CMPSCI 711: More Advanced Algorithms

Graphs 8: Submodular Maximization

Andrew McGregor

- A $1/2 - \epsilon$ approx for monotone submodular maximization.

  [Badanidiyuru et al., KDD 14]
- A $1/3 - \epsilon$ approx for non-monotone submodular maximization.

  [Chekuri, Gupta, Quanrud, ICALP 15]

Last Compiled: March 8, 2018

Submodularity

Definition

A function f from subsets of U is submodular if $\forall A \subset B$ and $u \in U \setminus B$,

$$f_A(u) \geq f_B(u) \quad \text{where} \quad f_X(Y) = f(X \cup Y) - f(X)$$

i.e., there are "diminishing returns". Assume $f(\emptyset) = 0$. For example,

- Coverage: If $U = \{S_1, S_2, \ldots\}$ where each $S_i \subseteq [n]$ and for $A \subseteq U$

$$f(A) = |\cup_{i \in A} S_i|$$

- Cuts: If U is the vertices of a graph $G = (U, E)$ and for $A \subseteq U$,

$$f(A) = \delta(A)$$

  where $\delta(A)$ is size of cut $(A, U \setminus A)$. Submodular because for $u \notin B$,

$$f_A(u) = |\Gamma(u) \setminus A| - |\Gamma(u) \cap A| \geq |\Gamma(u) \setminus B| - |\Gamma(u) \cap B| = f_B(u)$$

$(1/2 - \epsilon)$-Approx for Monotone Submodular Functions

Problem: Find A the maximizes $f(A)$ subject to $|A| \leq k$.

Algorithm:

- Make guesses $z = 1, (1+\epsilon), (1+\epsilon)^2, \ldots$ for $\text{OPT} = \max_{A:|A|\leq k} f(A)$
- For guess z: Let $A = \emptyset$ and then add each $u \in U$ to A if

$$f_A(u) \geq \frac{z}{2k} \quad \text{and} \quad |A| < k$$

Theorem

Algorithm returns a $(1-\epsilon)/2$ approximation if f is monotone, i.e.,

$$f(A) \leq f(B) \text{ for any } A \subseteq B \ .$$

Analysis

- Let $O = \{o_1, o_2, \dots\} = \arg\max_{A:|A|\leq k} f(|A|)$. Consider guess z

$$f(O) \leq z \leq (1+\epsilon)f(O)$$

- If $|A| = k$ and $A = \{u_1, \dots, u_k\}$ then,

$$f(A) = \sum_{i=1}^{k} \left(f(\{u_i, u_{i-1}, \dots, u_1\}) - f(\{u_{i-1}, \dots, u_1\})\right) \geq \frac{z}{2} \geq \frac{f(O)}{2}$$

- If $|A| < k$. Suppose $o_i \notin A$ and we'd picked A' when o_i arrived. Then,

$$f(A \cup \{o_i\}) - f(A) \leq f(A' \cup \{o_i\}) - f(A') < \frac{z}{2k}$$

and so

$$f(O) \leq f(O \cup A) \leq k \times \frac{z}{2k} + f(A) = z/2 + f(A)$$

and therefore $f(A) \geq f(O) - z/2 \geq \frac{(1-\epsilon)}{2} \times f(O)$.

$(1/3 - \epsilon)$-Approx for Non-monotone Submodular Functions

Problem: Find A the maximizes $f(A)$ subject to $|A| \leq k$ where f.

Algorithm:

- Assume we have guess z such that $\text{OPT} \leq z \leq (1+\epsilon)\text{OPT}$
- $S_1 \leftarrow \emptyset$, $B \leftarrow \emptyset$
- For each element e in the stream:
  - If $|S_1| < k$ and $f_{S_1}(e) > \frac{z}{3k}$ then $B \leftarrow B \cup \{e\}$
  - If $|B| = k/\epsilon$:
    - Remove a random element from B and add it to S_1
    - Remove any element f from B such that $f_{S_1}(f) \leq \frac{z}{3k}$
- Post-Processing: Return $S = \arg\max_{Z \in \{S_1, S_2\}} f(Z)$ where S_2 is the best solution from B.

Theorem

Algorithm returns a $(1-\epsilon)/3$ approximation in expectation.

Analysis: Preliminary Lemmas

Lemma

If $|S_1| = k$*, then* $f(S_1) \geq z/3 \geq \text{OPT}/3$*.*

Proof: Immediate from sub-modularity since each element added increases f by $z/(3k)$.

Lemma

If $|S_1| < k$*, then for the optimum set* O*.*

$$f(S_1 \cup O) \leq f(S_1) + f(O \cap B) + z/3$$

Proof:

- For each element $o \in O \setminus B$,

  $$f_{S_1}(o) \leq z/(3k)$$

  and so $f_{S_1}(O \setminus B) \leq |O \setminus B| \times z/(3k) \leq z/3$.
- By submodularity

  $$\begin{aligned} f(S_1 \cup O) = f(S_1) + f_{S_1}(O) &\leq f(S_1) + f_{S_1}(O \setminus B) + f_{S_1}(O \cap B) \\ &\leq f(S_1) + z/3 + f(O \cap B) \end{aligned}$$

Analysis: A General Technical Lemma

Lemma (Buchbinder et al. SODA 2014)

Given sets $A = \{a_1, a_2, \ldots\}$ and B, let A' be random subset of A where $a_i \in A'$ with probability p_i. Sampling need not be independent. Then,

$$\mathbb{E}\left[f(A' \cup B)\right] \geq (1-p)f(B) \quad \textit{where} \quad p = \max p_i \ .$$

Proof:

- Assume $p_1 \geq p_2 \geq \ldots$. Let $A_i = \{a_1, \ldots, a_i\}$ and $A'_i = A_i \cap A'$.
- Let $X_i = 1$ if $a_i \in A$ and $X_i = 0$ otherwise. Then $\mathbb{E}\left[f_B(A')\right] =$

$$= \mathbb{E}\left[\sum_{i=1}^{|A|} X_i f_{B \cup A'_{i-1}}(a_i)\right] \geq \mathbb{E}\left[\sum_{i=1}^{|A|} X_i f_{B \cup A_{i-1}}(a_i)\right] = \sum_{i=1}^{|A|} p_i f_{B \cup A_{i-1}}(a_i)$$

$$= \sum_{i=1}^{|A|} p_i \left(f_B(A_i) - f_B(A_{i-1})\right) = p_{|A|} f_B(A) + \sum_{i=1}^{|A|-1} (p_i - p_{i+1}) f_B(A_i)$$

$$= p_{|A|}(f(B \cup A) - f(B)) + \sum_{i=1}^{|A|-1} (p_i - p_{i+1})(f(B \cup A_i) - f(B)) \leq -pf(B)$$

Analysis: Algorithm Returns a $1/3 - \epsilon$ approximation

- If $|S_1| = k$, we're done. Hence assume,

$$f(S_1 \cup O) \leq f(S_1) + f(O \cap B) + z/3$$

- $f(S_2) \geq f(O \cap B)$ since we find opt solution amongst buffer. Hence,

$$f(S_1 \cup O) \leq f(S_1) + f(S_2) + z/3 \leq 2f(S) + z/3$$

- Since the probability an element in U gets added to S_1 is at most,

$$1 - \left(1 - \frac{1}{k/\epsilon}\right)^k ,$$

  the technical lemma gives

$$\mathbb{E}\left[f(S_1 \cup O)\right] \geq \left(1 - \frac{1}{k/\epsilon}\right)^k f(O) \geq (1-\epsilon)\mathrm{OPT}$$

- Therefore

$$\begin{aligned}\mathbb{E}\left[f(S)\right] &= \mathbb{E}\left[f(S_1 \cup O)\right]/2 - z/6 \\ &= (1-\epsilon)f(O)/2 - (1+\epsilon)f(O)/6 = f(O)(1-2\epsilon)/3\end{aligned}$$