Contribution ID: **26**

Type: **not specified**

Achieving Regular and Fair Learning in Combinatorial Multi-Armed Bandit

Friday, June 21, 2024 11:00 AM (30 minutes)

Combinatorial multi-armed bandit refers to the model that aims to maximize cumulative rewards in the presence of uncertainty. Motivated by two important wireless network applications, in addition to maximizing cumulative rewards, it is important to ensure fairness among arms (i.e., the minimum average reward required by each arm) and reward regularity (i.e., how often each arm receives the reward). In this paper, we develop a parameterized regular and fair learning algorithm to achieve these three objectives. In particular, the proposed algorithm linearly combines virtual queue-lengths (tracking the fairness violations), Time-Since-Last-Reward (TSLR) metrics, and Upper Confidence Bound (UCB) estimates in its weight measure. Here, TSLR is similar to age-of-information and measures the elapsed number of rounds since the last time an arm received a reward, capturing the reward regularity performance, and UCB estimates are utilized to balance the tradeoff between exploration and exploitation in online learning. Through capturing a key relationship between virtual queue-lengths and TSLR metrics and utilizing several non-trivial Lyapunov functions, we analytically characterize zero cumulative fairness violation, reward regularity, and cumulative regret performance under our proposed algorithm. These findings are corroborated by our extensive simulations.

Primary authors: WU, Xiaoyi; LI, Bin (Pennsylvania State University)

Presenter: LI, Bin (Pennsylvania State University)

Session Classification: Parallel session: Reinforcement learning for wireless scheduling