Max-Min Grouped Bandits

Puranjay Datta ,19D070048

28-11-2022

Problem Statement

Max-Min Grouped Bandits

- K groups of arms where each group has N arms
- Each round select a group and pull all the arms in that group
- Reward and identity of every arm pulled are revealed
- Goal: Group with highest mean reward subject to constraint: Minimum reward in the group exceeds the given threshold(mn).
- Assumption: Bernoulli rewards, Identifiability

Example 1: $[0.1, 0.2, 0.3, 0.2], [0.6, 0.6, 0.6, 0.6], [0.9, 0.8, 0.7, 0.6]$
Example 2: $[0.4, 0.5, 0.4, 0.5], [0.6, 0.6, 0.6, 0.6], [0.9, 0.8, 0.7, 0.6]$
Example 3: $[0.1, 0.2, 0.3, 0.2], [0.1, 0.1, 0.1, 0.1], [0.9, 0.8, 0.7, 0.6]$
Example 4: $[0.4, 0.5, 0.4, 0.5], [0.3, 0.3, 0.4, 0.4], [0.9, 0.8, 0.7, 0.1]$

Problem Statement

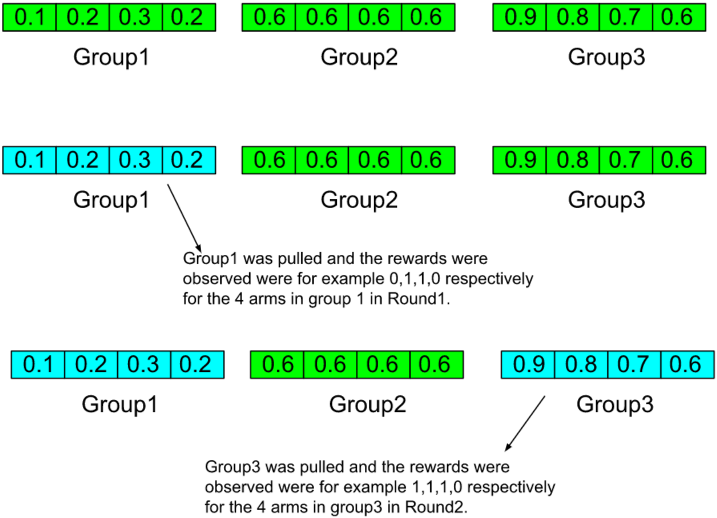
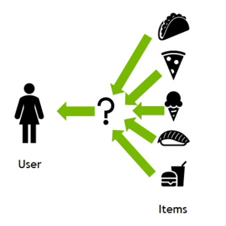


Figure: Simulation

Motivation

- Recommendation Systems-packages and goods that can be sold together
- Resource Allocation-select best computing device
- Service Centres



(a) Recommendation System[1]

(b) Resource Allocation[2]

[1] https://developer.nvidia.com/blog/how-to-build-a-winning-recommendation-system-part-1/
[2] https://www.proofhub.com/articles/resource-allocation-tips

Successive Elimination-Vanilla Case

- Player maintains an active set of arms S.
- At every round player first samples from the reward distribution of every arm in the active set.
- Player then removes all arms in the active set with estimated rewards that are outside the anytime confidence interval around the highest estimated reward in the active set.
- When the active set has 1 arm, the player identifies this arm with high probability as the best arm.

Successive Elimination-Vanilla Case

Successive Elimination($\{1, 2, 3..., n\}, \delta$)
$S \leftarrow \{1, 2, 3..., n\}$
while $1 \leq t \leq \infty$ **do**
    Pull arms in S
    Update all j in S: $U_j(t, \delta) = \sqrt{\frac{log(\frac{4t^2}{\delta})}{2pulls[j]}}$
    $S \leftarrow S - \{i \in S; \exists j \in S : \hat{\mu}_{j,t} - U_j(t, \frac{\delta}{n}) \geq \hat{\mu}_{i,t} + U_i(t, \frac{\delta}{n})\}$
    Stop when $|S| = 1$
end while
return S
end procedure

Group Successive Elimination-Removing Min-Violating Group

- Player maintains an active set of Groups of arms S.
- At every round player first samples from the reward distribution of every arm in the active set.
- Player Marks the Group as dormant(eligible to proceed to next step) and removes it from active set S and transfers it to set S' if the LCB of all the arms in that group is above min_threshold.
- Player marks the group inactive(discard the group/ineligible to proceed to next step) and removes it from active set S if any of the arms' UCB goes below the min_threshold.
- Continue until all the groups are either marked as dormant/inactive.

Group Successive Elimination-Identifying best group

- Player maintains an active set of Groups of arms S'.
- At every round player first samples from the reward distribution of every arm in the active set.
- Player then removes all the groups whose UCB is less than the maximum LCB amongst all the groups.
- When the active set has 1 group, the player identifies this group with high probability as the best group.

Group Successive Elimination-4 CASES

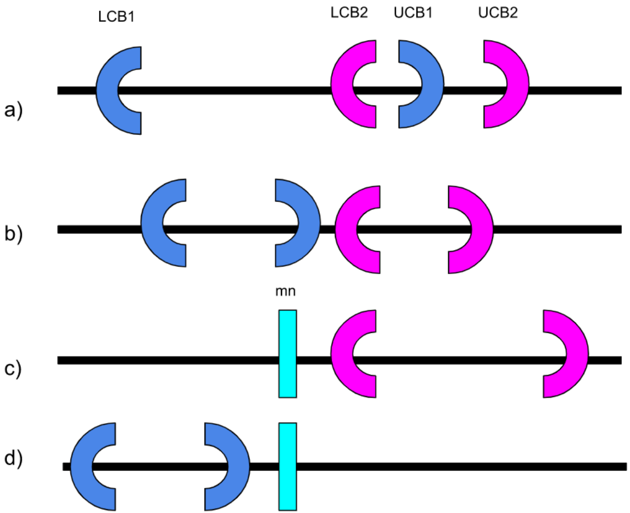


Figure: 4 Cases encountered in Successive Elimination

Group Successive Elimination-Algorithm

Group Successive Elimination($\{1, 2, 3..., K \ groups\}, \delta, mn$)
$S \leftarrow \{1, 2, 3..., K\}$,$S' \leftarrow \{\}$
while $1 \leq t \leq \infty$ **do**
    Pull arms in S (all K arms in every group)
    $S \leftarrow S - \{i \in S; \exists j \in i : \hat{\mu}_{i,j,t} + U_{i,j}(t, \frac{\delta}{NK}) \leq mn$
    $S' \leftarrow S' + \{i \in S; \forall j \in i : \hat{\mu}_{i,j,t} - U_{i,j}(t, \frac{\delta}{NK}) \geq mn$
    Stop when $|S| = 0$
end while
while $1 \leq t \leq \infty$ **do**
    Pull arms in S' (all N arms in every group)
    $S' \leftarrow S' - \{i \in S; \exists j \in S : \sum_{k=1}^{N}(\hat{\mu}_{i,k,t} + U_{i,k}(t, \frac{\delta}{NK})) \leq \sum_{k=1}^{N}(\hat{\mu}_{j,k,t} - U_{j,k}(t, \frac{\delta}{NK}))$
    Stop when $|S'| = 1$
end while
return S'
end procedure

Group Successive Elimination-Results

Successive Elimination

Result: For the Successive Elimination-vanilla case with n arms, we find the best arm with $1-\delta$ probability with a sample complexity T given by:

$T \leq \sum_{i\neq i^*}^{n} c.\Delta_i^{-2} log(\frac{n.log(\Delta_i^{-2})}{\delta})$ where $\Delta_i = \mu_i - \mu^*$, for some constant c

Group Successive Elimination

Result: For our case i.e with K groups and N arms per group we find the best group subject to minimum threshold condition with $1-\delta$ probability and sample complexity T given by

$T_{i,1} = c.\Delta_{G_i}^{-2} log(\frac{NK.log(\Delta_{G_i}^{-2})}{\delta})$ where $\Delta_i = \mu_{G_i} - \mu^*$, for some c

$T_{i,2} = min_l\ c.\Delta_{th,i,l}^{-2} log(\frac{NK.log(\Delta_{th,i,l}^{-2})}{\delta})$ where $\Delta_{th,i,l} = \mu_{i,l} - mn$,for some c

$T_{i,3} = max_l\ c.\Delta_{th,i,l}^{-2} log(\frac{NK.log(\Delta_{th,i,l}^{-2})}{\delta})$

$T_i = min(T_{i,2}, max(T_{i,1}, T_{i,3})),\ T = \sum_{i\neq i^*}^{K} T_i$

Median Elimination-Vanilla Case

- Best Arm identification problem with two constraints ϵ, δ
- ϵ refers to epsilon optimality which means an arm whose true mean is epsilon less than optimal arm can be chosen with $1 - \delta$ probability.
- Sample every arm in S for $\frac{1}{(\frac{\epsilon_l}{2})^2} \log(\frac{3}{\delta_l})$
- removes all the arms less than the median hence the median elimination ends within $log_2(n)$ rounds.

Median Elimination Algorithm-Vanilla Case

Median Elimination($\{1, 2, 3..., n\}, \epsilon, \delta$)
$S \leftarrow \{1, 2, 3..., n\}, \epsilon_1 = \frac{\epsilon}{4}, \delta_1 = \frac{\delta}{2}, l = 1$
while $1 \leq l \leq \infty$ **do**
    Sample every arm a in S_l for $\frac{1}{(\frac{\epsilon_l}{2})^2} \log(\frac{3}{\delta_l})$
    Let $\hat{p}_a$ denote its total received reward
    Find the median(m_l) of received rewards $\hat{p}_a$ of all arms a in S
    $S_{l+1} = S_l - \{a : \hat{p}_a < m_l\}$
    $\epsilon_{l+1} = \frac{3}{4}\epsilon_l, \delta_{l+1} = \frac{\delta_l}{4}, l = l + 1$
    Stop when $|S_l| = 1$
end while
return S_l
end procedure

Group Median Elimination Algorithm

Median Elimination($\{1, 2, 3..., K\}, \epsilon, \delta$)
$S \leftarrow \{1, 2, 3..., K\}, \epsilon_1 = \frac{\epsilon}{4}, \delta_1 = \frac{\delta}{2}, l = 1$
while $1 \leq l \leq \infty$ **do**
Sample every arm a in every group in S for $\frac{1}{(\frac{\epsilon_l}{2})^2} \log(\frac{3}{\delta_l})$
Let $\hat{p}_{i,j}$ denote its total received reward of (i^{th} group,j^{th} arm)
Find the median(m_l) $g_i = \sum_{j=1}^{N} \hat{p}_{i,j}$ of all groups in S.
$\epsilon_{l+1} = \frac{3}{4}\epsilon_l, \delta_{l+1} = \frac{\delta_l}{4}, l = l + 1$
Eliminate all the groups which have a min-violating arm ie
$\hat{p}_{i,j} \leq \mu_{th} - \frac{\epsilon_l}{2}$
From the remaining arms eliminate until $\frac{n_l}{2}$ arms remain :
$S_{l+1} = S_l - \{i : \hat{g}_i < m_l\}$
Stop when $|S_l| = 1$
end while
return S_l
end procedure

Group Median Elimination-Results

Median Elimination

Let ϵ be the epsilon optimality and δ be the probability of not picking the optimal arm and n be the number of arms,then
Sample complexity $= O(\frac{n}{\epsilon^2}\log(\frac{1}{\delta}))$

Group Median Elimination

Let ϵ be the epsilon optimal group (instead of arms) and δ be the probability of not picking the optimal group and N be the number of arms per group, K be the number of groups,then
Sample complexity $= O(\frac{NK}{\epsilon^2}\log(\frac{2N}{\delta}))$

LUCB-vanilla case

- Sample all arms once
  for each t>n sample arms h_t, l_t **do**
  $h_t = argmax\ \mu_i, l_t = argmax\ \mu_i + U(T_i(t), \frac{\delta}{n})$
  Eliminate arms based on Successive Elimination condition
  end for
- h_t: Arm with highest empirical mean, l_t: Arm with highest UCB, $Ti(t)$:Number of pulls of arm i till time t.
- KL-LUCB bounds
  $u_a(t) := max\{q \in [\hat{p}_a(t), 1] : N_a(t)d(\hat{p}_a(t), q) \leq \beta(t, \delta)\}$
  $l_a(t) := min\{q \in [0, \hat{p}_a(t)] : N_a(t)d(\hat{p}_a(t), q) \leq \beta(t, \delta)\}$ where,
  $\beta(t, \delta) = ln(\frac{t}{\delta}) + 4ln(ln(\frac{t}{\delta}))$
- LUCB has same bounds as Successive Elimination$= T \leq \sum_{i \neq i^*}^{n} c.\Delta_i^{-2} log(\frac{n.log(\Delta_i^{-2})}{\delta})$

Group LUCB

Group LUCB($\{1, 2, 3..., K \ groups\}, \delta, mn$)

$S \leftarrow \{1, 2, 3..., K\}$,$S' \leftarrow \{\}$

while $1 \leq t \leq \infty$ **do**

Pull arms $h_t = argmin \ \hat{\mu}_{i,j} \ \ s.t \ \ \hat{\mu}_{i,j} \geq mn$

$l_t = argmax \ \hat{\mu}_{i,j} \ \ s.t \ \ \hat{\mu}_{i,j} \leq mn$

$S \leftarrow S - \{i \in S; \exists j \in i : \hat{\mu}_{i,j,t} + U_{i,k}(t, \frac{\delta}{NK}) \leq mn$

$S' \leftarrow S' + \{i \in S; \forall j \in i : \hat{\mu}_{i,j,t} - U_{j,k}(t, \frac{\delta}{NK}) \geq mn$

Stop when $|S| = 0$

end while

while $1 \leq t \leq \infty$ **do**

Pull arms $h_t = argmax \ \sum_{j=1}^{N} \hat{\mu}_{i,j}$

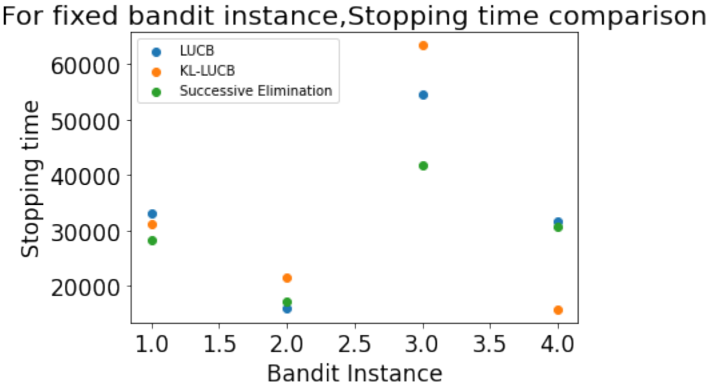
$l_t = argmax \ \sum_{j=1}^{N} \hat{\mu}_{i,j} + U_{i,j}(t, \frac{\delta}{NK})$

$S' \leftarrow S' - \{i \in S; \exists j \in S : \sum_{k=1}^{N}(\hat{\mu}_{i,k,t} + U_{i,k}(t, \frac{\delta}{NK})) \leq \sum_{k=1}^{N}(\hat{\mu}_{j,k,t} - U_{j,k}(t, \frac{\delta}{NK}))$

Stop when $|S'| = 1$

end while

Simulation-Results



Example 1:[[0.1,0.2,0.3,0.2],[0.6,0.6,0.6,0.6],[0.9,0.8,0.7,0.6]]
Example 2:[[0.4,0.5,0.4,0.5],[0.6,0.6,0.6,0.6],[0.9,0.8,0.7,0.6]]
Example 3:[[0.4,0.5,0.4,0.5],[0.3,0.3,0.4,0.4],[0.9,0.8,0.7,0.1]]
Example 4:[[0.1,0.2,0.3,0.2],[0.1,0.1,0.1,0.1],[0.9,0.8,0.7,0.6]]

Concluding Remarks

- Bounds for sample complexity scale by NK in the number of arms N for Group Successive Elimination.
- For Median elimination sample complexity scale by NK in number of arms except for the term inside the log term which can be further explored.
- Vanilla LUCB has bounds the same as that of Successive Elimination but the authors of the paper say that the bounds can be improved in terms of analysis.
- Future work will involve exploiting the minimum and best group conditions simultaneously instead of doing it in two steps.

Thank You

References

Recommendation Systems,[Online].Available:https://developer.nvidia.com/blog/how-to-build-a-winning-recommendation-system-part-1/

Resource Allocation,[Online].Available:https://www.proofhub.com/articles/resource-allocation-tips

E. Even-Dar, S. Mannor, and Y. Mansour, "Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems." [Online].Available:https://jmlr.csail.mit.edu/papers/volume7/evendar06a/evendar06a.pdf

E. Kaufmann and S. Kalyanakrishnan, "Information complexity in bandit subset selection." [Online].Available: http://proceedings.mlr.press/v30/Kaufmann13.pdf

S. Gupta, G. Joshi, and O. Yağan, "Best-arm identification in correlated multi-armed bandits," Sep 2021. [Online]. Available: https://arxiv.org/abs/2109.04941

K. Jamieson and R. Nowak, "Best arm survey - homes.cs.washington.edu." [Online].Available:https://homes.cs.washington.edu/ jamieson/resources/bestArmSurvey.pdf

Z. Wang and J. Scarlett, "Max-min grouped bandits - arxiv.org." [Online]. Available: https://arxiv.org/pdf/2111.08862.pdf