Greedy Algorithms 2

Jie Wang

University of Massachusetts Lowell
Department of Computer Science

Independent Systems Greedy

- Want to figure out exactly when a greedy strategy does produce optimal solution. We first look at maximization.
- **Independent system**: Let E be a finite set and $\mathcal{I}$ a collection of subsets of E. Call $(E, \mathcal{I})$ an *independent system* if E is closed under inclusion:

$$\text{If } A \in \mathcal{I} \text{ and } B \subset A \text{ then} B \in \mathcal{I}.$$

- Example 1: E is any set of vectors in some vector space; $\mathcal{I}$ is a collection of sets of linearly independent vectors.
- Example 2: E is a set of edges of an undirected graph; $\mathcal{I}$ is the collection of all acyclic sets of edges.
- Example 3: E is the edges of an undirected graph; $\mathcal{I}$ is the collection of sets of independent edges (i.e. no more than one edge sharing a vertex).

Activity Selection

Example 4:

- E is the set of all intervals $[s_i, f_i]$.
- $\mathcal{I}$ is a collection of subsets of non-overlapping intervals in E.
- Then $(E, \mathcal{I})$ is an independent system for permissable activities.

MAX-ISS and Greedy Algorithm

- Maximum Independent Subset (MAX-ISS):
  - Find $I \in \mathcal{I}$ such that $c(I)$ is maximum.
  - c: $E \to R^+$ is a profit measure on each element of E.
- Greedy Algorithm: Sort elements of E such that

$$c(e_1) \geq c(e_2) \geq \cdots \geq c(e_n).$$

$I = \emptyset$
for $i = 1$ **to** n
    if $I \cup \{e_i\} \in \mathcal{I}$
        $I = I \cup \{e_i\}$
return I

Approximation Ratio

- Let $F \subseteq E$ that contains an independent subset. An independent subset $I \subseteq F$ is *maximal independent subset* of F if no independent subset of F contains I as a proper subset. Let

$$\begin{aligned} u(F) &= \min\{|I| \mid I \text{ is a maximal indepdendent subset of } F\} \\ v(F) &= \max\{|I| \mid I \text{ is an independent subset of } F\} \end{aligned}$$

- **Theorem 1.** Let I_G be the solution obtained from the Greedy Algorithm and I^* a maximum solution to MAX-ISS . Then

$$1 \leq \frac{c(I^*)}{c(I_G)} \leq \max_{F \subseteq E} \frac{v(F)}{u(F)}.$$

Proof

- Let $E = \{e_1, e_2, \cdots, e_n\}$ with $c(e_1) \geq c(e_2) \geq \cdots \geq c(e_n)\}$.
- Let $E_i = \{e_1, \cdots, e_i\}$ and $I_i = E_i \cap I_G$. Then I_i is a maximum independent subset of E_i.
    - If not, then there exists $e_j \in E_i - I_G$ such that $I_i \cup \{e_j\}$ is independent.
    - Thus, $I_{j-1} \cup \{e_j\} \subseteq I_i \cup \{e_j\}$, and hence independent.
    - This implies that e_j must be included to I_G by the Greedy Algorithm at the j-th iteration, contradicting to $e_j \notin I_G$.
- Thus, $|I_i| \geq u(E_i)$.
- It follows from $E_i \cap I^*$ being independent that $|E_i \cap I^*| \leq v(E_i)$.
- For $i = 1, \cdots, n$,

$$|I_i| - |I_{i-1}| = \begin{cases} 1, \text{if } e_i \in I_G, \\ 0, \text{otherwise.} \end{cases}$$

Proof Continued

- Then

$$\begin{aligned} c(I_G) &= \sum_{e_i \in I_G} c(e_i) = \sum_{i=1}^{n} c(e_i)(|I_i| - |I_{i-1}|) \\ &= \sum_{i=1}^{n-1} |I_i|(c(e_i) - c(e_{i+1})) + |I_n| c(e_n) \\ &\geq \sum_{i=1}^{n-1} u(E_i)(c(e_i) - c(e_{i+1})) + u(E_n) c(e_n). \end{aligned}$$

- Likewise,

$$c(I^*) = \sum_{i=1}^{n-1} |I_i^*|(c(e_i) - c(e_{i+1})) + |I_n^*| c(e_n),$$

where $I_i^* = E_i \cap I^*$.

Proof Continued

- Let
$$r = \max_{F \subseteq E} \frac{v(F)}{u(F)},$$
then $r \geq 1$.
- It is evident that for any particular set E_i, we have $v(E_i)/u(E_i) \leq r$. Thus,

$$\begin{aligned}
c(I^*) &\leq \sum_{i=1}^{n-1} v(E_i)(c(e_i) - c(e_{i+1})) + v(E_n)c(e_n) \\
&\leq \sum_{i=1}^{n-1} r \cdot u(E_i)(c(e_i) - c(e_{i+1})) + r \cdot u(E_n)c(e_n) \\
&\leq r \cdot c(I_G). \quad \square
\end{aligned}$$

Matroids

- A matroid is an independent system $(E, \mathcal{I})$ such that for any $F \subseteq E$:

$$v(F) = u(F).$$

- A weighted matroid is a matroid where there is a weight for each element, and $|A|$ is defined to be

$$w(A) = \sum_{e \in A} w(e).$$

Matroid Example 1

Graph matroid: For a given graph G, let E be the set of edges and $\mathcal{I}$ the family of edge sets that form an acyclic subgraphs of G. Then $(E, \mathcal{I})$ is a matroid.

- **Proof**. Let $F \subseteq E$. Suppose that the subgraph (V_F, F) has m connected components. Then the maximal acyclic subgraph for each connected component $C = (V_C, E_C)$ is simply a spanning tree of C with $|V_C| - 1$ edges. Thus, every maximal acyclic subgraph of (V_F, F) has exactly $|V| - m$ edges. $\square$

Matroid Example 2

Example 2. Let E be a finite set. Given a subset $A \subseteq E$, let

$$C_A = \{B \subseteq E \mid A \not\subset B\}.$$

Then (E, C_A) is a matroid.
Proof. Let $F \subseteq E$.

- If $A \not\subset F$, then $F \in C_A$ and so F is independent, which is also maximal. Hence, $u(F) = v(F) = |F|$.
- If $A \subseteq F$, then every maximal independent subset of F must be in the form of $F - \{e\}$ for some $e \in A$. Thus, $u(A) = v(A) = |F| - 1$. $\square$

Activity Selections

The following independent system $(E, \mathcal{I})$ for activity selections is not a matroid, where

- E is the set of all intervals $[s_i, f_i]$.
- $\mathcal{I}$ is a family of subsets of non-overlapping intervals in E.
  - **Proof**. Let $E = \{[1,6],[2,3],[3,4],[4,5]\}$, $A = \{[1,6]\}$, $B = \{[2,3],[3,4],[4,5]\}$.
  - Both $A \in \mathcal{I}$ and $B \in \mathcal{I}$ and both are maximal.
  - But $|A| < |B|$.

Greedy Algorithm for MAX-ISS and Matroid

Theorem 2. An independent system $(E, \mathcal{I})$ is a matroid iff for any cost function c, the greedy algorithm for MAX-ISS produces an optimal solution for any instance.

Proof. Theorem 1 provides the "only if" part. For the "if" part, supposed that $(E, \mathcal{I})$ is not a matroid.

- There is a subset $F \subseteq E$ such that F has two maximal independent subsets I and I' with $|I| > |I'|$.
- Let $\epsilon \in (0, 1/|I'|)$. Construct a cost function c as follows:

$$c(e) = \begin{cases} 1+\epsilon, & \text{if } e \in I', \\ 1, & \text{if } e \in I - I', \\ 0, & \text{if } e \in E - (I \cup I'). \end{cases}$$

  Clearly, $c(I) > c(I')$.
- Since $c(e') > c(e)$ for $e' \in I'$ and $e \in I$, the greedy algorithm for MAX-ISS produces the solution set I'. A contradiction. $\square$

Matroids

- By the definition of matroids, an independent system is a matroid iff all maximal independent subsets have the same size. Thus,
- An independent system $(E, \mathcal{I})$ is a matroid if it satisfies the following *exchange property*:
  - If $A \in \mathcal{I}$, $B \in \mathcal{I}$, and $|A| < |B|$, then there exists $e \in B - A$ such that $A \cup \{e\} \in \mathcal{I}$.

Submodularity Greedy For MIN

- For maximization MAX greedy there are independent subset systems and matroids.
- For MIN greedy there is submodularity.
- Let f be a function: $2^E \to R^+$. Function f is *submodular* if for any $A, B \in 2^E$:

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

Minimum Set Cover

- Given a collection C of subsets of a set E, find a subcollection C' of C with minimum cardinality such that every element of E appears in a subset in C'.
- For a subcollection A of C, define

$$f(A) = \Big|\bigcup_{S \in A} S\Big|.$$

- Then $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$. Thus, function f is submodular.
- Greedy Algorithm

```
A = ∅
while |E| > f(A)
    Choose S ∈ C such that f(A ∪ {S}) is maximum
    A = A ∪ {S}
```

Analysis

- Suppose that $S_1, S_2, \cdots, S_k$ are selected by the greedy algorithm in order.
- Let $C_i = \{S_1, \cdots, S_i\}$. Then

$$f(C_{i+1}) - f(C_i) \geq \frac{|E| - f(C_i)}{opt}.$$

- What is the upper bound of k?
- It can be shown that, if $A \subseteq B$ implies that $\Delta_X f(A) \geq \Delta_X f(B)$, then $k \leq opt \cdot (1 + \ln \gamma)$, where $\gamma = \max_{1 \leq i \leq k}\{|S_i|\}$, and $\Delta_X f(A) = f(A \cup \{X\}) - f(A)$
- It can be shown that, $A \subseteq B \Rightarrow \Delta_X f(A) \geq \Delta_X f(B)$ iff f is submodular and f is monotone increasing (i.e., $A \subseteq B \Rightarrow f(A) \leq f(B)$).
- **Theorem 3**. The greedy algorithm produces an approximation within $(\ln n + 1)$ from the optimal.

Meaning of Submodularity

- The earlier the better!
- Monotone increasing gain!

Well, so much for greedy algorithms for now. We will revisit this issue later.