Thursday, February 8, 2018 3:18 PM

Approximation Algorithms for Set Cover

→ Greedy algorithm & analysis

$U = \{e_1, e_2, \dots e_n\}$

$\mathcal{S} = \{S_1, S_2, \dots S_m\}$

goal: pick min # subsets from $\mathcal{S}$ to cover $\bar{U}$.

Greedy Algo for Set Cover

Initially $Y = \bar{U}$

While $Y \neq \bar{\Phi}$

choose $\bar{S} \in \mathcal{S}$ with maximum $|S \cap Y|$

set $Y \leftarrow Y \setminus S$

Theorem [Johnson 1971]: Greedy is $(1+\ln n)$-approximation for set cover.

Independent of # of sets

$n_0 = n$

$n_1 = $ # uncovered elements after 1st set

$n_2 = $ # " " 2nd set

$n_i = $ " ith set

let $\boxed{k = |OPT|}$, i.e $\exists$ ~~k~~ sets to cover all elements.

let $\boxed{k = |OPT| \text{ , i.e } \exists \ k \text{ sets to cover elements}}$

1^{st} set we pick covers $\geq \frac{n}{k}$ elements.

$S_1^*\ S_2^* \dots S_k^* \subseteq \mathcal{S}$ cover U together

at least 1 set $\geq \frac{n}{k}$ elements from U

$\Rightarrow \quad n_1 \leq n\left(1 - \frac{1}{k}\right)$ ✓

Using same argument @ time 2, greedy set covers $\geq \frac{n_1}{k}$ elements from uncovered elts.

$\Rightarrow \quad n_2 \leq n_1\left(1 - \frac{1}{k}\right) \leq n\left(1 - \frac{1}{k}\right)^2$ ✓

After 't' steps,

$$n_t \leq n\left(1 - \frac{1}{k}\right)^t$$

At what value of t will $n_t < 1$

$$\boxed{n\left(1 - \frac{1}{k}\right)^t < 1}$$

$$\left(1 - \frac{1}{k}\right)^t < \frac{1}{n}$$

$$t = k \ln n + 1$$

$$\left(\left(1 - \frac{1}{k}\right)^k\right)^{\ln n} \approx \left(\frac{1}{e}\right)^{\ln n} = \frac{1}{n}$$

Opt $= k$ ... $\ln n + 1$

$$\left.\begin{array}{l} Opt = k \\ Us = k \ln n + 1 \end{array}\right\} \quad \text{Apx Ratio} \simeq \ln n + 1$$

To summarize

- LP based f - approximation ← deterministic
- LP based $O(\ln n)$ - approximation ← randomized
- Greedy ~~Off~~ $\ln n$ - approx (deterministic)

Q: Is there a 2 - approx for set cover?

A: No! in fact, any polytime $\ln n (1-\varepsilon)$ approx $\Rightarrow P = NP$ ← $\boxed{\text{Hardness of Approx.}}$

Related question :-

Max k coverage

Given $(U, \mathcal{S})$, & a parameter 'k', choose k sets to maximize # elements covered

$\boxed{\text{Can implement same greedy algorithm just stop after } k \text{ rounds!}}$

Proof (can use same ideas) to show that we will cover at least

$$\underline{OPT}\left(1 - \left(1 - \frac{1}{k}\right)^k\right)$$

$$\simeq OPT\left(1 - \frac{1}{e}\right)$$

$$\simeq 0.63 \cdot OPT$$

(0.63 approx for Max k Coverage)

Hardness 1

(Max k coverage)

Theorem: Any $(0.63-\epsilon)$ - approx $\Rightarrow$ $P = NP$

Hardness of approx!!

Detour

Submodular functions

$f : 2^{\mathcal{S}} \rightarrow \mathbb{N}$

$U = \{1\ 2\ 3\ 4\ 5\}$

$\mathcal{S} = \{ S_1 = \{1\ 2\ 3\},$
$S_2 = \{1\ 4\}$
$S_3 = \{1\ 3\ 5\}$
$S_4 = \{4\}$
$S_5 = \{4\ 5\} \}$

eg

$f(\{S_1\}) = 3$

$f(\{S_1, S_2\}) = 4$

$f(\{S_1, S_5\}) = 5$

Ξ

$f(\{S_1\ S_2\}) \leq f(\{S_1\ S_2\ S_3\})$

$\hookrightarrow$ monotone function

$f(X) \geq f(Y)$ if $X \supseteq Y$

Diminishing returns property

$X \subseteq 2^{\mathcal{S}}$ $\quad$ $Y \subseteq 2^{\mathcal{S}}$ $\quad$ $X \subseteq Y$

$S_i \notin X$ & $S_i \notin Y$

$f(X \cup S_i) - f(X) \geq f(Y \cup S_i) - f(Y)$

Any function over a domain 2^D which satisfies diminishing returns property is called a $\boxed{\text{Submodular function}}$

Game theory, ML, economics, Computer vision, etc.

If f satisfies both (P1) & (P2) then its called MONOTONE SUBMODULAR fn.

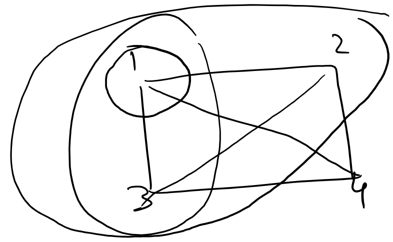
<u>Coverage fn</u> $\rightarrow$ Monotone & Submodular

$G = (V, E)$

$f : 2^V \rightarrow \mathbb{N}$

$f(S) = $ # edges crossing S

HW: show that f is <u>submodular</u> !!



$f(\{1\}) = 3$

$f(\{1, 3\}) = 4$

$f(\{1, 2, 3\}) = 3$

$f(\{1, 2, 3, 4\}) = 0$

For <u>monotone</u> submodular fns,

Greedy Algo $= \left(1 - \frac{1}{e}\right)$ - approximation

Given Monotone Submodular fn f & a target k, pick ~~set~~ X of $|X| = k$ to maximize $f(X)$

also the best possible algorithm !!

<u>MIN SUM SET COVER PROBLEM</u>

given U, $\mathcal{S}$

Find an <u>ordering of sets</u> to minimize

$$\sum_{e \in U} \text{cover}(e)$$

$U = \{1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\}$

$\mathcal{S} = \{\{1\ 2\ 3\ 4\}, \{5\ 6\ 7\ 8\ 9\ 10\},$
$\{1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\}, \{9\}, \{10\}\}$

Set Cover $=$ S_1, S_2 size 2

$O = S_1\ S_2\ S_3\ S_4\ S_5$

Acc / O, $\text{Cover}(1) = 1$, $\text{Cover}(5) = 2$
$2 = 1$, $\text{Cover}(6) = 2$
$3 = 1$, $\vdots$
$4 = 1$, $10 = 2$

Obj Value $= 4\times1 + 6\times2 = 16$

$O' = S_2\ S_1\ S_3\ S_4\ S_5$
$6\times1 + 4\times2 = 14$

$S_3\ S_4\ S_5\ S_1\ S_2$
$= 8\times1 + 2 + 3 = 13$

$S_3\ S_2\ S_1\ S_4\ S_5$
$8\times1 + 2\times2 = 12$

Turns out,
greedy is very good for MSSC
$\mapsto$ 4-approximation !!

Recall

Find an ordering to minimize $\max_e \text{cover}(e)$ $\Big\}$ = Set cover

Any $(4-\varepsilon)$-approx in polytime $\Rightarrow P = NP$

Proof (will show $O(1)$-approx, 4 is much more clever proof)

A_1 = # uncovered elts after time 1

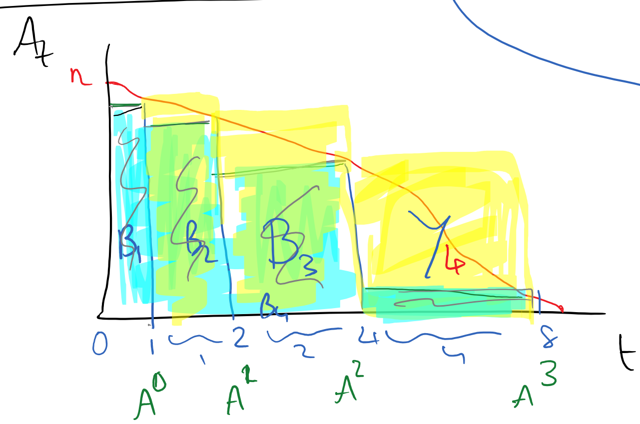
A_2 = " " " " time 2

A_t = " " time t

O_t = # uncovered elts in opt after time t

Claim:

$$\sum_e \text{cover}(e) = \sum_t A_t \quad (\text{for algo soln}) \qquad \sum_e \text{cover}^*(e) = \sum_t O_t$$



$\sum A_t$ = area under this ~~the~~ curve.

$$\frac{1}{2}\sum_{t\geq 1} A_t \leq \sum_{j\geq 0} A^j\cdot\left(2^j - 2^{j-1}\right) \leq \sum_{t\geq 1} A_t$$

$2B_3 \geq Y_4 \geq$ Area under red curve

$$\sum_{j\geq 0} A^j \cdot 2^j$$

(A^j = # uncovered elts @ time 2^j)

Goal

$$\boxed{\sum_{j\geq 0} A^j \cdot 2^j \leq O(1) \sum_{j\geq 0} O^j \cdot 2^j}$$

One attempt

s.t. $\forall j,\ A^j \leq O(1) \cdot O^j$??

We'll show

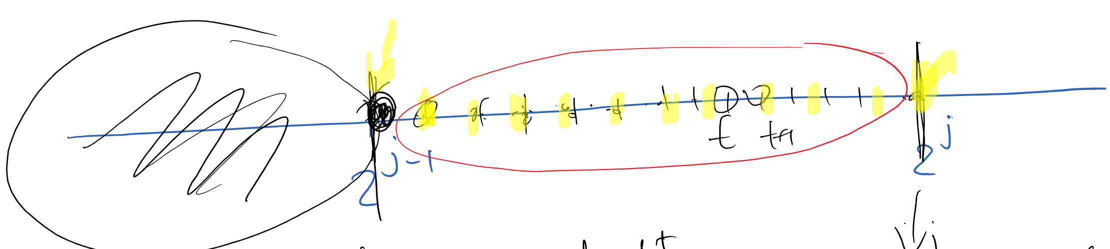
either $A^j \leq 10 \cdot O^j$ (or) $A^j \leq \frac{1}{3} \cdot A^{j-1}$

$$\boxed{A^j \leq \frac{1}{3} A^{j-1} + 10 \cdot O^j}$$

$$\sum_j 2^j \cdot A^j \leq \frac{2}{3} \sum_j 2^{j-1} A^{j-1} + 10 \sum_j 2^j O^j$$

$$\text{Alg} \leq \frac{2}{3} \text{Alg} + 10 \cdot \text{OPT}$$

$$\Rightarrow \text{Alg} \leq 30 \cdot \text{OPT}.$$



$X = \frac{9}{10} A_j$ elts

O^{j-1} = Uncovered elts in OPT @

A^j = uncovered elts @ 2^j

O^{j-1} = Uncovered elts in OPT @ 2^{j-1}

A^j = uncovered elts @ 2^j

either OPT sol'n covers $\geq \frac{9}{10}$ of elts of A^j by 2^{j-1} OR

if leaves $\geq \frac{1}{10}$ of A^j uncovered

$\frac{X}{2^{j-1}}$ coverage

In case ② $A^j \leq 10\, O^{j-1}$

Case ①

Opt in time 2^{j-1} has covered $\geq \frac{9}{10}$th of my uncovered elts @ time 2^j

Main Claim

$\forall\ 2^{j-1} \leq t \leq 2^t$,

$$\boxed{A_{t-1} - A_{t} \geq \frac{\frac{9}{10} \cdot A^j}{2^{j-1}}}$$

Greedy Rule

Focus on

$X = \overline{O^{j-1}} \cap A^j$ = set of elts opt covers in 2^{j-1} time which Alg still has uncovered at 2^j time

at time $t = 2^j$

$\exists$ set in OPT which covers $\geq \frac{|X|}{2^{j-1}}$ elts $\geq \frac{\frac{9}{10} \cdot A^j}{2^{j-1}}$

$$A^{j-1} - A^{j} \geq \frac{9}{10} \cdot A^j$$

$\frac{19}{10} \hat{A}^j \le A^{j-1}$

$\Rightarrow \hat{A}^j \le \left(\frac{10}{19}\right) \cdot A^{j-1}$

different parameters will give me

$\hat{A}^j \le \frac{1}{3} A^{j-1}$, completing the proof

Recap

Greedy

→ ln n set cover

→ $1-\frac{1}{e}$ for Max k Coverage

→ It extends to Submodular Max

→ O(1) approx for MSSC problem

Primal Dual Technique

algos are combinatorial

analysis will ~~not~~ use LP's extensively

rakri.github.io

www.cs.cmu.edu/~ravishan

2018