Multi-Armed Bandits with Cost Subsidy

Deeksha Sinha, Karthik Abinav Sankararaman,
Abbas Kazerouni, Vashist Avadhanula

Motivation: Intelligent SMS routing

- Many businesses send text messages (SMSes) to their customers for two-factor authentication, order confirmations, transaction alerts or for marketing purposes.
- Typically, they partner with multiple cellular *aggregators* to send these SMSes.
- Aggregators differ in their delivery rate and cost per SMS. Further, delivery rate of aggregators fluctuates with time due to congestion, network outages etc.
- Businesses would want to dynamically choose an aggregator that incurs low cost and currently has a good delivery rate (quality).
- This leads to an MAB formulation with multiple objectives.

Problem Formulation

Identify cheap aggregators with sufficient quality

Each aggregator (i) is modeled as an arm with known cost (c_i) and unknown quality (μ_i). Arms having qualities within a specified threshold $(1-\alpha)$ of the highest quality arm (μ_{max}) are of *sufficient quality*. Ideally, pick lowest cost arm with sufficient quality (i_*).

Consider two types of regret – quality and cost:

- Incur quality regret when chosen arm is not of sufficient quality.
  Quality Regret: $\sum_{t=1}^{T} \max\{(1-\alpha)\mu_{max} - \mu_{i_t}, 0\}$.
- Incur cost regret when chosen arm has cost higher than i_*.
  Cost Regret: $\sum_{t=1}^{T} \max\{c_{i_t} - c_{i_*}, 0\}$.

Minimize Cost Regret + Quality Regret

Challenges with Existing Approaches

Naïve generalizations of TS/UCB do not work well

Both TS and UCB algorithms can be naturally generalized for this problem (to get CS-TS and CS-UCB algorithms) as follows:

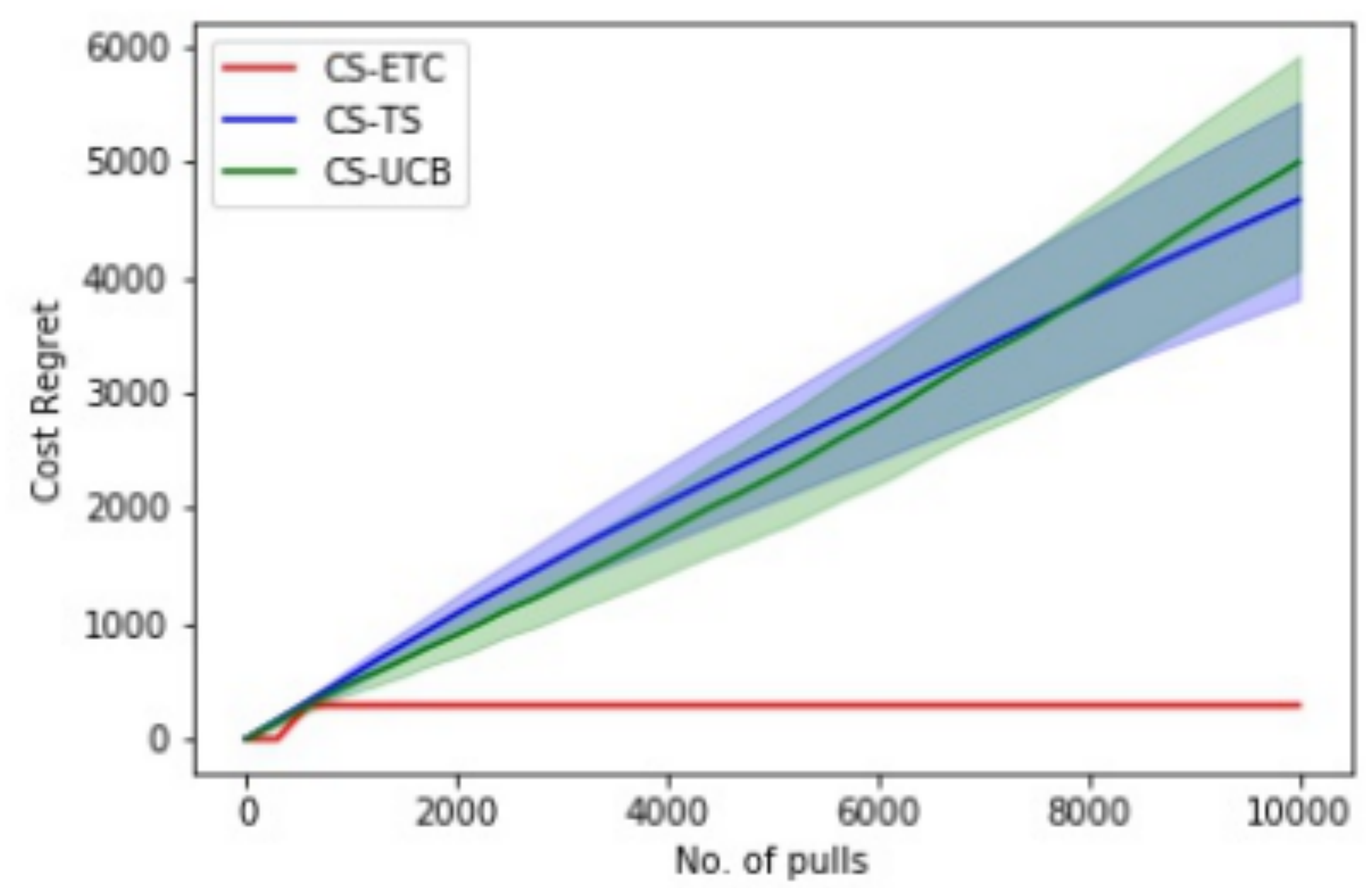
1. Estimate set of sufficient quality arms

$$\mathcal{F} = \{i | \mu_i^{score} \geq (1-\alpha)\mu_j^{score}\ \forall j\}$$

where μ_i^{score} is μ_i^{UCB} for CS-UCB and a sample from the posterior reward distribution of arm i for CS-TS.

2. Play cheapest arm in $\mathcal{F}$.

We prove that this generalization of TS (CS-TS) can incur linear regret. Empirically, we show the same for CS-UCB.



Lower Bound

Theorem: For any algorithm and setup with K arms, there exists an instance such that cost regret + quality regret is

$$\Omega((1-\alpha)^2 K^{\frac{1}{3}} T^{\frac{2}{3}}).$$

This suggests that our problem is harder than the standard MAB problem where the lower bound is $\sqrt{KT}$.

Proposed Explore-Then-Commit Algorithm (CS-ETC)

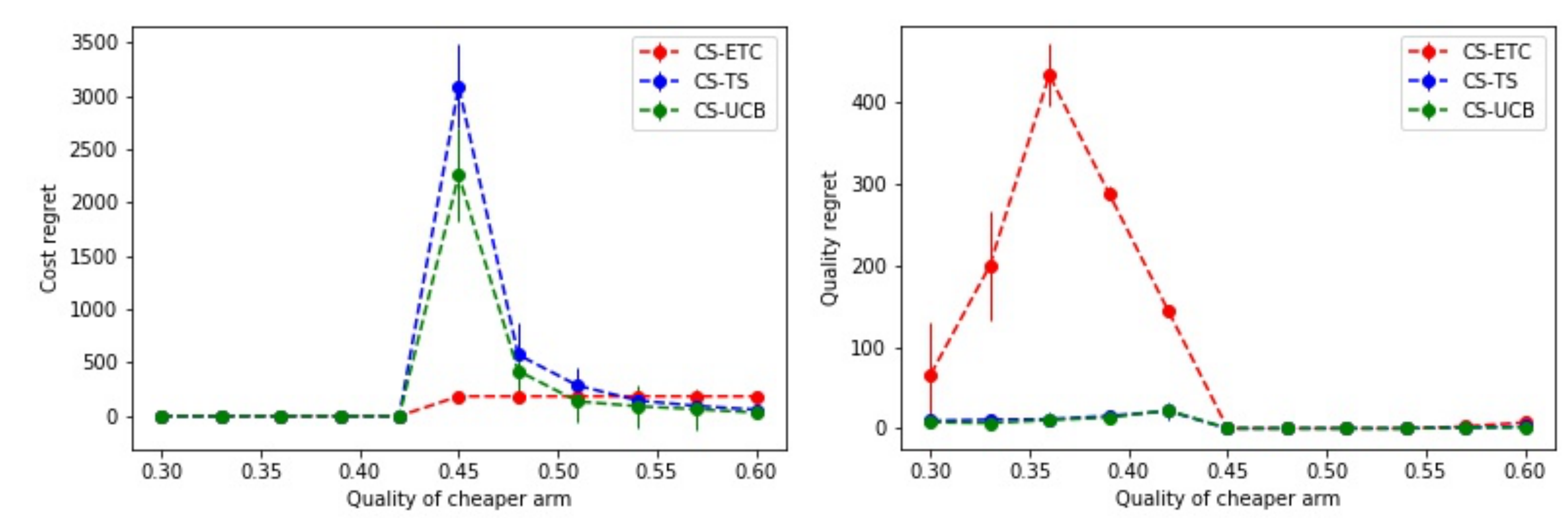
1. Pure Exploration Phase: Pull each arm $O(K^{\frac{1}{3}} T^{\frac{2}{3}})$ times.
2. UCB Phase:
   a. Maintain UCB and LCB estimates for each arm.
   b. Estimate set of sufficient quality arms
   $$\mathcal{F} = \{i | \mu_i^{UCB} \geq (1-\alpha)\mu_j^{LCB}\ \forall j\}$$
   c. Pick cheapest arm in $\mathcal{F}$.

Theorem: For any instance, the cost regret + quality regret incurred by the CS-ETC algorithm is

$$O(K^{\frac{1}{3}} T^{\frac{2}{3}} \sqrt{\log T}).$$

Numerical Experiments

Consider a two-arm setting with $\mu_1 = 0.5, c_1 = 1, c_2 = 0, \alpha = 0.1$, T= 5000 and μ_2 is varied.



CS-UCB/CS-TS work well when the mean rewards of arms are well differentiated. CS-ETC is ideal when the mean rewards are close to one another.