

Reinforcement learning for telecommunication network: from Opportunistic Spectrum Access to IoTs

Speaker: Raphaël Féraud

Abstract

In reinforcement learning, an agent chooses actions in order to maximize the rewards given by a dynamic environment. As the environment is initially unknown, the agent has to interact with it to gather information. Moreover, only the reward of the chosen actions is revealed. That is why the agent faces the exploration/exploitation dilemma: she has to explore loosely estimated actions in order to build a better estimate, and she would like to maximize her cumulated reward by playing the empirically best actions. The mathematical framework which handles the exploration/exploitation dilemma is called the multi-armed bandit problem. In comparison to the general reinforcement framework, the main difference is that the state of the environment does not depend on the actions chosen previously. This simplification makes it possible to build efficient algorithms that can be proven optimal. Although the multi-armed bandits have been designed for clinical trials, they are used in many fields such as advertising, recommendation, marketing optimization, web site optimization, and since a decade in telecommunication networks. The first use case was the Opportunistic Spectrum Access (OPA), where a Secondary User aims to find a channel free from Primary User. With the development of IoTs networks and 5G networks, new use cases have emerged for instance for configuring Self Organized Network, or for optimizing energy consumption in IoT and sensor networks.