Ph.D. in Information Technology: Thesis Defenses

March 11th, 2021
online by Teams – at 14.00

Alberto Maria METELLI – XXXIII Cycle

Exploiting Environment Configurability in Reinforcement Learning

Supervisor: Prof. **Marcello Restelli**

Abstract:

In the last decades, Reinforcement Learning (RL) has emerged as an effective approach to address complex control tasks. The formalism typically employed to model the sequential interaction between the artificial agent and the environment is the Markov Decision Process (MDP). In an MDP, the agent perceives the state of the environment and performs actions. As a consequence, the environment transitions to a new state and generates a reward signal. The goal of the agent consists of learning a policy, i.e., a prescription of actions, that maximizes the long-term reward.

In the traditional setting, the environment is assumed to be a fixed entity that cannot be altered externally. However, there exist several real-world scenarios in which the environment can be modified to a limited extent and, therefore, it might be beneficial to act on some of its features. We call this activity environment configuration, that can be carried out by the agent itself or by an external entity, such as a configurator. Although environment configuration arises quite often in real applications, this topic is very little explored in the literature.

In this dissertation, we aim at formalizing and studying the diverse aspects of environment configuration. The contributions are theoretical, algorithmic, and experimental and can be broadly subdivided into three parts.

The first part of the dissertation introduces the novel formalism of Configurable Markov Decision Processes (Conf-MDPs) to model the configuration opportunities offered by the environment. At an intuitive level, there exists a tight connection between environment, policy, and learning process. We explore the different nuances of environment configuration, based on whether the configuration is fully auxiliary to the agent's learning process (cooperative setting) or guided by a configurator having an objective that possibly conflicts with the agent's one (non-cooperative setting).

In the second part, we focus on the cooperative Conf-MDP setting and we investigate the learning problem consisting of finding an agent policy and an environment configuration that jointly optimize the long-term reward. We provide algorithms for solving finite and continuous Conf-MDPs and experimental evaluations are conducted on both synthetic and realistic domains.

The third part addresses two specific applications of the Conf-MDP framework: policy space identification and control frequency adaptation. In the former, we employ environment configurability to improve the identification of the agent's perception and actuation capabilities. In the latter, instead, we analyze how a specific configurable environmental parameter, the control frequency, can affect the performance of the batch RL algorithms

Alessandro NUARA – XXXIII Cycle

Machine Learning Algorithms for the Optimization of Internet Advertising Campaigns

Supervisor: Prof. **Nicola Gatti**

Abstract:

Online advertising revenue grew to 124:6 billion dollars in 2019, showing a 15% increase over the previous year. The opportunities provided by this market attracted wide attention of the scientific community as well as the industry that requires automatic tools to manage the main processes involved in this market. For this purpose, Artificial Intelligence can play a crucial role in providing techniques to support both publishers and advertisers in their tasks. In this thesis, we design and analyze Machine Learning algorithms for the optimization of Internet advertising campaigns from the advertiser's point-of-view. An advertising campaign is composed of a set of sub-campaigns that differ for the ad (e.g., including text or images), target (e.g., keywords, age, interests), or channel (e.g., search, social, display). In pay-per-click advertising, to get an ad impressed, the advertisers take part in an auction carried out by the publisher, in which they set a bid and a daily budget for each sub-campaign. The bid represents the maximum amount of money the advertisers are willing to pay for a click, whereas the daily budget is the maximum spend in a day for a sub-campaign. The advertisers' goal is to create a set of sub-campaigns and, for each of them, set the bid/daily budget values that maximize the revenue under budget or return-on-investment constraints. This optimization problem is particularly challenging, as it includes many intricate subproblems.

In this work, we study four different settings, and we propose algorithms specifically crafted for each of them. For all the problems, we provide a theoretical analysis and an empirical evaluation of our approaches in both synthetic and real-world scenarios showing their superiority if compared with baselines and with the human campaign management.

Matteo PAPINI – XXXIII Cycle

Safe Policy Optimization

Supervisor: Prof. **Marcello Restelli**

Abstract:

Policy Optimization is a family of reinforcement learning algorithms that is particularly suited to real-world control tasks due to its ability of managing high-dimensional decision variables and noisy signals. This also makes Policy Optimization one of the most pressing targets of safety concerns. Outside of simulation, the trial-and-error behavior typical of learning agents can have concrete, potentially catastrophic consequences. The design of reliable adaptive agents for real-world settings requires, first of all, a better theoretical understanding of the learning algorithms used to train them. In this dissertation, we highlight the potential and limitations of existing policy optimization techniques, with a special focus on policy gradient algorithms. We study theoretical properties of policy gradients that are relevant to safety. We establish novel guarantees of monotonic performance improvement and convergence. We also study the trade-offs that safety requirements inevitably engage with sample complexity and exploration. Besides improving the theoretical understanding of policy gradient methods, we design new algorithms with more desirable properties, and evaluate them on simulated continuous control tasks.

Andrea TIRINZONI – XXXIII Cycle

Exploiting Structure for Transfer in Reinforcement Learning

Supervisor: Prof. **Marcello RESTELLI**

Abstract:

Recent advancements have allowed reinforcement learning algorithms to achieve outstanding results in a variety of complex sequential decision-making problems, from playing board and video games to the control of sophisticated robotic systems. However, current techniques are still very inefficient, in the sense that they require a huge amount of experience before learning near-optimal behavior. One solution to mitigate this limitation is knowledge transfer, i.e., the process of reusing experience obtained while facing previous tasks to speed-up the learning process of new related problems. In this thesis, we offer a number of contributions to the field of transfer in reinforcement learning, from practical to theoretical aspects. We do so in the context of structured domains, a concept that we introduce to model problems with similarities that enable knowledge transfer. We start by studying how to reuse old experience from a set of source tasks to reduce the sample complexity for learning a target task. For this problem, we derive two novel algorithms for batch and online settings, respectively. We then study the problem of generating new experience, i.e, of exploration in the target task given knowledge from previous tasks. We first design a practical algorithm that explores the target task driven by a prior distribution over its solution that is learned from the source tasks. We then study this problem from a theoretical perspective under the assumption that the underlying task structure, or an approximation of it, is known. For both multi-armed bandits and Markov decision processes, we design different algorithms for which we formally establish the benefits of exploiting structure, while ensuring optimality in specific cases. All together, these results advance our understanding of knowledge transfer, one of the key components towards the deployment of reinforcement learning agents to the real world.

PhD Committee
Prof. **Andrea Bonarini**, DEIB
Prof. **Alessandro Lazaric,** Facebook AI Research
Prof. **Gergely Neu,** Universitat Pompeu Fabra
Prof. **Carmine Ventre**, King's College