A Proof for Approximation Ratio

We prove Theorem 1 in this section. First, we prove the following technical lemma:

Lemma 1 *For any positive integer $K = 1, 2, \ldots$, and any real numbers $b_1, \ldots, b_K \in [0, B]$, where B is a real number in $[0, 1]$, we have the following bounds*

$$\max\left\{\frac{1}{K}, 1 - \frac{K-1}{2}B\right\}\sum_{k=1}^{K} b_k \leq 1 - \prod_{k=1}^{K}(1 - b_k) \leq \sum_{k=1}^{K} b_k.$$

Proof 1 *First, We prove that $1 - \prod_{k=1}^{K}(1 - b_k) \leq \sum_{k=1}^{K} b_k$ by induction. Notice that when $K = 1$, this inequality trivially holds. Assume that this inequality holds for K, we prove that it also holds for $K + 1$. Note that*

$$\begin{aligned} 1 - \prod_{k=1}^{K+1}(1 - b_k) &= \left[1 - \prod_{k=1}^{K}(1 - b_k)\right][1 - b_{K+1}] + b_{K+1} \\ &\overset{(a)}{\leq} \left[\sum_{k=1}^{K} b_k\right][1 - b_{K+1}] + b_{K+1} \\ &\leq \sum_{k=1}^{K+1} b_k\,, \end{aligned} \tag{16}$$

where (a) follows from the induction hypothesis. This concludes the proof for the upper bound $1 - \prod_{k=1}^{K}(1 - b_k) \leq \sum_{k=1}^{K} b_k$.

Second, we prove that $1 - \prod_{k=1}^{K}(1 - b_k) \geq \frac{1}{K}\sum_{k=1}^{K} b_k$. Notice that this trivially follows from the fact that

$$1 - \prod_{k=1}^{K}(1 - b_k) \geq \max_k b_k \geq \frac{1}{K}\sum_{k=1}^{K} b_k\,. \tag{17}$$

Finally, we prove the lower bound $1 - \prod_{k=1}^{K}(1 - b_k) \geq \left[1 - \frac{K-1}{2}B\right]\sum_{k=1}^{K} b_k$ by induction.

Base Case: *Notice that when $K = 1$, we have*

$$1 - \prod_{k=1}^{K}(1 - b_k) = b_1 = \left[1 - \frac{K-1}{2}B\right]\sum_{k=1}^{K} b_k.$$

That is, the lower bound trivially holds for the case with $K = 1$.

Induction: *Assume that the lower bound holds for K, we prove that it also holds for $K+1$. Notice that if $1 - \frac{K}{2}B \leq 0$,*

then this lower bound holds trivially. For the non-trivial case with $1 - \frac{K}{2}B > 0$*, we have*

$$
\begin{aligned}
1 - \prod_{k=1}^{K+1}(1-b_k) &= \frac{1}{K+1}\sum_{i=1}^{K+1}\left\{(1-b_i)\left[1-\prod_{k\neq i}(1-b_k)\right]+b_i\right\} \\
&\overset{(a)}{\geq} \frac{1}{K+1}\sum_{i=1}^{K+1}\left\{(1-b_i)\left[1-\frac{K-1}{2}B\right]\sum_{k\neq i}b_k+b_i\right\} \\
&\overset{(b)}{\geq} \frac{1}{K+1}\sum_{i=1}^{K+1}\left\{(1-B)\left[1-\frac{K-1}{2}B\right]\sum_{k\neq i}b_k+b_i\right\} \\
&\overset{(c)}{=} \frac{K}{K+1}\left\{(1-B)\left[1-\frac{K-1}{2}B\right]\sum_{k=1}^{K+1}b_k\right\}+\frac{1}{K+1}\sum_{k=1}^{K+1}b_k \\
&= \left\{1-\frac{K}{2}B+\frac{K(K-1)}{2(K+1)}B^2\right\}\sum_{k=1}^{K+1}b_k \geq \left\{1-\frac{K}{2}B\right\}\sum_{k=1}^{K+1}b_k\,, \qquad (18)
\end{aligned}
$$

where (a) follows from the induction hypothesis, (b) follows from the fact that $b_i \leq B$ *for all* i *and* $1 - \frac{K-1}{2}B > 0$*, and (c) follows from the fact that* $\sum_{i=1}^{K+1}\sum_{k\neq i}b_k = K\sum_{k=1}^{K+1}b_k$*. This concludes the proof.*

We have the following remarks on the results of Lemma 1:

Remark 1 *Notice that the lower bound* $1 - \prod_{k=1}^{K}(1-b_k) \geq \frac{1}{K}\sum_{k=1}^{K}b_k$ *is tight when* $b_1 = b_2 = \ldots = b_K = 1$*. So we cannot further improve this lower bound without imposing additional constraints on* b_k*'s.*

Remark 2 *From Lemma 1, we have*

$$
1 - \frac{K-1}{2}B \leq \frac{1-\prod_{k=1}^{K}(1-b_k)}{\sum_{k=1}^{K}b_k} \leq 1\,.
$$

Thus, if $B(K-1) \ll 1$*, then* $1 - \prod_{k=1}^{K}(1-b_k) \approx \sum_{k=1}^{K}b_k$*. Moreover, for any fixed* K*, we have* $\lim_{B\downarrow 0}\frac{1-\prod_{k=1}^{K}(1-b_k)}{\sum_{k=1}^{K}b_k} = 1$*.*

We now prove Theorem 1 based on Lemma 1. Notice that by definition of $c_{\max}$, we have $\langle \Delta(a_k \mid \{a_1, \ldots, a_{k-1}\}), \theta^* \rangle \leq c_{\max}$. From Lemma 1, for any $A \in \Pi_K(E)$, we have

$$
\max\left\{\frac{1}{K}, 1 - \frac{K-1}{2}c_{\max}\right\}\langle c(A), \theta^*\rangle \leq f(A, \theta^*) \leq \langle c(A), \theta^*\rangle\,. \qquad (19)
$$

Consequently, we have

$$
\begin{aligned}
f(A^{\text{greedy}}, \theta^*) &\overset{(a)}{\geq} \max\left\{\frac{1}{K}, 1 - \frac{K-1}{2}c_{\max}\right\}\langle c(A^{\text{greedy}}), \theta^*\rangle \\
&\overset{(b)}{\geq} (1-\mathbf{e}^{-1})\max\left\{\frac{1}{K}, 1 - \frac{K-1}{2}c_{\max}\right\}\max_{A\in\Pi_K(E)}\langle c(A), \theta^*\rangle \\
&\overset{(c)}{\geq} (1-\mathbf{e}^{-1})\max\left\{\frac{1}{K}, 1 - \frac{K-1}{2}c_{\max}\right\}\langle c(A^*), \theta^*\rangle \\
&\overset{(d)}{\geq} (1-\mathbf{e}^{-1})\max\left\{\frac{1}{K}, 1 - \frac{K-1}{2}c_{\max}\right\} f(A^*, \theta^*)\,, \qquad (20)
\end{aligned}
$$

where (a) and (d) follow from (19); (b) follows from the facts that $\langle c(A), \theta^*\rangle$ is a monotone and submodular set function in A and A^{greedy} is computed based on the greedy algorithm; and (c) trivially follows from the fact that $A^* \in \Pi_K(E)$. This concludes the proof for Theorem 1.

B Proof for Regret Bound

We start by defining some useful notations. Let $\Pi(E) = \bigcup_{k=1}^{L} \Pi_k(E)$ be the set of all (ordered) lists of set E with cardinality 1 to L, and $w : \Pi(E) \to [0, 1]$ be an arbitrary weight function for lists. For any $A \in \Pi(E)$ and any w, we define

$$h(A, w) = 1 - \prod_{k=1}^{|A|} \left[1 - w(A^k)\right], \tag{21}$$

where A^k is the prefix of A with length k. With a little bit abuse of notation, we also define the feature $\Delta(A)$ for list $A = (a_1, \ldots, a_{|A|})$ as $\Delta(A) = \Delta(a_{|A|}|\{a_1, \ldots, a_{|A|-1}\})$. Then, we define the weight function $\bar{w}$, its high-probability upper bound U_t, and its high-probability lower bound L_t as

$$\begin{aligned}
\bar{w}(A) &= \Delta(A)^T \theta^*, \\
U_t(A) &= \mathrm{Proj}_{[0,1]}\left[\Delta(A)^T \bar{\theta}_t + \alpha\sqrt{\Delta(A)^T M_t^{-1} \Delta(A)}\right], \\
L_t(A) &= \mathrm{Proj}_{[0,1]}\left[\Delta(A)^T \bar{\theta}_t - \alpha\sqrt{\Delta(A)^T M_t^{-1} \Delta(A)}\right]
\end{aligned} \tag{22}$$

for any ordered list A and any time t. Note that $\mathrm{Proj}_{[0,1]}[\cdot]$ projects a real number onto interval $[0, 1]$, and based on Equation 5, 21, and 22, we have $h(A, \bar{w}) = f(A, \theta^*)$ for all ordered list A. We also use $\mathcal{H}_t$ to denote the history of past actions and observations by the end of time period t. Note that U_{t-1}, L_{t-1} and A_t are all deterministic conditioned on $\mathcal{H}_{t-1}$. For all time t, we define the "good event" as $\mathcal{E}_t = \{L_t(A) \le \bar{w}(A) \le U_t(A), \forall A \in \Pi(E)\}$, and $\bar{\mathcal{E}}_t$ as the complement of $\mathcal{E}_t$. Notice that both $\mathcal{E}_{t-1}$ and $\bar{\mathcal{E}}_{t-1}$ are also deterministic conditioned on $\mathcal{H}_{t-1}$. Hence, we have

$$\begin{aligned}
&\mathbb{E}\left[f(A^*, \theta^*) - f(A_t, \theta^*)/\gamma\right] = \mathbb{E}\left[h(A^*, \bar{w}) - h(A_t, \bar{w})/\gamma\right] \\
\le\, &P(\mathcal{E}_{t-1})\mathbb{E}\left[h(A^*, \bar{w}) - h(A_t, \bar{w})/\gamma | \mathcal{E}_{t-1}\right] + P(\bar{\mathcal{E}}_{t-1}),
\end{aligned} \tag{23}$$

where the above inequality follows from the naive bound that $h(A^*, \bar{w}) - h(A_t, \bar{w})/\gamma \le 1$. Notice that under event $\mathcal{E}_{t-1}$, we have $h(A, L_{t-1}) \le h(A, \bar{w}) \le h(A, U_{t-1})$ for all ordered list A. Thus, we have $h(A^*, \bar{w}) \le h(A^*, U_{t-1})$. On the other hand, since A_t is computed based on a γ-approximation algorithm, by definition

$$h(A^*, U_{t-1}) \le \max_{A \in \Pi_K(E)} h(A, U_{t-1}) \le h(A_t, U_{t-1})/\gamma.$$

Combining the above inequalities, under event $\mathcal{E}_{t-1}$, we have

$$h(A^*, \bar{w}) - h(A_t, \bar{w})/\gamma \le \frac{1}{\gamma}\left[h(A_t, U_{t-1}) - h(A_t, \bar{w})\right].$$

Recall that A_t^k is the prefix of A_t with length k, then we have

$$\begin{aligned}
&h(A_t, U_{t-1}) - h(A_t, \bar{w}) = \prod_{k=1}^{K}(1 - \bar{w}(A_t^k)) - \prod_{k=1}^{K}(1 - U_{t-1}(A_t^k)) \\
=\, &\sum_{k=1}^{K}\left[\prod_{i=1}^{k-1}(1 - \bar{w}(A_t^i))\right]\left(U_{t-1}(A_t^k) - \bar{w}(A_t^k)\right)\left[\prod_{j=k+1}^{K}(1 - U_{t-1}(A_t^j))\right] \\
\le\, &\sum_{k=1}^{K}\left[\prod_{i=1}^{k-1}(1 - \bar{w}(A_t^i))\right]\left(U_{t-1}(A_t^k) - \bar{w}(A_t^k)\right),
\end{aligned}$$

where the last inequality follows from the fact that $0 \le U_{t-1}(A_t^j) \le 1$. Let $\mathcal{G}_{tk}$ be the event that item a_k^t is examined at time t, then we have $\mathbb{E}\left[\mathbf{1}\left[\mathcal{G}_{tk}\right] | \mathcal{H}_{t-1}\right] = \prod_{i=1}^{k-1}(1 - \bar{w}(A_t^i))$. Moreover, since $\bar{w}(A_t^k) \ge L_{t-1}(A_t^k)$ under event $\mathcal{E}_{t-1}$

and $\mathcal{E}_{t-1}$ is deterministic conditioned on $\mathcal{H}_{t-1}$, for any $\mathcal{H}_{t-1}$ s.t. $\mathcal{E}_{t-1}$ holds, we have

$$\begin{aligned}
&\mathbb{E}\left[h(A_t, U_{t-1}) - h(A_t, \bar{w})|\mathcal{H}_{t-1}\right] \\
\leq& \sum_{k=1}^{K} \mathbb{E}\left[\mathbf{1}\left[\mathcal{G}_{tk}\right]|\mathcal{H}_{t-1}\right]\left[U_{t-1}(A_t^k) - L_{t-1}(A_t^k)\right] \\
\overset{(a)}{\leq}& 2\alpha\mathbb{E}\left[\sum_{k=1}^{K}\mathbf{1}\left[\mathcal{G}_{tk}\right]\sqrt{\Delta(A_t^k)^T M_{t-1}^{-1}\Delta(A_t^k)}\middle|\mathcal{H}_{t-1}\right] \\
\overset{(b)}{=}& 2\alpha\mathbb{E}\left[\sum_{k=1}^{\min\{C_t,K\}}\sqrt{\Delta(A_t^k)^T M_{t-1}^{-1}\Delta(A_t^k)}\middle|\mathcal{H}_{t-1}\right],
\end{aligned}$$

where (a) follows from the definitions of U_{t-1} and L_{t-1} (see Equation 22), and (b) follows from the definitions of C_t and $\mathcal{G}_{tk}$. Plug the above inequality into Equation 23, we have

$$\begin{aligned}
&\mathbb{E}\left[f(A^*, \theta^*) - f(A_t, \theta^*)/\gamma\right] \\
\leq& P(\mathcal{E}_{t-1})\frac{2\alpha}{\gamma}\mathbb{E}\left[\sum_{k=1}^{\min\{C_t,K\}}\sqrt{\Delta(A_t^k)^T M_{t-1}^{-1}\Delta(A_t^k)}\middle|\mathcal{E}_{t-1}\right] + P(\bar{\mathcal{E}}_{t-1}) \\
\leq& \frac{2\alpha}{\gamma}\mathbb{E}\left[\sum_{k=1}^{\min\{C_t,K\}}\sqrt{\Delta(A_t^k)^T M_{t-1}^{-1}\Delta(A_t^k)}\right] + P(\bar{\mathcal{E}}_{t-1}).
\end{aligned}$$

So we have

$$R^{\gamma}(n) \leq \frac{2\alpha}{\gamma}\mathbb{E}\left[\sum_{t=1}^{n}\sum_{k=1}^{\min\{C_t,K\}}\sqrt{\Delta(A_t^k)^T M_{t-1}^{-1}\Delta(A_t^k)}\right] + \sum_{t=1}^{n}P(\bar{\mathcal{E}}_{t-1}).$$

The regret bound can be obtained based on a worst-case bound on $\sum_{t=1}^{n}\sum_{k=1}^{\min\{C_t,K\}}\sqrt{\Delta(A_t^k)^T M_{t-1}^{-1}\Delta(A_t^k)}$, and a bound on $P(\bar{\mathcal{E}}_{t-1})$. The derivations of these two bounds are the same as in Zong *et al.* Zong et al. (2016). Specifically, we have

Lemma 2 *The following worst-case bound holds*

$$\sum_{t=1}^{n}\sum_{k=1}^{\min\{C_t,K\}}\sqrt{\Delta(A_t^k)^T M_{t-1}^{-1}\Delta(A_t^k)} \leq K\sqrt{\frac{dn\log\left[1+\frac{nK}{d\sigma^2}\right]}{\log\left(1+\frac{1}{\sigma^2}\right)}}.$$

Please refer to Lemma 2 in Zong *et al.* Zong et al. (2016) for the derivation of Lemma 2. We also have the following bound on $P(\bar{\mathcal{E}}_t)$:

Lemma 3 *For any t, $\sigma > 0$, $\delta \in (0,1)$, and*

$$\alpha \geq \frac{1}{\sigma}\sqrt{d\log\left(1+\frac{nK}{d\sigma^2}\right) + 2\log\left(\frac{1}{\delta}\right)} + \|\theta^*\|_2,$$

we have $P(\bar{\mathcal{E}}_t) \leq \delta$.

Please refer to Lemma 3 in Zong *et al.* Zong et al. (2016) for the derivation of Lemma 3. Based on the above two lemmas, if we choose

$$\alpha \geq \frac{1}{\sigma}\sqrt{d\log\left(1+\frac{nK}{d\sigma^2}\right) + 2\log(n)} + \|\theta^*\|_2,$$

we have $P(\bar{\mathcal{E}}_t) \leq 1/n$ for all t and hence

$$R^{\gamma}(n) \leq \frac{2\alpha K}{\gamma}\sqrt{\frac{dn\log\left[1+\frac{nK}{d\sigma^2}\right]}{\log\left(1+\frac{1}{\sigma^2}\right)}} + 1.$$

This concludes the proof for Theorem 2.

C Extension to Special Case of Dependent Click Model

In this section, we discuss a special case of *dependent click model (DCM)* (Guo et al., 2009), which establishes the deficiency of existing online diversity-driven approaches that learn from penalizing the *unexamined* items even in the multiple click scenario.

C.1 DCM Background

The following description follows from the formulation discussed in Katariya et al. (2016). The DCM is an extension of the cascade model (Craswell et al., 2008) to multiple clicks. The model assumes that the user scans a list of K items $A = (a_1, \ldots, a_K) \in \Pi_K(E)$ from the first item a_1 to the last a_K. The DCM is parameterized by L item-dependent attraction probabilities $\bar{w} \in [0,1]^L$ and K position-dependent termination probabilities $\bar{v} \in [0,1]^K$. After the user examines item a_k, the item attracts the user with probability $\bar{w}(a_k)$. If the user is attracted by the item a_k, then the user clicks on the item and terminates scanning the list with probability $\bar{v}(k)$. If this happens, the user is satisfied with item a_k and does not examine any of the remaining items. If item a_k is not attractive or the user does not terminate, the user examines item a_{k+1}. The first item is examined with probability one. The probability $\bar{w}(a_k)$ is conditioned on the event that the user examines item at position k. The probability $\bar{v}(k)$ is conditioned on the event that the user is attracted to the item at position k.

Under this model, the probability that the user is satisfied with an item in list A is:

$$f_{DCM}(A, \bar{v}, \bar{w}) = 1 - \textstyle\prod_{k=1}^{K} (1 - \bar{v}(k)\bar{w}(a_k)). \tag{24}$$

Clearly, the list which places k-th most attractive item at the k-th position maximizes the objective function in (24).

C.2 Special Case of DCM and the Associated Bandit Setting

In order to extend `CascadeLSB` to the multiple-click scenario, we assume the following. Similar to model in Section 2.2, the diversity in this model is over d topics, such as movie genres or restaurant types. The preferences of the user are a distribution over these topics represented by a vector $\theta^* = (\theta^*_1, \ldots, \theta^*_d)$.

The attraction probability of item a_k is defined in (3), i.e,

$$\bar{w}(a_k) = \langle \Delta(a_k \mid \{a_1, \ldots, a_{k-1}\}), \theta^* \rangle. \tag{25}$$

The quantity in (3) is the gain in topic coverage after item a_k is added to the first $k-1$ items weighted by the preferences of the user θ^* over the topics. The termination probability is kept constant to be $\bar{v}(k) = \zeta$ for all positions $k \in [K]$. Notice that ζ is the probability that the user terminates scanning the list after getting attracted to (i.e. clicked) an item. $\zeta = 1$ will lead to the CDCM in Section 3. Any $\zeta < 1$ will lead to multiple clicks until the user is satisfied (i.e. the user clicks on an item and then decides to terminate). Under this assumption, the probability that at least one item in A is satisfactory is $f_{DCM}(A, \theta^*, \zeta)$, where

$$f_{DCM}(A, \theta, \zeta) = 1 - \prod_{k=1}^{K} (1 - \zeta \langle \Delta(a_k \mid \{a_1, \ldots, a_{k-1}\}), \theta \rangle) \tag{26}$$

for any list A, preferences θ, topic coverage c, and termination probability ζ. Let us denote the list that maximizes (26) under user preferences θ^* as

$$A^*_{DCM} = \arg\max_{A \in \Pi_K(E)} f_{DCM}(A, \theta^*, \zeta). \tag{27}$$

We again follow the greedy algorithm of (7) that maximizes $f_{DCM}(A, \theta^*, \zeta)$ approximately. The algorithm chooses K items sequentially. The k-th item a_k is chosen such that it maximizes its gain over previously chosen items $a_1, \ldots, a_{k-1}$. In particular, for any $k \in [K]$,

$$a_k = \arg\max_{e \in E \setminus \{a_1, \ldots, a_{k-1}\}} \langle \Delta(e \mid \{a_1, \ldots, a_{k-1}\}) \theta^* \rangle. \tag{28}$$

Deriving approximation guarantees for the above algorithm is a part of our future work. We conjecture that the approximation factor, say γ_{DCM}, would be very similar to γ in Theorem 1, with a multiplicative factor of ζ coming in at least for this special case of DCM.

The bandit setting in this case follows from Section 4. We call it *dependent click linear submodular bandit*. An instance of this problem is a tuple $(E, c, \theta^*, K, \zeta)$, where $E = [L]$ represents a ground set of L items, c is the topic coverage function in Section 2.2, θ^* are user preferences in Section 3, $K \leq L$ is the number of recommended items, and ζ is the termination probability which is assumed to be same for all the positions in this special case. The preferences θ^* are unknown to the learning agent.

Our learning agent interacts with the user as follows. At time t, the agent recommends a list of K items $A_t = (a_1^t, \ldots, a_K^t) \in \Pi_K(E)$. The attractiveness of item a_k at time t, $w_t(a_k^t)$, is a realization of an independent Bernoulli random variable with mean $\langle \Delta(a_k^t \mid \{a_1^t, \ldots, a_{k-1}^t\}), \theta^* \rangle$. The termination at the k-th position at time t, $v_t(k)$, is a realization of an independent Bernoulli random variable with mean ζ. The user examines the list from the first item a_1^t to the last a_K^t and clicks on all attractive items till the user is satisfied. i.e. the realization $v_t(k)$ is one after a click. The *feedback* is the sequence of clicks and no-clicks till the index where the user gets satisfied, $\{C_{1_t}, \ldots, C_{T_t}\}$, where $T_t = \min\{k \in [K] : w_t(a_k^t) = 1 \text{ and } v_t(k) = 1\}$. We assume that $\min \emptyset = \infty$. That is, if the user clicks on an item and decides to terminate, then $T_t \leq K$; and if the user does not click or decide not to terminate on any item, then $T_t = \infty$. We say that item e is *examined* at time t if $e = a_k^t$ for some $k \in [\min\{T_t, K\}]$. Note that the attractiveness of all examined items at time t can be computed from $\{C_{1_t}, \ldots C_{T_t}\}$. In particular, $w_t(a_k^t) = \mathbb{1}\{C_{k_t} : k \in [min\{T_t, K\}]\}$. The *reward* is defined as $r_t = \mathbb{1}\{T_t \leq K\}$. That is, the reward is one if the user is satisfied (clicks and decides to terminate) by at least one item in A_t; and zero otherwise.

The goal of the learning agent is to maximize its expected cumulative reward. This is equivalent to minimizing the expected cumulative regret with respect to the *optimal list* in (27). The regret is defined analogously to (9), i.e.,

$$R^{\gamma_{DCM}}(n) = \sum_{t=1}^{n} \mathbb{E}\left[f_{DCM}(A^*_{DCM}, \theta^*, \zeta) - f_{DCM}(A_t, \theta^*, \zeta)/\gamma_{DCM}\right]. \tag{29}$$

C.3 Algorithm `dcmLSB`

Our algorithm for solving dependent click linear submodular bandit is provided in Algorithm 2. We call it `dcmLSB`. `dcmLSB` is almost same as `CascadeLSB`. The only difference is that the *feedback* is a sequence of clicks and no-clicks till the index where the user gets satisfied, $\{C_{1_t}, \ldots, C_{T_t}\}$ with $T_t = \min\{k \in [K] : w_t(a_k^t) = 1 \text{ and } v_t(k) = 1\}$, instead of just the index of the first click $C_t = \min\{k \in [K] : w_t(a_k^t) = 1\}$ as in `CascadeLSB`.

C.4 Experiments

We compare `dcmLSB` with `LSBGreedy`, which assumes feedback on the entire recommended list even after the user is satisfied and terminates scanning the list. We discuss two experiments in this section – first, on a simulated setting, and second, on the MovieLens dataset. We work with topic coverage and the parameters as described in Section 8.4 and use $\gamma_{DCM} = 1$ during regret computation in (29) .

C.4.1 Synthetic Experiments

This experiment illustrates the need for modeling both diversity and position bias via partial-click feedback even when one considers multiple clicks on the recommended list. The setting is an extension to the synthetic experiment in Section 8.3, providing more room for multiple clicks through larger d, K, and L.

We simulate a problem with $L = 61$ items and $d = 5$ topics. We recommend $K = 8$ items and simulate a single user whose preferences are $\theta^* = (0.300, 0.275, 0.225, 0.200, 0)$. The attractiveness of items $1, 2$, and 3 in topic 1 is 0.5, and 0 in all other topics. The attractiveness of items $4, 5$, and 6 in topic 2 is 0.5, and 0 in all other topics. The attractiveness of items $7, 8$, and 9 in topic 3 is 0.5, and 0 in all other topics. The attractiveness of items 10 and 11 in topic 4 is 0.5, and 0 in all other topics. The attractiveness of remaining 50 items in topic 5 is 1, and 0 in other topics.

The optimal recommended list is $A^* = (1, 4, 7, 10, 2, 5, 8, 11)$ in that order. The n-step regret of both the algorithms is shown in Figure 6(a). We observe that the regret of `dcmLSB` flattens and does not increase with the number of steps

Algorithm 2 `dcmLSB`

Inputs: Parameters $\sigma > 0$ and $\alpha > 0$ (Section 6)
$M_0 \leftarrow I_d,\ B_0 \leftarrow \mathbf{0}$ ▷ Initialization
for $t = 1, \ldots, n$ **do**
  $\bar{\theta}_{t-1} \leftarrow \sigma^{-2} M_{t-1}^{-1} B_{t-1}$ ▷ Regression estimate
  $S \leftarrow \emptyset$ ▷ Recommend list and receive feedback
  for $k = 1, \ldots, K$ **do**
    for all $e \in E \setminus S$ **do**
      $x_e \leftarrow \Delta(e \mid S)$
    end for
    $a_k^t \leftarrow \arg\max_{e \in E \setminus S} \left[x_e^\intercal \bar{\theta}_{t-1} + \alpha \sqrt{x_e^\intercal M_{t-1}^{-1} x_e} \right]$
    $S \leftarrow S \cup \{a_k^t\}$
  end for
  Recommend list $A_t \leftarrow (a_1^t, \ldots, a_K^t)$
  Observe click sequence $\{C_{k_t} \in \{0,1\} \forall\, k \in [min\{T_t, K\}]$.
  $M_t \leftarrow M_{t-1},\ B_t \leftarrow B_{t-1}$ ▷ Update statistics
  for $k = 1, \ldots, \min\{T_t, K\}$ **do**
    $x_e \leftarrow \Delta\left(a_k^t \mid \{a_1^t, \ldots, a_{k-1}^t\}\right)$
    $M_t \leftarrow M_t + \sigma^{-2} x_e x_e^\intercal,\ B_t \leftarrow B_t + x_e C_{k_t}$
  end for
end for

n. This means that `dcmLSB` learns the optimal solution. The regret of `LSBGreedy` grows linearly with the number of steps n, which means `LSBGreedy` does not learn the optimal solution. When `LSBGreedy` recommends A^*, it severely underestimates the preference for topic 3 or topic 4, because it assumes feedback at the lower positions even if the first few positions are clicked and the user becomes satisfied in between the list. Because of this, `LSBGreedy` switches to recommending other suboptimal items at some point in time. After some time, `LSBGreedy` swiches back to recommending item A^*, and then it oscillates between optimal and suboptimal items similar to the experiment in Section 8.3. Therefore, `LSBGreedy` has a linear regret and performs poorly.

C.4.2 Experiments on MovieLens

We experimented the Movielens dataset on a similar setting as described in Section 8.5 to compare `dcmLSB` and `LSBGreedy`. We work with $d = 10$ and $K = 8$. The regret is shown in Figure 6(b). We observe that the regret of `dcmLSB` is sublinear and much lower than `LSBGreedy`. This shows that the current diversity-driven online approaches are insufficient to handle cases with partial-feedback albeit in the form of multiple-clicks.

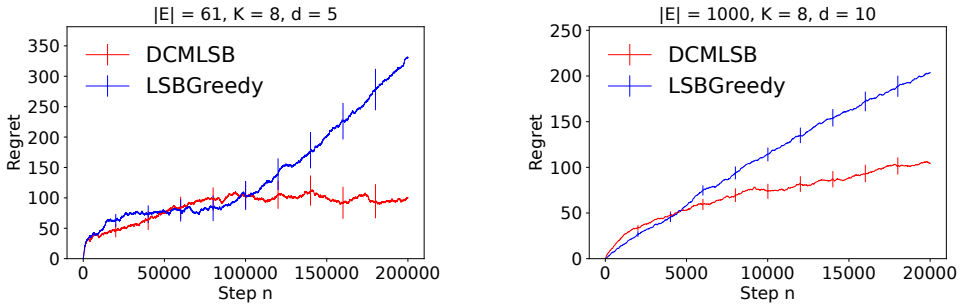

Figure 6: Regret on synthetic (left) and Movielens (right) dataset in the special case of DCM model.