Combinatorial Multi-armed Bandits & α, β - Approximation

Lightning Day
September 15, 2023
Sharif University of Technology

Reza Pishkoo

Presentation Overview

1. Multi armed bandit (MAB)
2. Combinatorial multi armed bandit (CMAB)
3. General CMAB Framework
4. Main theorem
5. Comparing to classical MAB

Multi armed bandit

- m arms
- reward
- optimal arm
- regret
- tradeoff between exploration and exploitation

real world application

- online advertising scenario
- website contains a set of web pages
- advertiser can select at most k web pages

definition of CMAB

- super arms : any set of arms
- offline computation oracle computes the optimal super arm
- (α, β)-approximation oracle
- $\alpha, \beta \leq 1$
- α : oracle outputs a super arm whose expected reward is at least α fraction of the optimal expected reward
- β : success probability
- (α, β)-approximation regret

definition of CMAB framework

- m arms associated with a set of random variables $X_{i,t}$ for $1 \leq i \leq m$ and $t \geq 1$
- super arms : every set of arms
- $X_{i,t}$: a random variable with bounded support on $[0, 1]$
- $X_{i,t}$: random outcome of the i-th arm in its t-th trial
- $\mu = (\mu_1, \mu_2, ..., \mu_m)$: vector of expectations of all arms
- $T_{i,t}$: number of times the outcome of arm i is revealed after the first t rounds
- $R_t(S)$: non-negative random variable denoting the reward of super arm s

Reward

- $\mathbb{E}\left[R_t(S)\right]$: a function of only the set of arms S and the expectation vector μ
- $r_\mu = \mathbb{E}\left[R_t(S)\right]$
- Monotonicity : if for all $i \in [m]$, $\mu_i \leq \mu'_i$, we have $r_\mu(S) \leq r_{\mu'}(S)$ for all S
- Bounded smoothness : there exists a stricly increasing function $f(.)$ such that, we have $|r_\mu(S) - r_{\mu'}(S)| \leq f(\Lambda)$ if $\max_{i \in S} |\mu_i - \mu'_i| \leq \Delta$

CMAB algorithm

- Algorithm A selects S_t^A
- maximize $\mathbb{E}_{S,R}\left[\sum_{t=1}^{n} R_t(S_t^A)\right] = \mathbb{E}_S\left[\sum_{t=1}^{n} r_\mu(S_t^A)\right]$
- $opt_\mu = \max_S r_\mu(S)$
- $S_\mu^* = \max_S r_\mu(S)$
- (α, β)-approximation oracle : $\mathbb{P}[r_\mu(S) \geq \alpha.opt_\mu] \geq \beta$
- $Reg_{\mu,\alpha,\beta}^A(n) = n.\alpha.\beta.opt_\mu - \mathbb{E}_S\left[\sum_{t=1}^{n} r_\mu(S_t^A)\right]$

CMAB algorithm

- empirical mean :
  $\hat{\mu}_i = (\sum_{j=1}^{s} X_{i,j})/s$
- adjusted mean :
  $\bar{\mu}_i = \hat{\mu}_i + \sqrt{\frac{3ln(t)}{2T_i}}$

For each arm i, maintain: (1) variable T_i as the total number of times arm i is played so far; (2) variable $\hat{\mu}_i$ as the mean of all outcomes $X_{i,*}$'s of arm i observed so far.
For each arm i, play an arbitrary super arm $S \in \mathcal{S}$ such that $i \in S$ and update variables T_i and $\hat{\mu}_i$.
$t \leftarrow m$.
while true do
  $t \leftarrow t + 1$.
  For each arm i, set $\bar{\mu}_i = \hat{\mu}_i + \sqrt{\frac{3 \ln t}{2T_i}}$.
  $S = \text{Oracle}(\bar{\mu}_1, \bar{\mu}_2, \ldots, \bar{\mu}_m)$.
  Play S and update all T_i's and $\hat{\mu}_i$'s.
end while

Algorithm 1: CUCB with computation oracle

1. Multi armed bandit (MAB)
2. Combinatorial multi armed bandit (CMAB)
3. General CMAB Framework
4. **Main theorem**
5. Comparing to classical MAB

Preliminaries

- Super arm S is bad if $r_\mu(S) < \alpha.opt_\mu$
- $S_B = \{S | r_\mu(S) < \alpha.opt_\mu\}$
- $\Delta^i_{min} = \alpha.opt_\mu - \max\{r_\mu(S) | S \in S_B, i \in S\}$
- $\Delta^i_{max} = \alpha.opt_\mu - \min\{r_\mu(S) | S \in S_B, i \in S\}$
- $\Delta_{max} = \max_{i \in [m]} \Delta^i_{max}$
- $\Delta_{min} = \min_{i \in [m]} \Delta^i_{min}$

Theorem

Theorem

The (α, β)-approximation regret of the CUCB algorithm in n rounds using an (α, β)-approximation oracle is at most

$$\sum_{i \in [m], \Delta^i_{min} > 0} \left(\frac{6ln(n)\Delta^i_{min}}{\left(f^{-1}(\Delta^i_{min})\right)^2} + \int_{\Delta^i_{min}}^{\Delta^i_{max}} \frac{6ln(n)}{\left(f^{-1}(x)\right)^2} dx \right) + \left(\frac{\pi^2}{3} + 1 \right) .m.\Delta_{max}$$

simplified form :

$$\left(\frac{6ln(n)}{\left(f^{-1}(\Delta_{min})\right)^2} + \frac{\pi^2}{3} + 1 \right) .m.\Delta_{max}$$

Comparing to classical MAB

- every super arm is a simple arm
- $f(x) = x$
- $\alpha = \beta = 1$
- $\Delta^i_{max} = \Delta^i_{min}$
- $\Delta^i = \max_{j \in [m]} \mu_j - \mu_i$
  the regret bound of the classical MAB

$$\sum_{i \in [m], \Delta^i > 0} \frac{6 ln(n)}{\Delta^i} + \left(\frac{\pi^2}{3} + 1 \right) .m.\Delta_{max}$$

- the coefficient of the regret bound in (Auer et al., 2002) : $\sum_{i \in [m], \Delta^i > 0} \frac{8}{\Delta^i}$