Internship project, master level, 2023

Title: Emergent Communication through Intrinsically-Motivated Goal-Conditioned Reinforcement Learning in Multi-Agent Environments

Supervision: Clément Moulin-Frier, Eleni Nisioti and Gautier Hamon

Team: Flowers team, Inria Bordeaux

Duration: 6 months, around February 2023 (with flexibility). **Level:** Master 2 research internship

Keywords: multi-agent reinforcement learning, emergent communication, intrinsic motivation, energy-based learning, contrastive learning, simulation environments, scientific programming with Python.

How to apply: contact clement.moulin-frier@inria.fr ; gautier.hamon@inria.fr ; eleni.nisioti@inria.fr with a CV and letter of motivation, ideally by the end of November 2023. We also recommend sending documents or reports describing previous projects you have been working on (even if they are not directly related to the topic), as well as your grades and links to some of your code repositories.

Requirements: We are looking for highly motivated MSc students (Master II). Programming skills and prior experience with Python and deep learning frameworks (Pytorch, Tensorflow) are expected.

Project

How can a population of reinforcement learning agents autonomously learn a diversity of cooperative skills in a shared environment? In the single-agent paradigm, goal-conditioned policies have been combined with intrinsic motivation mechanisms to endow agents with the ability to master a wide diversity of autonomously discovered goals. Over the last 15 years, the FLOWERS team has made important advances in this direction through the framework of Intrinsically Motivated Goal Exploration Process (IMGEPs, see (Colas et al., 2021) for a recent review). In an IMGEP, the agent is considered as *autotelic*, in the sense that it self-generates its own goals (by opposition to extrinsically motivated approaches, where the goals are provided through external supervision -- e.g. (Andrychowicz et al., 2017)).

Transferring this idea to cooperative Multi-Agent Reinforcement Learning (MARL, (Leibo et al., 2017; Littman, 1994)) entails a challenge: intrinsically motivated agents that sample goals independently focus on a shared cooperative goal with low probability, impairing their learning performance. In a recent contribution, we propose a new learning paradigm for modeling such settings, the Decentralized Intrinsically Motivated Skill Acquisition Problem (Dec-IMSAP, Figure 1, Top-Left), and employ it to solve cooperative navigation tasks (Masquil et al., 2022). Agents in a Dec-IMSAP are trained in a fully decentralized way, which comes in contrast to previous contributions in multi-goal MARL that consider a centralized goal-selection mechanism (e.g. (Mordatch & Abbeel, 2017; Yang et al., 2019)). Our empirical analysis indicates that a sufficient condition for efficiently learning a diversity of cooperative tasks is to ensure that a group *aligns* its goals, i.e. the agents pursue the same cooperative goal and learn to coordinate their actions to achieve them. We introduce the Goal-coordination game, a fully-decentralized emergent communication algorithm, where goal alignment emerges from the maximization of individual rewards in multi-goal cooperative environments and show that it is able to reach equal performance to a centralized training baseline that guarantees aligned goals.

In this internship, the objective will be to scale up the previous work mentioned above to more complex goals and environments. For this aim, we will rely on methods for emergent communication in large spaces we have recently developed (Lemesle et al., 2022), based on a combination of energy-based learning and contrastive learning (Figure 1, bottom). This will allow considering high-dimensional goal spaces and more complex communication systems. In terms of environment, we will consider recent simulation environments such as Grafter (*Grafter*, 2021/2022, Figure 1, top-right), where multiple agents interact in an open-ended environment to craft objects and build complex structures (in the spirit of Minecraft).

The details of the project can be adapted according to the specific skills and scientific interests of the student.

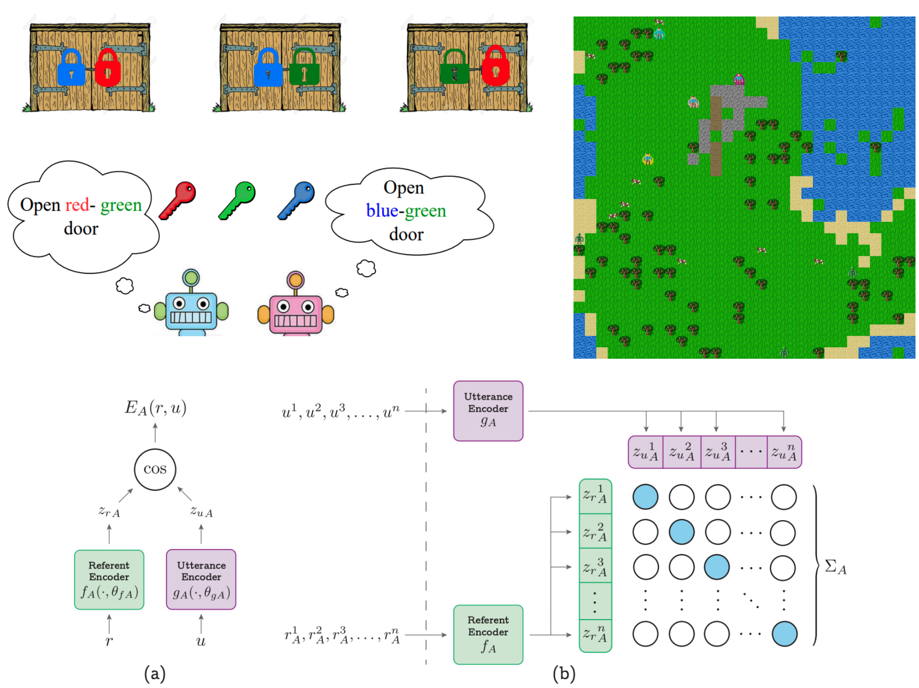

Figure 1. Top-Left: Illustrative example of a DecIMSAP, from (Masquil et al., 2022): two agents are in a shared environment where goals have the form of doors that can be opened when all locks have been matched with the key of the same color. Assuming that an agent can carry at most one key, it takes two to open a door, which makes the goals cooperative. How can we ensure that the two agents will learn how to open all doors? Prior to this evaluation, the agents are allowed to explore the environment autonomously, setting their own goals and pursuing them. **Top-Right**: A global view of the Grafter environment (*Grafter*, 2021/2022) with players highlighted. **Bottom:** Dual encoder architecture for emergent communication proposed in (Lemesle et al., 2022), where referents and utterances are mapped to a shared latent space.

References

Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., McGrew, B., Tobin, J., Abbeel, O. P., & Zaremba, W. (2017). Hindsight experience replay. *Advances in Neural Information Processing Systems*, 5048–5058.

Colas, C., Karch, T., Sigaud, O., & Oudeyer, P.-Y. (2021). Intrinsically Motivated Goal-Conditioned Reinforcement Learning: A Short Survey. *ArXiv:2012.09830 [Cs]*. http://arxiv.org/abs/2012.09830

Grafter. (2022). [Python]. Griddly. https://github.com/GriddlyAI/grafter (Original work published 2021)

Leibo, J. Z., Zambaldi, V., Lanctot, M., Marecki, J., & Graepel, T. (2017). Multi-agent Reinforcement Learning in Sequential Social Dilemmas. *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 464–473.

Lemesle, Y., Karch, T., Laroche, R., Moulin-Frier, C., & Oudeyer, P.-Y. (2022). *Emergence of Shared Sensory-motor Graphical Language from Visual Input* (arXiv:2210.06468). arXiv. https://doi.org/10.48550/arXiv.2210.06468

Littman, M. L. (1994). Markov games as a framework for multi-agent reinforcement learning. *Machine Learning Proceedings 1994*, 157–163. https://doi.org/10.1016/B978-1-55860-335-6.50027-1

Masquil, E., Hamon, G., Nisioti, E., & Moulin-Frier, C. (2022). *Intrinsically-Motivated*

Goal-Conditioned Reinforcement Learning in Multi-Agent Environments (arXiv:2211.06082). arXiv. https://doi.org/10.48550/arXiv.2211.06082

Mordatch, I., & Abbeel, P. (2017, March 14). Emergence of Grounded Compositional Language in Multi-Agent Populations. *Thirty-Second AAAI Conference on Artificial Intelligence*. http://arxiv.org/abs/1703.04908

Yang, J., Nakhaei, A., Isele, D., Fujimura, K., & Zha, H. (2019). *CM3: Cooperative Multi-goal Multi-stage Multi-agent Reinforcement Learning*. International Conference on Learning Representations. https://openreview.net/forum?id=S1lEX04tPr