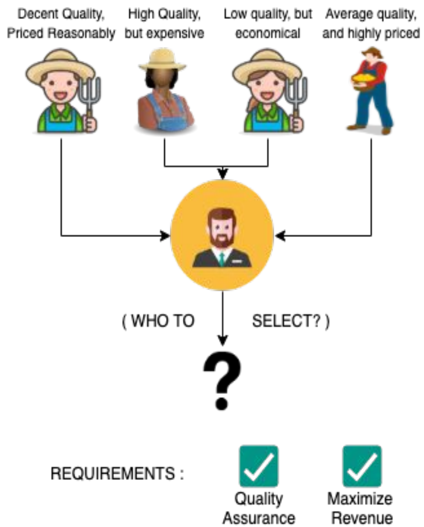
A Multi-Arm Bandit Approach To Subset Selection Under Constraints

Ayush Deva, Kumar Abhishek and Sujit Gujar



We model our problem as an Integer Linear Program (ILP) where a planner needs to select a subset of agents, each with its own quality and cost, so as to maximize revenue whilst ensuring that the average quality is above a threshold.

We propose a dynamic programming based algorithm, DPSS to solve for it.

But What If Qualities Are Unknown ?

We consider a setting where the qualities of the agents are unknown to the planner beforehand and needs to estimated through sequential selection. We model this as a Multi Arm Bandit problem and leverage the popular UCB algorithm to design an abstract algorithm SS-UCB.

The algorithm takes in input the available agents, their costs, quality threshold (α), tolerance parameter (ε_2) and a suitable offline subset selection algorithm, SSA.

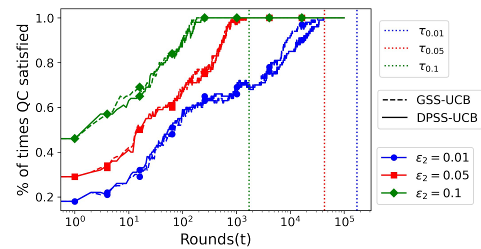
Algorithm 2 SS-UCB

Inputs: N, α, ϵ_2, R, costs $\boldsymbol{c} = \{c_i\}_{i \in N}$
For each agent i, maintain: w_i^t, q_i^t, $(\hat{q}_i^t)^+$
$\tau \leftarrow \frac{3 \ln T}{2\epsilon_2^2}$; $t = 0$
while $t \leq \tau$ **(Explore Phase) do**
  Play a super-arm $S^t = N$
  Observe qualities X_i^j, $\forall i \in S^t$ and update w_i^t, $\hat{q}_i^t$
  $t \leftarrow t + 1$
while $t \leq T$ **(Explore-Exploit Phase) do**
  For each agent i, set $(\hat{q}_i^t)^+ = \hat{q}_i^t + \sqrt{\frac{3 \ln t}{2 w_i^t}}$
  S^t = SSA $(\{(\hat{q}_i^t)^+\}_{i \in N}, c, \alpha + \epsilon_2, \mathrm{R})$
  Observe qualities X_i^j, $\forall i \in S^t$ and update w_i^t, $\hat{q}_i^t$
  $t \leftarrow t + 1$

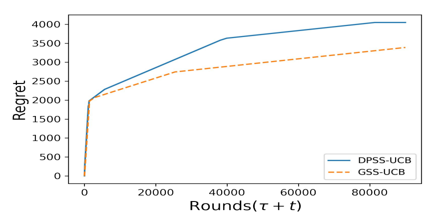
Key Results

Using DPSS as our SSA in Algorithm 2 (DPSS-UCB), we show that :

1. DPSS-UCB returns a subset which approximately satisfies the quality constraint with a high probability after **τ** rounds, where **τ** ~ O(ln *T*)

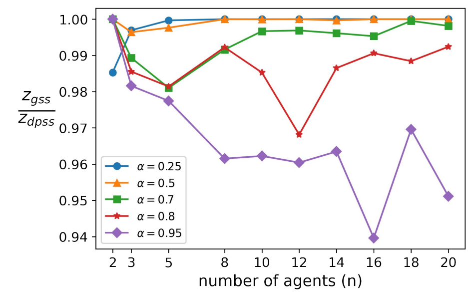


2.. The algorithm incurs a regret of O(ln *T*) after **τ** rounds



Approximate but Faster Solution

The time complexity of DPSS is of O(2^n), which makes it difficult to scale when n is large. We propose an approximate, greedy-based, polynomial time, O(n logn), algorithm, GSS, to our ILP. Further, we empirically show that by using GSS as the SSA in Algorithm 2 (GSS-UCB), we achieve similar results to DPSS-UCB.



Applications

References

1. Shweta Jain, Sujit Gujar, Satyanath Bhat, Onno Zoeter, and Y Narahari. 2018. A quality assuring, cost optimal multi-armed bandit mechanism for expertsourcing. Artificial Intelligence 254 (2018), 44–63.
2. Wei Chen, Yajun Wang, and Yang Yuan. 2013. Combinatorial Multi-Armed Bandit:General Framework and Applications. In Proceedings of the 30th International Conference on Machine Learning (ICML), 151-159.

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

AAMAS2021 LONDON, UK