Contextual Combinatorial Cascading Bandits

Abstract

The purpose of this document is to provide both the basic paper template and submission guidelines.

1. Introduction

2. Related Works

3. Contextual Combinatorial Cascading Bandits

3.1. Setting

We model our problem as a contextual combinatorial cascading bandit. Suppose we have $E = \{1, ..., L\}$ a finite set of L ground items. Let $\prod^k = \{(a_1, ..., a_k) : a_1, ..., a_k \in E, a_i \neq a_j \text{ for any } i \neq j\}$ be the set of all k-tuples of distinct items from E. Let $\mathcal{S} \subset \prod^{\leq K}(E)$ consist of feasible actions with length no more than K.

At time t, the learning agent is revealed with feature vectors $X_{t,a} \in \mathbb{R}^d$ for every basic arm $a \in E$, where we have $\|X_{t,a}\|_2 \leq 1$; this feature vector combines both information of the user and the corresponding basic arm. We use Bernoulli random variable $\mathbf{w}_t(a) \in \{0, 1\}$, which is called weight for arm a at time t, to indicate whether the user on round t will click on the item a or not. Assume $\mathbf{w}_{t,a}$ are mutually independent and satisfy

$$\mathbb{E}[\mathbf{w}_{t,a} | X_{t,a}] = \theta_*^\top X_{t,a} \tag{1}$$

where θ_* is an unknown d-dimensional vector with the assumption that $\|\theta_*\|_2 \leq 1$ and $0 < \theta_*^\top x_{t,a} < 1$ for all t, a. At time t, the learning agent chooses a solution $\mathbf{A}_t = (\mathbf{a}_1^t, ..., \mathbf{a}_{|\mathbf{A}_t|}^t) \in \mathcal{S}$ based on its past observations. The user then checks from the first item and stops if she has checked all items or click on one interesting. Suppose we add position discount $1 \geq \gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_K > 0$. If the user clicks on the k-th item of $\mathbf{A}_t$, the reward we receive is γ_k. At the end of time t, the agent observes $\mathbf{O}_t$ items in $\mathbf{A}_t$ and receive the reward

$$\mathbf{r}_t = \max_{1 \leq k \leq |\mathbf{A}_t|} \gamma_k \mathbf{w}_t(\mathbf{a}_k^t) = \bigvee_{k=1}^{|A_t|} \gamma_k \mathbf{w}_t(\mathbf{a}_k^t),$$

where we use the notation that $\bigvee_{1 \leq i \leq n} a_i = \max_{1 \leq i \leq n} a_i$. Note that every time t, we have observed $\mathbf{w}_t(\mathbf{a}_k^t), 1 \leq k \leq \mathbf{O}_t$.

If we define a function f on $A = (a_1, ..., a_{|A|}) \in \mathcal{S}, w = (w(1), ..., w(L))$ by

$$f : \mathcal{S} \times [0,1]^E \to [0,1]$$
$$f(A, w) = \sum_{k=1}^{|A|} \gamma_k (\prod_{i=1}^{k-1} (1 - w(a_i))) w(a_k),$$

then we have $\mathbf{r}_t = f(\mathbf{A}_t, \mathbf{w}_t)$ and $\mathbb{E}[r_t] = f(\mathbf{A}_t, \theta_*^\top X_t)$ where $X_t = (X_{t,1} \cdots X_{t,L}) \in \mathbb{R}^{d \times L}$. Let

$$A_t^* = \text{argmax}_{A \in \mathcal{S}} f(A, \theta_*^\top X_t).$$

The goal of our learning algorithm is to minimize the expected cumulative regret in T steps

$$R(T) = \mathbb{E}[\sum_{t=1}^T f(A_t^*, \theta_*^\top X_t) - f(\mathbf{A}_t, \theta_*^\top X_t)].$$

In the rest of this paper, we denote $\Delta_{t,\mathbf{A}_t} = f(A_t^*, \theta_*^\top X_t) - f(\mathbf{A}_t, \theta_*^\top X_t)$ and $R(T) = \mathbb{E}[\sum_{t=1}^T \Delta_{t,\mathbf{A}_t}]$.

Let $\mathcal{H}_t$ denote all history at the end of time t; $\mathcal{H}_t$ consists of $\{X_s, \mathbf{A}_s = (\mathbf{a}_1^s, ..., \mathbf{a}_{|\mathbf{A}_s|}^s), \mathbf{O}_s, \mathbf{w}_s(\mathbf{a}_k^s) : k \in [\mathbf{O}_s], s \in [t]\}$ and $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot | \mathcal{H}_t]$. By equation (1), we have $\mathbb{E}[(\gamma_k \mathbf{w}_{s,\mathbf{a}_k^s}) | \mathcal{H}_{s-1}] = \theta_*^\top (\gamma_k X_{s,\mathbf{a}_k^s})$. By ridge regression of data

$$\{(\gamma_k X_{s,\mathbf{a}_k^s}, \gamma_k \mathbf{w}_{s,\mathbf{a}_k^s})\}_{k \in [\mathbf{O}_s], s \in [t]}$$

let $\hat{\theta}_t$ be the l^2-regularized least-squares estimate of θ_* with regularization parameter $\lambda > 0$:

$$\hat{\theta}_t = (X^{t,\top} X^t + \lambda I)^{-1} X^{t,\top} \mathbf{Y}^t \tag{2}$$

where $X^t \in \mathbb{R}^{(\sum_{s=1}^t \mathbf{O}_s) \times d}$ is the matrix whose rows are $\gamma_k X_{s,\mathbf{a}_k^s}^\top$ and $\mathbf{Y}^t$ is the column vector whose elements are $\gamma_k \mathbf{w}_s(\mathbf{a}_k^s)$, $k \in [\mathbf{O}_s], s \in [t]$. Let

$$\mathbf{V}_t = X^{t,\top} X^t + \lambda I = \lambda I + \sum_{s=1}^t \sum_{k=1}^{\mathbf{O}_s} \gamma^{2k-2} X_{s,\mathbf{a}_k^s} X_{s,\mathbf{a}_k^s}^\top.$$

Then $\mathbf{V}_t \in \mathbb{R}^{d \times d}$ is a positive invertible matrix.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

3.2. Algorithm

Theorem 3.1 (Theorem 2 in (Abbasi-Yadkori et al., 2011)) *Let*

$$\beta_t(\delta) = \sqrt{\log(\det(\mathbf{V}_t)) + 2\log(\frac{1}{\delta})} + \sqrt{\lambda}.$$

Then for any $\delta > 0$, with probability at least $1-\delta$, for all $t > 0$, we have

$$\|\hat{\theta}_t - \theta_*\|_{\mathbf{V}_t} \leq \beta_t(\delta). \tag{3}$$

Our proposed algorithm, ConComCascade, is described in Algorithm 1. First, it computes the upper confidence bounds (UCBs) $\mathbf{U}_t \in [0,1]^E$ on the expected weights of all items in E. The UCB of item a at time t is defined as:

$$\mathbf{U}_t(a) = \min\{\hat{\theta}_{t-1}^\top X_{t,a} + \beta_{t-1}(\delta)\|X_{t,a}\|_{\mathbf{V}_{t-1}^{-1}}, 1\}. \tag{4}$$

By theorem 3.1, it holds that

Lemma 3.2 *For any $\delta > 0$, with high probability at least $1-\delta$, for any $t > 0$,*

$$0 \leq \mathbf{U}_t(a) - \theta_*^\top X_{t,a} \leq 2\beta_{t-1}(\delta)\|X_{t,a}\|_{\mathbf{V}_{t-1}^{-1}}.$$

Proof.

$$\begin{aligned}\left|\hat{\theta}_{t-1}^\top X_{t,a} - \theta_*^\top X_{t,a}\right| &\leq \|\hat{\theta}_{t-1} - \theta_*\|_{\mathbf{V}_{t-1}}\|X_{t,a}\|_{\mathbf{V}_{t-1}^{-1}} \\ &\leq \beta_t(\delta)\|X_{t,a}\|_{\mathbf{V}_{t-1}^{-1}}.\end{aligned}$$

□

Let

$$C_\gamma = \sum_{k=1}^{K} \gamma_k^2. \tag{5}$$

Theorem 3.3 *For any $\delta > 0$, with probability at least $1-\delta$, we have*

$$R(T) = O(\frac{d}{f^*}\log(C_\gamma T)\sqrt{KT}) \tag{6}$$

Algorithm 1 ConComCascade

```
//Initialization
Parameters: $\delta > 0, \lambda > 0, 1 \geq \gamma_1 \geq \cdots \geq \gamma_K > 0$
$\hat{\theta}_0 = 0, \beta_0(\delta) = 0, \mathbf{V}_0 = \lambda I$

for all t=1,2,... do
    //Compute UCBs
    $\forall a \in E$ :
    $\mathbf{U}_t(a) \leftarrow \min\{\hat{\theta}_{t-1}^\top X_{t,a} + \beta_{t-1}(\delta)\|X_{t,a}\|_{\mathbf{V}_{t-1}^{-1}}, 1\}$

    //Choose action $\mathbf{A}_t$ using UCBs $\mathbf{U}_t$
    $\mathbf{A}_t = (\mathbf{a}_1^t, ..., \mathbf{a}_{|\mathbf{A}_t|}^t) \leftarrow \text{argmax}_{A \in \mathcal{S}} f(A, \mathbf{U}_t)$
    Observe $\mathbf{O}_t, \mathbf{w}_t(\mathbf{a}_k^t), k \in [\mathbf{O}_t]$

    //Update statistics
    $\mathbf{V}_t \leftarrow \mathbf{V}_{t-1} + \sum_{k=1}^{\mathbf{O}_t} \gamma_k^2 X_{t,\mathbf{a}_k^t} X_{t,\mathbf{a}_k^t}^\top$
    $X^t \leftarrow (X^{t-1,\top}, \gamma_1 X_{t,\mathbf{a}_1^t}, ..., \gamma_{\mathbf{O}_t} X_{t,\mathbf{a}_{\mathbf{O}_t}^t})^\top$
    $Y^t \leftarrow (Y^{t-1,\top}, \gamma_1 \mathbf{w}(\mathbf{a}_1^t), ..., \gamma_{\mathbf{O}_t} \mathbf{w}_t(\mathbf{a}_{\mathbf{O}_t}^t))^\top$
    $\hat{\theta}_t \leftarrow (X^{t,\top} X^t + \lambda I)^{-1} X^{t,\top} \mathbf{Y}^t$
    $\beta_t(\delta) \leftarrow \sqrt{\log(\det(\mathbf{V}_t)) + 2\log(\frac{1}{\delta})} + \sqrt{\lambda}$
end for  t
```

Proof. With probability at least $1-\delta$,

$$\begin{aligned} R_T &= \sum_{t=1}^{T} \mathbb{E}_t[\Delta_{\mathbf{A}_t}] \\ &\leq \sum_{t=1}^{T} \frac{8}{f_t^*}\beta_{t-1}(\delta) \sum_{k=1}^{\mathbf{O}_t} \|\gamma^{k-1} X_{t,\mathbf{a}_k^t}\|_{\mathbf{V}_{t-1}^{-1}} \\ &\leq \frac{8}{f^*}\beta_T(\delta)\mathbb{E}[\sum_{t=1}^{T}\sum_{k=1}^{\mathbf{O}_t} \|\gamma^{k-1} X_{t,\mathbf{a}_k^t}\|_{\mathbf{V}_{t-1}^{-1}}] \\ &\leq \frac{8}{f^*}\beta_T(\delta)\sqrt{(\sum_{t=1}^{T}\mathbf{O}_t)\mathbb{E}[\sum_{t=1}^{T}\sum_{k=1}^{\mathbf{O}_t} \gamma^{2k-2}\|X_{t,\mathbf{a}_k^t}\|_{\mathbf{V}_{t-1}^{-1}}^2]} \\ &\leq O(\frac{d}{f^*}\log(C_\gamma T)\sqrt{TK}) \end{aligned} \tag{7}$$

□

Lemma 3.4 $1 \geq \gamma_1 \geq \cdots \geq \gamma_K > 0$. *If we have $w \leq w'$, then for any A,*

$$f(A, w) \leq f(A, w').$$

Proof. It suffices to prove when $A = \{1, ..., m\}, m \leq K$. First,

$$\begin{aligned} f(A, w) &= \sum_{k=1}^{m} \gamma_k \prod_{i=1}^{k-1}(1 - w_i)w_k \\ &\geq f(K, w_1, ..., w_{K-1}, w_K'). \end{aligned}$$

For $f(K, w_1, ..., w_k, w'_{k+1}, ..., w'_K) \geq f(K; w_1, ..., w'_k, ..., w'_K)$, it is equivalent to prove that $(w_k - w'_k)(1 - \gamma(w'_{k+1} + (1 - w'_{k+1})w'_{k+2} + ...)) > 0$. □

Lemma 3.5 $\sum_{k=1}^{K} \gamma^{k-1} \prod_{i=1}^{k-1} (1 - (a_i + c_i))(a_k + c_k) \leq \sum_{k=1}^{K} \gamma^{k-1} \prod_{i=1}^{k-1} (1 - a_i)a_k + \sum_{k=1}^{K} \gamma^{k-1} c_k$.

This lemma can be proved by induction. Need to change definition of Ut to min(Ut,1). So similarly, we can have

Lemma 3.6 $E[\Delta_{\mathbf{A}_t} | \mathcal{H}_t] \leq \frac{2}{f_t^*} \beta_{t-1}(\delta) \sum_{k=1}^{\min\{\mathbf{O}_t, |\mathbf{A}_t|\}} \|\gamma^{k-1} X_{t,\mathbf{a}_k^t}\|_{\mathbf{V}_{t-1}^{-1}}$

References

Abbasi-Yadkori, Yasin, Pál, Dávid, and Szepesvári, Csaba. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pp. 2312–2320, 2011.