Online Learning to Approach a Person with No-Regret: Supplementary Material

Hyemin Ahn, Yoonseon Oh, Sungjoon Choi, Claire J. Tomlin, and Songhwa Oh

I. PROOF OF THEOREM 1

Lemma 1 (Lemma 5.1 in [1]): For $\delta \in (0, 1)$, if $\beta_k = 2 \log(|Q|/\pi_k/\delta)$, where $\sum \pi_k = 1$ and $\pi_k = \pi_k^2 k^2/6$, $|\mathcal{P}(q) - \mu_{k-1}(q)| \leq \beta_k^{1/2} \sigma_{k-1}(q)$ $\forall q \in Q$, with probability $1 - \delta$.

Lemma 2: If $|\mathcal{P}(q) - \mu_{k-1}(q)| \leq \beta_k^{1/2} \sigma_{k-1}(q)$ $\forall q \in Q$, $r_k \leq \sum_{t=1}^{T_k} 2\beta_k^{1/2} \sigma_{k-1}(\xi_k(t))$, where $T_k = |\xi_k|$.

Proof: For $\xi_k$ chosen at the $k$th round, the GP-UCB algorithm is applied such that:

$$\xi_k = \arg \max_{\xi \in \Xi} \sum_{t=1}^{t^*} \left( \mu_{k-1}(\xi(t)) + \beta_k^{1/2} \sigma_{k-1}(\xi(t)) \right).$$

Therefore, it is clear that

$$\sum_{t=1}^{T_k} (\mu_{k-1}(\xi_k(t)) + \beta_k^{1/2} \sigma_{k-1}(\xi_k(t)))$$

$$\geq \sum_{t=1}^{t^*} (\mu_{k-1}(\xi^*(t)) + \beta_k^{1/2} \sigma_{k-1}(\xi^*(t))) \geq f(\xi^*)$$

Hence, we have

$$r_k = f(\xi^*) - f(\xi_k)$$

$$\leq \sum_{t=1}^{T_k} (\mu_{k-1}(\xi_k(t)) + \beta_k^{1/2} \sigma_{k-1}(\xi_k(t))) - f(\xi_k)$$

$$\leq \sum_{t=1}^{T_k} (\mu_{k-1}(\xi_k(t)) - \mathcal{P}(\xi_k(t))) + \beta_k^{1/2} \sigma_{k-1}(\xi_k(t))$$

$$\leq \sum_{t=1}^{T_k} 2\beta_k^{1/2} \sigma_{k-1}(\xi_k(t))$$

Learning 3: For $\delta \in (0, 1)$ and $\beta_k$ defined as in Lemma 1, with probability at least $1 - \delta$,

$$\sum_{k=1}^{K} r_k^2 \leq C_1 \beta_K \gamma_K$$

(1)

where $C_1 = 8T_{\max}/\log(1 + \sigma^2)$ and $\gamma_K = \max_{A \in \Xi} \mathbb{I}(p_{q(A)}; \mathcal{P}_{q(A)})$ is the maximum information gain after $K$ rounds. Here, $\mathcal{P}_{q(A)}$ and $p_{q(A)}$ are sets of comfort scores and corresponding observations at states in $A$, respectively.

Proof: From Lemma 2, we have

$$r_k^2 \leq \left( \sum_{t=1}^{T_k} 2\beta_k^{1/2} \sigma_{k-1}(\xi_k(t)) \right)^2$$

$$\leq 4\beta_k \left( \sum_{t=1}^{T_k} \sigma_{k-1}(\xi_k(t)) \right)^2 \leq 4\beta_k T_k \sum_{t=1}^{T_k} \sigma_{k-1}^2(\xi_k(t))$$

since $\beta_k$ is non-decreasing. The last inequality is due to the Cauchy-Schwarz inequality. By defining $C_2 = \sigma^2/\log(1 + \sigma^2)$ as done in [1], we have

$$r_k^2 \leq 4\beta_k T_k \sigma^2 \sum_{t=1}^{T_k} \sigma^2_{k-1}(\xi_k(t))$$

$$\leq 4\beta_k T_k \sigma^2 \left( \sum_{t=1}^{T_k} C_2 \log(1 + \sigma^2_{k-1}(\xi_k(t))) \right)$$

$$= 8\sigma^2 C_2 T_k \beta_K \left( \frac{1}{2} \sum_{t=1}^{T_k} \log(1 + \sigma^2_{k-1}(\xi_k(t))) \right)$$

Using Lemma 5.3 in [1], for $A_k \in \Xi_k$, we have

$$\mathbb{I}(p_{q(A_k)}; \mathcal{P}_{q(A_k)}) = \sum_{\xi \in \Xi_k} \left( \frac{1}{2} \sum_{t=1}^{T_k} \log(1 + \sigma^2_{k-1}(\xi(t))) \right)$$

Noting that $|A_k| = k$, we arrive at

$$\sum_{k=1}^{K} r_k^2 \leq 8\sigma^2 C_2 T_{\max} \beta_K \mathbb{I}(p_{q(A_K)}; \mathcal{P}_{q(A_K)}) \leq C_1 \beta_K \gamma_K$$

Lastly, $C_1$ can be simplified to $C_1 = 8T_{\max}/\log(1 + \sigma^2)$.

Since $R_K^2 \leq K \sum_{k=1}^{K} r_k^2$ using the Cauchy-Schwarz inequality, Theorem 1 has been proven.

REFERENCES