Online Bayesian Recommendation with No Regret

Yiding Feng (MSR – New England), Wei Tang (WashU), Haifeng Xu (UChicago)

Microsoft Research · Washington University in St. Louis · The University of Chicago · EC 2022

Online Bayesian Recommendation

Setup:

- Video state $\theta \sim \lambda \in \Delta([m])$, λ is public prior information
- User has binary action $\mathcal{A} = \{0, 1\}$
- User's utility function ρ: $[m] \times \mathcal{A} \to \mathbb{R}$; Platform's state-independent utility $\xi: \mathcal{A} \to \mathbb{R}$ and $\xi(a) = a$

Information Asymmetry & Signaling Scheme:

- Only the platform can observe the realized video state
- Signal Space Σ, and signaling scheme π: $[m] \to \Delta(\Sigma)$

Online Setting: User's preference ρ is *unknown* to the platform!

Online Bayesian Recommendation:

For each round $t = 1, \dots, T$

- Platform commits Σ_t and π_t: $[m] \to \Delta(\Sigma_t)$;
- Video $\theta \sim \lambda$ is realized, and a signal $\sigma_t \sim \{\pi_t(\theta, \sigma)\}_{\sigma \in \Sigma_t}$ is sent to the user;
- Upon seeing σ_t, user forms a Bayesian posterior $\mu_t(\sigma_t, i) \triangleq \frac{\lambda(i)\pi_t(i,\sigma_t)}{\sum_{j\in[m]} \lambda(j)\pi_t(j,\sigma_t)}$
- With posterior, user selects action $a_t \in \operatorname{argmax}_{a\in\mathcal{A}} \mathbb{E}_{\theta\sim\mu_t}[\rho(\theta, a)]$;
- Platform gets utility $\xi(a_t)$

Applications: REELS · TikTok

Stackelberg Regret: $\mathrm{Reg}(T) = \sum_t U(\pi^*) - \sum_{\pi_1,\dots,\pi_T} U(\pi_t)$

where $U(\pi)$ denotes platform's expected payoff of π

Benchmark: Let $\omega(i) \triangleq \big(\rho(i,1) - \rho(i,0)\big) \cdot \lambda(i)$

Optimal π^*: $\Sigma^* = \mathcal{A}$, and a **threshold structure**: $\exists i^\dagger \in [m]$ s.t

(a) $\forall i \neq i^\dagger, \pi^*(i,1) = \mathbb{I}\left\{\frac{\omega(i)}{\lambda(i)} \geq \frac{\omega(i^\dagger)}{\lambda(i^\dagger)}\right\}$; (b) $\pi^*(i^\dagger, 1) = \frac{-\sum_{i\neq i^\dagger} \omega(i)\pi^*(i)}{\omega(i^\dagger)}$.

Persuasiveness of π: Whether $\sum_{i\in[m]} \omega(i)\pi(i,1) \geq 0$

Lower Bound and Challenge

Lower Bound: No online policy can achieve an expected regret better than $\Omega(\mathrm{loglog}\, T)$, even when $m = 2$.

Challenges to Have a Good Algorithm

The platform's feedback is *limited* and *probabilistic!: The platform knows whether π is persuasive or not only when (a) the realized signal σ_t is 1; (b) $\sigma_t = 0$, and $a_t = 1$.*

- Challenge 1: difficult to identify the signs of $\{\omega(i)\}$
- Challenge 2: difficult to determining if $U(\pi^*) \geq C$

	Feedback ($m = 2$)$^{(*)}$	Oracle Access	Query
Our Problem	$\mathbb{I}\{p \geq v\}$ w.p. p	MemberOracle w. no initial interior point	Vector $\boldsymbol{\pi}$
(Contextual) Dynamic Pricing [1, 2]	$\mathbb{I}\{p \geq v\}$	Separation Oracle	Scaler p_t

$(*)$: Here, $p \leftarrow \pi(i^\dagger, 1), v \leftarrow \pi^*(i^\dagger, 1)$

Two Algorithms

Algorithm 1 to Achieve $O(m! \cdot \mathrm{loglog}\, T)$:

- Enumerating all possible total order over the states
- There must exists a total order r^* over the states w.r.t. true $\frac{\lambda(i)}{\omega(i)}$
- Once r^* is pinned down, a mechanism similar to [2] to search $\pi^*(i^\dagger, 1)$ suffices

High-level Descriptions:

- Exploring phase 1:
  A binary search to identify $\underline{U}$ such that $\underline{U} \leq U(\pi^*) \leq 2\underline{U}$
- Exploring phase 2:
  Use a mechanism similar to [2] to identify a *persuasive* $\pi^\dagger$ such that $U(\pi^\dagger) \geq U(\pi^*) - 1/T$

Algorithm 2 to Achieve $O(\mathrm{poly}(m)\log T)$:

Platform's problem is optimizing a linear program with membership oracle access: known $\boldsymbol{\lambda} = (\lambda(i))_{i\in[m]}$ but unknown $\boldsymbol{\omega} = (\omega(i))_{i\in[m]}$

$$\max_{\boldsymbol{\pi}} \ \langle \boldsymbol{\pi}, \boldsymbol{\lambda} \rangle$$
$$\text{s.t. } \langle \boldsymbol{\pi}, \boldsymbol{\omega} \rangle \geq 0$$
$$\pi(i,1) \in [0,1], \quad \forall i \in [m]$$

Find initial point of the feasible region of $\boldsymbol{\omega}$ is not trivial!

Future Directions

Algorithm to achieve $O(\mathrm{poly}(m) \cdot \mathrm{loglog}\, T)$ (ongoing work)

[1]. Renato Paes Leme and Jon Schneider. Contextual search via intrinsic volumes. FOCS'18

[2]. Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. FOCS'03