Maximizing and Satisficing in Bandits with Graph Information

Parth Kashyap Thaker [1] Mohit Malu [1] Nikhil Rao [2] Gautam Dasarathy [1]

[1]Arizona State University [2]Microsoft

Bandits with Graph Information

Multi-armed bandits (MAB) has emerged as an important framework to model decision making and search under uncertainty. However, in many modern-day applications such as drug discovery, recommender systems and policy evaluations, traditional MAB methods are rendered ineffective due to following challenges

Challenges:

- Enormous options (arms) to evaluate.
- Split second decision making.
- Limited methods to incorporate structure.
- Inaccurate structural information.

This work incorporates *approximate graph information* in existing MAB framework to tackle aforementioned challenges.

Potential sources of approximate graph information include but are not limited to social networks, preference similarity, etc.

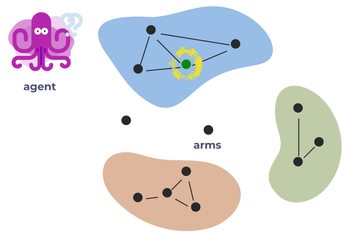


Figure 1. Bandits with graph information

Problem Framework and Statement

Problem Framework:
Consider a n-armed bandit problem with rewards as follows:

$$r_t^i = \mu_i + \eta_t, \quad \forall i \in [n], \ \eta_t \text{ is } \sigma\text{-sub-Gaussian}, \forall t \in \mathbb{N}$$

Further, consider the availability of graph information in the form of similarity graph G such that,

$$\|\boldsymbol{\mu}\|_G^2 \triangleq \langle \boldsymbol{\mu}, L_G \boldsymbol{\mu} \rangle = \sum_{\{i,j\} \in E_G} A_{ij}(\mu_i - \mu_j)^2 \leq \epsilon$$

If the above is satisfied, we say that the arm rewards are ϵ-smooth with respect to a graph G.

Problem Statement:
Design a sampling policy $\pi_t : t \to [n]$ based on the past measurements to find the following:

- **P1:** The *best arm* i^* such that $i^* = \arg\max_{i \in [n]} \mu_i$.
- **P2:** An ζ-*approximate best arm* i' such that $\mu_{i^*} - \mu_{i'} \leq \zeta$.

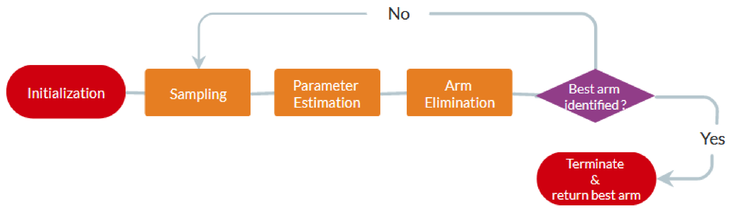


Figure 2. Bandit Flowchart

GRUB Algorithm

- **GRUB** (**GR**aph based **U**pper confidence **B**ound) algorithm incorporates graph information in to UCB strategy.
- Graph information helps in forming mean and confidence estimates of the arms that have not been sampled.

The GRUB algorithm proceeds in three major steps as shown in the flow graph:

1. **Parameter Estimation:** At each step GRUB computes an estimate of mean and confidence bounds of all arms. The mean estimate at any time T is given by:

$$\hat{\boldsymbol{\mu}}_T = \arg\min_{\boldsymbol{\mu} \in \mathbb{R}^n} \left\{ \left[\sum_{t=1}^{T} (r_{t,\pi_t} - \mu_{\pi_t})^2 \right] + \rho \langle \boldsymbol{\mu}, L_G \boldsymbol{\mu} \rangle \right\}, \tag{1}$$

2. **Arm Elimination:** At any time t, arm a is retained only if it's upper confidence bound is greater than the best lower confidence bound.
3. **Sampling Strategy:** Use intelligent (incorporating graph information), random or cyclic policy to sample the next arms.

Theoretical Results

Theorem 1 *[GRUB Sample Complexity (Informal)]*

Consider n-armed bandit problem with ϵ-smooth mean vector $\boldsymbol{\mu}$ w.r.t. graph G. Then, GRUB succeeds in finding the best arm with high probability after no more than $T_{\text{sufficient}}$ rounds given as follows

$$T_{\text{sufficient}} = \sum_{j \in \text{clusters}} \left[\sum_{i \in \mathcal{H}_j} \mathcal{O}\left(\frac{1}{\Delta_i^2}\right) + \max_{i \in \mathcal{N}_j} \mathcal{O}\left(\frac{1}{\Delta_i^2}\right) \right], \tag{2}$$

where $\mathcal{H}_j$ and $\mathcal{N}_j$ indicate Competitive and Non-competitive arms in cluster j.

Novelty

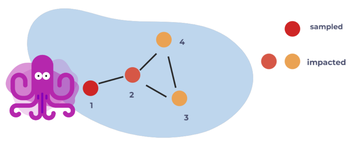


Figure 3. Impact of sampling an arm

Sampling any arm provides additional insights into connected arms due to the presence of graph structure. This is quantified using the notion of **Resistance Distance** $r(\cdot,\cdot)$ on graph G, where resistance distance between (i,j) w.r.t. graph G is denoted by,

$$r_{\delta,G}(i,j) = R_{ii} + R_{jj} - R_{ij} - R_{ji}$$

where $R \triangleq \left(L_G + \delta \mathbb{1}\mathbb{1}^T\right)^{\dagger}$, $\dagger$ denotes the Moore-Penrose inverse, L_G is the Laplacian of graph G, and $\mathbb{1} \in \mathbb{R}^n$ is the vector of all 1's.

Theorem 2 *[Lower Bound (Informal)]*

Consider n-armed bandit problem with ϵ-smooth mean vector $\boldsymbol{\mu}$ w.r.t. graph G consisting only of isolated cliques. Then any δ-PAC algorithm will need at least $T_{\text{necessary}}$ steps to terminate given as follows, provided $\delta \leq 0.1$

$$T_{\text{necessary}} = \sum_{C \in \mathcal{C}_G / C^*} \min_{j \in C} \left\{ \frac{4\sigma^2 \log 5}{(\Delta_j - \sqrt{\epsilon})^2} \right\} + \sum_{j \in C^*/1} \frac{4\sigma^2 \log 5}{\Delta_j^2} \tag{3}$$

Intuitive examples

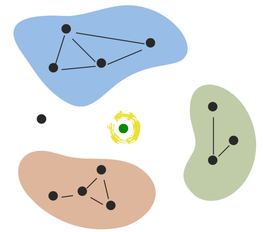


Figure 4. Graph A

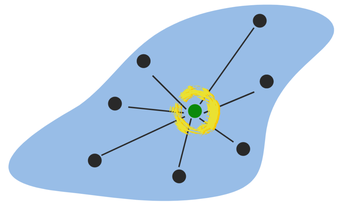
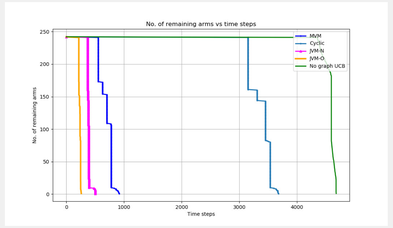


Figure 5. Graph B

- **Graph A** : Clustered graph with isolated optimal arm. In the best case scenario, sample complexity of the bandit problem scales as $\mathcal{O}(\#\text{ clusters})$.
- **Graph B** : Star graph with optimal arm at the center. In the best case scenario, sample complexity of the bandit problem scales as $\mathcal{O}(\#\text{ arms})$.

Experimental Results



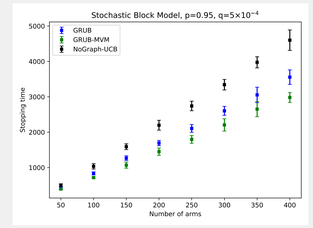


Figure 6. Performance of GRUB with using various sampling protocols for Github social graph [Left] and for SBM ($m = 2$) [Right]. The UCB method without graph information is significantly slower compared to the graph-based variants.

Future Work

- Extension of theoretical results to account for improved sampling policies.
- Misspecifications with respect to graph G and smoothness parameters ϵ.
- Faster mean estimation by matrix inversion coupled with spectral sparsification of the graph

References

[1] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz.
Pure exploration in multi-armed bandits problems.
In *International conference on Algorithmic learning theory*, pages 23–37. Springer, 2009.

[2] Tomáš Kocák and Aurélien Garivier.
Best arm identification in spectral bandits.
arXiv preprint arXiv:2005.09841, 2020.