Notice and Invitation

ECE Seminar

Topic

Self-Unaware Bandits with Switching Costs

Presented by

Amir Alipour-Fanid

Advisor

Dr. Kai Zeng

Time: April 6, 2021 02:00 PM Eastern Time (US and Canada)

Zoom Meeting Link: https://gmu.zoom.us/j/93117572011
Meeting ID: 931 1757 2011

Abstract:

We study a family of multi-armed bandit (MAB) problems, wherein, not only the player cannot observe the reward on the played arm (*self-unaware player*), but also it incurs switching costs when shifting to a new arm. We study two cases: In Case 1, at each round, the player is able to either *play* or *observe* the chosen arm but not both. In Case 2, the player can choose an arm to play, and at the same round, choose another arm to observe. In both cases, the player incurs a cost for consecutive arm switching due to playing or observing the arms. We propose two novel online learning-based algorithms each addressing one of the aforementioned MAB problems. We theoretically prove that the proposed algorithms for Case 1 and Case 2, achieve sublinear order-optimal regret of $Q(\sqrt[4]{KT^3 \ln K})$ and $Q(\sqrt[3]{(K-1)T^2 \ln K})$, respectively, where K is the number of arms and T is the total number of rounds. For Case 2, we extend the player's capability to multiple $m > 1$ observations and show that more observations do not necessarily improve the regret bound due to incurring switching costs. However, we derive an upper bound for switching cost as $c < 1/\sqrt[3]{m^2}$ for which the regret bound is improved as the number of observations increases. Finally, through this study we found that a generalized version of our approach gives an interesting sublinear regret upper bound result of $Q(T^{\frac{s+1}{s+2}})$ for any self-unaware bandit player with s number of binary decision dilemma before taking the action. To validate and complement the theoretical findings, we construct an adversarial environment and conduct extensive simulations on the proposed algorithms to provide empirical evaluation results.