CMPSCI 711: More Advanced Algorithms

Graphs 9: Set Cover and Max Coverage

Andrew McGregor

- A $1/2 - \epsilon$ approx for max k coverage in $\tilde{O}(k/\epsilon)$ space.

  [McGregor, Vu, ICDT 17]

- A $2/\delta$ approx for set cover in $2/\delta$ passes and $\tilde{O}(mn^{\delta})$ space.

  [Har-Peled, Indyk, Mahabadi, Vakilian, PODS 16]

- A $pn^{1/p}$ approx for set cover in p passes and $\tilde{O}(n)$ space.

  [Chakrabarti, Wirth SODA 16]

Last Compiled: March 20, 2018

MaxCoverage

Can use submodular maximization algorithm to get $1/2-\epsilon$ approx for MaxCoverage with $\tilde{O}(\epsilon^{-1}\mathrm{OPT})$ space. Will reduce this to $\tilde{O}(\epsilon^{-3}k)$ space.

- Idea: Maximize coverage of elements in random subset $R \subseteq [n]$.
- Defining R: Assume we have guess v satisfying $\mathrm{OPT}/2 \leq v \leq \mathrm{OPT}$. Let $R = \{e \in [n] : h(e) = 1\}$ where $h : [n] \to \{0,1\}$ is a 2λ-wise independent hash function such that

$$p := \mathbb{P}[h(e) = 1] = \lambda/v$$

  for $\lambda = c\epsilon^{-2}k\log m$ and c a large constant.
- Lemma: With high probability, for any k sets $S_1, \ldots, S_k$,

$$|S_1' \cup \ldots \cup S_k'| = |S_1 \cup \ldots \cup S_k|p \pm \epsilon vp \quad \text{where } S_i' = S_i \cap R \ .$$

- If sets $A_1, \ldots, A_k$ α-approx max coverage in R and $O_1, \ldots, O_k$ give optimum coverage in $[n]$ then

$$|\cup A_i| \geq \frac{|\cup A_i'|}{p} - \epsilon v \geq \alpha \cdot \frac{|\cup O_i'|}{p} - \epsilon v \geq \alpha \cdot |\cup O_i| - 2\epsilon v \geq (\alpha - 2\epsilon) \cdot \mathrm{OPT}$$

Proof of Lemma

- Fix collection of k sets and let D be their union and let $X = |D'|$.

Theorem (Chernoff with Limited Independence)

Let $X_1, \ldots, X_n$ be boolean random variables. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[X]$ where $\mu \leq n/2$. If X_i are $\lceil \gamma\mu \rceil$-wise independent, then

$$\mathbb{P}[|X - \mu| \geq \gamma\mu] \leq \exp(-\lfloor \min(\gamma, \gamma^2) \cdot \mu/3 \rfloor) .$$

- $\mathbb{E}[X] = p|D| \leq \lambda$ and using the above theorem, $\gamma = \epsilon v/|D|$,

  $$\mathbb{P}[|X - \mu| \geq \epsilon vp] = \mathbb{P}[|X - \mu| \geq \gamma|D|p] \leq \exp\left(-\lfloor \min(\gamma, \gamma^2) \cdot \mu/3 \rfloor\right)$$

  since hash function is $\lceil \gamma\mu \rceil = \lceil \epsilon vp \rceil$-wise independent.
- $|D| \leq \textsc{opt} \leq 2v$ implies $\gamma = \epsilon v/|D| \geq \epsilon/2$ and so,

  $$\exp\left(-\left\lfloor \min(\gamma, \gamma^2) \cdot \frac{\mu}{3} \right\rfloor\right) = \exp\left(-\left\lfloor \min(1, \gamma) \cdot \frac{\epsilon vp}{3} \right\rfloor\right) \leq \exp\left(-\left\lfloor \frac{1}{2} \cdot \frac{ck \log m}{3} \right\rfloor\right) \leq \frac{1}{m^{10k}}$$
- Lemma follows by union bound over all $\binom{m}{k}$ collections of k sets.

Set-Cover: Algorithm 1

Theorem

A $2/\delta$-pass, $O(mn^{\delta})$-space algorithm returning a $2/\delta$-approximation for the minimum set cover.

Algorithm (Cover Random Subset): Assume $\text{OPT}/2 \leq k \leq \text{OPT}$

1. $U \leftarrow [n]$ tracks uncovered elements. $I \leftarrow \emptyset$ stores set IDs in cover.
2. Repeat $1/\delta$ times:
   - 2.1 Sample $\ell = 10kn^{\delta}\log(mn)$ elements R from U at random.
   - 2.2 In one-pass: Compute a set cover of R
     - 2.2.1 Add any set to I if it covers $|R|/k$ uncovered elements in R. For any set S not added, temporally store set of uncovered elements in $S \cap R$.
     - 2.2.2 Add temporary stored sets that cover of uncovered elements in R.
   - 2.3 In another pass: Update U

Space is $\tilde{O}(mn^{\delta})$: storing U and R requires $\tilde{O}(n)$ bits and each of the, at most m, sets that are temporarily stored each has size $< |R|/k = \tilde{O}(n^{\delta})$.

Approximation: At most $(k + \text{OPT})/\delta \leq 2\text{OPT}/\delta$ sets at chosen. Just remains to show that chosen sets are a set cover...

Analysis

Lemma

Let U_i be the set of elements uncovered before the ith round of the algorithm. Then, with high probability

$$|U_{i+1}| < |U_i|/n^\delta$$

for all rounds. Hence, all elements covered after $1/\delta$ rounds.

Proof:

- Consider any collection of k sets whose coverage C satisfies

$$|U_i \setminus C| \geq |U_i|/n^\delta \ .$$

- This collection of sets can't be chosen if R intersects $U_i \setminus C$ and

$$\mathbb{P}\left[R \cap (U_i \setminus C)\right] = \left(1 - \frac{|U_i \setminus C|}{|U_i|}\right)^\ell \leq (1 - n^{-\delta})^\ell \leq e^{-10k \log(mn)} = \frac{1}{(mn)^{10k}}$$

- Hence, the lemma follows by the union bound over all $\binom{m}{k}$ collections of sets and (at most) n rounds of the algorithm.

Set-Cover: Algorithm 2

Theorem

A p-pass, $\tilde{O}(n)$-space algorithm returning a $pn^{1/p}$-approximation for the minimum set cover.

Algorithm (Decreasing Threshold):

1. For $i = 1$ to $\log_\alpha n$ where $\alpha = n^{1/p}$:
   - In pass i: add any set that covers more than n/α^i new elements.

Space/Pass Analysis: Space is $\tilde{O}(n)$ and setting $\alpha = n^{1/p}$ gives p passes.
Exercise: Can combine last two passes to get $p-1$ pass algorithm.

Approximation: Let U_i be the set of uncovered elements before ith pass. Then, we know $\mathrm{OPT} \cdot n/\alpha^{i-1} \geq |U_i|$ and so in ith pass we add at most

$$\frac{|U_i|}{n/\alpha^i} \leq \alpha \mathrm{OPT}$$

sets. In total, we add $p\alpha = pn^{1/p}$ sets.