Online Appendix for Context-Adaptive Big Data Stream Mining

Cem Tekin, Luca Canzian, Mihaela van der Schaar
Electrical Engineering Department, University of California, Los Angeles
Email: cmtkn@ucla.edu, luca.canzian@gmail.com, mihaela@ee.ucla.edu

Abstract

This online appendix includes the proofs that are omitted from the submission due to space limitations.

We start with a simple lemma which gives an upper bound on the highest level hypercube that is active at any time t.

Lemma 1: **A bound on the level of active hypercubes.** All the active hypercubes $\mathcal{A}_i^d(t)$ for type-d contexts at time t have at most a level of $(\log_2 t)/p + 1$.

Proof: Let $l+1$ be the level of the highest level active hypercube. We must have $A\sum_{j=0}^{l} 2^{pj} < t$, otherwise the highest level active hypercube will be less than $l+1$. We have for $t/A > 1$, $A\frac{2^{p(l+1)}-1}{2^p-1} < t \Rightarrow 2^{pl} < \frac{t}{A} \Rightarrow l < \frac{\log_2 t}{p}$. ■

A. Proof of Lemma 1

This directly follows from the number of trainings and explorations that are required before any arm can be exploited (see definition of $S^i_{\boldsymbol{C}_i(t)}(t)$). If the prediction at any training or exploration step is incorrect or a high cost arm is chosen, learner i loses at most 2 from the highest realized reward it could get at that time slot, due to the fact an incorrect prediction will result in one unit of loss and the cost of an action can at most be one.

B. Proof of Lemma 2

Let Ω denote the space of all possible outcomes, and w be a sample path. The event that the ACAP exploits when $\boldsymbol{x}_i(t) \in \boldsymbol{C}$ is given by $\mathcal{W}^i_{\boldsymbol{C}}(t) := \{w : S^i_{\boldsymbol{C}}(t) = \emptyset, \boldsymbol{x}_i(t) \in \boldsymbol{C}, \boldsymbol{C} \in \mathcal{A}_i(t)\}$. We will bound the probability that ACAP chooses a suboptimal arm for learner i in an exploitation step when i's context vector is in the set of active hypercubes $\boldsymbol{C}$ for any $\boldsymbol{C}$, and then use this to bound the expected number of times a suboptimal arm is chosen by learner i in exploitation steps using ACAP. Recall that reward loss in every step in which a suboptimal arm is chosen can be at most 2.

Let $\mathcal{V}^i_{k,\boldsymbol{C}}(t)$ be the event that a suboptimal arm k is chosen for the set of hypercubes $\boldsymbol{C}$ by learner i at time t. For $k \in \mathcal{K}_i \cap \mathcal{F}_i$, let $\mathcal{E}^i_{k,\boldsymbol{C}}(t)$ be the set of rewards collected by learner i from arm k in time slots when the context vector of learner i is in the active set $\boldsymbol{C}$ by time t. For $j_i \in \mathcal{K}_i \cap \mathcal{M}_{-i}$, let $\mathcal{E}^i_{j_i,\boldsymbol{C}}(t)$ be the set of rewards collected from selections of learner j_i in time slots $t' \in \{1,\dots,t\}$ for which $N^i_{1,j_i,l}(t') > D_2(t')$ and the context

vector of learner i is in the active set $\boldsymbol{C}$ by time t. Let $\mathcal{B}^i_{j_i,\boldsymbol{C}}(t)$ be the event that at most t^ϕ samples in $\mathcal{E}^i_{j_i,\boldsymbol{C}}(t)$ are collected from suboptimal arms of learner j_i. For $k \in \mathcal{K}_i \cap \mathcal{F}_i$ let $\mathcal{B}^i_{k,\boldsymbol{C}}(t) := \Omega$. In order to facilitate our analysis of the regret, we generate two different artificial independent and identically distributed (i.i.d.) processes to bound the probabilities related to deviation of sample mean reward estimates $\bar{r}^{i,d}_{k,C^d}(t)$, $k \in \mathcal{K}_i$, $d \in \mathcal{D}$ from the expected rewards, which will be used to bound the probability of choosing a suboptimal arm. The first one is the *best* process in which rewards are generated according to a bounded i.i.d. process with expected reward $\overline{\mu}^d_{k,C^d}$, the other one is the *worst* process in which the rewards are generated according to a bounded i.i.d. process with expected reward $\underline{\mu}^d_{k,C^d}$. Let $\bar{r}^{b,i,d}_{k,C^d}(t)$ denote the sample mean of the t samples from the best process and $\bar{r}^{w,i,d}_{k,C^d}(t)$ denote the sample mean of the t samples from the worst process. We have for any $k \in \mathcal{L}^i_{\boldsymbol{C},B}$

$$
\begin{aligned}
&P\left(\mathcal{V}^i_{k,\boldsymbol{C}}(t), \mathcal{W}^i_{\boldsymbol{C}}(t)\right) \\
&\leq P\left(\max_{d\in\mathcal{D}} \bar{r}^{b,i,d}_{k,C^d}(N^{i,d}_{k,C^d}(t)) \geq \overline{\mu}_{k,\boldsymbol{C}} + H_t, \mathcal{W}^i_{\boldsymbol{C}}(t)\right) \\
&+ P\left(\max_{d\in\mathcal{D}} \bar{r}^{b,i,d}_{k,C^d}(N^{i,d}_{k,C^d}(t)) \geq \bar{r}^{w,i,d^*(\boldsymbol{C})}_{k^*(\boldsymbol{C}),C^{d^*(\boldsymbol{C})}}(N^{i,d^*(\boldsymbol{C})}_{k^*(\boldsymbol{C}),C^{d^*(\boldsymbol{C})}}(t))\right. \\
&-2t^{\phi-1}, \max_{d\in\mathcal{D}} \bar{r}^{b,i,d}_{k,C^d}(N^{i,d}_{k,C^d}(t)) < \overline{\mu}_{k,\boldsymbol{C}} + L2^{-l_{\max}(\boldsymbol{C})\alpha} \\
&+H_t + 2t^{\phi-1}, \bar{r}^{w,i,d^*(\boldsymbol{C})}_{k^*(\boldsymbol{C}),C^{d^*(\boldsymbol{C})}}(N^{i,d^*(\boldsymbol{C})}_{k^*(\boldsymbol{C}),C^{d^*(\boldsymbol{C})}}(t)) \\
&\left.> \underline{\mu}_{k^*(\boldsymbol{C}),\boldsymbol{C}} - L2^{-l_{\max}(\boldsymbol{C})\alpha} - H_t, \mathcal{W}^i_{\boldsymbol{C}}(t)\right) \qquad (1)
\end{aligned}
$$

$$
\begin{aligned}
&+ P\left(\bar{r}^{w,i,d^*(\boldsymbol{C})}_{k^*(\boldsymbol{C}),C^{d^*(\boldsymbol{C})}}(N^{i,d^*(\boldsymbol{C})}_{k^*(\boldsymbol{C}),C^{d^*(\boldsymbol{C})}}(t)) \leq \underline{\mu}_{k^*(\boldsymbol{C}),\boldsymbol{C}} - H_t\right. \\
&\left.+2t^{\phi-1}, \mathcal{W}^i_{\boldsymbol{C}}(t)\right) + P((\mathcal{B}^i_{k,\boldsymbol{C}}(t))^c),
\end{aligned}
$$

where $H_t > 0$. In order to make the probability in (1) equal to 0, we need

$$
4t^{\phi-1} + 2H_t \leq (B-2)L2^{-l_{\max}(\boldsymbol{C})\alpha}. \qquad (2)
$$

By Lemma 1, (2) holds when

$$
4t^{\phi-1} + 2H_t \leq (B-2)L2^{-\alpha}t^{-\alpha/p}. \qquad (3)
$$

For $H_t = 4t^{\phi-1}$, $\phi = 1 - z/2$, $z \geq 2\alpha/p$ and $B = 12/(L2^{-\alpha}) + 2$, (3) holds by which (1) is equal to zero. Also by using a Chernoff-Hoeffding bound we can show that

$$
P\left(\max_{d\in\mathcal{D}} \bar{r}^{b,i,d}_{k,C^d}(N^{i,d}_{k,C^d}(t)) \geq \overline{\mu}_{k,\boldsymbol{C}} + H_t, \mathcal{W}^i_{\boldsymbol{C}}(t)\right) \leq D/t^2,
$$

and

$$
\begin{aligned}
&+ P\left(\bar{r}^{w,i,d^*(\boldsymbol{C})}_{k^*(\boldsymbol{C}),C^{d^*(\boldsymbol{C})}}(N^{i,d^*(\boldsymbol{C})}_{k^*(\boldsymbol{C}),C^{d^*(\boldsymbol{C})}}(t)) \leq \underline{\mu}_{k^*(\boldsymbol{C}),\boldsymbol{C}} - H_t\right. \\
&\left.+2t^{\phi-1}, \mathcal{W}^i_{\boldsymbol{C}}(t)\right) \leq 1/t^2.
\end{aligned}
$$

We also have $P(\mathcal{B}^i_{k,\boldsymbol{C}}(t)^c) = 0$ for $k \in \mathcal{F}_i$ and $P(\mathcal{B}^i_{j_i,\boldsymbol{C}}(t)^c) \leq E[X^i_{j_i,\boldsymbol{C}}(t)]/t^\phi \leq 2F_{\max}\beta_2 t^{z/2-1}$. for $j_i \in \mathcal{M}_{-i}$, where $X^i_{j_i,\boldsymbol{C}}(t)$ is the number of times a suboptimal arm of learner j_i is selected when learner i calls j_i in exploration and exploitation phases in time slots when the context vector of i is in the set of hypercubes $\boldsymbol{C}$ which are active by time t. Combining all of these we get $P\left(\mathcal{V}^i_{k_i,\boldsymbol{C}}(t), \mathcal{W}^i_{\boldsymbol{C}}(t)\right) \leq (1+D)/t^2$, for $k \in \mathcal{F}_i$ and $P\left(\mathcal{V}^i_{j_i,\boldsymbol{C}}(t), \mathcal{W}^i_{\boldsymbol{C}}(t)\right) \leq (1+D)/t^2 + 2F_{\max}\beta_2 t^{z/2-1}$, for $j_i \in \mathcal{M}_{-i}$. We get the final bound by summing these probabilities from $t=1$ to T.

C. Proof of Lemma 3

Let $X^i_{j_i,\boldsymbol{C}}(T)$ denote the random variable which is the number of times a suboptimal arm for learner $j_i \in \mathcal{M}_{-i}$ is chosen in exploitation steps of i when $\boldsymbol{x}_i(t')$ is in set $\boldsymbol{C} \in \mathcal{A}_i(t')$ for $t' \in \{1,\ldots,T\}$. It can be shown that $E[X^i_{j_i,\boldsymbol{C}}(T)] \leq 2F_{\max}\beta_2$. Thus, the contribution to the regret from suboptimal arms of j_i is bounded by $4F_{\max}\beta_2$. We get the final result by considering the regret from all $M-1$ other learners.

D. Proof of Lemma 4

The following lemma bounds the one-step regret to learner i from choosing near optimal arms. This lemma is used later to bound the total regret from near optimal arms.

Lemma 2: **One-step regret due to near-optimal arms for a set of hypercubes.** Let $\mathcal{L}^i_{\boldsymbol{C},B}$, $B = 12/(L2^{-\alpha})+2$ denote the set of suboptimal actions for set of hypercubes $\boldsymbol{C}$. When ACAP is run with parameters $p > 0$, $2\alpha/p \leq z < 1$, $D_1(t) = D_3(t) = t^z \log t$ and $D_2(t) = F_{\max}t^z \log t$, for any set of hypercubes $\boldsymbol{C}$, the one-step regret of learner i from choosing one of its near optimal classifiers is bounded above by $BL2^{-l_{\max}(\boldsymbol{C})\alpha}$, while the one-step regret of learner i from choosing a near optimal learner which chooses one of its near optimal classifiers is bounded above by $2BL2^{-l_{\max}(\boldsymbol{C})\alpha}$.

Proof: At time t if $\boldsymbol{x}_i(t) \in \boldsymbol{C} \in \mathcal{A}_i(t)$, the one-step regret of any near optimal arm of any near optimal learner $j_i \in \mathcal{M}_{-i}$ is bounded by $2BL2^{-l_{\max}(\boldsymbol{C})\alpha}$ by the definition of $\mathcal{L}^i_{\boldsymbol{C},B}$. Similarly, the one-step regret of any near optimal arm $k \in \mathcal{F}_i$ is bounded by $BL2^{-l_{\max}(\boldsymbol{C})\alpha}$. ■

At any time t for the set of active hypercubes $\boldsymbol{C}_i(t)$ that the context vector of i belongs to, $l_{\max}(\boldsymbol{C}_i(t))$ is at least the level of the active hypercube $x^d_i(t) \in C^d_i(t)$ for some type-d context. Since a near optimal arm's one-step regret at time t is upper bounded by $2BL2^{-l_{\max}(\boldsymbol{C}_i(t))\alpha}$, the total regret due to near optimal arms by time T is upper bounded by

$$2BL\sum_{t=1}^{T} 2^{-l_{\max}(\boldsymbol{C}_i(t))\alpha} \leq 2BL\sum_{t=1}^{T} 2^{-l(C^d_i(t))\alpha}.$$

Let $l_{\max,u}$ be the maximum level type-d hypercube when type-d contexts are uniformly distributed by time T. We must have

$$A\sum_{l=1}^{l_{\max,u}-1} 2^l 2^{pl} < T \tag{4}$$

otherwise the highest level hypercube by time T will be $l_{\max,u}-1$. Solving (4) for $l_{\max,u}$, we get $l_{\max,u} < 1+\log_2(T)/(1+p)$. $\sum_{t=1}^{T} 2^{-l(C^d_i(t))\alpha}$ takes its greatest value when type-d context arrivals by time T is uniformly

distributed in $\mathcal{X}_d$. Therefore we have

$$\sum_{t=1}^{T} 2^{-l(C_i^d(t))\alpha} \leq \sum_{l=0}^{l_{\max,u}} 2^l A 2^{pl} 2^{-\alpha l} < \frac{A 2^{2(1+p-\alpha)}}{2^{1+p-\alpha}-1} T^{\frac{1+p-\alpha}{1+p}}.$$

E. Proof of Theorem 1

For each hypercube of each type-d context, the regret due to trainings and explorations is bounded by Lemma 1. It can be shown that for each type-d context there can be at most $4T^{1/(1+p)}$ hypercubes that is activated by time T. Using this we get a $O(T^{z+1/(1+p)} \log T)$ upper bound on the regret due to explorations and trainings for a type-d context. Then we sum over all types of contexts $d \in \mathcal{D}$. We show in Lemma 4 that the regret due to near optimal arm selections in exploitation steps is $O(T^{\frac{1+p-\alpha}{1+p}})$. In order to balance the time order of regret due to explorations, trainings and near optimal arm selections in exploitations, while at the same time minimizing the number of explorations and trainings, we set $z = 2\alpha/p$, and $p = \frac{3\alpha+\sqrt{9\alpha^2+8\alpha}}{2}$, which is the value which balances these two terms. Notice that we do not need to balance the order of regret due to suboptimal arm selections since its order is always less than the order of trainings and explorations. We get the final result by summing these two terms together with the regret due to suboptimal arm selections in exploitation steps which is given in Lemma 2.