Learning &
Adaptive Systems

ETH zürich

# Information-directed Goal Selection for Reinforcement Learning

Project Proposal for Master Thesis

## Motivation

Goal-conditioned reinforcement learning augments the standard MDP formulation by conditioning the reward signal on a *goal*, which represents a task of interest [Schaul et al., 2015]. For instance, in the case of quadruped locomotion, the goal could represent velocity commands; for a manipulation task, it might involve 6DoF object positions. This slight reformulation naturally extends to policies, which can be in turn goal-conditioned, and thus instructed to perform specific tasks directly during inference.

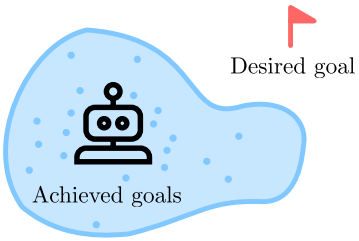


Figure 1: The problem of goal-selection: the agents needs to select a goal, such that attempting to achieve it is helpful in learning to achieve a desired goal (in red). Goals that have been achieved in the past (in blue) are usually good candidates.

While goals can be provided externally by humans, a curriculum of goals can also be self-supervised. This is a particularly promising approach when the agent interacts with an unknown environment, leading to exploration methods such as GoExplore [Ecoffet et al., 2021]. At a high level, the problem of goal selection also involves an exploration-exploitation dilemma: should the agent pursue goals that have not been achieved in the past [Pitis et al., 2020], or rather goals which are relevant, i.e. likely to be instructed at inference?

## Scope of the Project

The goal of this thesis is to tackle this problem from an information-directed perspective [Nikolov et al., 2019]: an agent should select goals resulting in trajectories which are maximally informative of the optimal policy for relevant tasks. Such an approach should address the exploration-exploitation tradeoff by only selecting relevant goals once the agent is capable of making significant progress towards them.

This project will mostly investigate principled approaches for self-supervised goal selection, along the lines of the one described above. While some grasp of first principles will be required for reviewing existing approaches and designing new ones, a significant part of the project will be empirical in nature. In particular, it will involve scaling methods to complex environments requiring the application of goal-conditioned deep reinforcement learning algorithms [Andrychowicz et al., 2017].

### Tasks

- reviewing related literature and formalizing the goal selection problem
- designing several practical criteria for goal selection
- benchmarking the designed approaches when coupled with deep RL algorithms

## Contact

Please, contact Marco Bagatella (mbagatella@ethz.ch) or Jonas Hübotter (jhuebotter@ethz.ch) with your CV and a short description of why you find this project interesting.

## References


[Andrychowicz et al., 2017] Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., McGrew, B., Tobin, J., Pieter Abbeel, O., and Zaremba, W. (2017). Hindsight experience replay. *Advances in Neural Information Processing Systems*.

[Ecoffet et al., 2021] Ecoffet, A., Huizinga, J., Lehman, J., Stanley, K. O., and Clune, J. (2021). First return, then explore. *Nature*, 590(7847):580–586.

[Nikolov et al., 2019] Nikolov, N., Kirschner, J., Berkenkamp, F., and Krause, A. (2019). Information-directed exploration for deep reinforcement learning. In *International Conference on Learning Representations*.

[Pitis et al., 2020] Pitis, S., Chan, H., Zhao, S., Stadie, B., and Ba, J. (2020). Maximum entropy gain exploration for long horizon multi-goal reinforcement learning. In *International Conference on Machine Learning*.

[Schaul et al., 2015] Schaul, T., Horgan, D., Gregor, K., and Silver, D. (2015). Universal value function approximators. In *International Conference on Machine Learning*.