Representation-driven Option Discovery in Reinforcement Learning

Marlos C. Machado

Department of Computing Science, University of Alberta
Alberta Machine Intelligence Institute (Amii)
Canada CIFAR AI Chair
machado@ualberta.ca

Talk Overview

The ability to reason at multiple levels of temporal abstraction is a fundamental aspect of intelligence. In reinforcement learning (RL), this attribute is often modelled through temporally extended courses of actions called options (Sutton et al. 1999). Many approaches within the options framework assume a well-defined set of options is already known. Without such a set, determining which options to consider remains an open question. In this talk, I will introduce a general framework for option discovery, which uses the agent's representation to discover useful options (Machado et al. 2023). By leveraging these options to generate a rich stream of experience, the agent can improve its representations and learn more effectively. This representation-driven option discovery approach creates a virtuous cycle of refinement, continuously improving both the representation and options, and it is particularly effective for problems where agents need to operate at varying levels of abstraction to succeed.

I will begin by discussing the representation-driven option discovery (ROD) cycle at a high level before presenting a concrete instantiation of it. At its core, the ROD cycle involves the agent interacting with the environment to collect data, which is then used to learn a representation. This representation defines subtasks for the agent, which solves them and leverages the learned policies to define options. These options enable the agent to act in new ways, altering its interaction with the environment and generating a different data stream. This, in turn, refines the learned representation, creating a continuous, constructivist cycle of increasingly complex temporal abstractions (see Figure 1).

The concrete instantiation of the ROD cycle I will describe addresses the exploration problem in RL. Here, the agent starts by selecting actions uniformly at random to learn the successor representation (Dayan 1993). Inspired by eigenoptions (Machado, Bellemare, and Bowling 2017), the agent discovers options through the successor representation's eigenvectors, which are added to its option set. This significantly reduces the time required to cover the environment and learn an effective reward-maximizing policy.

For clarity, the representation-driven option discovery framework and its different instantiations will be first presented in the tabular case. After establishing the proper intuitions, I will discuss how one can scale such an approach to powerful nonlinear function approximators like neural networks. Such an approach relies on recent results on how to properly estimate the eigenfunctions of the graph Laplacian with neural networks (Gomez, Bowling, and Machado 2024), being able to achieve state-of-the-art performance in many challenging benchmarks such as Atari 2600 games and 3D navigation tasks (Klissarov and Machado 2023).

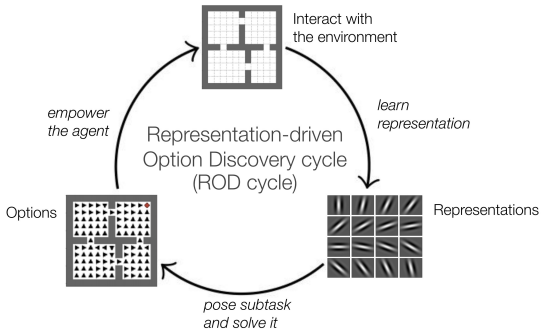

Figure 1: The representation-driven option discovery cycle.

References

Dayan, P. 1993. Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, 5(4): 613–624.

Gomez, D.; Bowling, M.; and Machado, M. C. 2024. Proper Laplacian representation learning. In *ICLR*.

Klissarov, M.; and Machado, M. C. 2023. Deep Laplacian-based options for temporally-extended exploration. In *ICML*.

Machado, M. C.; Barreto, A.; Precup, D.; and Bowling, M. 2023. Temporal abstraction in reinforcement learning with the successor representation. *Journal of Machine Learning Research*, 24(80): 1–69.

Machado, M. C.; Bellemare, M. G.; and Bowling, M. 2017. A Laplacian framework for option discovery in reinforcement learning. In *ICML*.

Sutton, R. S.; Precup, D.; and Singh, S. 1999. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artif. Intell.*, 112: 181–211.