Submodular functions and random k-subsets

Yuval Filmus

April 11, 2012

Abstract

We relate the average value of a submodular function on k-subsets to various quantities, including its value on the empty set, on the universe and on the k-subset maximizing its value. As a result, we deduce that on non-negative functions, the random set algorithm has an approximation ratio of $k(n-k)/n(n-1)$.

1 Introduction

A set-function f defined on subsets of a set U is *submodular* if

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$

for any two subsets A, B of U. Denote all subsets of U of size k by $\binom{U}{k}$. We define

$$E_k(f) = \mathop{\mathbb{E}}_{A \in \binom{U}{k}} f(A), \qquad M_k(f) = \max_{A \in \binom{U}{k}} f(A).$$

How small can $E_k(f)$ be, compared to other prominent values of f? We answer this question in the two theorems below.

Theorem 1. *Let f be a submodular function on a set U of size n. For $0 \leq k \leq n$,*

$$E_k(f) \geq \frac{k}{n} f(U) + \frac{n-k}{n} f(\emptyset).$$

This theorem is tight for linear functions.

Theorem 2. *Let f be a submodular function on a set U of size n. For $0 \leq k \leq n$ and any set $O \in \binom{U}{k}$,*

$$E_k(f) \geq \frac{k(n-k)}{n(n-1)}(f(O) + f(U \setminus O)) + \frac{k(k-1)}{n(n-1)} f(U) + \frac{(n-k)(n-k-1)}{n(n-1)} f(\emptyset).$$

This theorem is tight for the directed cut function on the directed complete bipartite graph $K_{k,n-k}$, taking as O one side of the bipartition.

Corollary 3. *Let f be a submodular function on a set U of size n. For $0 \leq k \leq n$,*

$$E_k(f) \geq \frac{k(n-k)}{n(n-1)} M_k(f) + \frac{k(k-1)}{n(n-1)} f(U) + \frac{(n-k)(n-k-1)}{n(n-1)} f(\emptyset).$$

If f is non-negative, then Corollary 3 implies that

$$\frac{E_k(f)}{M_k(f)} \geq \frac{k(n-k)}{n(n-1)}.$$

This is the approximation ratio of the random set algorithm for the problem of maximizing f over $\binom{U}{k}$.

If f is symmetric ($f(A) = f(U \setminus A)$), then the same argument used to deduce the corollary from the theorem shows that the approximation ratio of the random set algorithm for the problem of maximizing f over $\binom{U}{k}$ is

$$\frac{E_k(f)}{M_k(f)} \geq \frac{2k(n-k)}{n(n-1)}.$$

2 Proofs

Theorem 1. *Let f be a submodular function on a set U of size n. For $0 \leq k \leq n$,*

$$E_k(f) \geq \frac{k}{n} f(U) + \frac{n-k}{n} f(\emptyset).$$

Proof. The proof is by induction on k. When $k = 0$, there is nothing to prove. Assuming the theorem is true for some k, we prove it for $k+1$. Let $A \in \binom{U}{k}$. Repeated application of submodularity gives

$$\sum_{x \in U \setminus A} f(A \cup \{x\}) \geq (n-k-1) f(A) + f(U).$$

Averaging over all choices of A, we deduce

$$(n-k) E_{k+1}(f) \geq (n-k-1) E_k(f) + f(U).$$

The induction hypothesis gives

$$\begin{aligned} E_{k+1}(f) &\geq \frac{n-k-1}{n-k} E_k(f) + \frac{1}{n-k} f(U) \\ &\geq \frac{n-k-1}{n} f(\emptyset) + \frac{(n-k-1)k}{n(n-k)} f(U) + \frac{1}{n-k} f(U) \\ &= \frac{n-k-1}{n} f(\emptyset) + \frac{k+1}{n} f(U). \end{aligned}$$

□

Theorem 2. *Let f be a submodular function on a set U of size n. For $0 \leq k \leq n$ and any set $O \in \binom{U}{k}$,*

$$E_k(f) \geq \frac{k(n-k)}{n(n-1)} (f(O) + f(U \setminus O)) + \frac{k(k-1)}{n(n-1)} f(U) + \frac{(n-k)(n-k-1)}{n(n-1)} f(\emptyset).$$

Proof. For subsets $A \subseteq U$, we can define a function f_A on $U \setminus A$ by $f_A(B) = f(A \cup B)$. It is easy to check that f_A is also submodular. We will use Theorem 1 repeatedly on such functions. We denote the restriction of a set-function g to domain X by $g|_X$.

Let $\overline{O} = U \setminus O$. We have

$$\binom{n}{k} E_k(f) = \sum_{A \in \binom{U}{k}} f(A) = \sum_{B \subseteq O} \sum_{C \in \binom{\overline{O}}{k-|B|}} f(B \cup C).$$

We can interpret the sum on C as an average:

$$\binom{n}{k}E_k(f) = \sum_{B\subseteq O}\binom{n-k}{k-|B|}E_{k-|B|}(f_B|_{\overline{O}}).$$

Using Theorem 1, we deduce

$$\begin{aligned}\binom{n}{k}E_k(f) &\geq \sum_{B\subseteq O}\binom{n-k}{k-|B|}\left(\frac{k-|B|}{n-k}f_B|_{\overline{O}}(\overline{O}) + \frac{n-2k+|B|}{n-k}f_B|_{\overline{O}}(\emptyset)\right)\\ &= \sum_{B\subseteq O}\binom{n-k-1}{k-|B|-1}f_{\overline{O}}(B) + \binom{n-k-1}{k-|B|}f(B).\end{aligned}$$

Conditioning on the size of B, we can interpret the latter sum as a linear combination of $E_t(f_{\overline{O}})$ and $E_t(f|_O)$:

$$\binom{n}{k}E_k(f) \geq \sum_{t=0}^{k}\binom{k}{t}\left(\binom{n-k-1}{k-t-1}E_t(f_{\overline{O}}) + \binom{n-k-1}{k-t}E_t(f|_O)\right).$$

Using Theorem 1 again,

$$\begin{aligned}\binom{n}{k}E_k(f) &\geq \sum_{t=0}^{k}\binom{k}{t}\binom{n-k-1}{k-t-1}\left(\frac{t}{k}f_{\overline{O}}(O) + \frac{k-t}{k}f_{\overline{O}}(\emptyset)\right)\\ &+ \binom{k}{t}\binom{n-k-1}{k-t}\left(\frac{t}{k}f|_O(O) + \frac{k-t}{k}f|_O(\emptyset)\right)\\ &= \sum_{t=0}^{k}\binom{k-1}{t-1}\left(\binom{n-k-1}{k-t-1}f(U) + \binom{n-k-1}{k-t}f(O)\right)\\ &+ \binom{k-1}{t}\left(\binom{n-k-1}{k-t-1}f(\overline{O}) + \binom{n-k-1}{k-t}f(\emptyset)\right)\\ &= \binom{n-2}{k-2}f(U) + \binom{n-2}{k-1}f(O) + \binom{n-2}{k-1}f(\overline{O}) + \binom{n-2}{k}f(\emptyset).\end{aligned}$$

□