Appendix

I. Theoretical Analysis

In this section, we provide a theoretical analysis of the regret bond using our CCBM approach. The upper bound is derived under the principle that arms belonging to similar context space should have similar expected reward values.

Assumption 1. *(Lipschitz-continuous) There exists $C > 0$ such that for any arm a, a' with arm context $O_a, O_{a'} \in \mathcal{X}$, we have $|r_a - r_{a'}| \leq L||O_a - O_{a'}||_1$.*

Assumption 2. *(Bounded Reward) The reward of each arm is bounded by $0 < r < r^{max}$.*

We set the $h_T = \lceil T^{\frac{1}{4}} \rceil$ for the arm context partition and $K(n_x) = n_x^{\frac{1}{2}} log(n_x)$ as the control function to identify the under-explored arm hypercubes. Then, the regret can be bounded as follow:

Theorem 1. *Let $h_T = \lceil T^{\frac{1}{4}} \rceil$ and $K(n_x) = n_x^{\frac{1}{2}} log(n_x)$, if Assumptions 1 and 2 hold true, the regret $R(T)$ is bounded by:*

$R(T) \leq (1-\frac{1}{e})Br^{max}2M(M^{\frac{1}{2}}log(MT)T^{\frac{3}{4}} + T^{\frac{1}{4}}) + (1-\frac{1}{e})Br^{max}M\begin{pmatrix} |\mathcal{A}_m^t| \\ B \end{pmatrix}\frac{\pi^2}{3} + (3BL + \frac{8}{3}B(r^{max}+L))MT^{\frac{3}{4}}$.

Proof. The regret $R(T)$ can be divided into the following summands:

$$\mathbb{E}[R(T)] = \mathbb{E}[R_{explore}(T)] + \mathbb{E}[R_{exploit}(T)],$$

where the term $\mathbb{E}[R_{explore}(T)]$ is the regret due to the exploration process, and the term $\mathbb{E}[R_{exploit}(T)]$ corresponds to the regret in the exploitation phase. We first derive a bound on $\mathbb{E}[R_{explore}(T)]$. According to Algorithm 2, the set of under-explored hypercubes $P_T^{\text{ue},t}$ is non-empty during the exploration phase, which implies that there exists at least one hypercube p with $C^t(p|x) \leq K(n_x) = n_x^{\frac{1}{2}}\log(n_x)$. Because we only explore in the first t_τ rounds, $n_x < Mt_\tau < MT$ holds. Certainly, there can be a maximum of $\lceil (MT)^{\frac{1}{2}}\log(MT)\rceil$ exploration phases in which p is under-explored. Given h_T hypercubes in the partition and a total of M users, the upper limit for exploration phases is $h_T M\lceil (MT)^{\frac{1}{2}}\log(MT)\rceil$. Owing to the submodularity of reward function and its bounded nature, the maximum regret for an incorrect selection in one exploration phase is constrained by $(1-1/e)Br^{\max}$. Therefore, we have

$$\begin{aligned}\mathbb{E}[R_{explore}(T)] &\leq (1-\frac{1}{e})Br^{\max}h_T M\lceil (MT)^{\frac{1}{2}}\log(TM)\rceil \\ &= (1-\frac{1}{e})Br^{\max}M\lceil T^{\frac{1}{4}}\rceil\lceil (MT)^{\frac{1}{2}}\log(TM)\rceil.\end{aligned}$$

Given $\lceil T^{\frac{1}{4}}\rceil \leq 2T^{\frac{1}{4}}$, we can further bound the maximum regret as:

$$\mathbb{E}[R_{explore}(T)] \leq (1-\frac{1}{e})Br^{\max}2M(M^{\frac{1}{2}}T^{\frac{3}{4}}\log(MT) + T^{\frac{1}{4}}).$$

Before deriving the bound on $\mathbb{E}[R_{exploit}(T)]$, we first define some auxiliary functions. For each hypercube p, we define $\bar{\mu}(p) = \sup_{O\in p}\mu(O)$ and $\underline{\mu}(p) = \inf_{O\in p}\mu(O)$ be the best and worst expected quality over all contexts $O \in p$. Also, the context at center of a hypercube p is defined as $\hat{O}_p$ and its expected quality $\hat{\mu}(p) = \mu(\hat{O}_p)$. Let $\bar{\mu}_p^t = [\bar{\mu}(p_1^t), \ldots, \bar{\mu}(p_{h_T}^t)]$, $\underline{\mu}_p^t = [\underline{\mu}(p_1^t), \ldots, \underline{\mu}(p_{h_T}^t)], \tilde{\mu}_p^t = [\tilde{\mu}(p_1^t), \ldots, \tilde{\mu}(p_{h_T}^t)]$, and define $\tilde{S}^{*,t}(p^t)$

$$\tilde{S}^{*,t}(p^t) = \underset{S\subseteq\mathcal{A}_m^t, |S|\leq B}{\arg\max} R(S, \tilde{\mu}_p^t)$$

Let $\tilde{S}^{*,t}(p^t)$ be the optimal set and $\tilde{S}^t(p^t)$ be the set that is chosen by Algorithm 1 and we will have $R(\tilde{S}^t(p^t), \tilde{\mu}_p^t) \geq (1-1/e)\cdot R(S^{*,t}(p^t), \tilde{\mu}_p^t)$. The arm set $\tilde{S}^t(p^t)$ is used to identify subsets of arms which are sub-optimal. Let

$$\mathcal{L}^t(p^t) = \{G \subseteq \mathcal{A}_m^t, |G| = B : R(\tilde{S}^t(p^t), \underline{\mu}_p^t) - R(G, \bar{\mu}_p^t) \geq An_x^\theta\}$$

be the set of suboptimal subsets of arms for hypercubes p^t, where $A > 0$ and $\theta < 0$. We call a subset G of arms in $\mathcal{L}^t(p^t)$ suboptimal for p^t, as the sum of the worst expected reward in $\tilde{S}^t(p^t)$ is at least an amount An_x^θ higher than the sum of the best expected reward for subset G. We call subsets in $S_B^t\backslash\mathcal{L}^t(p^t)$ near-optimal for p^t. Here, S_B^t denotes the set of all B-element subsets of arm set M^t. Then, $\mathbb{E}[R_{exploit}(T)]$ can be divided into

$$\mathbb{E}[R_{exploit}(T)] = \mathbb{E}[R_s(T)] + \mathbb{E}[R_n(T)]$$

where $\mathbb{E}[R_s(T)]$ is the regret due to suboptimal choices, i.e., the subsets of arms from $\mathcal{L}^t(p^t)$ are selected; $\mathbb{E}[R_n(T)]$ is the regret due to near-optimal choices, i.e., the subsets of arms from $S_B^t\backslash\mathcal{L}^t(p^t)$ are selected. In the following, we prove the bound of each term. We first give the bound for $\mathbb{E}[R_s(T)]$.

For $1 \leq t \leq T$, let $W^t = \{\mathcal{P}^{ue,t} = \emptyset\}$ be the event that slot t is an exploitation phase. By the definition of $\mathcal{P}^{ue,t}$, in this case, it holds that $C^t(p_m^t) > K(n_x) = n_x^{\frac{1}{2}}\log(n_x), \forall p \in p^t$. Let V_G^t be the event that subset $G \in \mathcal{L}^t(p^t)$ is selected at time slot t. Then, it holds that

$$\begin{aligned}R_s(T) = \sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{G\in\mathcal{L}^t(p^t)} I_{\{V_G^t, W^t\}}\times \\ \left((1-\frac{1}{e})R(S^{*,t}(x^t), r^t) - R(G, r^t)\right)\end{aligned}$$

where, in each time slot, the loss due to selecting a suboptimal subset $G \in \mathcal{L}^t(p^t)$ is considered. Since the maximum regret of selecting G is bounded by $(1-\frac{1}{e})Br^{\max}$, we have

$$R_s(T) \leq (1-\frac{1}{e})Br^{\max}\sum_{t=1}^{T}\sum_{G\in\mathcal{L}^t(p^t)} I_{\{V_G^t,W^t\}}$$

and taking the exception, the regret is hence bounded by

$$\mathbb{E}[R_s(T)] \leq (1-\frac{1}{e})Br^{\max}\sum_{t=1}^{T}\sum_{G\in\mathcal{L}^t(p^t)} \mathbb{E}[I_{\{V_G^t,W^t\}}]$$

$$= (1-\frac{1}{e})Br^{\max}\sum_{t=1}^{T}\sum_{G\in\mathcal{L}^t(p^t)} \text{Prob}\{V_G^t,W^t\}$$

In the event of V_G^t, by the design of the algorithm, this means that with the estimated arm quality, the rewards of selecting arms in G is at least as high as the reward of selecting arms in $\tilde{S}^t(p^t)$, i.e., $R(G,\hat{r}_p^t) \geq R(\tilde{S}^t(p^t)),\hat{r}_p^t$. Thus, we have:

$$\text{Prob}\{V_G^t,W^t\} \leq \text{Prob}\left\{R(G,\hat{r}_p^t) \geq R(\tilde{S}^t(p^t),\hat{r}_p^t)\right\}$$

The event in the right-hand side implies at lease one of the three following events for any $H(n_x) > 0$:

$$E_1 = \{R(G,\hat{r}_p^t) \geq R(G,\bar{\mu}_p^t,) + H(n_x), W^t\}$$

$$E_2 = \{R(\hat{S}^t(p^t),\hat{r}_p^t) \leq R(\hat{S}^t(p^t),\underline{\mu}_p^t) - H(n_x), W^t\}$$

Hence, we have for the original event in (19)

$$\left\{R(G,\hat{r}_p^t) \geq R(\hat{S}^t(p^t),\hat{r}_p^t)\right\} \subseteq E_1 \cup E_2$$

The probability of the three event E_1, E_2 will be bounded separately. Let start by bounding E_1. Recall that the best expected quality of arms in set p is $\bar{\mu}(p) = \sup_{O\in p}\mu(O)$. Therefore, the expected quality of arm m in G is bounded by

$$\begin{aligned}
E[\hat{r}(p_m^t)] &= \mathbb{E}\left[\frac{1}{|\mathcal{E}^t(p_m^t)|}\sum_{(\tau,k):O_k^T\in p_m^T, n\in\mathcal{S}^T} r\left(O_k^\tau\right)\right] \\
&= \frac{1}{|\mathcal{E}^t(p_m^t)|}\underbrace{\sum_{(\tau,k):O_k^T\in p_m^T, n\in\mathcal{S}^T}}_{|\mathcal{E}^t(p_m^t)|summands}\underbrace{\mu\left(O_k^\tau\right)}_{\leq\bar{\mu}(p_m^t)} \\
&\leq \bar{\mu}(p_m^t)
\end{aligned}$$

This implies

$$\begin{aligned}
\text{Prob}\{E_1\} &= \text{Prob}\{R(G,\hat{r}_p^t) \geq R(G,\bar{\mu}_p^t) + H(n_x), W^t\} \\
&\leq \text{Prob}\{\hat{r}(p_m^t) \geq \bar{\mu}(p_m^t) + \frac{H(n_x)}{B}, \exists m\in G, W^t\} \\
&\leq \text{Prob}\{\hat{r}(p_m^t) \geq \mathbb{E}[\hat{r}(p_m^t)] + \frac{H(n_x)}{B}, \exists m\in G, W^t\} \\
&= \sum_{m\in G}\text{Prob}\{\hat{r}(p_m^t) \geq \mathbb{E}[\hat{r}(p_m^t)] + \frac{H(n_x)}{B}, W^t\}
\end{aligned}$$

The first inequality comes from the fact that $\{R(G,\tilde{r}_p^t) \geq R(G,\bar{\mu}_p^t) + H(n_x)\} \subseteq \{\hat{r}(p_m^t) \geq \bar{\mu}(p_m^t) + \frac{H(n_x)}{B}, \exists m\in G\}$, which can be easily verified by *reductio ad absurdum* and submodularity property. Now, applying Chernoff-Hoeffding bound (note that for each arm, the estimated quality is bounded by $r^{\max}$ and exploiting that event W^t implies that at least $n_x^{\frac{1}{2}}\log(n_x)$ samples were drawn, we get

$$\begin{aligned}
\text{Prob}\{E_1\} &\leq \sum_{m\in G}\text{Prob}\{\hat{r}(p_m^t) - \mathbb{E}[\hat{r}(p_m^t)] \geq \frac{H(n_x)}{B}, W^t\} \\
&\leq \sum_{m\in G}\exp\left(\frac{-2|\mathcal{E}^t(p_m^t)|H(n_x)^2}{B^2(r^{\max})^2}\right) \\
&\leq \sum_{m\in G}\exp\left(\frac{-2H(n_x)^2 n_x^{\frac{1}{2}}\log(n_x)}{B^2(r^{\max})^2}\right)
\end{aligned}$$

Analogously, it can be proven for event E_2, that

$$\begin{aligned}
\text{Prob}\{E_2\} &= \text{Prob}\{R(\tilde{S}^t(p^t),\hat{r}_p^t) \geq R(\tilde{S}^t(p^t),\underline{\mu}_p^t) - H(n_x), W^t\} \\
&\leq \sum_{m\in S^t(p^t)}\exp\left(\frac{-2H(n_x)^2 n_x^{\frac{1}{2}}\log(n_x)}{B^2(r^{\max})^2}\right)
\end{aligned}$$

So far, the analysis was performed with respected to an arbitrary $H(n_x) > 0$. In the remainder of the proof, we choose $H(n_x) = Br^{\max}n_x^{-\frac{1}{4}}\sqrt{1-\frac{log(B)}{2log(n_x)}}$. Then, we have

$$\begin{aligned}
&\text{Prob}\{E_1\} \leq B\exp\left(-\frac{2H(n_x)^2 n_x^{\frac{1}{2}}\log(n_x)}{B^2(r^{\max})^2}\right) \leq B\exp(-2\log(n_x)) \\
&\leq Bn_x^{-2} \leq Bt^{-2}
\end{aligned}$$

and analogously

$$\text{Prob}\{E_2\} \leq Bn_x^{-2} \leq Bt^{-2} \tag{1}$$

To sum up,

$$\begin{aligned}
\text{Prob}\{V_G^t,W^t\} &\leq \text{Prob}\{E_1\cup E_2\} \\
&\leq \text{Prob}\{E_1\} + \text{Prob}\{E_2\} \leq 2Bt^{-2}
\end{aligned}$$

Given this we have:

$$\begin{aligned}
E[R_s(T)] &\leq \left(1-\frac{1}{e}\right)Br^{\max}\times\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{G\in\mathcal{L}(p^t)}\text{Prob}\{V_G^t,W^t\} \\
&\leq \left(1-\frac{1}{e}\right)Br^{\max}\begin{pmatrix}|\mathcal{A}_m^t| \\ B\end{pmatrix}M\sum_{t=1}^{T}2Bt^{-2} \\
&\leq \left(1-\frac{1}{e}\right)B^2(r^{\max})\begin{pmatrix}|\mathcal{A}_m^t| \\ B\end{pmatrix}\cdot 2M\sum_{t=1}^{\infty}t^{-2} \\
&\leq \left(1-\frac{1}{e}\right)B^2(r^{\max})M\begin{pmatrix}|\mathcal{A}_m^t| \\ B\end{pmatrix}\frac{\pi^2}{3}
\end{aligned}$$

where $\begin{pmatrix}|\mathcal{A}_m^t| \\ B\end{pmatrix}$ is maximum possible number of subsets of size B in each time slot.

Now we give a bound for $E[R_n(T)]$.

For $1 \leq t \leq T$, consider the event W^t as in the previous proof, the regret due to near-optimal subsets can be written as

$$R_n(T) = \sum_{t=1}^{T} I_{\{W^t, S^t \in S_b \backslash \mathcal{L}^t(p^t)\}} \times \left(\left(1 - \frac{1}{e}\right) \cdot R(S^{*,t}(x^t), r^t) - R(S^t, r^t) \right)$$

where in each time slot in which the selected subset S^t is near-optimal, i.e., $S^t \in S_b \backslash \mathcal{L}^t(p^t)$, the regret is considered for selecting S^t instead of the $S^{*,t}(x^t)$. Let $Q_t = W^t \cap \{S^t \in S_b \backslash \mathcal{L}^t(p^t)\}$ denotes the event of selecting a near-optimal arm set. Then, we have

$$E[R_n(T)] = \sum_{t=1}^{T} E[I_{\{Q_t\}} \times \left((1 - \frac{1}{e}) \cdot R(S^{*,t}(x^t), r^t) - R(S^t, r^t) \right)]$$

By the definition of conditional expectation, this is equivalent to

$$\begin{aligned} & E[R_n(T)] \\ &= \sum_{t=1}^{T} \text{Prob}\{Q(t)\} \cdot \\ & E\left[\left(1 - \frac{1}{e}\right) \cdot R(S^{*,t}(x^t), r^t) - R(S^t, r^t) \middle| Q(t) \right] \\ &\leq \sum_{t=1}^{T} E\left[\left(1 - \frac{1}{e}\right) \cdot R(S^{*,t}(x^t), r^t) - R(S^t, r^t) \middle| Q(t) \right] \end{aligned}$$

Now, let t be the time slot, where $Q(t)$ holds true, i.e., the algorithm enters an exploitation phase and $J \in S_b \backslash \mathcal{L}^t(p^t)$. By the definition of $\mathcal{P}^{ue,t}$, it holds that $C^t(p_m^t) > K(n_x) = n_x^{\frac{1}{2}} \log(n_x)$ for all $p_m^t \in p^t$. In addition, since $J \in S_b \backslash \mathcal{L}^t(p^t)$, it holds

$$R(\tilde{S}^t(p^t), \underline{\mu}_p^t) - R(J, \bar{\mu}_p^t) < A n_x^{\theta}$$

To bound the regret, we have to give an upper bound on

$$\sum_{t=1}^{T} \mathbb{E}\left[\left(1 - \frac{1}{e}\right) \cdot R(S^{*,t}(x^t), r^t) - R(J, r^t) \middle| Q(t) \right]$$

$$= \sum_{t=1}^{T} \left((1 - \frac{1}{e}) \cdot R(S^{*,t}(x^t), \mu^t) - R(J, \mu^t) \right)$$

Applying Lipschitz-continuous condition several times yields:

$$\begin{aligned} & \left(1 - \frac{1}{e}\right) \cdot R(S^{*,t}(x^t), \mu^t) - R(J, \mu^t) \\ &\leq \left(1 - \frac{1}{e}\right) \cdot R(S^{*,t}(x^t), \tilde{\mu}_p^t) + BLh_T^{-1} - R(J, \mu_x^t) \\ &\leq \left(1 - \frac{1}{e}\right) \cdot R(\tilde{\mu}_p^t, S^{*,t}(p^t)) + BLh_T^{-1} - R(J, \mu_x^t) \\ &\leq R(S^t(\tilde{p^t}), \underset{p}{\mu^t}) + BLh_T^{-1} - R(J, \mu_x^t) \\ &\leq R(S^t(\tilde{p^t}), \underset{p}{\mu^t}) + 2BLh_T^{-1} - R(J, \mu_x^t) \\ &\leq R(S^t(\tilde{p^t}), \underset{p}{\mu^t}) + 3BLh_T^{-1} - R(J, \mu_x^t) \\ &\leq 3BLh_T^{-1} + An_x^{\theta} \leq 3BLh_T^{-1} + At^{\theta} \end{aligned}$$

where the third inequality follows the definition of $\tilde{S}^{*,t}(p^t)$ and $\tilde{S}^t(p^t)$. Using $h_T^{-1} = \lceil T^{\frac{1}{4}} \rceil^{-1} \leq T^{-\frac{1}{4}}$, we further have

$$\mathbb{E}\left[R(S^{*,t}(x^t), r^t) - R(J, r^t) | Q(t) \right] \leq 3BLT^{-\frac{1}{4}} + At^{\theta}$$

Therefore, the regret can be bounded by

$$\begin{aligned} \mathbb{E}\left[R_n(T)\right] &\leq \sum_{t=1}^{T} \left(3BLT^{-\frac{1}{4}} + At^{\theta} \right) \\ &\leq 3BLT^{\frac{3}{4}} + \frac{A}{1+\theta} T^{1+\theta}. \end{aligned}$$

Combining the above results, the regret $R(T)$ is bounded by

$$\begin{aligned} R(T) \leq (1 - \frac{1}{e}) Br^{\max} 2M(M^{\frac{1}{2}} T^{\frac{3}{4}} \log(MT) + T^{\frac{1}{4}}) + \\ \left(1 - \frac{1}{e}\right) B^2 r^{\max} M \left(\begin{array}{c} |\mathcal{A}_m^t| \\ B \end{array} \right) \frac{\pi^2}{3} \\ + 3BLT^{\frac{3}{4}} + \frac{A}{1+\theta} T^{1+\theta} \end{aligned}$$

In order to balance the leading orders, we select the parameters z, γ, A, θ as following values $z = \frac{1}{2}, \gamma = \frac{1}{4}, \theta = -\frac{1}{4}$, and $A = 2Br^{\max} + 2BL$. The regret $R(T)$ reduces to

$$\begin{aligned} R(T) \leq & (1 - \frac{1}{e}) Br^{\max} 2M(M^{\frac{1}{2}} T^{\frac{3}{4}} \log(MT) + T^{\frac{1}{4}}) \\ & + \left(1 - \frac{1}{e}\right) \cdot MB^2 r^{\max} \left(\frac{M^{\max}}{B} \right) \frac{\pi^2}{3} \\ & + \left(3BL + \frac{8}{3} B(r^{\max} + L) \right) T^{\frac{3}{4}} \end{aligned}$$

□

In summary, the leading order of the cumulative regret is $O(T^{\frac{3}{4}} log(T))$, indicating a sublinear growth over the time horizon T. This implies that our CCBM scheme exhibits asymptotic optimality and converges toward optimal strategy.