Bicriteria Approximation Algorithms for the Submodular Cover Problem

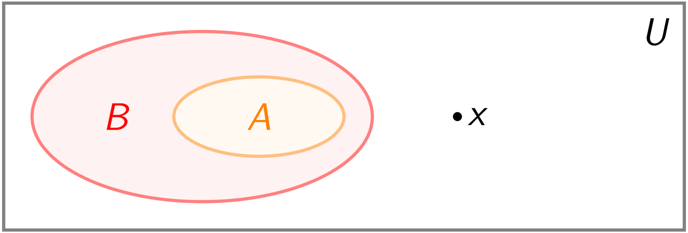
Wenjing Chen Victoria G. Crawford

Department of Computer Science&Engineering
Texas A&M University

Submodular Functions

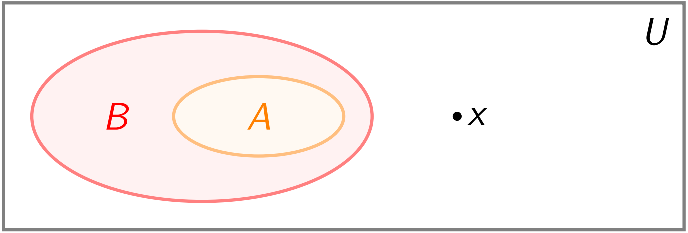
A function $f : 2^U \to \mathbb{R}$ defined on subsets of a ground set U of size n is submodular if for all $A \subseteq B \subseteq U$ and $x \notin B$,

$$f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B).$$



Monotone Functions

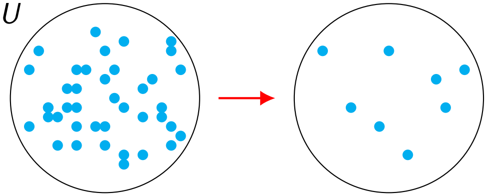
f is monotone if for all $A \subseteq B \subseteq U$, $f(A) \leq f(B)$.



Data Summarization

Condense data set U into a relatively small summary, i.e., find a subset of U

- Remove repetitive information
- Easier to run algorithms on, understand, load into memory, etc.



Data Summarization Example

Define f to measure the quality of a summary,

$$f(X) = \overbrace{|\cup_{x\in X} t(X)| - \gamma \sum_{x\in X} \sum_{x\in X, y\in X} \frac{|t(x)\cup t(y)|}{|t(x)\cap t(y)|}}^{\text{“summary quality”}}.$$

- The objective is submodular.
- If $\gamma = 0$, f is monotone.

Data Summarization: Optimization Setting

A "good" summary X should satisfy

- f should achieve a certain value
- $|X|$ should be as small as possible.

Problem Setup

Definition

Submodular Cover problem (SCP) is to find

$$\begin{aligned} & \arg\min |X| \\ s.t. \quad & f(X) \geq \tau. \end{aligned}$$

The optimal solution is denoted as OPT.

- ▷ The problem is NP-hard.
- ▷ We propose bicriteria approximation algorithms for this problem.

Problem Setup: Submodular Maximization

Definition

The Submodular Maximization Problem (SMP) is

$$\begin{aligned} & \arg\max f(X) \\ s.t. \quad & |X| \leq \kappa \end{aligned}$$

.

Problem Setup: Bicriteria Algorithm

- An (α, β)-bicriteria approximation algorithm for SCP returns a solution X such that $|X| \leq \alpha |OPT|$ and $f(X) \geq \beta\tau$.
- An (α, β)-bicriteria approximation algorithm for SMP returns a solution X such that $f(X) \geq \alpha f(OPT)$ and $|X| \leq \beta\kappa$.

Summarizing the Results

Table: Theoretical guarantees of a subset of algorithms in this paper

Alg name	soln size	number of queries
stoch-greedy-c	$(1+\alpha)\ln(3/\epsilon)$	$O(\frac{\alpha}{1+\alpha} n \ln(1/\delta)\ln^2(1/\epsilon)\log_{1+\alpha}(\lvert OPT\rvert))$
thresh-greedy-c	$\ln(2/\epsilon)+1$	$O(\frac{n}{\epsilon}\ln(\frac{\lvert OPT\rvert}{\epsilon}))$
stream-c ▲	$(1+\alpha)(2/\epsilon+1)\lvert OPT\rvert$	$O(\log(\lvert OPT\rvert)(\frac{n}{\epsilon}+\mathcal{T}((1+\alpha)\lvert OPT\rvert/\epsilon^2))$

▲: f is non-monotone

- All the algorithms in the table archieve $f(S) \geq (1-\epsilon)\tau$.
- Other results include the Regularized Submodular Cover Problem (RSCP).

MSCP

▷ Results for monotone objective: f is further assumed to be monotone. The problem is denoted as MSCP.

Monotone objective: Converting Theorem

Any randomized (γ, β)-bicriteria approximation algorithm for MSMP can be converted into a $((1+\alpha)\beta, \gamma - \epsilon)$-bicriteria approximation algorithm for MSCP w.h.p.

Query complexity:
$O(\log_{1+\alpha}(|OPT|) \ln(1/\delta) \mathcal{T}(n) / \ln(\frac{1-\gamma+\epsilon}{1-\gamma}))$.

Input: A MSCP instance with threshold τ, a (γ, β)-bicriteria approximation algorithm for MSMP where γ is in expectation, $\alpha > 0$, $\epsilon > 0$
Output: $S \subseteq U$

```
S_i ← ∅, ∀i ∈ {1, ..., ln(1/δ)/ ln((1−γ+ϵ)/(1−γ))}
g ← (1 + α)
while f(S_i) < (γ − ϵ)τ ∀i do
    for i ∈ {1, ..., ln(1/δ)/ ln((1−γ+ϵ)/(1−γ))} do
        S_i ← (γ, β)-bicriteria approximation for MSMP with objective function f_τ and budget g
    g ← (1 + α)g
return S
```

Monotone objective: Stochastic Greedy for cover

W.h.p, `stoch-greedy-c` is a $((1+\alpha)\lceil \ln(3/\epsilon) \rceil, 1-\epsilon)$-bicriteria approximation algorithm for MSCP.

Query complexity:
$O\left(\alpha n \ln(1/\delta) \ln^2(3/\epsilon) \log_{1+\alpha}(|OPT|)\right)$.

Input: ϵ, α, δ
Output: $S \subseteq U$

$S_i \leftarrow \emptyset \; \forall i \in \{1, ..., \ln(1/\delta)/\ln(2)\}$
$r \leftarrow 1, g \leftarrow 1+\alpha$
while $f(S_i) < (1-\epsilon)\tau \; \forall i$ **do**
  for $i \in \{1, ..., \ln(1/\delta)/\ln(2)\}$ **do**
    $R \leftarrow$ sample $\min\{n, n\ln(3/\epsilon)/g\}$ elements from U
    $u \leftarrow \text{argmax}_{x \in R} \Delta f_\tau(S_i, x)$
    $S_i \leftarrow S_i \cup \{u\}$
  $r \leftarrow r+1$
  if $r > \ln(3/\epsilon)g$ **then** $g \leftarrow (1+\alpha)g$
return $\text{argmin}\{|S_i| : f(S_i) \geq (1-\epsilon)\tau\}$

Nonmonotone case

- Results for general objective: f can be nonmonotone.

Non-Monotone Objective:

stream-c returns S for SCP such that $f(S) \geq (1-\epsilon)\tau$ and $|S| \leq (1+\alpha)(2/\epsilon)|OPT|$.

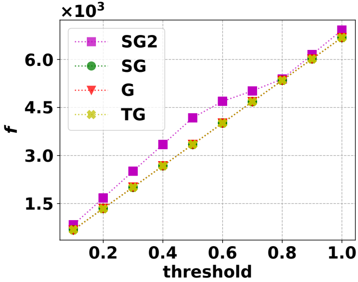
Query complexity:

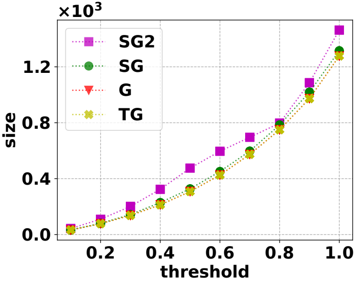
$\log_{1+\alpha}(|OPT|)\left(\frac{2n}{\epsilon} + \mathcal{T}\left((1+\alpha)\left(\frac{4}{\epsilon^2}|OPT|\right)\right)\right)$.

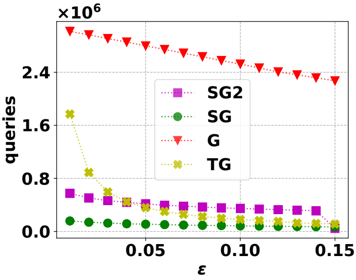
Input: ϵ, α
Output: $S \subseteq U$

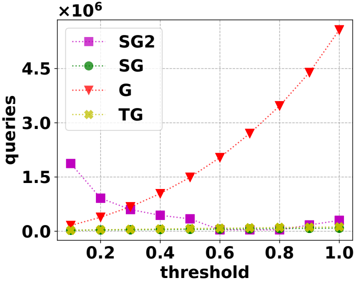
```
1: S ← ∅, S_1 ← ∅, ..., S_{2/ε} ← ∅
2: g ← 1 + α
3: while f(S) < (1 − ε)τ do
4:     for u ∈ U do
5:         if ∃j s.t. Δf(S_j, u) ≥ ετ/(2g) and |S_j| < 2g/ε then
6:             S_j ← S_j ∪ {u}
7:     S ← argmax{f(X) : X ⊆ ∪_{i=1}^{2/ε} S_i, |X| ≤ 2g/ε}
8:     g = (1 + α)g
9: return S
```

Monotone Experiments









Nonmonotone Experiments