Internship project, master level, 2022

Title: Emergent communication in multi-agent reinforcement learning

Supervision: Clément Moulin-Frier, Julius Taylor and Eleni Nisioti

Team: Flowers team, Inria Bordeaux

Duration: 6 months, around February July 2022 (with some flexibility).

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

How to apply: contact clement.moulin-frier@inria.fr ; julius.taylor@inria.fr ; eleni.nisioti@inria.fr with a CV and letter of motivation.

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

Context

What are the conditions for complex communication systems to emerge in populations of artificial agents? How emergent communication systems can in turn support the acquisition of an open-ended repertoire of cooperative skills? These questions are currently gaining considerable interest in the Artificial Intelligence (AI) community due to recent advances in machine learning. Deep Reinforcement Learning (DRL) algorithms (Mnih et al., 2015) have highly improved the abilities of artificial agents to acquire complex behavior by learning through experience an action policy maximizing a long-term reward. In a multi-agent setting, i.e. when multiple agents interact in a shared environment and co-acquire action policies for maximizing their own reward functions, the coupled learning processes can converge to complex cooperative or competitive strategies (Multi-Agent Reinforcement Learning, MARL, (Leibo et al., 2017; Littman, 1994)).
Among these strategies, communication systems can spontaneously emerge at the population level as a way to optimize the realization of complex tasks. However, recent contributions in Emergent Communication through MARL are still limited to relatively simple environments where agents learn how to solve a single predefined cooperative task (Foerster et al., 2016; Jaques et al., 2019; Mordatch & Abbeel, 2017).

Projects

The objective of the internship is to implement computational mechanisms supporting the emergence of a grounded, shared and compositional communication system in populations of interacting agents and to study how emergent communication systems can support the realization of complex cooperative tasks in multi-agent environments.

Below we describe several possible research directions that can be explored during the internship.

Project 1: Emergent communication through multi-agent intrinsically motivated goal exploration processes. - Building autonomous machines that can explore open-ended environments, discover possible interactions and autonomously build repertoires of skills is a general objective of artificial intelligence. 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)).
We would like to generalize the IMGEP framework to MARL. For this aim, we will design a complex multi-agent environment filled with various types of objects offering a large goal space to agents, some of these goals requiring certain levels of cooperation to be achieved. We will study how agents can self-generate their own goals and align them at the population level, learning how to collectively achieve them in a fully autonomous fashion for a given time budget, without any external supervision.
We identified the following desiderata:

- Goals should be generated by agents themselves, there should be no external supervision (contrarily to recent work in multi-goal MARL, e.g. (Yang et al., 2019))
- Agents should express some degree of shared intentionality (Tomasello & Carpenter, 2007), allowing two or more agents to focus on the same goal for some period of time.
- Communication should emerge in the agent population as a way to converge towards shared goals and to coordinate for achieving them.

Project 2: Socially-supervised representation learning in an embodied multi-agent environment. - In recent work (Taylor et al., 2021), we have proposed that aligning internal subjective representations, which naturally arise in a multi-agent setup where agents receive partial observations of the same underlying environmental state, can lead to the learning of more data-efficient internal representations. We proposed that multi-agent environments, where agents do not have access to the observations of others but can communicate within a limited range, guarantees a common context that can be leveraged in individual representation learning. The reason is that subjective observations necessarily refer to the same subset of the underlying environmental states and that communication about these states can freely offer a supervised signal. We referred to our setting as *socially supervised representation learning*. We

showed that our proposed architecture allows the emergence of aligned and efficient representations across an agent population.
However, in our preliminary experiments, we used a database of labeled images as a simplification of a multi-agent environment. We now want to study socially supervised representation learning in embodied agents truly interacting through perception and action in a shared dynamic environment. For this aim, we will design a simulation environment filled with different objects, where mobile agents can freely navigate and send communication signals to each other in a close range. We will study how subjectivity introduced when agents observe distinct perspectives of the same environment state contributes to learning abstract and aligned representations that can be leveraged for solving downstream MARL tasks, opening promising perspectives at the intersection of representation learning, emergent communication and reinforcement learning.

Project 3: Solving the non-stationarity issue in MARL through iterated learning and recurrent action policies. One of the main issues in decentralized MARL is that, from the perspective of a single agent, the environment appears as highly non-stationary due to the presence of other learning agents that update their own action policies over time (Hernandez-Leal et al., 2019). This non-stationary aspect can strongly impair the convergence of learning algorithms toward stable action policies across the agent population. Most recent approaches addressing this issue are based on the so-called "*centralized learning decentralized execution*" paradigm (Lowe et al., 2017), where agents have access to each other's observations and actions during training, yet can execute the learned policies based only on individual observations. However, such centralized training is not realistic from a biological point of view, where agents cannot directly access each other’s observations.
We believe that the non-stationarity issue in MARL can be solved by combining two mechanisms: iterated learning and recurrent world models. On the one hand, iterated learning (Kirby et al., 2014; Li & Bowling, 2019) consists in regularly introducing new untrained agents in the population (in the same way that new infants are regularly born in human populations). This allows to maintain a diversity of action policies within the whole agent population, each agent having the possibility to be trained with other agents at different skill levels. On the other hand, recurrent world models (e.g. (Ha & Schmidhuber, 2018)) allow each agent to select their action at a given time step of an episode conditioned on representations integrating the observations from all previous time steps. In principle, such recurrent world models can therefore capture relevant features of the other's action policies (e.g. capturing their respective skill levels and act accordingly see e.g. (Ndousse et al., 2021)). We will computationally study how these two mechanisms could solve the non-stationarity issue in MARL by learning action policies that are robust to a diversity of other agents' policies.

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
Foerster, J., Assael, Y. M., de Freitas, N., & Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 2137–2145.
Ha, D., & Schmidhuber, J. (2018). Recurrent World Models Facilitate Policy Evolution. *Advances in Neural Information Processing Systems*, *31*. https://papers.nips.cc/paper/2018/hash/2de5d16682c3c35007e4e92982f1a2ba-Abstract.html
Hernandez-Leal, P., Kaisers, M., Baarslag, T., & de Cote, E. M. (2019). A Survey of Learning in Multiagent Environments: Dealing with Non-Stationarity. *ArXiv:1707.09183 [Cs]*. http://arxiv.org/abs/1707.09183
Jaques, N., Lazaridou, A., Hughes, E., Gulcehre, C., Ortega, P. A., Strouse, D., Leibo, J. Z., & de Freitas, N. (2019). Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning. *Proceedings of the 35 Th International Conference on Machine Learning, Stockholm, Sweden*. https://openreview.net/forum?id=B1lG42C9Km
Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114. https://doi.org/10.1016/j.conb.2014.07.014
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.
Li, F., & Bowling, M. (2019). *Ease-of-Teaching and Language Structure from Emergent Communication* (pp. 15851–15861).
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
Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, O. P., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in Neural Information Processing Systems*, 6379–6390.
Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D., … Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236
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
Ndousse, K., Eck, D., Levine, S., & Jaques, N. (2021). Emergent Social Learning via Multi-agent Reinforcement Learning. *ArXiv:2010.00581 [Cs, Stat]*. http://arxiv.org/abs/2010.00581
Taylor, J., Nisioti, E., & Moulin-Frier, C. (2021). Socially Supervised Representation Learning: The Role of Subjectivity in Learning Efficient Representations. *ArXiv:2109.09390 [Cs]*. http://arxiv.org/abs/2109.09390
Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, *10*(1), 121–125. https://doi.org/10.1111/j.1467-7687.2007.00573.x
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