Projet de recherche doctoral

Title: Unsupervised pre-training of flexible multitask agents by learning to achieve a large diversity of goals

Context and objectives

With the advent of novel architectures (diffusion policies, transformers, VQ-VAEs, …) and of large foundational models pre-trained from data at the internet scale (LLMs, VLMs, …) research in reinforcement learning is currently making progress towards more widely applicable agents at a very fast pace. This is the case for instance in robotics with the emergence of Vision Language Action (VLA) models such as OpenVLA [Kim et al., 2024] and several others.

However, pre-training such large models for robotics tasks generally leverages imitation learning methods that require a large dataset of expert demonstrations collected from many robotic platforms. Collecting this dataset is costly and is currently the bottleneck of these methods.

The alternative approach that we will consider in this PhD project does not rely on imitation learning. It casts the pre-training problem as an unsupervised reinforcement learning problem in which an initially naive agent generates a very wide diversity of goals and learns to achieve these goals using simple goal-dependent reward functions. Such a pre-trained policy can then be readily used or quickly fine-tuned to achieve more sophisticated user-defined goals.

This PhD topic lies at the crossroad between unsupervised learning, multitask transfer, goal-conditioned reinforcement learning (GCRL) [Colas et al., 2018] and quality-diversity (QD) methods [Pugh et al., 2016]. The PhD advisors, Olivier Sigaud and Nicolas Perrin-Gilbert, have a significant expertise in these domains [Cachet et al., 2024, Doncieux et al., 2018, Péré et al., 2018, Macé et al., 2023]. This PhD project will contribute to a larger effort at ISIR towards combining recent learning architectures, foundational models and RL to build general purpose policies for robots.

Specific challenges

Robotic policy pre-training can be envisioned from various perspectives, which rely on different frameworks: unsupervised learning, multitask transfer, GCRL and QD methods all come with their specific assumptions and limitations. A central effort in the PhD will consist in extracting a common core from these various perspectives to work towards their unification.

For instance, the literature in unsupervised learning and GCRL generally considers an agent that sequentially learns to achieve more and more difficult goals according to some curriculum. In contrast, quality-diversity methods and recent related approaches consider a more parallel goal sampling dynamics [Frans et al., 2024]. The former approach is generally more sample efficient but can get trapped in restricted goal regions, whereas the latter may suffer from non sustainable training costs. So a unifying perspective should make it possible to address this trade-off between generality and cost.

Along another line, the thesis will look for adequate learning architectures for policy pre-training. Until recently, most GCRL and QD methods have relied so far on simple policy representations and the agents generally learn from user-defined compact state representations. In this PhD project we envision to use richer policy representations (diffusion policies, transformers, …) to build agents that can learn from multiple sensors, such as visual or tactile information.

If the thesis is successful, it will provide an original building block in the more general effort to design the next generation of flexible autonomous agents to control robots.

References :

[Cachet et al., 2024] Cachet, T., Dance, C. R., & Sigaud, O. (2024). Bridging environments and language with rendering functions and vision-language models. *arXiv preprint arXiv:2409.16024*.

[Colas et al., 2018] Cédric Colas, Tristan Karch, Olivier Sigaud, and Pierre-Yves Oudeyer. Autotelic agents with intrinsically motivated goal-conditioned reinforcement learning: a short survey. Journal of Artificial Intelligence Research, 74:1159–1199, 2022.

[Doncieux et al., 2018] Doncieux, S., Filliat, D., ... & Sigaud, O. (2018). Open-Ended Learning: A Conceptual Framework Based on Representational Redescription. *Frontiers in neurorobotics*, *12*.

[Frans et al., 2024] Frans, K., Park, S., Abbeel, P., & Levine, S. (2024). Unsupervised Zero-Shot Reinforcement Learning via Functional Reward Encodings. *arXiv preprint arXiv:2402.17135*.

[Kim et al., 2024] Kim, M. J., Pertsch, K., Karamcheti, S., Xiao, T., Balakrishna, A., Nair, S., ... & Finn, C. (2024). OpenVLA: An Open-Source Vision-Language-Action Model. *arXiv preprint arXiv:2406.09246*.

[Macé et al., 2023] Macé, V., Boige, R., Chalumeau, F., Pierrot, T., Richard, G., & Perrin-Gilbert, N. (2023, July). The quality-diversity transformer: Generating behavior-conditioned trajectories with decision transformers. In *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 1221-1229).

[Péré et al., 2018] Péré, A., Forestier, S., Sigaud, O., & Oudeyer, P. Y. (2018). Unsupervised learning of goal spaces for intrinsically motivated goal exploration. *arXiv preprint arXiv:1803.00781*.

[Pugh et al., 2016] Pugh, J. K., Soros, L. B., & Stanley, K. O. (2016). Quality diversity: A new frontier for evolutionary computation. *Frontiers in Robotics and AI*, *3*, 202845.