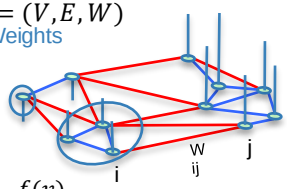
Cheap Bandits

Manjesh K Hanawal, Venkatesh Saligrama (BU, Boston, USA)
Michal Valko, Remi Munos (INRIA, Lille, France)

BOSTON UNIVERSITY

Problem Statement

- Undirected Graph: $G = (V, E, W)$
  - N Nodes, $W = \{w_{ij}\}$: Weights
- Signal on Graph
  - Reward Function $f: V \to R$
- Locate maxima $u^* = argmax_{v \in V} f(v)$
- Actions:
  - Noisy Cluster Averages; Differentiated Costs

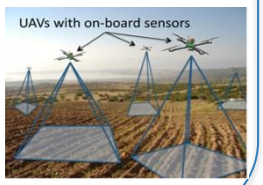


Goal: Locate u* with min Cost?

Motivation and Application

Surveillance/Geography
Forest Cover Dataset: Labeled samples on 30m² region
Nodes: Forest Regions; **Edge weights**: feature similarity;
Rewards: Density of species. Locate highest density.
Actions: Zoom-in (high cost); Zoom-out (low cost).

Other Examples:
Sensor Netw orks
Radar Search
Online advertisements



Large Number of nodes
Few samples to observe. $T \ll N$

Reward Function

- Linear Reward: $f = Q\alpha^*$
  - Q is the eigenvectors of the graph Laplacian
  - Linear bandits, w ith parameter α^*
- Smooth Reward
  - Neighboring nodes have similar rew ard

$$(u, v) \in E \Rightarrow f(u) \sim f(v):$$

$$|Lf|_2^2 = \sum_{(u,v)} w_{uv}\big(f(u) - f(v)\big)^2$$

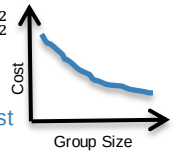
  - Assume $|Lf|_2^2 \le c$ [Valko et. al. ICML'15]

Actions and Costs

- Actions set is subset of simplex : $S \subset \Delta^N$
  - Node actions: Sample nodes
    For a node $u \in V$ use $s(u) = \delta(u - v)$
  - Group actions: Sample Average of a subset
    For a subset $A \in V$ use $s(u) = \frac{1}{|A|}\sum_{v \in V}\delta(u - v)$
- Cost of actions
  - For all $s \in S$

$$C(s) = \sum_{u,v}\big(s(u) - s(v)\big)^2 = |Ls|_2^2$$

- Why this cost function
  - Larger the group size smaller the cost
  - Probing Nodes has high cost
  - In Fourier domain: Energy of s



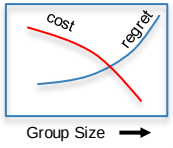
Learning Setting and Objective

- Policy (π): In each round t, select action $s_t \in S$
  - Observed Reward at rounds t:
    $r(s_t) = \sum_{u \in V} s_t(u) f(u) + \epsilon_t(noise)$
  - Regret of policy π:
    $R_T(\pi) = Tf(u^*) - E[\sum r(s_t)]$
  - Cost of policy π:
    $C_T(\pi) = \sum_t^T C(s_t)$
- **Goal :**

$$min_{(\pi,S)} C_T(\pi)$$
$$s.t \quad R_T(\pi) \le R_T^*$$

- Conflicting goals:
  - Node actions giv e better estimates, but costly
  - Group actions giv e poor estimates, but cheaper

What is the best Regret Constraint R_T^* ?



CheapUCB achieves at least 25% cost savings

Lower Bounds

- No smoothness constraints ($c \to \infty$)
  - Arbitrary bounded set of action
    $R_T = \Omega(N\sqrt{T})$ [Dani et. al. COLT'08]
  - Finite set of actions
    $R_T = \Omega(\sqrt{NT})$ [Chu et. al. AISTATS'08]
- For smooth rew ard

$$R_T = \Omega\big(\sqrt{dT}\big): \; d \ll N$$

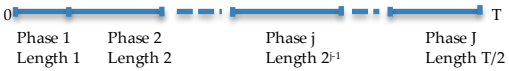
d= $\max\{i: \lambda_i(i-1) \le \frac{T}{\log T}\}$ [Valko et. al. ICML'14]

  - d-sparsely connected clusters
  - Need at least one sample from each cluster

We aim for $R_T^* = O(d\sqrt{T})$

CheapUCB Algorithm

- Main idea:use similarity of neighborhood
  - Group actions provide good node information
    $u \in A : f(u) \sim \frac{1}{|A|}\sum_{v \in A} f(v) + const$
- CheapUCB: UCB based algorithm. Selects arms optimistically like SpectralUCB [Valko et. al. ICML'15]



- **Phases:** Split the T into J= ⌊log T⌋ phases
- **Length:** Phase j=1,2,…J is of 2^{j-1} rounds
- **Select action:** In phase j select groups of size J-j+1

Zoom-in slowly using progressively costly actions

Algorithm	Regret bound	Cost
SpectralUCB (ICML'14)	$O(d\sqrt{T})$	T
CheapUCB (This paper)	$O(d\sqrt{T})$	3T/4

Experiments