Lecture Notes: Greedy Set Cover

Palash Dey
palash.dey@cse.iitkgp.ac.in
Indian Institute of Technology Kharagpur

November 6, 2025

Definition 1 (Weighted Set Cover) *Given a universe $\mathcal{U}$ of size n, a collection $S_1, \ldots, S_m$ of subsets of $\mathcal{U}$ with costs $w_1, \ldots, w_m$ respectively, compute $\mathcal{J} \subset [m]$ such that $\cup_{j \in \mathcal{J}} S_j = \mathcal{U}$ and $\sum_{j \in \mathcal{J}} w_j$ is minimized.*

Description of the Greedy set cover algorithm

It is an iterative algorithm. In every iteration, we pick a set to be included in our solution until we have a set cover.

In the first iteration, let $j^* = \arg\min_{j \in [m]} \frac{w_j}{|S_j|}$. We include S_{j^*} in our solution, remove S_{j^*} from the collection of input sets and the elements covered by S_{j^*} from the universe $\mathcal{U}$ and every set of the input collection of sets. Notice that, we now have another instance of the set cover problem with strictly smaller universe size. We repeat the same in every iteration.

Runtime of the Greedy set cover algorithm

The algorithm makes at most $\min\{m, n\}$ iterations and every iteration clearly runs in polynomial-time. So this is a polynomial-time algorithm.

Approximation Factor of the Greedy set cover algorithm

Let OPT and ALG be the cost of respectively an optimal set cover and a set cover output by the greedy algorithm. Suppose the greedy algorithm runs for k iterations. We can assume without loss of generality (by renaming the input collection of sets) that the greedy algorithm has picked the set S_i in the i-th iteration for $i \in [k]$. Then $\mathsf{ALG} = \sum_{i=1}^{k} w_i$.

Since we are updating the instance in every iteration (deleting elements and sets), let OPT_i be the cost of an optimal set cover of the instance we have at the beginning of iteration i for $i \in [k+1]$. Then we have

$$\mathsf{OPT} = \mathsf{OPT}_1 \geqslant \mathsf{OPT}_2 \geqslant \cdots \geqslant \mathsf{OPT}_{k+1}$$

Let n_i be the number of (new) elements covered by the set S_i picked in the i-th iteration of the greedy algorithm. We now make the following claim.

Claim 1 $w_1 \leqslant \frac{n_1}{n}\mathsf{OPT}$.

Proof: We use the following result: for all positive integers $a_1, \ldots, a_\ell, b_1, \ldots, b_\ell$ for every positive integer ℓ, we have

$$\min_{i=1}^{\ell} \frac{a_i}{b_i} \leqslant \frac{\sum_{i=1}^{\ell} a_i}{\sum_{i=1}^{\ell} b_i} \leqslant \max_{i=1}^{\ell} \frac{a_i}{b_i}.$$

Let $S'_1, \dots, S'_r$ be an optimal solution. Then we have the following.

$$\frac{w_1}{n_1} \leqslant \min_{i=1}^{r} \frac{w(S'_i)}{|S'_i|} \leqslant \frac{\sum_{i=1}^{\ell} w(S'_i)}{\sum_{i=1}^{\ell} |S'_i|} = \frac{\mathsf{OPT}}{\sum_{i=1}^{\ell} |S'_i|} \leqslant \frac{\mathsf{OPT}}{n}$$

$\square$

We now have

$$\begin{aligned}
\mathsf{ALG} &= \sum_{i=1}^{k} w_i \\
&\leqslant \sum_{i=1}^{k} \frac{n_i}{n - \sum_{j=1}^{i-1} n_j} \mathsf{OPT}_i \\
&\leqslant \sum_{i=1}^{k} \frac{n_i}{n - \sum_{j=1}^{i-1} n_j} \mathsf{OPT} \\
&\leqslant \sum_{i=1}^{n} \frac{1}{i} \\
&= H_n
\end{aligned}$$

where H_n is the n-th harmonic number which is at most $1 + \ln n$.