Github link

CS-328
HW-2

Date: / / Page no:

Sagar Bisen
18110149

(1)

(1) Suppose there are n number of users and let interest vectors of all the users are like:-

$$v_1 = \{ a_{11}, a_{12} \dots \}$$
$$v_2 = \{ a_{21}, a_{22} \dots \}$$
$$v_3 = \{ a_{31}, a_{32} \dots \}$$

As per the question, V (computed vectors of all users will be :-

$$V = \{ v_1, v_2, v_3, v_4 \dots v_n \}$$

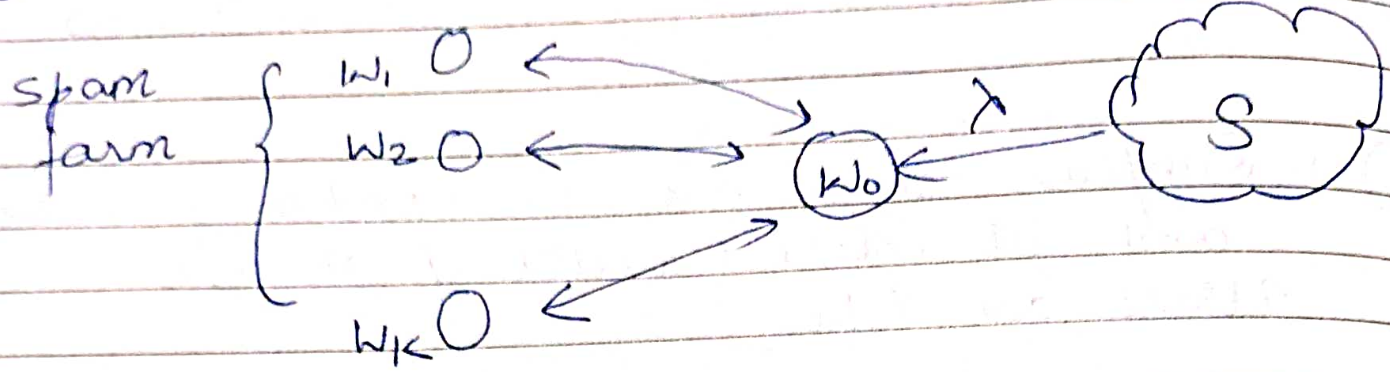
Since we know that any personalized Page Rank vector can be written in the form of linear combination of $\{ v_1, v_2 .. v_n \}$

$$I = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \dots \alpha_n v_n$$
$\alpha \in R$

Therefore $I \in \text{span}(V)$

Set of all personalized page rank vectors $= \text{span}(V)$

(2) Given :-



Parameters

N (total pages) = $|S| + K + 1$

$$\lambda \left(\begin{matrix}\text{pagerank flows} \\ \text{from } S \to w_0\end{matrix}\right) = \sum_{j \in S} \frac{p_j}{d_j}$$

$p_j \to$ pagerank of page j

$d_j \to$ outdegree

$p_0 \to$ page rank of page w_0

We need to calculate p_0 using λ, K, N and α (teleportation parameter)

Now Page rank of w_0 (p_0) depends on 3 factors:

1. Page rank recieved from S $\to$ $\underline{\alpha\lambda}$
2. Page ranks recieved from $\{w_0 \dots w_k\}$ $\to$ $\alpha \sum_{i=1}^{K} p_i$
3. Pagerank recieved from random jump $\to$ $(1-\alpha) \times \frac{1}{N}$

Date: / / Page no:

Combining all three factors

$$p_0 = \alpha\lambda + \alpha \sum_{i=1}^{K} p_i + \frac{(1-\alpha)}{N} \quad - \text{(i)}$$

(1) (2) (3)

Also

$1 \le i \le K$

$$p_i = \frac{\alpha p_0}{K} + \frac{(1-\alpha) \times 1}{N} \quad - \text{(ii)}$$

(uniform distribution of p_0 onto all p_i) (random jump)

Substitute (ii) in (i)

$$p_0 = \alpha\lambda + \alpha \sum_{i=1}^{K} \left(\frac{\alpha p_0}{K} + \frac{(1-\alpha)}{N} \right) + \frac{1-\alpha}{N}$$

$$p_0 = \alpha\lambda + \frac{\alpha(1-\alpha)K}{N} + \alpha^2 p_0 + \frac{1-\alpha}{N}$$

~~Ros~~

$$p_0(1-\alpha^2) = \alpha\lambda + \frac{(1-\alpha)}{N}(K\alpha+1)$$

$$p_0 = \frac{\alpha\lambda}{1-\alpha^2} + \frac{(\alpha K+1)}{(1+\alpha)N}$$

~~Got~~

(3) Given n distinct items.

no. of items with frequency $K = \frac{C}{K^3}$

no. of items with freq 1 $= \frac{C}{1^3}$

with freq 2 $= \frac{C}{2^3}$

with freq 3 $= \frac{C}{3^3}$

Hence $n = \frac{C}{1^3} + \frac{C}{2^3} + \frac{C}{3^3} \dots$

$$n = \sum_{k=1}^{\infty} \frac{C}{k^3}$$

$$n = C \sum_{k=1}^{\infty} \frac{1}{k^3}$$

Now ~~wit~~ we know that series $\left(\sum_{k=1}^{\infty} \frac{1}{k^3} \right)$

converges to a constant value and let the value be V_0.

Now replace $\sum_{k=1}^{\infty} \frac{1}{k^3}$ with V_0

Date: / / Page no:

$$n = c V_0$$

$$c = \frac{n}{V_0} \quad (V_0 \text{ is constant})$$

Therefore $c = O(n)$

Now to compare CM sketch & CS sketch which is better for the given distribution (w & d) are fixed.

(1) CM sketch = $\hat{f}_x - f_x \in [0, \varepsilon m]$

(2) CS sketch: $\hat{f}_x - f_x \in [-\varepsilon \|f\|_2, \varepsilon \|f\|_2]$

length of stream = m

$$m = \sum_{j=1}^{\infty} j \times f(j)$$

$$= \sum_{j=1}^{\infty} j \times \frac{c}{j^3} \qquad \left[f(i) = \frac{c}{i^3} \right]$$

$$= \sum_{j=1}^{\infty} \frac{c}{i^2}$$

Also $\frac{c}{i^2}$ converges to $\frac{\pi^2}{6}$

Hence $m = \frac{c\pi^2}{6}$

$$||f||_2 = \sqrt{\sum_{j=1}^{\infty} f_j^2}$$

no. of items with frequency $k = \frac{c}{k^3}$

$$\Rightarrow \sqrt{\sum_{k=1}^{\infty} \frac{c}{k^3}(k^2)} = \sqrt{\sum_{k=1}^{\infty} \frac{c}{k}}$$

But $\sum_{k=1}^{\infty} \frac{1}{k}$ series diverges

Hence $\sqrt{\sum_{k=1}^{\infty} \frac{c}{k}}$ will also diverge

$||f||_2 \rightarrow$ diverges. and does not bound.

Hence from ~~ab~~ 1 + 2

CM sketch $\rightarrow$ $\hat{f}_x - f_x \in [0, \varepsilon m]$ bound
But CS sketch does not as $||f||_2$ is not bound

Hence CM Sketch will be better for this distribution.

Discussed with Harshit Kumar (18110063)