Learning in Generalized Linear Contextual Bandits with Stochastic Delays

Renyuan Xu

Mathematical Institute, University of Oxford

Joint work with Zhengyuan Zhou (NYU) and Jose Blanchet (Stanford)

December 11, 2019

Personalized Recommendation with Delayed Feedback

Advertisement

Conversion

- Recommendation engine utilizes **user features** (gender, age, browsing behavior, shopping history, salary, and etc)
- User feedback/Conversion comes in a **delayed** manner
- **Question:** How to do recommendation?

Problem Set-Up

- T: the number of rounds
- K: the number of possible actions
- In each round $t \leq T$:
  - learner observes K feature vectors $x_{t,a} \in \mathbb{R}^d$, $a \in [K]$
  - learner takes action a_t
  - reward y_{t,a_t} will be observed in round $t + D_t$ (with a delay D_t)
- Delay D_t: stochastic, possibly correlated and unbounded
- Generalized Linear Model ($X_t = x_{t,a_t}$ and $Y_t = y_{t,a_t}$):

$$Y_t = g\left(\langle \theta^*, X_t \rangle\right) + \epsilon_t$$

  - θ^* unknown, ϵ_t noise, g inverse link function

Results

Algorithm

- Upper confidence bound (UCB) type of algorithm
- Confidence bound depends on delays
- Select a subset of samples to calculate the estimator for θ^* (MLE)

Our Regret Bound

$$R_T = O\left(d\sqrt{T}\log T + \sqrt{\mu_D + M_D}\sqrt{Td\log T} + \sqrt{\sigma_G}\sqrt{Td}\,(\log T)^{3/4}\right)$$

with high probability

- μ_D, M_D, σ_D: delay-dependent parameters
- Delays can be possibly heavy-tailed
- The highest order term $O(d\sqrt{T}\log(T))$ does not depend on delays
- Tighter bound in d: standard Base/Sup LinUCB Decomposition

Wed Dec 11th 5 – 7 PM @ East Exhibition Hall B + C #2