[13][14], -a metric that might not be indicative of emergence of communication [10]. To study emergence of communication between interacting agents, we must also understand the conditions under which language evolves [7]. Thus we need to be able to test for these conditions in a consistent manner, and evaluate the effect of the environment on the emerged protocol [2].

Classical works from Game Theory[1] [11] [4] have proved that certain incentive structures, and in particular those which fall under the purvey of general sum, do in fact necessitate some form of communication to reach equilibrium . Additionally, [12] show that in the mixed-motive setting, the level of required collaboration tends to determine the emergence of communication between agents. These works establish a strong relation between mixed-motive settings and the emergence of communication. However, a classical shortcoming is their reliance on referential games with simplified structured interactions between agents. We take inspiration from their demonstrations and intuitions to propose a versatile mixed-motive RL environment that is not only conducive for the emergence of communication [3], but also, allows for fine-grained evaluation of the impact of communication on agent behaviour.

2. HexaJungle

HexaJungle is a suite of environments to test interactions and evaluate the emergence of communication. The goal is for two agents to reach a jungle exit. The optimal outcome for agents is dependent on predefined incentive structures that require collaboration or communication.

The agents live on an hexagonal grid as seen in 1, where they can go from cell to cell by changing their orientation and moving forward. To emulate realistic partial observability, they perceive cells in a 120 degree field of view, with a simplified form of occlusion. Agents may execute an additional action that allows them to climb on the shoulders of an other agent, giving then an unobstructed field of view, and to move towards Boulders (see later). Agents interact with the cells of the hexagonal grid depending on the type of elements (detailed therafter, illustrated in 1). Obstacles (dark brown) are non-traversable elements. Trees (dark green) disappear when crossed by an agent, which collects one wood log. Rivers (blue) are not obstructing the view of the agent, but an agent drowns when entering a river cell. If both agents enter it at the same time and have collected enough logs, they build a bridge. Boulders (light brown) are obstacles that can

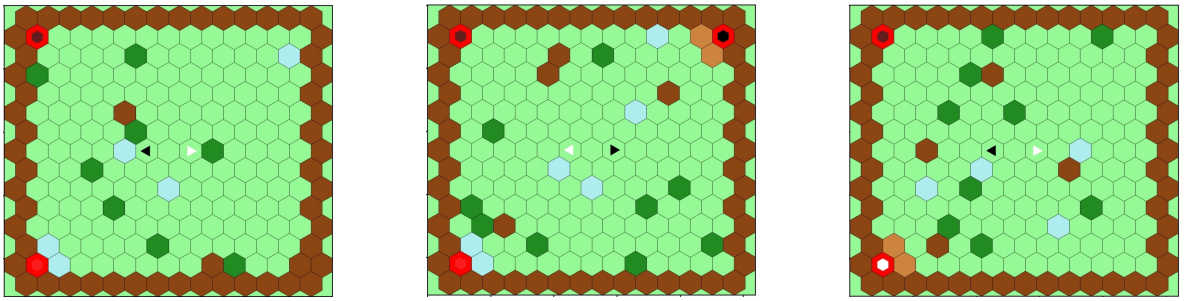

Figure 1. Left:At the top left is the Easy Exit accessible to both. To cross the river(bottom left) agents must collect enough logs. Middle:Agents may collaborate to reach the bottom-left RiverExit. In the top-right corner the exit is advantageous to Black, but neither agent has this information at the outset. Right:At the bottom left is an Exit surrounded by boulders which is advantageous only to Agent White.

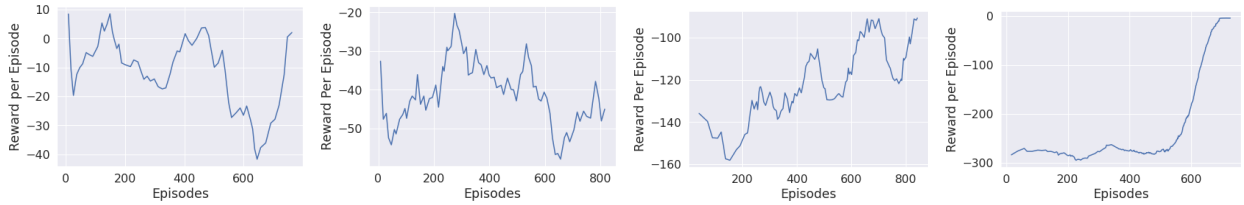

Figure 2. Performance: 1. Easy Exit 2. White and Easy Exit 3. River Exit 4. Black and River Exit

be climbed from an adjacent cell, by an agent standing on the shoulders of an other agent. Finally, we propose 4 different kind of Exits, which are essential in shaping the incentive structure. EasyExit terminates the episode for the agent entering the cell, providing a low reward. HardExit, instead, provides an average reward. BlackExit provides a reward only to the black agent, and symmetrically, WhiteExit provides a reward only to the white agent (the other agent can exit through these cells, but receive a low reward). An episode fully terminates when both agents have exited or after a time limit.

These reward structures allow us to create simple environments for mixed-motive MARL, and therefore test for the emergence of communication and collaboration in an embodied scenario. They allow for physical interaction with the world as well as between agents, which can collaborate to gain more information about their surrounding or access areas that they couldn't access alone. We illustrate different settings that exhibit these properties in a sample of environments presented in 1.

3. Experiments and Results

We carry out four experiments that reflect varying incentive structures. Utilizing RL Lib[9] we use PPO with LSTMS for 800 training episodes.

For Experiment(1) we set only an EasyExit, which means that agents neither need to collaborate nor communicate to get to it.For Experiment(2), we add an additional exit that carries a reward for Agent White alone(promoting self-interested behaviour in Agent White), and is surrounded by boulders. Its location is unknown to either agent until they cooperate. Experiment(3) is meant to test for collaboration. There are two exits, EasyExit with low reward for both agents and RiverExit, with high reward for both agents, iff they can collaborate to reach the exit together(after collecting logs from around the jungle) and build a bridge. Experiment(4) adds a higher degree of complexity as in addition to the RiverExit, we add an exit that incentivises Agent Black to behave selfishly.

The results as seen in 2 show clearly that it becomes harder for the agents to exit the jungle successfully (a) as the conflict of interest increases and (b) the need for collaboration increases. What is especially interesting is the results from Experiment(4) which indicate how agents may overcome the limitation posed by a selfish agent in the group and learn to work together to achieve a common goal.

4. Conclusion

We propose a simulator that is not only well suited to the emergence of communication, but allows for rigorous evaluation of said communication. Our code can be found at: https://github.com/kiranikram/HexaJungle

References

[1] Robert J Aumann. "Subjectivity and correlation in randomized strategies". In: Journal of mathematical Economics 1.1 (1974), pp. 67–96.

[2] Alexander I Cowen-Rivers and Jason Naradowsky. "Emergent Communication with World Models". In: arXiv preprint arXiv:2002.09604 (2020).

[3] Françoise Forges. "Equilibria with communication in a job market example". In: The Quarterly Journal of Economics 105.2 (1990), pp. 375–398.

[4] Francoise Forges. "An approach to communication equilibria". In: Econometrica: Journal of the Econometric Society (1986), pp. 1375–1385.

[5] Pablo Hernandez-Leal, Bilal Kartal, and Matthew E Taylor. "A very condensed survey and critique of multiagent deep reinforcement learning". In: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems. 2020, pp. 2146–2148.

[6] A Harry Klopf. Brain function and adaptive systems: a heterostatic theory. 133. Air Force Cambridge Research Laboratories, Air Force Systems Command, United ..., 1972.

[7] Angeliki Lazaridou and Marco Baroni. "Emergent Multi-Agent Communication in the Deep Learning Era". In: arXiv preprint arXiv:2006.02419 (2020).

[8] Angeliki Lazaridou, Anna Potapenko, and Olivier Tieleman. "Multi-agent Communication meets Natural Language: Synergies between Functional and Structural Language Learning". In: arXiv preprint arXiv:2005.07064 (2020).

[9] Eric Liang et al. "RLlib: Abstractions for distributed reinforcement learning". In: International Conference on Machine Learning. PMLR. 2018, pp. 3053–3062.

[10] Ryan Lowe et al. "On the pitfalls of measuring emergent communication". In: arXiv preprint arXiv:1903.05168 (2019).

[11] Roger B Myerson. "Multistage games with communication". In: Econometrica: Journal of the Econometric Society (1986), pp. 323–358.

[12] Michael Noukhovitch et al. "Emergent Communication under Competition". In: arXiv preprint arXiv:2101.10276 (2021).

[13] Yi Ren et al. "Compositional languages emerge in a neural iterated learning model". In: arXiv preprint arXiv:2002.01365 (2020).

[14] Cinjon Resnick et al. "Capacity, bandwidth, and compositionality in emergent language learning". In: arXiv preprint arXiv:1910.11424 (2019).