Community Exploration: From Offline Optimization to Online Learning

Xiaowei Chen1  Weiran Huang2  Wei Chen3  John C.S. Lui4

1The Chinese University of Hong Kong  2Huawei Noah’s Ark Lab  3Microsoft Research

Motivating Example: Online Advertising

Place Ads on different websites

Goal: Maximize the reward (# distinct visitors)

Challenges
- Limited Budget: How to allocate the budget on each website?
- Duplicate Visitors: The same visitor may visit the website several times.
- Unknown Possible Visitor Numbers: How to learn the parameters?

Model Descriptions

- Communities: disjoint sets \( C_1, \ldots, C_m \)
- \( \Phi \) Example: Visitors of different websites
- Explore Community \( C_i \) on Explore once, meet a member in \( C_i \) uniformly at random
- Reward: # of distinct members \( \in C_i \) \( U \rightarrow U \cap C_m \)
- Budget: explore communities at most \( K \) times

Our Results

Offline Problems (Non-adaptive, and adaptive exploration)
- Setting: Community sizes are known
- Solution: Greedy method/policy
- Conclusion: Greedy method/policy is optimal

Online Learning (Non-adaptive, and adaptive exploration)
- Setting: Community sizes are unknown
- Solution: Combinatorial Lower Confidence Bound (CLCB) algorithm
- Conclusion: Logarithmic/constant regret bound

Offsite Optimization

Key Assumption: The sizes of communities \( d = (d_1, \ldots, d_m) \) are known

Problem 1. Offline Non-adaptive

Non-adaptive Exploration
- Determine the budget allocation \( \pi = (\pi_1, \ldots, \pi_m) \) before the exploration
- Explore \( C_i \) for \( K_i \) times
- \( \sum_{i \in [m]} K_i = K \)

Goal: Find an optimal budget allocation \( \pi^* = (\pi_1^*, \ldots, \pi_m^*) \) to maximize the expected reward

Maximize

\[
\max_{\pi} \sum_{i \in [m]} \left( 1 - \frac{1}{K_i} \right) \sum_{t \in \pi} k_i \]

Expected Reward

Constraint

Greedy Method (time complexity \( O(m \log m) \))

Start from \( \pi = (\pi_1, \ldots, \pi_m) \) (good initial point)

At each step (stop when \( \sum_{i \in [m]} K_i = K \)), we choose community \( C_i \) such that

\[
i^* \in \arg \max_{i \in [m]} \frac{1}{\pi_i} - \frac{1}{K_i},
\]

Theorem: Greedy method obtains optimal budget allocation!

Problem 2. Offline Adaptive

Adaptive Exploration
- Step by step exploration
- Choose community to explore based on previous results

Goal: Find an optimal policy \( \pi^* \) to maximize the expected reward

Mapping function \( \phi \) (previous results) \( \rightarrow \) next community to explore

Greedy Policy \( \pi^* \)
- At each step, choose the community which has the largest percentage of unvisited members

Adaptive submodular

1 \(- 1/e\)

Theorem: Greedy policy is the optimal among all the policies!

- Proved by inductive reasoning
- Applied to the general reward function \( f(\#\text{distinct members}) \)

Online Learning

- The sizes of communities are unknown
- Online advertising in multiple rounds
- In each round, we try to maximize the reward

Problem 3. Online Non-adaptive

For each round \( t = 1, \ldots, T \)
- Choose "action" \( \pi_t = (\pi_{1t}, \ldots, \pi_{mt}) \) based on previous exploration results
- Explore community \( C_i \) for \( \pi_{it} \) times (non-adaptive exploration)

Problem 4. Online Adaptive

For each round \( t = 1, \ldots, T \)
- Choose an "action" \( \pi_t \) based on previous exploration results
- Explore community with policy \( \pi_t \) (adaptive exploration)

Goal: Maximize the cumulative rewards in \( T \) rounds

Bandeit Algorithm

At round \( t \) (\( \tilde{\pi}_i(t) \): unbiased estimator of \( 1/d_i \))

- Compute lower confidence bound
  - Radius \( \pi_t_k = \sqrt{\frac{2 \ln t}{d_k}} \)
  - \( \hat{\pi}_t_k = \max(0, \hat{\pi}_t_k - \pi_t_k) \)

- Play Oracle (\( \Gamma = (1/d_1, \ldots, 1/d_m) \))

- Online, Non-adaptive
  - At each step, choose \( i^* \in \arg \max_{i} (1 - \hat{\pi}_t_i)^b \)

- Online, Adaptive
  - At each step, choose \( i^* \in \arg \max_{i} (1 - \hat{\pi}_t_i)^b \)
  - \( c_i \): # members that is already met in \( C_i \)

- Update estimates
  - \( \pi_{it+1} = \pi_{it} \cdot \min(1, \gamma_{it+1}) \)
  - Collision within one round
    - \( \gamma_{it} = \frac{1}{d_i} \cdot \mathbb{1}[d_i(x \in [1:m] = \{d(x), \gamma_{it+1} + \gamma_{it+2} + \cdots + \gamma_{it+2} + \cdots + \gamma_{it+2} + \cdots + \gamma_{it+2} \}) \)
  - Full information feedback
    - \( \gamma_{it} = \mathbb{1}[d_i(x \in [1:m] = \{d(x), \gamma_{it+1} + \gamma_{it+2} + \cdots + \gamma_{it+2} + \cdots + \gamma_{it+2} \}) \)

- Regret Bound (Problem 3&4)

  - Collisions within one round: \( Reg(T) \sim O(\log T) \)
  - Tighter bound, leveraging existing analysis framework
    - Regret bound: problem dependent const. \( O(1) \)

  - Full information feedback: problem dependent const. \( O(1) \)

  - Full information feedback: problem dependent const. \( O(1) \)

  - Full information feedback: problem dependent const. \( O(1) \)