**Contextual Combinatorial Cascading Bandit Experiment**

# 1 Synthetic dataset

Let $I = \{e_1, e_2, \dots, e_L\}$ be a set of arms, each associated with a $d$-dimensional vector $\mu_i$ randomly drawn from $\{x \in \mathbb{R}^d : \|x\|_2 = 1\}$. At round $t$, the context corresponding to the $i$-th arm, denoted as $x_{t,i}$, is generated using $x_{t,i} = (\mu_i + h \cdot b_{t,i})/\|\mu_i + h \cdot b_{t,i}\|_2$, where $b_{t,i}$ is randomly drawn from $\{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ and $h$ is a constant throughout the experiment. The expectation of the Bernoulli realization of an arm $e \in I$ is $w_t(e) = \theta_*^T x_i + \epsilon_{i,t}$, where $\theta_*$ s.t. $\|\theta_*\|_2 = 1$ is randomly initialized and holds throughout the experiment, and $\epsilon_{i,t} \sim N(\mu_i, \sigma_i)$ i.i.d. be the fluctuation. In this experiment, the set of avaiable superarms $S$ is $\{A \subseteq I : |A| = k\}$. The following approaches were tested:

1. Li Shuai et al. Use linear regression on $\mathbf{O_t}$ observed arms and then UCB to select candidated.
2. 2014 Qin Lijing et al. Same w/ Li but takes more information (from $\mathbf{O_t}$ to $k$) with a full feedback.
3. 2015 Kveton Branislav et al. Combinatorial Cascading UCB which maintains its upper confidence bound purely by the historical payoff of the superarms selected. The contextual information $x_{i,t}$ is totally ignored so it can only catch the $\theta_*^T \mu$ part while suffers a lot from noisy.
4. Random.

# 2 Movielens

This section introduced Movielens dataset and its movie recommendation challenge. Let $L = \#\text{movies}$, Movielens is a matrix $A \in \{0,1\}^{\#\text{users}\cdot L}$ where each entry $A_{ij}$ is a boolean clickthrough indicator for the $i$-th user and the $j$-th movie. We split $A = A^1 + A^2$, by randomly putting hot entries in $A$ into $A^1$ and $A^2$ according to a bernulli distribution. At round $t$, a user indexed $i_t$ is randomly selected from a predefined set of users, and the agent is required to recommend movies using $A^1$ and $\mathcal{H}_t$ such that $A^2_{i_t j} = 1$ for as many those recommended $j$s as possible.

We formulate the movie recommendation problem into a combinatorial bandit problem. Let $A^1 = USV^T$ be the SVD decomposition and the movies $I = \{e_1, \dots, e_L\}$ be the set of arms, we define the context associated with $e_j$, at round $t$, as $x_{t,e_j} = u_{i_t}^T v_j$, where $u_l$, $v_l$ denote the $l$-th row of $U$, $V$, respectively. Upon received a super arm $A \in \{A \subseteq I : |A| = k\}$, the reward $r_t$ is calculated using its definition and $w_t(e_j) = A^2_{i_t e_j}$. Some notes follows:

1. Define $x_{t,e_j} = (u_{i_t}, v_j)$ is not applicable because the argmax over arms will ignore the stochastic part of the context.
2. Split users into training and testing, as 2014 Qin Lijing et al. did, is not applicable because the context is constant for each arm.
3. Measurement for Movielens is accuracy instead of regret because we have no access to the true $\theta_*$, if there is one.

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

| **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.**