Online Learning to Rank with Features

Shuai Li
Tor Lattimore
Csaba Szepesvari

Chinese University of Hong Kong

DeepMind

Setup

- Many items to be ranked
- E.g. Movies, news articles
- Online setting, limited modeling assumptions
- L items, K positions. $L \gg K$
- Each round the learner chooses an ordering of the items.
- Observes feedback in the form of "clicks"

Modeling Assumptions

- Hopeless to learn the value of all rankings
- Assume items are associated w. features
- $P(C_{ti} = 1 \mid A_t) = \alpha(A_t(i))\, \varphi_i(A_t)$
  - Attractiveness: $\alpha(A_t(i))$
  - Examination probability: $\varphi_i(A_t)$
- $\alpha(a) = \langle a, \theta \rangle$
  - unknown param in $\mathbb{R}^d$: θ
- $\varphi_i : \mathcal{A} \to [0,1]$ satisfies
  - Exam. prob. ind. order of items above

① $\boxed{\varphi_i(A) = \varphi_i(A') \text{ when } A([i-1]) = A'([i-1])}$

② $\boxed{\varphi_{i+1}(A) \leq \varphi_i(A)}$ ← Exam. prob. is decreasin

③ $\boxed{\varphi_i(A) \geq \varphi_i(A^*) \text{ when } A^* \text{ orders items by attractiveness.}}$

- Assumptions are satisfied by most standard click models
  - Exam prob. is smallest for A^*

Optimal Design

- let $\mathcal{A} \subset \mathbb{R}^d$
- Choose $x_1, \dots, x_n \in \mathcal{A}$
- Observe $y_1, \dots, y_n$ w. $y_t \sim \mathrm{Normal}(\langle x_t, \theta \rangle, 1)$
- LSE: $\hat{\theta} = G^{-1} \sum_{t=1}^{n} x_t y_t$ w. $G = \sum_{t=1}^{n} x_t x_t^T$
- $\mathbb{E}[\hat{\theta}] = \theta$, $\mathbb{E}[\langle \hat{\theta} - \theta, a \rangle^2] = \|a\|^2_{G^{-1}}$
- Should choose $(x_t)_{t=1}^{n}$ in prop. to

$$\pi^* = \operatorname*{argmin}_{\pi} \max_{a \in \mathcal{A}} \|a\|^2_{G_\pi^{-1}}$$

w. $G_\pi = \sum_a \pi(a)\, a a^T$

- Then $\mathbb{E}[\langle \hat{\theta} - \theta, a \rangle^2] = O\left(\frac{d}{n}\right)$

Algorithm

- Ranking ≈ Sorting with noise
- Algorithm maintains blocks of incomparable items

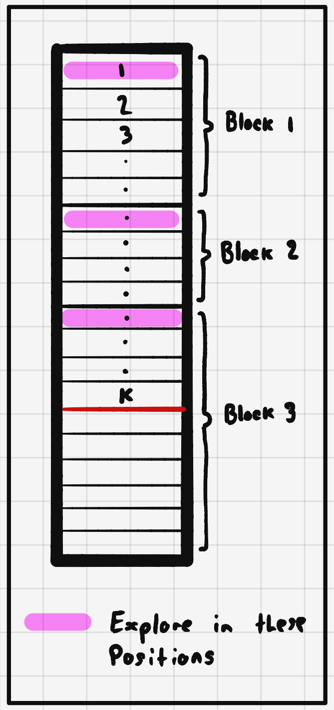

Figure 1

- First position in each block is used for exploration
- Remaining items are sorted according to empirical estimate of attractiveness
- Block is spit if the learner identifies a partition such that all items in one part are more attractive than the other with high probability

RecurRank

Theory

- Regret measured with respect to action that orders items from most to least attractive
- $R_n = \mathbb{E}\left[\sum_{t=1}^{n} \sum_{i=1}^{K} v_i(A^*) - v_i(A_t)\right]$

  $= O\left(K\sqrt{dn \log(nL)}\right)$

  - optimal action: A^*
  - dimension dipendence is optimal
  - only logarithmic in # items

Previous Work

- Either do not use features or,
- Depend on strong click model assumptions
- TopRank $R_n = O\left(K^{3/2}\sqrt{Ln \log(nL)}\right)$
- New algorithm improves on previous best for "generic" model, even in tabular case

Summary

- Ranking with guarantees
- Minimal Assumptions
- Efficient Algorithm
- Theory improves on prior work with fewer assumptions
- Limitation: Fixed features