Object Detection Using Deep Learning

Malyala Preethi Sravani (200070041)
Sakshi Heda (200070071)

Indian Institute of Technology, Bombay
Department of Electrical Engineering

EE6106
Online Learning and Optimization
April, 2023

Table of Contents

Introduction

- Conventional multi arm bandit (MAB) problem:
  - learner pulls one out of $K \in \mathbb{N}_+$ arms
  - reward is obtained (sampled from an unknown-mean distribution)
- Multi-play multi-armed bandits:
  - play multiple arms in a single time slot
- Shareable multi-play multi-armed bandits (*shareable* MP-MAB)
  - Each arm can be played any number of times

Problem Statement

- $K \in \mathbb{N}_+$ arms, indexed by $1, 2, ..., k$
- rewards according to the distribution X_k with means, μ_k (not known to the player)
- finite rewards capacity, m_k, (not known to the player)
- Reward from arm $k = \min\{a_k, m_k\} \times X_k$
- Without loss of generality $\mu_1 \geq \mu_2 \geq ... \geq \mu_k$

Problem Statement

N number of plays are assigned in each time slot, and the player chooses how these N plays are distributed among the K arms.

Objective: estimate the means of the reward distribution and capacity of these arms, while maximising the reward.

Theoretical Background

OrchExplore Algorithm:

Stage 1: Initial iterations: runs PIE in odd time slots, PUE in even time slots

Stage 2: After each run of PIE or PUE, update the upper and lower confidence bounds:

$$m^l_{k,t} := max\{\lceil \hat{\nu}_{k,t}/\hat{\mu}_{k,t} + \phi(\tau_t, \delta)\phi(l_t, \delta)\rceil, 1\}$$
$$m^u_{k,t} := min\{\lceil \hat{\nu}_{k,t}/\hat{\mu}_{k,t} + \phi(\tau_t, \delta)\phi(l_t, \delta)\rceil, N\}$$

Once the PUE set $\mathcal{Y}_t = \phi$, OrchExplore runs only PIE.

Stage 3: Once $\mathcal{Y}_t = \phi$ and $\mathcal{E}_t = \phi$, then pure *exploitation* - it allocates plays to empirical optimal arms according to these arms' reward capacities.

Parsimonious Individual Exploration (PIE)

The Oracle function: input - reward capacity's lower bounds m_t^l and empirical means $\hat{\mu}_k$ of the arms output - a_t^{IE}, assigns the highest mean to the arm with highest lower bound on capacity.

- $\mathcal{S}_t$, the set of all arms that will be played
- L_t to be the least favoured among these
- arms suct that KL-UCB index, $u_{k,t}$ is greater than or equal to the least favoured arm's mean $\hat{\mu}_{k,t} \longrightarrow \mathcal{E}_t$

$$u_{k,t} := \sup\{q \geq 0 : \hat{\tau}_{k,t}, \mathsf{kl}(\hat{\mu}_{k,t}, q) \leq \log t + 4 \log\log t\}$$

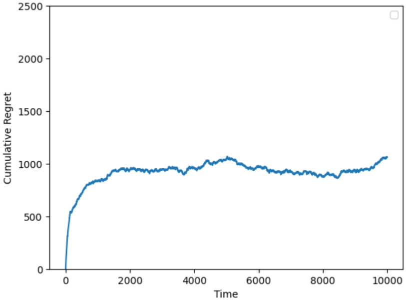
Parsimonious Individual Exploration (PIE)

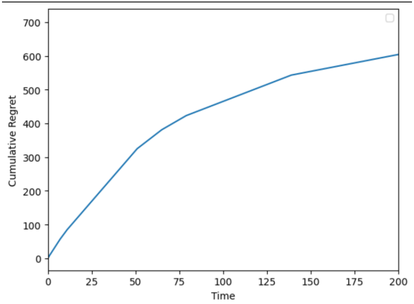
- With a probability of $1/2$, assign one play from L_t to one arm randomly uniformly selected from $\mathcal{E}_t$ The following values are updated:
  - empirical mean, $\hat{\mu}_t$
  - KL-UCB indexes, u_t
  - effective times of IE, $\hat{\tau}_t$
  - time slot index, t

Parsimonious United Exploration (PUE)

- $\mathcal{Y}_t$: all those arms in $\mathcal{S}_t - L_t$ for which $m^l_{k,t} \neq m^u_{k,t}$
- increase the empirical mean of these arms by a large positive value, M.
- Now, $\hat{\mu'}_t$ and the upper capacity bound, $m^u_{k,t}$ are given as input to the Oracle function and the output, a^{UE}_t is used to play the round
-
  - "full load" reward mean, $\hat{\nu}_t$
  - effective times of UE, $\hat{\text{ß}}_t$
  - time slot index, t

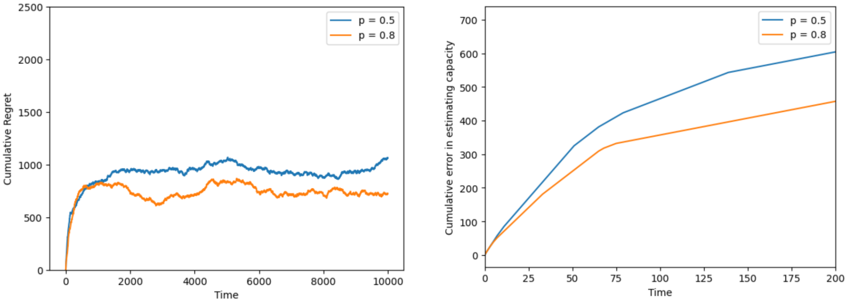
Results and Conclusions





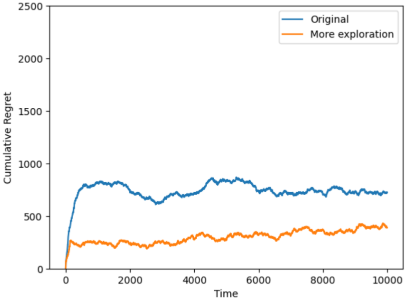
Results of original algorithm

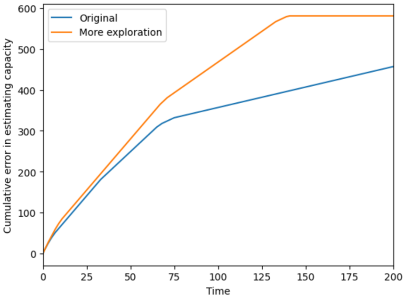
Results and Conclusions



Increasing the probability of random exploration

Results and Conclusions





Exploring more arms

References I

[1] Xuchuang Wang, Hong Xie, and John C. S. Lui. "Multiple-Play Stochastic Bandits with Shareable Finite-Capacity Arms". In: *Proceedings of the 39th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato. Vol. 162. Proceedings of Machine Learning Research. PMLR, 17–23 Jul 2022, pp. 23181–23212. URL: https://proceedings.mlr.press/v162/wang22af.html.