INSTITUT
DES SYSTÈMES
INTELLIGENTS
ET DE ROBOTIQUE

# Internship proposal: Co-evolution of policies and environments

## Overview

- **Supervisors:** Stéphane DONCIEUX, Alexandre CONINX (ISIR AMAC Team)
- **Duration:** 6 months

## Topic description

### Scientific context

Reinforcement learning methods allow to build a policy that maximizes a given reward in a particular environment. The generated policy heavily depends on the domain it has been tested on. It creates two different issues: (1) the domain may be too hard for the learning process to proceed efficiently (bootstrap problem) and (2) the policy may not generate the same expected behavior in different domains (generalization issue). These two challenges are particularly important when applying reinforcement learning to robotics as the obtained policies are expected to face different situations and behave accordingly without the need to restart learning from scratch each time a modification of the environment occurs (modification of lighting conditions, of object positions or shape, etc).

Both problems can be dealt with at the same time by approaches that vary the evaluation conditions. The generalization problem can be reduced by evaluating on multiple randomized environment conditions [TFR+17, TPA+18]. It reduces the sample efficiency of the process, as a policy has to be evaluated multiple times, but this limitation can be mitigated with approaches relying on surrogate models [PKMD11]. Likewise, the bootstrap problem can be dealt with approaches that adapt the domain during the learning process, an approach called curriculum learning in the machine learning field [BLCW09]. It has led to policies generating policies solving very challenging conditions of benchmark tasks as the BipedalWalker [WLCS19].

### Internship goals and roadmap

In the work cited previously, the evaluation conditions are randomly varied, but they are not selected. Evolutionary approaches allow to perform both policy parameter and evaluation parameter co-evolution to generate an 'arm race' effect in which evaluation conditions become progressively more challenging, thus forcing policies to bootstrap and generalize better [DJ04]. The goal of the internship will be to study such approaches in the context of policy generations and in the context of quality-diversity algorithms (QD-algorithms) [PSS16, CD17] that aim at generating diverse sets of solutions instead of single optimal solutions.

The proposed roadmap for the internship is described below:

1. compare POET to an approach based on IPCA (Incremental Pareto-Coevolution Archive) [DJ04];

2. propose new QD-algorithms including the generation of diverse evaluation conditions;
3. optionally, compare these approaches to their equivalent in the deep learning community (curriculum learning, adversarial approaches).

## Candidate profile

The candidate should have a strong interest in machine learning and be enrolled in a MSc or engineering school program in computer science, machine learning or related fields. Good development skills and proficiency in Python programming language are mandatory. Good development skills in C++ is appreciated. The project will require working in close cooperation with several PhD students and researchers and requires good teamwork abilities. A working knowledge of English is required; knowledge of French is appreciated but not necessary.

## How to apply

Send an e-mail to `stephane.doncieux@sorbonne-universite.fr` with [Co-evolution of policies and environments] in the topic with a CV and motivation letter.

## References

[BLCW09] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48. ACM, 2009.

[CD17] Antoine Cully and Yiannis Demiris. Quality and diversity optimization: A unifying modular framework. *IEEE Transactions on Evolutionary Computation*, 22(2):245–259, 2017.

[DJ04] Edwin D De Jong. The incremental pareto-coevolution archive. In *Genetic and Evolutionary Computation Conference*, pages 525–536. Springer, 2004.

[PKMD11] Tony Pinville, Sylvain Koos, Jean-Baptiste Mouret, and Stéphane Doncieux. How to promote generalisation in evolutionary robotics: the progab approach. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, pages 259–266. ACM, 2011.

[PSS16] Justin K Pugh, Lisa B Soros, and Kenneth O Stanley. Quality diversity: A new frontier for evolutionary computation. *Frontiers in Robotics and AI*, 3:40, 2016.

[TFR$^{+}$17] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 23–30. IEEE, 2017.

[TPA$^{+}$18] Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaad Boochoon, and Stan Birchfield. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 969–977, 2018.

[WLCS19] Rui Wang, Joel Lehman, Jeff Clune, and Kenneth O Stanley. Poet: open-ended coevolution of environments and their optimized solutions. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 142–151. ACM, 2019.