Thresholding Bandits with Augmented UCB

Subhojyoti Mukherjee [1]
K.P. Naveen [2]
Nandan Sudarsanam [1,3]
Balaraman Ravindran [1,3]

[1] Indian Institute of Technology Madras,
[2] Indian Institute of Technology Tirupati,
[3] Robert Bosch Centre for Data Science and AI

August 24, 2017

Overview

Problem Definition of TBP

- At every timestep the learner selects (pulls) a single action (arm) and receives an i.i.d random reward from the distribution attached to that arm.

Problem Definition of TBP

- At every timestep the learner selects (pulls) a single action (arm) and receives an i.i.d random reward from the distribution attached to that arm.
- **Primary aim:** Identify *all* the arms whose expected mean of the reward distribution (r_i) is above a particular threshold τ given as input.

Problem Definition of TBP

- At every timestep the learner selects (pulls) a single action (arm) and receives an i.i.d random reward from the distribution attached to that arm.
- **Primary aim:** Identify *all* the arms whose expected mean of the reward distribution (r_i) is above a particular threshold τ given as input.
- **Condition:** This has to be achieved within T timesteps of exploration and this is termed as a fixed-budget problem.

Problem Definition of TBP

- We define the set $S_\tau = \{i \in \mathcal{A} : r_i \geq \tau\}$.

Problem Definition of TBP

- We define the set $S_\tau = \{i \in \mathcal{A} : r_i \geq \tau\}$.
- S_τ^c denote the complement of S_τ, i.e., $S_\tau^c = \{i \in \mathcal{A} : r_i < \tau\}$.

Problem Definition of TBP

- We define the set $S_\tau = \{i \in \mathcal{A} : r_i \geq \tau\}$.
- S_τ^c denote the complement of S_τ, i.e., $S_\tau^c = \{i \in \mathcal{A} : r_i < \tau\}$.
- Let $\hat{S}_\tau$ denote the recommendation of a learning algorithm after T time units of exploration, while $\hat{S}_\tau^c$ denotes its complement.

Problem Definition of TBP

- We define the set $S_\tau = \{i \in \mathcal{A} : r_i \geq \tau\}$.
- S_τ^c denote the complement of S_τ, i.e., $S_\tau^c = \{i \in \mathcal{A} : r_i < \tau\}$.
- Let $\hat{S}_\tau$ denote the recommendation of a learning algorithm after T time units of exploration, while $\hat{S}_\tau^c$ denotes its complement.
- The goal of the learning agent is to minimize the **expected** loss at the end of budget T:

$$\mathbb{E}[\mathcal{L}(T)] = \mathbb{P}\Big(\underbrace{\{S_\tau \cap \hat{S}_\tau^c \neq \emptyset\}}_{\textbf{Rejected good arms}} \cup \underbrace{\{\hat{S}_\tau \cap S_\tau^c \neq \emptyset\}}_{\textbf{Accepted bad arms}} \Big)$$

Challenges in the TBP Settings

- Closer an arm's reward mean to $\tau \Rightarrow$ Harder the problem.

Challenges in the TBP Settings

- Closer an arm's reward mean to $\tau \Rightarrow$ Harder the problem.
- Lesser the budget $\Rightarrow$ Harder the problem.

Challenges in the TBP Settings

- Closer an arm's reward mean to $\tau \Rightarrow$ Harder the problem.
- Lesser the budget $\Rightarrow$ Harder the problem.
- Higher the variance of an arm's reward distribution $\Rightarrow$ Harder the problem.

Contributions

- We propose the **Augmented UCB (AugUCB)** algorithm for the fixed-budget TBP setting.

Contributions

- We propose the **Augmented UCB (AugUCB)** algorithm for the fixed-budget TBP setting.
- AugUCB takes into account the **empirical variances** of the arms along with mean estimates.

Contributions

- We propose the **Augmented UCB (AugUCB)** algorithm for the fixed-budget TBP setting.
- AugUCB takes into account the **empirical variances** of the arms along with mean estimates.
- It is the **first variance-based arm elimination algorithm** for the considered TBP settings.

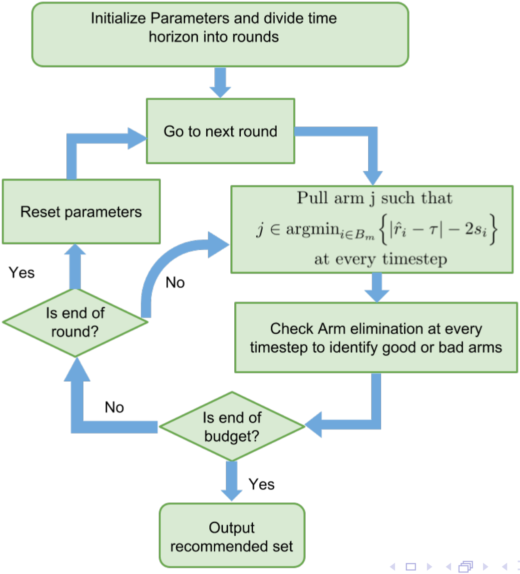
Contributions

- We propose the **Augmented UCB (AugUCB)** algorithm for the fixed-budget TBP setting.
- AugUCB takes into account the **empirical variances** of the arms along with mean estimates.
- It is the **first variance-based arm elimination algorithm** for the considered TBP settings.
- It **addresses an open problem** discussed in Auer and Ortner (2010) of designing an algorithm that can eliminate arms based on variance estimates.

Contributions

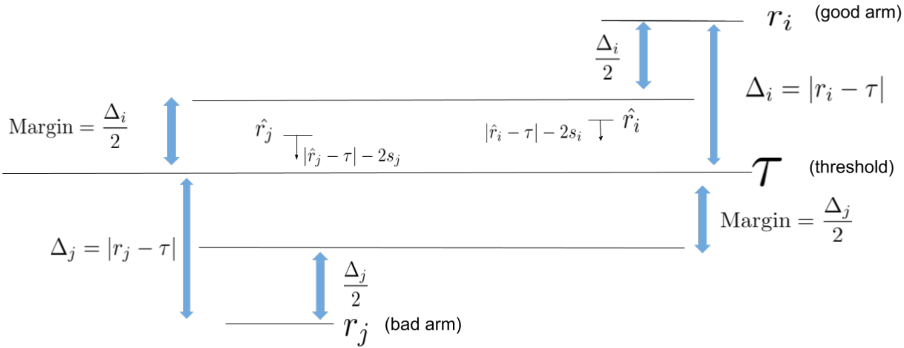
- We propose the **Augmented UCB (AugUCB)** algorithm for the fixed-budget TBP setting.
- AugUCB takes into account the **empirical variances** of the arms along with mean estimates.
- It is the **first variance-based arm elimination algorithm** for the considered TBP settings.
- It **addresses an open problem** discussed in Auer and Ortner (2010) of designing an algorithm that can eliminate arms based on variance estimates.
- We also define a **new problem complexity** which uses empirical variance estimates along with arm's mean for giving the theoretical bound.

AugUCB algorithm



AugUCB algorithm (Intuition, Arm pulling)

- Like UCB-Imp, AugUCB also divides the time budget T into rounds.
- At every timestep we pull arm j s.t. $j \in \arg\min_{i \in B_m} \left\{ |\hat{r}_i - \tau| - 2s_i \right\}$ (like APT).



Arm pull

- We pull the arm that minimizes $j \in \arg\min_{i \in B_m} \left\{ |\hat{r}_i - \tau| - 2s_i \right\}$

Arm pull

- We pull the arm that minimizes $j \in \arg\min_{i \in B_m} \left\{ |\hat{r}_i - \tau| - 2s_i \right\}$
- We define the confidence interval $s_i = \sqrt{\frac{\rho\psi_m(\hat{v}_i + 1)\log(T\epsilon_m)}{4n_i}}$.

Arm pull

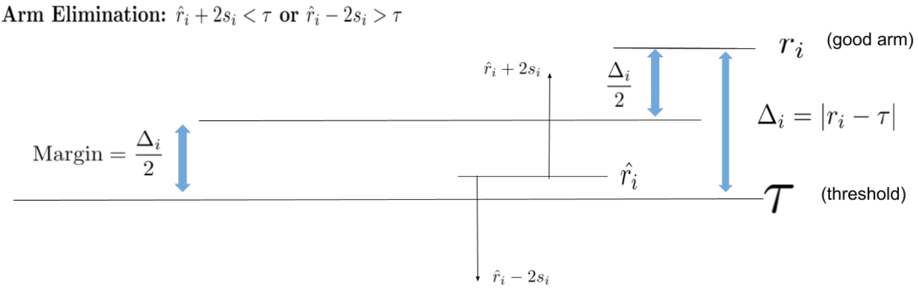
- We pull the arm that minimizes $j \in \arg\min_{i \in B_m} \left\{ |\hat{r}_i - \tau| - 2s_i \right\}$
- We define the confidence interval $s_i = \sqrt{\frac{\rho\psi_m(\hat{v}_i+1)\log(T\epsilon_m)}{4n_i}}$.
- s_i decreases with more n_i and ψ_m and ρ ensures that it decreases at a correct rate.

Arm pull

- We pull the arm that minimizes $j \in \arg\min_{i \in B_m} \left\{ |\hat{r}_i - \tau| - 2s_i \right\}$
- We define the confidence interval $s_i = \sqrt{\frac{\rho\psi_m(\hat{v}_i + 1)\log(T\epsilon_m)}{4n_i}}$.
- s_i decreases with more n_i and ψ_m and ρ ensures that it decreases at a correct rate.
- Note that $\hat{v}_i$ estimated variance in s_i makes the algorithm pull the arm which shows more variance.

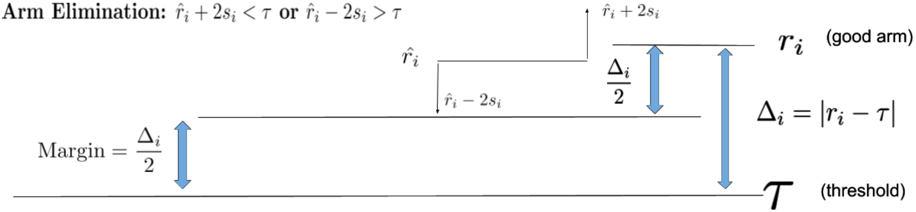
AugUCB algorithm (Intuition, Arm Elimination)

- We eliminate an arm when we are sure that $\hat{r}_i$ is close to r_i with high probability and hence identify it as good or bad arm.
- It's risky to eliminate an arm when $\hat{r}_i$ is inside *Margin*.
- Confidence interval s_i will make sure arm i is not eliminated while inside Margin with a high probability.



AugUCB algorithm (Intuition, Arm Elimination)

- Now we see that $\hat{r}_i$ has moved close to its true estimate r_i.
- We eliminate i and re-allocate the remaining budget to pull arms close to the threshold



Expected Loss of AugUCB

Theorem

For $K \geq 4$ and $\rho = 1/3$, the expected loss of the AugUCB algorithm is given by,

$$\mathbb{E}[\mathcal{L}(T)] \leq 2KT \exp\left(-\frac{T}{4096 \log(K \log K) H_{\sigma,2}}\right).$$

Table: AugUCB vs. State of the art

Algorithm	Upper Bound on Expected Loss	Oracle
AugUCB	$\exp\left(-\frac{T}{4096 \log(K \log K) H_{\sigma,2}} + \log(2KT)\right)$	No
UCBEV	$\exp\left(-\frac{1}{512}\frac{T-2K}{H_{\sigma,1}} + \log(6KT)\right)$	Yes
APT	$\exp\left(-\frac{T}{64H_1} + 2\log((\log(T)+1)K)\right)$	No
UCBE	$\exp\left(-\frac{T-K}{18H_1} - 2\log(\log(T)K\right)$	Yes

Experiments

- We compare with APT, AugUCB, UCBE, UCBEV, UA.

Experiments

- We compare with APT, AugUCB, UCBE, UCBEV, UA.
- Note that UCBE and UCBEV require access to H_1 and $H_{\sigma,1}$ as input and hence not implementable in real life.
- By access we mean that an oracle supplies them the H_1 or $H_{\sigma,1}$. They do not have access to individual means and variances.

Experiments

- We compare with APT, AugUCB, UCBE, UCBEV, UA.
- Note that UCBE and UCBEV require access to H_1 and $H_{\sigma,1}$ as input and hence not implementable in real life.
- By access we mean that an oracle supplies them the H_1 or $H_{\sigma,1}$. They do not have access to individual means and variances.
- APT, AugUCB, UA do not require access to H_1 or $H_{\sigma,1}$.

Experimental Setup

- This setup involves Gaussian reward distributions with $K = 100$, $T = 10000$ and $\tau = 0.5$ with the reward means set in two groups.

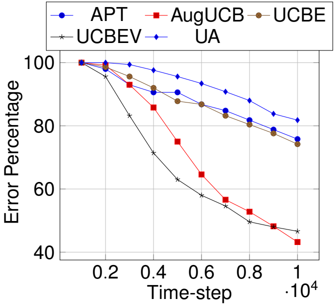
Experimental Setup

- This setup involves Gaussian reward distributions with $K = 100$, $T = 10000$ and $\tau = 0.5$ with the reward means set in two groups.
- The first 10 arms partitioned into two groups; the respective means are $r_{1:5} = 0.45$, $r_{6:10} = 0.55$.

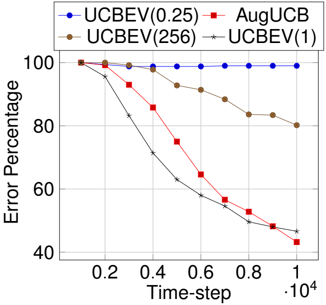
Experimental Setup

- This setup involves Gaussian reward distributions with $K = 100$, $T = 10000$ and $\tau = 0.5$ with the reward means set in two groups.
- The first 10 arms partitioned into two groups; the respective means are $r_{1:5} = 0.45$, $r_{6:10} = 0.55$.
- The means of arms $i = 11 : 100$ are chosen same as $r_{11:100} = 0.4$.
- Variances are set as $\sigma^2_{1:5} = 0.3$ and $\sigma^2_{6:10} = 0.8$; $\sigma^2_{11:100}$ are independently and uniformly chosen in the interval $[0.2, 0.3]$.

Experimental Results

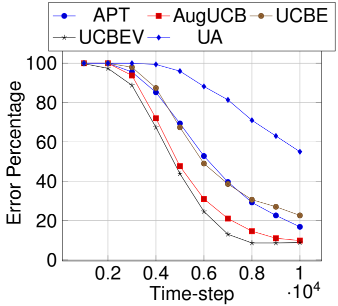


(a) Expt-1: Two Group Setting (Advance)

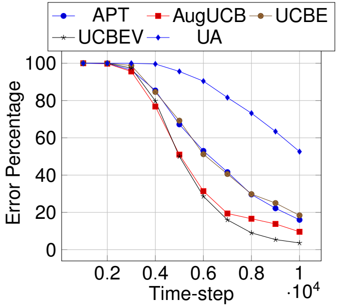


(b) Expt-2: Two Group Setting (Advance)

Experimental Results



(c) Expt-1: Arithmetic Progression (Gaussian)



(d) Expt-2: Geometric Progression (Gaussian)

Conclusions

- We proposed the AugUCB algorithm for the fixed budget TBP which uses variance estimation and arm elimination to give an improved theoretical and experimental guarantees than APT.

Conclusions

- We proposed the AugUCB algorithm for the fixed budget TBP which uses variance estimation and arm elimination to give an improved theoretical and experimental guarantees than APT.
- Further studies are required to establish a lower bound on the expected loss of AugUCB.

Conclusions

- We proposed the AugUCB algorithm for the fixed budget TBP which uses variance estimation and arm elimination to give an improved theoretical and experimental guarantees than APT.
- Further studies are required to establish a lower bound on the expected loss of AugUCB.
- A more detailed analysis of the non-uniform arm selection and parameter selection is also required.

Thank You