Maximum k-Coverage

Vasilis Margonis

Combinatorial Optimization

$\propto\wedge\mu\forall$

November 14, 2017

Definition

Input:

- A universe of m elements: $\mathcal{U} = \{e_1, e_2, \ldots, e_m\}$.
- n subsets of $\mathcal{U}$: $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$.
- An integer **k**.

Goal:

- Find a subset S' of $\mathcal{S}$, such that $|S'| \leq k$ and the number of covered elements $\left| \bigcup_{S_i \in S'} S_i \right|$ is maximized.

ILP Formulation

$$
\begin{aligned}
\text{Maximize} \quad & \sum_{e_j \in \mathcal{U}} y_j \\
\text{Subject to} \quad & \sum_{S_i \in \mathcal{S}} x_i \leq k \\
& \sum_{e_j \in S_i} x_i \geq y_j \ , \quad \forall e_j \in \mathcal{U} \\
& y_j \in \{0, 1\}, \qquad \forall e_j \in \mathcal{U} \\
& x_i \in \{0, 1\}, \qquad \forall S_i \in \mathcal{S}
\end{aligned}
$$

Greedy Algorithm

Algorithm 1 GREEDY COVERING $(\mathcal{U}, \mathcal{S}, k)$

$C \leftarrow \varnothing$
while $|C| \leq k$ **do**
  Select $S_i \in \mathcal{S}$ such that $|S_i \cap \mathcal{U}|$ is maximized
  $\mathcal{U} \leftarrow \mathcal{U} - S_i$
  $C \leftarrow C \cup \{S_i\}$
end while
return C

Approximation Ratio

Theorem (1)

GREEDY COVERING is a $(1 - \frac{1}{e}) \simeq 0.632$*-approximation for* MAXIMUM k-COVERAGE.

- $OPT :=$ The value of an optimal solution.
- $x_i :=$ Number of elements covered in the i-th iteration.
- $y_i := \sum_{j=1}^{i} x_j$.
- $z_i := OPT - y_i$.
- $y_k = SOL :=$ Number of elements covered by GREEDY COVERING.

Analysis

Lemma (2)

For every $i = 1, \ldots, k$, $x_i \geq \frac{z_{i-1}}{k}$.

Proof:

- At each iteration, GREEDY COVERING selects the subset S_j which covers the maximum number of uncovered elements.
- The optimal solution uses k sets to cover OPT elements.
- Then, in the i-th iteration, there is a set that covers at least $\frac{z_{i-1}}{k}$ elements.
- Hence, $x_i \geq \frac{z_{i-1}}{k}$. □

Analysis

Lemma (3)

For every $i = 0, \dots, k$, $z_i \leq (1 - \frac{1}{k})^i \cdot OPT$.

Proof:

- The claim holds for $i = 0$, since $z_0 = OPT$.
- Inductively, we assume that $z_{i-1} \leq (1 - \frac{1}{k})^{i-1} \cdot OPT$, $i > 1$. Then

$$\begin{aligned} z_i &\leq z_{i-1} - x_i \overset{(2)}{\leq} z_{i-1} - \frac{z_{i-1}}{k} \\ &\leq z_{i-1} \cdot \left(1 - \frac{1}{k}\right) \\ &\leq \left(1 - \frac{1}{k}\right)^i \cdot OPT \qquad \square \end{aligned}$$

Analysis

Proof of Theorem (1).

- From lemma (3), it follows that $z_k \leq (1 - \frac{1}{k})^k \cdot OPT$.
- Also, for every $k \geq 1$, $(1 - \frac{1}{k})^k \leq 1/e$.
- Hence, $z_k \leq \frac{1}{e} \cdot OPT$.
- Then

$$SOL = y_k = OPT - z_k \geq OPT - \frac{1}{e} \cdot OPT = \left(1 - \frac{1}{e}\right) \cdot OPT \quad \square$$

Theorem (Feige, 1998)

For every $\varepsilon > 0$, there is no $(1 - \frac{1}{e} + \varepsilon)$-approximation algorithm for MAXIMUM *k*-COVERAGE, *unless* **P** = **NP**.