Greedy Algorithms and Submodularity[1]

In this note we generalize the set cover (and related) problems using the abstraction of *submodular functions*. Submodularity is a fundamental concept in discrete optimization, and arises in many surprising domains. Therefore, these generalizations have many applications.

Let V be a finite universe. A set function $f : 2^V \to \mathbb{R}$ assigns values to every subset of V. We call a set function non-negative if its range is non-negative reals. We call a set function monotone if for any two $A \subseteq B$, we have $f(A) \leq f(B)$. We call a set function submodular if it satifies the following

$$\text{For any two subsets } A \text{ and } B, \quad f(A) + f(B) \geq f(A \cap B) + f(A \cup B) \qquad \text{(Submodularity)}$$

There is an equivalent definition. Function f is submodular if

$$\text{For any } A \subseteq B \text{ and } i \in V \setminus B, \quad f(A \cup i) - f(A) \geq f(B \cup i) - f(B) \qquad (1)$$

This latter definition captures the "diminishing marginal utilities" dictum. If $f(S)$ indicates once happiness on obtaining a subset S, then the above states that the marginal increase in happiness upon receiving an item i only goes down as one has more.

Exercise: *Prove that two notions are equivalent.*

Submodular Set Cover

- There are many examples of submodular functions, but the one that connects to set cover is the following. Suppose the universe $V = [m]$ the indices of the sets in a set cover instance. Given any subset $I \subseteq [m]$, define $f(I) := |\bigcup_{j \in I} S_j|$. It is not too hard to show f is a non-negative, monotone submodular function. And, the set cover problem asks : find $I \subseteq [m]$ with the minimum cost such that $f(I) = f([m])$. The *submodular set cover* problem generalizes this for any non-negative monotone submodular function. More precisely, given *oracle* access to a non-negative, monotone submodular function over a universe V, costs $c(i)$ for $i \in V$, find $I \subseteq V$ such that $f(I) \geq f(V)$ and $c(I) := \sum_{j \in I} c(j)$ is minimized.

- **Greedy Algorithm.** The following greedy algorithm and its analysis generalizes the Set Cover greedy algorithm and analysis. We assume the range of the function is non-negative integers.

  procedure SUBMODULAR SET COVER($f : 2^V \to \mathbb{Z}_{\geq 0}, c(i) \ : \ i \in V$):
      Initialize $X \leftarrow \emptyset$.
      while $f(X) < f(V)$ **do**:
          Pick $j \in V$ which minimizes $\frac{c(j)}{f(X \cup j) - f(X)}$.
          $X \leftarrow X \cup j$.
      return X.

- **Analysis.** The analysis is very similar to the set-cover analysis we did.

[1] *Lecture notes by Deeparnab Chakrabarty. Last modified : 14th Dec, 2021*
These have not gone through scrutiny and may contain errors. If you find any, or have any other comments, please email me at deeparnab@dartmouth.edu. Highly appreciated!

Theorem 1. When the range of f is integer valued, SUBMODULAR SET COVER is a H_N-approximation, where $N = f(V)$.

Proof. Let us rename the elements such that we pick $\{1, 2, \ldots, r\}$ in that order into the set X. For simplicity, we let $Z_i := \{1, 2, \ldots, i-1\}$ be the set just before the ith loop. The greedy rule tells us that for all $1 \leq i \leq r$,

$$\forall j \in V, \ \ \frac{c(i)}{f(Z_{i-1} \cup \{i\}) - f(Z_{i-1})} \leq \frac{c(j)}{f(Z_{i-1} \cup \{j\}) - f(Z_{i-1})} \tag{2}$$

If we let O be the optimal solution, we get by an "averaging argument"

$$c(i) \leq (f(Z_i) - f(Z_{i-1})) \cdot \frac{\texttt{opt}}{\sum_{j \in O} f(Z_{i-1} \cup \{j\}) - f(Z_{i-1})} \tag{3}$$

Now we use submodularity of f to argue

$$\sum_{j \in O} f(Z_{i-1} \cup \{j\}) - f(Z_{i-1}) \geq f(Z_{i-1} \cup O) - f(Z_{i-1}) \tag{4}$$

Indeed, the right hand side is the sum of $|O|$ different marginals by adding the elements of O to Z_{i-1} in any order you want. Submodularity implies that the LHS term-wise dominates this sum.

Monotonicity of f then implies

$$\sum_{j \in O} f(Z_{i-1} \cup \{j\}) - f(Z_{i-1}) \geq f(O) - f(Z_{i-1}) \geq f(V) - f(Z_{i-1})$$

since O is a valid solution. Substituting in (3), we get that for all $i \in I$,

$$c(i) \leq \texttt{opt} \cdot \frac{f(Z_i) - f(Z_{i-1})}{f(V) - f(Z_{i-1})} \ \ \Rightarrow \ \ \mathsf{alg} \leq \texttt{opt} \cdot \sum_{i=1}^{r} \left(\frac{f(Z_i) - f(Z_{i-1})}{f(V) - f(Z_{i-1})} \right)$$

Now we use the integrality of f to argue that the summation is maximized when the marginals in the numerator are exactly 1, and in that case the summation becomes H_N, where $N = f(V)$. □

Exercise: *Generalize the charging argument for set-cover to prove that* SUBMODULAR SET COVER *is in fact an H_d approximation where $d = \max_{i \in V} f(\{i\})$.*

Constrained Submodular Maximization

- We want to solve the following problem: given an integer k, a non-negative, monotone submodular function $f : 2^V \to \mathbb{R}_{\geq 0}$, find a set $I \subseteq V$ with $|S| = k$ which maximizes $f(S)$. Here is the natural greedy algorithm.

```
procedure GREEDY SUBMODULAR MAXIMIZATION(f : 2^V → ℝ_{≥0}, k ∈ ℤ):
    Initialize X ← ∅.
    for i = 1 to k do:
        Find element i maximizing f(X ∪ i) − f(X).
        X ← X ∪ i.
    return X.
```

- **Analysis Sketch.** Let X_i be the set the algorithm maintains at the beginning i. So $X_1 = \emptyset$ and the algorithm returns X_{k+1}. Let the optimal set be O. By the greedy property, we have that for any $1 \le i \le k$ and any element $e \in V$

$$f(X_{i+1}) - f(X_i) \geq f(X_i \cup o) - f(X_i)$$

If we average over all $o \in O$, we get

$$\begin{aligned} f(X_{i+1}) - f(X_i) &\geq \frac{1}{k}\sum_{o \in O}\left(f(X_i \cup o) - f(X_i)\right) && (5) \\ &\geq \frac{1}{k}\left(f(X_i \cup O) - f(X_i)\right) \qquad \text{because of submodularity.} && (6) \\ &\geq \frac{1}{k}\left(f(OPT) - f(X_i)\right) \qquad \text{because of monotonicity.} && (7) \end{aligned}$$

(6) is a result of submodularity, as in (4). Now, with a little bit of arithmetic, one can get the following theorem. To rewrite (7), we get

$$f(X_{i+1}) \geq \frac{1}{k} f(OPT) + \left(1 - \frac{1}{k}\right) f(X_i)$$

The rest is arithmetic

$$\begin{aligned} f(X_k) &\geq f(OPT)\frac{1}{k}\left(1 + \left(1 - \frac{1}{k}\right) + \cdots + \left(1 - \frac{1}{k}\right)^{k-1}\right) + \left(1 - \frac{1}{k}\right)^k f(\emptyset) \\ &\geq f(OPT)\left(1 - (1 - 1/k)^k\right) \end{aligned}$$

since $f(\emptyset) \geq 0$. Using the fact that $(1 - 1/k)^k < 1/e$ where $e = 2.71...$, we get that the greedy algorithm is a $(1 - 1/e)$-factor approximation algorithm.

Theorem 2. GREEDY SUBMODULAR MAXIMIZATION gives an $(1 - 1/e)$-approximation.

Notes

The result on submodular set cover is from the paper [3] by Wolsey. The result on constrained submodular maximization is from the paper [2] by Fisher, Nemhauser, and Wolsey. Both these are extremely influential papers, and started off research on submodular optimization which has really flourished in the last 20 years. An excellent survey can be found in the paper [1] by Buchbinder and Feldman.

References

[1] Niv Buchbinder and Moran Feldman. Submodular functions maximization problems. In *Handbook of Approximation Algorithms and Metaheuristics, Second Edition*, pages 753–788. 2018.

[2] M. L. Fisher, G. L. Nemhauser, and L. A. Wolsey. An analysis of approximations for finding a maximum weight Hamiltonian circuit. *Oper. Res.*, 27(4):799–809, 1979.

[3] L. Wolsey. An Analysis of the Greedy Algorithm for the Submodular Set Covering Problem. *Combinatorica*, 2(4):385–393, 1982.