| | |
|---|---|
| **Algorithms for Massive Data Science** | **Lecture date: April 1 and 3, 2019** |
| **Instructor: Ben Moseley** | **Scribe: Arash Haddadan** |

# 1 Partitioning based submodular optimization

In this lecture we introduce a partitioning based greedy algorithm for submodular maximization. First, we review the preliminaries of submodular function maximization.

## 1.1 Preliminaries on submodular maximization

For the whole lecture let $U$ be a finite ground set, or the universe, where $|U| = n$. Let $f : 2^U \to \mathbb{R}$.

**Definition 1** *Function $f$ is submodular if*

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \textit{for } A, B \subseteq U. \tag{1}$$

For $i \in U$ and $A \subseteq U$, let $f_A(i) = f(A \cup \{i\}) - f(A)$. Then $f$ is submodular if and only if $f_A(i) \geq f_B(i)$ for all $A \subseteq B \subseteq U$ and $i \in U \setminus B$. A submodular function $f$ is *monotone* if $f_A(i) \geq 0$ for all $A \subseteq U$ and $i \in U$.

**Problem 1 (Submodular coverage)** *Given universe $U$ and monotone submodular function $f : 2^U \to \mathbb{R}_+$, find a set $S \subseteq U$ with $k$ elements such that $f(S)$ is maximized.*

Nemhauser, Wolsey and Fisher [2] showed a $(1 - \frac{1}{e})$-approximation for this problem. Their approximation algorithm is the following greedy procedure.

**Algorithm 1** GREEDY-SUB($U$)

```
1: X ← ∅
2: while |X| < k do
3:     u ← arg max_{u∈U} f_X(u)
4:     X ← X ∪ u
```

**Theorem 1 ([2])** GREEDY-SUB *is a $(1 - \frac{1}{e})$-approximation algorithm for the submodular coverage problem.*

## 1.2 A distributed algorithm for submodular coverage

Our goal in this lecture is to prove the following theorem.

**Theorem 2 (Barbosa et. al [1])** *There is 2-round MPC approximation algorithm which achieves approximation ratio of $\frac{1}{2}(1 - \frac{1}{e})$ for submodular coverage. The algorithm uses $O(\max\{\frac{n}{m}, mk\})$ memory per machine and $O(mk)$ communication.*

In this section we introduce the algorithm and prove a simple observation about the GREEDY-SUB algorithm that will become handy later. Next, we introduce continuous extensions of submodular functions. In particular we introduce Lovász extension and its properties. We finish this lecture with the proof of Theorem 2.

The distributed partitioning algorithm for submodular coverage is a follows:

- Randomly partition $U$ into sets $U_1, \ldots, U_m$, with $|U_i| = \frac{n}{m}$. Sub-universe $U_i$ is the set of elements on machine $i$.
- Each machine $i$ computes $G_i = \text{GREEDY-SUB}(U_i)$.
- One machine computes $T = \text{GREEDY-SUB}(\bigcup_{i=1}^{m} G_i)$.
- Let $G^* = \arg\max_{i=1,\ldots,m}\{f(G_i)\}$.
- Output $\arg\max\{f(G^*), f(T)\}$.

In fact, it is essential to distribute the data randomly between the machines.

**Exercise 1** *Prove that if the data is distributed deterministically the algorithm above is not an $O(1)$-approximation algorithm.*

The following lemma will be useful in proving Theorem 2.

**Lemma 3** *Let $A, B \subset U$ and $A \cap B = \emptyset$. Suppose that for each $e \in B$ we have* GREEDY-SUB$(A \cup \{e\}) =$ GREEDY-SUB$(A)$. *Then* GREEDY-SUB$(A \cup B) =$ GREEDY-SUB$(A)$.

**Proof:** Suppose not, and choose the smallest $i$ such that the $i$-th element $e_i^*$ added by GREEDY-SUB$(A \cup B)$ is different from the $i$-th element $e_i$ added by GREEDY-SUB$(A)$. The additive value of $e_i^*$ is more than $e_i$, so GREEDY-SUB$(A \cup \{e_i^*\})$ would pick $e_i^*$. This is a contradiction. □

### 1.3 Lovász extension

Submodular functions are discrete set functions. A common technique is to extend the definition to the continuous setting for algorithm analysis. There are two common continuous extensions: (1) Lovász extension and (2) multi-linear extention.

For the purpose of this lecture we only need Lovász extension.

**Definition 2** *Let $f : 2^U \to \mathbb{R}_+$ be a submodular function. The Lovász extension of $f$ is function $f^- : [0,1]^U \to \mathbb{R}_+$ such that*

$$f^-(X) = \mathbb{E}_{\theta \sim Uniform(0,1)}[f(\{i : X_i \geq \theta\})], \quad \text{for } X \in [0,1]^U. \tag{2}$$

For a set $S \subset U$, let $\mathbf{1}_S \in \mathbb{R}^U$ be the characteristic vector of $S$. Lovász extension has the following properties:

1. $f^-(\mathbf{1}_S) = f(S)$,
2. $f^-$ is a convex function,
3. $f^-(c \cdot X) \geq c \cdot f^-(X)$ for a scalar $c \in [0,1]$.

### 1.4 Proof of Theorem 2

Let OPT be the optimal solution to the submodular coverage problem. Let $\mathcal{V}(\frac{1}{m})$ be a random subset of $U$ were each element in $U$ is in $\mathcal{V}(\frac{1}{m})$ with probably $\frac{1}{m}$ independently at random. Also, let $p \in [0,1]^U$ be the vector defined as follows:

$$p_e = \Pr_{A \sim \mathcal{V}(\frac{1}{m})}[e \in \textsc{Greedy-Sub}(A \cup \{e\})] \text{ for } e \in \mathrm{OPT} \text{ and } p_e = 0 \text{ otherwise.} \tag{3}$$

**Lemma 4** *Let $S$ be a random subset of $U$ such that $\mathbb{E}[\boldsymbol{1}_S] = c \cdot p$. Then, $\mathbb{E}[f(S)] \geq c \cdot f^-(p)$.*

**Proof:** We have

$$\begin{aligned}
\mathbb{E}[f(S)] &= \mathbb{E}[f^-(1_S)] && \text{(by first property of Losász ext.)} \\
&\geq f^-(\mathbb{E}[1_S]) && \text{($f^-$ is convex by second property, and Jensen's inequality)} \\
&= f^-(c \cdot p) && (\mathbb{E}[\mathbf{1}_S] = c \cdot p) \\
&\geq c \cdot f^-(p) && \text{(Third property of Lovász ext.)}
\end{aligned}$$

□

The next lemma characterizes the quality of the solution $G_i$ constructed on machine $i$ using the Lovász extension.

**Lemma 5** *For each machine $i = 1, \ldots, m$ we have $\mathbb{E}(f(G_i))] \geq (1 - \frac{1}{e}) f^-(\boldsymbol{1}_{\mathrm{OPT}} - p)$.*

**Proof:** Let $O_i = \{e \ : \ e \in \mathrm{OPT} \text{ and } e \notin \textsc{Greedy-Sub}(U_i \cup \{e\})\}$ for $i = 1, \ldots, m$. By Lemma 3 we have $\textsc{Greedy-Sub}(U_i) = \textsc{Greedy-Sub}(U_i \cup O_i)$. Moreover, we have $f(G_i) \geq (1 - \frac{1}{e}) f(O_i)$ by Theorem 1. Notice that $U_i \sim \mathcal{V}(\frac{1}{m})$. Hence, we have

$$\Pr[e \in O_i] = 1 - \Pr[e \notin O_i] = 1 - p_e \text{ for } e \in U.$$

This implies $\mathbb{E}[\mathbf{1}_{O_i}] = \mathbf{1}_{\mathrm{OPT}} - p$. Therefore,

$$\begin{aligned}
\mathbb{E}(f(G_i)) &\geq (1 - \frac{1}{e}) \mathbb{E}(f(O_i)) \\
&\geq (1 - \frac{1}{e}) f^-(\mathbf{1}_{\mathrm{OPT}} - p) && \text{(By Lemma 4)}
\end{aligned}$$

□

The following lemma bounds the quality of the final solution obtained (which we denoted earlier by $T$) in terms of $f^-(p)$.

**Lemma 6** *We have $\mathbb{E}[\textsc{Greedy-Sub}(\bigcup_{i=1}^m G_i)] \geq (1 - \frac{1}{e}) f^-(p)$.*

**Proof:** Let $S = \bigcup_{i=1}^m G_i$. By Theorem 1 we have $f(\textsc{Greedy-Sub}(S)) \geq (1 - \frac{1}{e}) f(\mathrm{OPT} \cap S)$, since $\mathrm{OPT} \cap S$ is a feasible solution for the submodular coverage problem on sub-universe $S$. Fix edge $e \in \mathrm{OPT}$. We have

$$\begin{aligned}
\Pr[e \in S] &= \Pr[e \in \textsc{Greedy-Sub}(U_i) | e \in U_i] \\
&= \Pr_{A \sim \mathcal{V}(\frac{1}{m})}[e \in \textsc{Greedy-Sub}(A \cup \{e\})] \\
&= p_e.
\end{aligned}$$

This implies that $\mathbb{E}[\mathbf{1}_{\mathrm{OPT}\cap S}] = p$. By Lemma 4 we can conclude

$$\begin{aligned}\mathbb{E}[f(\textsc{Greedy-Sub}(S))] &\geq (1-\frac{1}{e})\mathbb{E}[f(S\cap \mathrm{OPT})] \\ &\geq (1-\frac{1}{e})f^-(p).\end{aligned}$$

□

We can now prove the main result.

**Proof of Theorem 2.** Let $D$ be the output of the distributed partitioning algorithm described in Section 1.2. We have

$$f(D) \geq \max\{f(T), f(G_1), \ldots, f(G_m)\}. \tag{4}$$

By Lemmas 5 and 6 we have

$$2\mathbb{E}[f(D)] \geq (1-\frac{1}{e})(f^-(p) + f^-(1_{\mathrm{OPT}} - p)).$$

**Claim 7** *We have $f^-(p) + f^-(1_{\mathrm{OPT}} - p) \geq f(\mathrm{OPT})$.*

**Proof:** Let $\mathrm{OPT} = \{e_1, \ldots, e_k\}$. For $i = 1, \ldots, k$, let $S_i = \{e_1, \ldots, e_i\}$, and let $S_0 = \emptyset$. For vector $q \in [0,1]^U$ denote by $X(q,\theta)$ the of elements $e_i$ in OPT where $q_{e_i} \geq \theta$, and $X_i(q,\theta) = X(q,\theta)\cap S_i$.

By submodularity of $f$, for $i = 1, \ldots, k$ we have $f_{X_{i-1}(q,\theta)}(e_i) \geq f_{S_{i-1}}(e_i)$. Moreover

$$\begin{aligned}f^-(q) &= \int_{\theta=0}^{1} \sum_{i=1}^{k} f_{X_{i-1}(q,\theta)}(e_i) \\ &\geq \int_{\theta=0}^{1} \sum_{i=1}^{k} f_{S_{i-1}}(e_i) \\ &\geq \sum_{i=1}^{k} q_{e_i} f_{S_{i-1}}(e_i).\end{aligned}$$

Let $q = p$, then we have $f^-(q) \geq \sum_{i=1}^{k} p_{e_i} f_{S_{i-1}}(e_i)$. Now, let $q = \mathbf{1}_{\mathrm{OPT}} - p$. We get $f^-(\mathbf{1}_{\mathrm{OPT}} - p) \geq \sum_{i=1}^{k}(1-p_{e_i})f_{S_{i-1}}(e_i)$. Combining the two we get

$$f^-(p) + f^-(\mathbf{1}_{\mathrm{OPT}} - p) \geq \sum_{i=1}^{k} f_{S_{i-1}}(e_i) = f(\mathrm{OPT})$$

□

This conclude the proof of Theorem 2. □

## References


[1] Rafael Barbosa, Alina Ene, Huy Le Nguyen, and Justin Ward. The power of randomization: Distributed submodular maximization on massive datasets. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 1236–1244. JMLR.org, 2015.

[2] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions–i. *Math. Program.*, 14(1):265–294, December 1978.