Abstract

In this talk, I will build the connection between Hamilton-Jacobi-Bellman equations and the multi-armed bandit (MAB) problems. MAB is a widely used paradigm for studying the exploration-exploitation trade-off in sequential decision making under uncertainty. Specifically, I will focus on the Bayesian optimal policy and its challenges due to the intractability of computing the optimal policy, especially for problems with large horizons. I will present an efficient algorithm that approximates the optimal policy by utilizing PDE solutions. Unlike traditional methods, the computational cost of our algorithm does not increase with the length of the horizon.