Ph.D. in Information Technology
Thesis Defense

July 8th, 2025
at 11:00 am
Room BIO1– building 21

Gianluca DRAPPO – XXXVII Cycle

Theoretical Analysis Of Hierarchical Reinforcement Learning And Its Application For Autonomous Mission Planning

Supervisor: Prof. Marcello Restelli

Abstract:

Hierarchical Reinforcement Learning (HRL) approaches have demonstrated success in solving a wide range of complex, structured, and long-horizon problems. However, a complete theoretical understanding of this empirical success is still lacking. In the context of the *option* framework, prior research has developed efficient algorithms for scenarios where the options are *fixed*, and only the high-level policy that selects among these options needs to be learned. Surprisingly, the more realistic scenario where *both* the high-level and low-level policies are learned has been largely overlooked from a theoretical perspective.

This dissertation takes a step toward addressing this gap. Focusing on the finite-horizon setting, we present three provably efficient algorithms to advance the theoretical understanding of this scenario. We examine a specific family of hierarchies defined by two levels, formalising the high-level problem as a Semi-Markov Decision Process (SMDP), while the low-level problem involves learning in a set of finite-horizon MDPs, structured according to the hierarchical definition used.

In the first two approaches, the hierarchical structure is defined by a set of options, characterised by their initiation sets and termination conditions. The third approach considers goal-based MDPs with sparse rewards, where the structure is predefined and subdivides the original MDP into a high-level MDP and a set of low-level MDPs, each associated with specific sub-goals.

We initially propose a method inspired by the Explore-Then-Commit approach from bandit literature. This method first learns the options' policies and then exploits them to learn the high-level policy, which selects among these options. The second algorithm addresses the simultaneous learning of both levels, mitigating the inherent non-stationarity by continuously alternating between two phases: in each phase, one level learns while the other is kept fixed. Finally, the third method leverages a low-level regret minimiser in each episode to learn the low-level policies for the sub-goals assigned by the high-level policy. The high-level policy plans over the high-level MDP, using the uncertainty propagated by the value function of the low-level policies.

The derived bounds are compared with the lower bounds for non-hierarchical finite-horizon problems. This allows us to identify problem structures where hierarchical approaches are provably preferable to standard methods that ignore hierarchical structure, even when no pre-trained low-level policies are available.

Lastly, we compare a customised Deep Reinforcement Learning approach with a hierarchical one in a practical, realistic problem: mission planning for a team of autonomous aircraft. We demonstrate how, consistent with the theoretical findings, the hierarchical approach, which exploits structural information, scales effectively with the complexity of the scenario.

PhD Committee

Prof. **Francesco Trovò, Politecnico di Milano**

Prof. **Alessandro Farinelli, Università degli Studi di Verona**

Prof. **Anders Jonsson, Universitat Pompeu Fabra**