Max Coverage—Randomized LP Rounding

CS 4820—April 2014

David Steuerer

Maximum Coverage Problem

- given: subsets $S_1, \dots, S_m$ of a ground set $U = \{u_1, \dots, u_n\}$, parameter $k \in \mathbb{N}$
- find: collection C of k sets so as to maximize the number of covered elements

This problem is NP-complete. In this lecture, we will develop an approximation algorithm for this problem based on linear programming.

LP relaxation for Max Coverage

Given an instance $\mathcal{I}$ of Max Coverage, we construct the following LP instance LP($\mathcal{I}$).

- *variables:* $x_1, \dots, x_m$ and $y_1, \dots, y_n$
- *constraints:*

$$\begin{aligned} &\sum_{i=1}^{m} x_i = k\,, && \text{(cardinality constraint)} \\ &\sum_{i:\, u_j \in S_i} x_i \geq y_j \quad \text{for } j \in \{1, \dots, n\}\,, && \text{(coverage constraints)} \\ &0 \leq x_i \leq 1 \quad \text{for } i \in \{1, \dots, m\}\,, \\ &0 \leq y_j \leq 1 \quad \text{for } j \in \{1, \dots, n\}\,. \end{aligned}$$

- *objective:* maximize $\sum_{j=1}^{n} y_j$

Claim 1. $\operatorname{Opt} \mathrm{LP}(\mathcal{I}) \geq \operatorname{Opt} \mathcal{I}$.

Proof. Let C be an optimal solution for $\mathcal{I}$, that is, a collection of k sets that cover $\operatorname{Opt} \mathcal{I}$ elements of U. We are to construct a solution of LP($\mathcal{I}$) with objectie value at least $\operatorname{Opt} \mathcal{I}$.

Consider the following solution to LP($\mathcal{I}$),

$$x_i = \begin{cases} 1 & \text{if } S_i \in C \\ 0 & \text{otherwise,} \end{cases} \quad \text{and } y_j = \begin{cases} 1 & \text{if } u_j \in \bigcup_{S_i \in C} S_i \\ 0 & \text{otherwise.} \end{cases}$$

This solution satisfies the cardinality constraint because exactly k of the variables $x_1, \dots, x_m$ are set to 1 and the rest are set to 0. The solution also satisfies the coverage constraints for all $j \in \{1, \dots, n\}$. If $y_j = 0$, then the corresponding coverage constraint is satisfied because all x_i values are nonnegative. Otherwise, if $y_j = 1$, then u_j is covered by C, which means that one of the sets $S_i \in C$ contains u_j. Therefore, at least one of terms of the sum $\sum_{i:\, u_j \in S_i} x_i$ is equal to 1, which is enough to satisfy the inequality.

Claim 2 (Randomized rounding algorithm). There exists a randomized algorithm that, given an optimal solution to LP($\mathcal{I}$), outputs a collection C of k sets that in expectation covers at least $(1 - 1/e) \cdot \operatorname{Opt} \mathrm{LP}(\mathcal{I})$ elements of U.

We defer the proof of Claim 2 to the next section. At this point, let us note that Claim 1 and Claim 2 together give a $(1 - 1/e)$-approximation algorithm for Max Coverage.

> **LP-based Approximation Algorithm for Max Coverage.** Given an instance $\mathcal{I}$ of Max Coverage, we construct the linear-programming instance $\mathrm{LP}(\mathcal{I})$. Using a polynomial-time algorithm for linear-programming, we compute an optimal solution for $\mathrm{LP}(\mathcal{I})$. We apply the randomized rounding algorithm from Claim 2 to this LP solution to obtain a solution for the original problem instance $\mathcal{I}$ that covers at least $(1 - 1/e) \cdot \mathrm{Opt}\,\mathrm{LP}(\mathcal{I})$ elements, which, by Claim 1, is within a $1 - 1/e$ factor of $\mathrm{Opt}(\mathcal{I})$—the maximum number of elements that k sets can cover.

Randomized Rounding—Proof of Claim 2

Let $\mathcal{I}$ be an instance of MC and let $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ be a solution to $\mathrm{LP}(\mathcal{I})$ with value $\mathrm{Opt}\,\mathrm{LP}(\mathcal{I})$. The following efficient randomized algorithm turns this LP solution into a collection C of k sets that covers at least $(1 - 1/e) \cdot \mathrm{Opt}\,\mathrm{LP}(\mathcal{I})$ elements in expectation.

- Interpret the numbers $x_1/k, \ldots, x_m/k$ as probabilities for the sets $S_1, \ldots, S_m$. (Notice that these numbers are nonnegative and add up to 1 according to the constraints of $\mathrm{LP}(\mathcal{I})$.)
- Choose k sets independently at random according to these probabilities.
- Output the collection C consisting of the k chosen sets.

Claim. For every element $u_j \in U$, the probability that the collection C produced by the rounding algorithm covers u_j is at least $(1 - 1/e) \cdot y_j$

Proof. If we choose a random set according to the probabilities $x_1/k, \ldots, x_m/k$, it covers element u_j with probability $\sum_{i:\, u_j \in S_i} x_i/k \geq y_j/k$. (Here, we use the coverage constraints.) Therefore, the probability that none of the k sets chosen by the rounding algorithm covers u_j is at most $(1 - y_j/k)^k$. Thus, the element u_j is covered by the collection C with probability at least $1 - (1 - y_j/k)^k$. It remains to verify that $1 - (1 - y_j/k)^k \geq (1 - 1/e) \cdot y_j$. In the interval $[0, 1]$, the function on the left is concave and the function on the right is linear. Since the inequality is satisfied at the end points of the interval (i.e., $y_j = 0$ and $y_j = 1$), it follows that the inequality holds in the entire interval. (A good way to verify this argument is to plot the two functions in the interval $[0, 1]$.)

Claim. The expected number of elements covered by C is at least $(1 - 1/e) \cdot \mathrm{Opt}\,\mathrm{LP}(\mathcal{I})$.

Proof. Let Z_j be the 0/1-valued random variable such $Z_j = 1$ indicates the event that C covers u_j. Then, the number of elements that C covers is equal to $\sum_{j=1}^{n} Z_j$. Therefore, by linearity of expectation, the expected number of elements covered by C is equal to

$$\mathbb{E} \sum_{j=1}^{n} Z_j = \sum_{j=1}^{n} \mathbb{E}\, Z_j$$

Since Z_j is a 0/1-valued random variable, the expectation of Z_j is equal to the probability that $Z_j = 1$. Hence, the expected number of elements covered by C is equal to

$$\sum_j \Pr\{Z_j = 1\} = \sum_j \Pr\{C \text{ covers } u_j\} \geq (1 - 1/e) \sum_j y_j = (1 - 1/e) \cdot \mathrm{Opt}\,\mathrm{LP}(\mathcal{I})\,.$$