Submodular Cost Submodular Cover with an Approximate Oracle

Victoria G. Crawford[1], Alan Kuhnle[2], My T. Thai[1]

[1]University of Florida

[2]Florida State University

Submodular Cost Submodular Cover (SCSC)

Definition (Submodular Cost Submodular Cover (SCSC))

Let $f, c : 2^S \to \mathbb{R}_{\geq 0}$ be monotone submodular functions defined on subsets of a ground set S of size n. Given threshold $\tau \leq f(S)$, SCSC is to find

$$\text{argmin}\{c(X) | X \subseteq S, f(X) \geq \tau\}.$$

- SCSC arises in many applications
  - Influence in a social network
  - Data summarization
- NP-hard

The Greedy Algorithm

The greedy algorithm has an approximation ratio of

$$\rho \left(\ln \left(\frac{\alpha}{\beta} \right) + 1 \right)$$

(Soma & Yoshida 2015).

Algorithm 1: $\texttt{greedy}(f, c, \tau)$

$f_\tau = \min\{f, \tau\}$
$i = 0$, $A_i = \emptyset$;
while $f(A_i) < \tau$ **do**
  $u = \mathrm{argmax}_{x \in S \setminus A_i} \frac{\Delta f_\tau(A_i, x)}{c(x)}$;
  $i = i + 1, A_i = A_{i-1} \cup \{u\}$;
end while
return A_i

Approximate Oracle

- **We analyse the greedy algorithm for SCSC given an approximate oracle to f**
  - Sketch of f
  - Noisy evaluations of f

Definition (ϵ-Approximate Oracle)

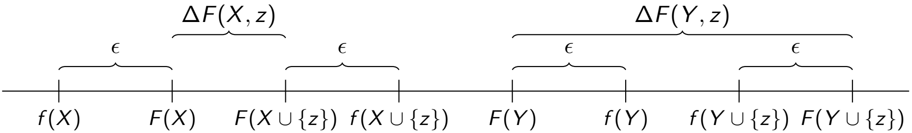
A function $F : 2^S \to \mathbb{R}_{\geq 0}$ is ϵ-approximate to $f : 2^S \to \mathbb{R}_{\geq 0}$ if for all $X \subseteq S$,

$$|f(X) - F(X)| \leq \epsilon.$$

Approximate Oracle

- F is not necessarily monotone submodular
  - Existing guarantees don't hold

Let $X \subseteq Y$, and $z \notin Y$.



Approximation Ratios

Theorem

Let A be the set returned by the greedy algorithm with a value oracle to ϵ-approximate oracle F. Then $f(A) \geq \tau - \epsilon$. And if $\mu > 4\epsilon c_{max}\rho / c_{min}$,

$$c(A) \leq \frac{\rho}{1 - \frac{4\epsilon c_{max}\rho}{c_{min}\mu}} \left(\ln \left(\frac{\alpha}{\beta} \right) + 2 \right) c(A^*).$$

- If $\epsilon = 0$, nearly reduces to existing result; $\rho \left(\ln \left(\frac{\alpha}{\beta} \right) + 1 \right)$ (Soma & Yoshida 2015)
- β can be very small

Approximation Ratios

Theorem

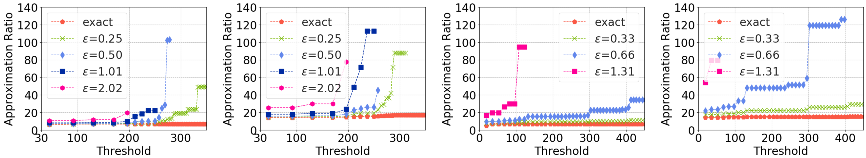
Let A be the set returned by the greedy algorithm with a value oracle to ϵ-approximate oracle F. Then $f(A) \geq \tau - \epsilon$. And if $\mu > 4\epsilon c_{max}\rho/c_{min}$, then for any $\gamma \in (0, 1 - 4\epsilon c_{max}\rho/c_{min}\mu)$,

$$c(A) \leq \frac{\rho}{1 - \frac{4\epsilon c_{max}\rho}{c_{min}\mu} - \gamma}\left(\ln\left(\frac{n\alpha\rho}{\gamma\mu}\right) + 2\right) c(A^*).$$

- No more β
- Incomparable

Application: Influence Threshold

- Find seed set of minimum cost such that expected propagation from seed set is at least τ
- Scalable influence estimator of Cohen et al. (2014)
  - Not submodular
  - ϵ-approximate
- Computed our approximation ratios



Thank you! Poster #168