Ph.D. in Information Technology
Thesis Defenses

March 14th, 2022
at 14:00
online by Teams

Lorenzo BISI – XXXIV Cycle

Algorithms for Risk-Averse Reinforcement Learning

Supervisor: Prof. **Marcello Restelli**

Abstract:

Keeping risk under control is a primary concern in many critical real-world domains, including finance and healthcare. The literature on risk-averse reinforcement learning (RL) has mostly focused on designing ad-hoc algorithms for specific risk measures. As such, most of these algorithms do not easily generalize to measures other than the one they are designed for. Furthermore, it is often unclear whether state-of-the-art risk-neutral RL algorithms can be extended to reduce risk. In this dissertation, we take a step towards overcoming these limitations, by following two different paths. The first one consists in proposing a single framework to optimize some of the most popular risk measures, including conditional value-at-risk, utility functions, and mean-variance. Leveraging theoretical results on state augmentation, we transform the decision-making process so that optimizing the chosen risk measure in the original environment is equivalent to optimizing the expected return in the transformed one. We then present a risk-sensitive meta-algorithm that transforms the trajectories it collects from the environment and feeds these into any risk-neutral policy optimization method. The second path we follow consists in considering, for the first time, risk-measures connected to the state-action occupancy distribution, instead of the return one. We define a novel measure of risk, which we call reward volatility, consisting of the variance of the rewards under the state-occupancy measure, and we study the optimization of a trade-off objective called mean-volatility. We provide a monotonic improvement for this objective, which allows then to derive a TRPO-like algorithm for risk-averse optimization. Finally, in order to understand the impact of mean-volatility optimization on sample-complexity, we study the convergence rate of an actor-critic approach optimizing this criterion. Thus, we extend recent analyses in the risk-neutral actor-critic setting to the mean-volatility case, in order to establish the sample-complexity required to attain an epsilon-accurate stationary point. All contributions are empirically validated with extensive experimental analyses on challenging benchmarks.

Edoardo VITTORI – XXXIV Cycle

Augmenting Traders with Learning Machines

Supervisor: Prof. **Marcello Restelli**

Abstract:

The financial markets are comprised of several participants with diverse roles and objectives. Asset management firms optimize the portfolios of pension funds, institutions, and private individuals; market makers offer liquidity by continuously pricing and hedging their risks; proprietary traders invest their own capital with sophisticated methodologies. The approaches adopted by these actors are either manual or expert systems that rely on the experience of traders, and thus are subject to human bias and error. This dissertation proposes innovative techniques to address the limitations of the current trading strategies. Specifically, we explore the use of algorithms capable of autonomously learning the aforementioned sequential decision-making processes. The development of these algorithms entails a careful reproduction of realistic environments, as well as the observance of trading objectives, i.e., maximizing returns while maintaining a low risk profile and minimizing costs. These algorithms all share a common core structure, that is making a trading decision conditional on the current state of the financial markets. Our main theoretical and algorithmic contributions include the extension of the online learning field, as we introduce transaction costs and conservativeness in online portfolio optimization, and the enhancement of Monte Carlo Tree Search algorithms to account for the stochasticity and high noise typical of the financial markets. In terms of experimental contributions, we apply Reinforcement Learning to learn profitable quantitative trading strategies and option hedging approaches superior to the standard Black & Scholes hedge. We also find that Reinforcement Learning combined with Mean Field Games enables the development of competitive bond market making strategies. Finally, we demonstrate that dynamic optimal execution methods can be learned through Thompson Sampling with Reinforcement Learning. The use of such advanced techniques in a production environment may allow the achievement of a competitive advantage that will translate into economic benefits.

PhD Committee

Prof. **Nicola Gatti**, Politecnico di Milano

Prof. **Petter Kolm**, New York University

Prof. **Aviv Tamar**, Technion - Israel Institute of Technology