IE613: Online Machine Learning **Jan-Apr 2018**

Lecture 26: Best Arm Identification Problem

Lecturer: M. K. Hanawal *Scribes: Sravan Patchala*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

26.1 Introduction

We looked at algorithms with uniform sampling and elimination, which led to unnecessary sampling and information, which can be avoided using adaptive algorithms. Earlier, our objective was to find the arm I_t such that

$$Pr(\mu_{I_t} > \mu_m - \epsilon) \geq 1 - \delta$$

which is the (ϵ, δ) - PAC strategy. Now we refine the objective and try to find the set S^*_m such that, it returns the best m arms. That is,

$$S^*_{m,\epsilon} = \{i : \mu_i \geq \mu_m - \epsilon\}.$$

Let S_δ be the set returned by the strategy. We would want that

$$Pr(S_\delta \subseteq S^*_{m,\epsilon}) \geq 1 - \delta$$

26.2 LUCB Algorithm

```
Data: m, $\epsilon, \delta$, U and L (Confidence Bounds)
Result: Set of arms J(t)
Initialization : B(1) = $\infty$, t = 1 ;
while $B(t) > \epsilon$ do
    Draw arm $u_t$ and $l_t$, t= t+1;
    J(t) := set of m arms with the highest empirical means ;
    $u_t = \operatorname{argmax}_{j \notin J(t)} U_j(t)$ ;
    $l_t = \operatorname{argmin}_{j \in J(t)} L_j(t)$ ;
    $B(t) = U_{u_t}(t) - L_{l_t}(t)$ ;
end
return J(t) ;
```

Algorithm 1: The general LUCB algorithm

The intuition behind the algorithm is that, the termination occurs when the gap between the lowest lower-confidence bounds of the m-highest arms and the highest upper-confidence bound of the remaining arms is less than ϵ i.e. the m arms with the higher empirical means are sufficiently separated from all the other arms.

The quantities $U_a(t)$ and $L_a(t)$ are defined as follows:

- For the case where the Hoeffding's inequality is used for calculating the bounds, we have

$$U_a(t) = \hat{\mu}_a(t) + \sqrt{\frac{\alpha log(t)}{N_a(t)}}$$

and

$$L_a(t) = \hat{\mu}_a(t) - \sqrt{\frac{\alpha log(t)}{N_a(t)}}$$

- For the LUCB1 case, we have the following confidence bounds

$$U_a(u,t) = \hat{\mu}_a(t) + \sqrt{\frac{log(\frac{k_1 K t^\alpha}{\delta})}{2u}}$$

and

$$L_a(u,t) = \hat{\mu}_a(t) - \sqrt{\frac{log(\frac{k_1 K t^\alpha}{\delta})}{2u}}$$

where u denotes the times the arm is played in that round.

We define a quantity H^γ

$$H^\gamma = \sum_a \frac{1}{(max\{\Delta_a, \gamma\})^2}$$

where

$$\Delta_a = \begin{cases} \mu_a - \mu_{m+1} & if\ a \leq m \\ \mu_m - \mu_a & if\ a > m \end{cases}$$

With these definitions in mind, the following theorems are stated without any proof:

Theorem 26.1 *The expected sample complexity of LUCB1 is* $O(H^{\epsilon/2}\ log(\frac{H^{\epsilon/2}}{\delta}))$

Theorem 26.2 *With probability at least* $1 - \delta$*, LUCB1 terminates after* $O(H^{\epsilon/2}\ log(\frac{H^{\epsilon/2}}{\delta}))$

where $H^{\epsilon/2}$ is given by

$$H^{\epsilon/2} = \sum_a \frac{1}{(max\{\Delta_a, \frac{\epsilon}{2}\})^2}$$

26.3 KL-LUCB Algorithm

The KL-LUCB algorithm has an important modification. The confidence intervals are given as follows:

$$U_a(t) = max\{q \in [\hat{\mu}_a, 1] : N_a(t)\ d(\hat{\mu}_a, q) \leq \beta(t, \delta)\}$$

and

$$L_a(t) = min\{q \in [0, \hat{\mu}_a] : N_a(t)\ d(\hat{\mu}_a, q) \leq \beta(t, \delta)\}$$

where $\beta(t, \delta)$ is known as the exploration sequence.

It can be shown that if $U_a(t)$ is given by $U_a(t) = \hat{\mu}_a + \sqrt{\frac{\beta(t,\delta)}{2N_a(t)}}$, for the UCB algorithm, then the $U_a(t)|_{LUCB} \geq U_a(t)|_{KL-LUCB}$. And once the lower concentration bound $L_a(t)$ is defined appropriately, we can show that $L_a(t)|_{LUCB} \leq L_a(t)|_{KL-LUCB}$. Thus, the KL-LUCB gives better/tighter concentration bounds compared to LUCB. Since the concentration bounds are tighter, we will thus have a lower sample complexity, since the algorithm will terminate quicker.