Contextual Combinatorial Cascading Bandit Experiment

1 Preliminary

1.1 Synthetic dataset

Let $\mathcal{S} = \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ be the unit ball of $\mathbb{R}^d$. Let $E = \{1, ..., L\}$ be the set of all base arms. We randomly choose θ_* with $\|\theta_*\|_2 = 1$ and randomly assign a $\mu_i \in \mathcal{S}$ to i for any $i \in E$. At each time t, we choose $b_{t,i} \in \mathcal{S}$ randomly for any base arm i. Also we fix a constant h to balance weights of μ_i and disturbance $b_{t,i}$. Let $x_{t,i} = \frac{\mu_i + hb_{t,i}}{\|\mu_i + h \cdot b_{t,i}\|_2}$ be the context of base arm i at time t. And the weight for base arm i at time t is $w_t(i) = \theta_*^\top x_{t,i} + \epsilon_{t,i}$ where $\epsilon_{t,i} \sim N(\mu_i, \sigma_i)$ for fixed σ_i.

1.2 MovieLens

Let L be the number of all movies and let M be the number of all users. The MovieLens dataset is a big matrix $A \in \mathbb{R}^{M \times L}$ where $A(i, j) \in \{0, 1\}$ denotes whether user i has watched movie j or not. We split A to be $H + F$ by putting entry-1 of A to H and F with probability $\sim \mathrm{Ber}(p)$ for some fixed p. We can regard H as know information about history 'What users have watched' and regard F as future criterion. We use H to derive feature vectors of both users and movies by SVD decomposition $H = USV^\top$ where $U = (u_1; ...; u_M)$ and $V = (v_1; ...; v_L)$. At every time t, use $x_{t,i} = u_i v_j^\top$ as the context information of base arm i and randomly choose a user I_t. And use $w_t(j) = F(I_t, j)$ as the weight of base arm j.

Notice that for this case, fixed number of base arms, it might have problem if we use (u_{I_t}, v_j) as context information. Since to find the best arm, it is equivalent to find the best one with highest weights sum, so is equivalent to the best one with highest $\theta_v^\top x$.

The measurement for MovieLens is accuracy because we don't know the true θ_*.

2 Disjunctive case

2.1 Need to involve Contextual information

We experiment both on synthetic data and MovieLens. We compare our method with $\gamma_k = 1$ to the algorithm in Cascading Bandits(ICML'2015) with $L =, K =,$

2.2 Need to involve position discount parameter γ

We experiment both on synthetic data and MovieLens. We compare our method with $\gamma_k = \gamma^{k-1}$ to the algorithm in Cascading Bandits(ICML'2015) with $L =, K =, \gamma_k = 1$.

2.3 Cascading Information

We experiment both on synthetic data and MovieLens. We compare our method with $\gamma_k = \gamma^{k-1}$ to the algorithm in Qin Lijing(2014) with $L =, K =, \gamma_k = 1$.

Algorithm	**Cumulative Reward** $\sum_{t=1}^{T} r_i$
Li	4342.73
Qin	4323.26
Monkey	1765.41
Kveton	1787.81
Perfect Play	N/A

Table 1: **Cumulative reward w.r.t different baselines, under Movielens setting.**

3 Conjunctive case

3.1 Need to involve Contextual information

We experiment on synthetic data. We compare our method with $\gamma_k = 1$ to the algorithm in Cascading Bandits(ICML'2015) with $L =, K =,$

3.2 Need to involve position discount parameter γ

We experiment on synthetic data. We compare our method with $\gamma_k = \gamma^{k-1}$ to the algorithm in Cascading Bandits(ICML'2015) with $L =, K =, \gamma_k = 1$.

3.3 Cascading Information

We experiment on synthetic data. We compare our method with $\gamma_k = \gamma^{k-1}$ to the algorithm in Qin Lijing(2014) with $L =, K =, \gamma_k = 1$.

4 Results

An example output, with $T = 10000$, is listed below.