STATS 710 – Seq. Dec. Making with mHealth Applications Dec 1, 2016

Lecture 22: Online ranking with top-1 feedback(Continued)

Instructors: Susan Murphy and Ambuj Tewari *Scribe: April Cho*

The online ranking problem with top-1 feedback [CT15] was introduced in the last lecture. This lecture is about the overall algorithm of Online ranking and proof on its regret bound.

1 Parameters

- T : time
- K : number of objects
- L : number of episodes
- ϵ : FTPL randomness
- $B_i = ((i-1)\frac{T}{L}+1, \cdots, i\frac{T}{L})$ is the ith block

2 Overall algorithm

Algorithm 1

for $i = 1, 2, \cdots, L$ **do**
  $\hat{s}_{i-1} = \hat{s}_{i-2} + \hat{e}_{i-1}$
  Select $I_1^{(i)}, \cdots, I_k^{(i)}$ unif at random without replacement
  for $t \in B_i$ **do**
    if $t = I_j^{(i)}$ for some j **then** # Explore
      Output a permutation that places object j on top
      Receive $\rho_{I_j}(j) \in \{0, 1\}$
    else # Exploit
      Draw $Z_t \in [0, \frac{1}{\epsilon}]^k$ with coordinates drawn indep from $Unif[0, \frac{1}{\epsilon}]$
      Output $\sigma_t = \text{argmin}_{\sigma \in S_k} \sigma^T(\hat{s}_{i-1} + Z_t)$
    end if
  end for
  $\hat{\rho}_i = (\rho_{I_1^{(i)}}(1), \rho_{I_2^{(i)}}(2), \cdots)^T$
end for

Overall feedback needed $= LK = T^{2/3}K$.

3 Sketch of regret analysis

$$\mathbb{E}\left[\sum_{i=1}^{L} \hat{\rho}_i^T \sigma_i^{FTPL}\right] - \min_\sigma \sum_{i=1}^{L} \hat{\rho}_i^T \sigma \leq 2\sqrt{BL} \qquad \text{where } B = B_\sigma B_{loss} B_\rho$$

and $\|\sigma_t\|_1 \leq B_\sigma$, $\|\rho_t\|_1 \leq B_\rho$, $|\sigma_t^T \rho_t| \leq B_{loss}$

Taking expectation over $I_1^{(i)}, \cdots, I_k^{(i)}$, we have

$$\mathbb{E}\left[\sum_{i=1}^{L} (\frac{\sum_{t\in B_i} \rho_t}{T/L})^T \sigma_i^{FTPL}\right] - \mathbb{E}\left[\min_\sigma \sum_{i=1}^{L} \hat{\rho}^T \sigma\right]$$

$$\geq \mathbb{E}\left[\frac{\sum_{i=1}^{L}\sum_{t\in Bi} \rho_t^T \sigma_i^{FTPL}}{T/L}\right] - \min_{\sigma\in s_k}[\frac{\sum_{i=1}^{L}\sum_{t\in B_i} \hat{\rho}^T \sigma}{T/L}]$$

We have the above inequality since $\mathbb{E}[min(*)] \leq min\mathbb{E}[*]$.

Now, the distribution of σ_t within block i is the same as the distribution of σ_i^{FTPL} except on K rounds. Within a round, the regret incurred due to exploration is therefore at most $B_{loss}K$. Since there are L blocks, we have $\mathbb{E}\left[\frac{\sum_{i=1}^{L}\sum_{t\in Bi} \rho_t^T \sigma_i^{FTPL}}{T/L}\right] \geq \mathbb{E}\left[\frac{\sum_{i=1}^{L}\sum_{t\in Bi} \rho_t^T \sigma_t}{T/L}\right] - \frac{LB_{loss}K}{T/L}$,

We have shown that

$$\mathbb{E}\left[\frac{\sum_{i=1}^{L}\sum_{t\in Bi} \rho_t^T \sigma_t}{T/L}\right] - \frac{LB_{loss}K}{T/L} - \min_{\sigma\in s_k}[\frac{\sum_{i=1}^{L}\sum_{t\in B_i} \hat{\rho}^T \sigma}{T/L}] \leq 2\sqrt{BL}$$

By multiply each term by T/L and moving terms around, we get

$$\mathbb{E}\left[\sum_{i=1}^{L}\sum_{t\in Bi} \rho_t^T \sigma_t\right] - \min_{\sigma\in s_k}[\sum_{i=1}^{L}\sum_{t\in B_i} \hat{\rho}^T \sigma]$$

$$\leq 2\sqrt{BL}\frac{T}{L} + LB_{loss}K$$

$$= 2\sqrt{B}\frac{T}{\sqrt{L}} + LB_{loss}K$$

$$= O(\text{poly}(K)T^{2/3})$$

when you tune L properly, and upper bound B, B_{loss} in terms of K

References

[CT15] Sougata Chaudhuri and Ambuj Tewari. Online ranking with top-1 feedback. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, volume 38 of *JMLR Workshop and Conference Proceedings*, pages 129–137, 2015.