COMS W6998-5: Algorithms through Geometric Lens (Fall?17)	Sept 18, 2017

Lecture 1 – Count sketch, lp sampler and dynamic graphs

Instructor: *Alex Andoni, Ilya Razenshteyn* Scribes: *Aishwarya Srinivasan*

1 Count Sketch

The count sketch problem can be defined as to maintain a variable X, under the update $(i, +1)$ or $(i, -1)$ such that we can find:

$$\phi - HH : list\ all\ X_i \quad s.t \quad |X_i| \geq \phi \sum_{i=1}^{n} |X_i|$$

No $X_i s.t |X_i| < \phi(1 - 2\epsilon) \sum |X_i|$

Space: O$(1/(\epsilon)^2(\phi)^2 * \log(n))$

2 lo -sampler

Idea:
1. Suppose X has only 10 coordinates = +1. $\sum |X_i = 10$
Use CS with $\phi = 1/10 =¿\ X_i \neq 0 = |X_i| \geq 1/10 \sum |X_i|$
return j: $X_j \neq 0$

2. Suppose X has $\approx 10\sqrt{n}$ coordinates
- Pick (init) a set of I ∈ [n]
Pr[i ∈ I] =$1/\sqrt{n}$
- Consider CS on $X|_I$ (projected on I) will have roughly 10 non-zero coordinates at least in expectation.

Algorithm : Prepare for all levels: 2 parts
(a) F_2 sketches on X with $\epsilon = 0.6$, prob of success ≥ 1-1/n
(b) For each j=0..... log(n) Pr[i∈ $I_j = 2^(- j)$]
CS_j : Pick random set I_j : Pr i $\in I_j = \sum j$
- Do the CS on $X|_I(j)$, with $\phi = 0.01$ and $\epsilon = 1/3$

Output: Use F_2 sketch to get $\hat{F}_2 \approx \sum(X_i)^2$ = No. of zero coordinates of X.
- If $\hat{F}_2 \in [2^j, 2^(j + 1)]$ use CS$'_j$, j'= max0,j-3 to recover ϕ HH, output random one.

Theorem: With prob ≥ 0.7, output is uniform on j: $X_j \neq 0$
- With prob $\leq 2/n$, output is arbitrary
- Rest, output is FAIL.

Proof:

$$\hat{F}_2 \approx \sum |X_i|^2 \in 2^j, 2^j + 1$$

$(1+\epsilon) approximation \in F_2(1 \pm \epsilon)$

$$F_2 \in 2^(j-1), 2^(j+2)$$

$F_2 = \sum |X_i|^2$
Using CS$'_j$,
S = size $|I'_j|$ - No. of coordinates we are using
S = $i \in I'_j : X_j \neq 0$
E[S] = F_2 Pr[i $\in I'_j$]
= $F_2(2^(- j + 3) \in [4, 32]$
Var[S] $\leq F_2$ Pr[i $\in I'_j] \leq E[S]$

Apply Chebyshev,
Pr[—S-E[S]$\geq \lambda] \leq Var[S]/(\lambda)^2$
when S=0, λ= E[S], output is failed (no heavy hitter found)
set λ = E[S]
S$\in$ [1,2 E[S]] with Pr $\geq$ 1-1/$E[S] \geq$ 0.75
CS'_j outputs random i $\in$ i $\in I'_j : X_i \neq 0$

There is $\leq 1/n$ Prob that F_2 sketch fails,
$\leq 1/n$ Prob that CS'_j fails

Got lo sampler with success prpb 0.7, we need to boost the prob.

3 Full lo sampler - repeat basic sketch

[k=O(lg(n)) times]
-Use all of them in parallel
Output: Run them one by one until none fail.

$$Pr[Fulllosamplerfails] \leq (0.3)(0.3)(0.3)....(0.3)ktimes < 1/n$$

4 Dynamic Graphs

Primitive: Fix a node V, sketch edges incident on v; at the end output a random edge incident on V.

$X^v \in +1, 0, -1^p$, $p = \binom{n}{2}$
$X^v ij = 0$ if no (i,j) edge
= +1 if i=v
= -1 if j=v (for all i¡j)

Dynamic Sample: Use lo sampler on X^v will output a random edge incident on V.
Space: Space of lo sampler = O(lg^4 n)
k= lg(n) ; lg= no.of levels; lg^2(n) = CS and F^2;

Algorithm: Use dynamic sample for each node V
Space:O(n lg^4 n)

Property: Use linearity of lo sampler (due to linearity of ToW and CS)

lo-sampler: can be seen

Z= M.X
Z- vector that stores the sketch R^s
M- matrix depends on random choices of I'_j, randomness of CS
X- vector

$Z(X^1 + X^2) = M(X^1 + X^2)$
$= MX^1 + MX^2$
$=Z(X^1)+Z(X^2)$

Consider a set of vertices- S
$X^S \rightarrow \sum_{v \in S} X^v$

$$(X^S)_(i,j) = +1 i \in S \quad j \notin S \quad (i,j) edge \qquad = -1 j \in S \quad i \notin S \quad (i,j) edge \qquad = 0 otherwise$$

Given a set S, we can use dynamic sample on vertices to sample a random edge incident on S.

$$Z(X^S) = \sum_{v \in S} Z(X^v)$$

use estimated prod on $\sum_{v \in S} Z(X^v)$

Algorithm: Sample edge from each v using Z(X^v)
- let S1, S2,.... Sk be the connected components
- contrast S1, S2, ...Sk and compute sketch.

$$Z(Si) = \sum_{X \in Si} Z(X^v) \qquad i = 1, 2, ...k$$

-recurse on components S1,S2, ...Sk until done.
-think each S1, S2 and S3 as one node; each components now is a node.

Note:
1. No. of iterations $\leq$ O(lg(n))
2. Iterations are not independent; use lo sampler for each node and for iterations independently. Space: O(n $\lg^5$ n)