Introduction

1

Güneş Akbaş
2017205003

Modern search engines use two-step process. First is text proccessing and with this method search engine find related webpage based on semantic or key-word methods. Second step is about Markov chain process. When related data forms a directed graph by first process between web pages, process turns into ranking by the Pagerank algorithms. For pagerank we have a directed graph first we need to convert weighted graph so we divide probabilities uniformly.

Then we have $G=(V,E)$ turning into matrice P,

$$P_{ij} = \begin{cases} \frac{1}{\deg(i)} & (i,j)\in E \\ 0 & (i,j)\notin E \end{cases}$$

P is $n \times n$ matrice
V is a set of vertices and $|V| = n$

uniform distribution between edges.

p_1^T is first row of P.

To turn P into transition matrice by terms of Markov chain and it should be done irreducible and aperiodic to stationary matrix converge uniquely.

First we handle all zero rows:

$p_i^T = [0, 0, \dots 0]$ when we replace $\frac{1}{n} e^T$ instead of p_i^T

$e^T = [1, 1, \dots 1]$ we get rid of all zero row.

$$\bar{P} := p_i^T \rightarrow \frac{1}{n} e^T$$

Then to ensure irreducibility we connect all vertices each other by means of get rid of any 0 entry of matrix $\bar{P}$.

$$\bar{\bar{P}} := \alpha \bar{P} + (1-\alpha) e e^T . \frac{1}{n} \qquad e = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \text{ (column uniform vector)}$$

(2)

$$\bar{\bar{P}} = \alpha \bar{P} + (1-\alpha) e e^T \cdot \frac{1}{n} \qquad 0 \leq \alpha \leq 1$$

with this, $\bar{\bar{P}}$ is irreducible due to $P_{ij} > 0 \quad \forall i \forall j \in [n]$

For being aperiodic, in a directed graph it is said to be aperiodic if there is no integer $k > 1$ divides the length of every cycle of the graph.

We know that $\bar{\bar{P}}$ is strongly connected we have all length of cycle so no k can divide these cycle lengths. Hence $\bar{\bar{P}}$ is aperiodic.

Then by fundemental theorem of Markov Chain we know any starting distribution $x^{(0)T}$ $\lim_{k \to \infty} x^{(0)T} \bar{\bar{P}}^k = \pi^T$ and π^T is unique.

Storage Issues

Size of matrices are too big and we need to deal with a lot of computational processes so to upgrade efficiency there is various methods:

Decomposition: $P = D^{-1} G$

holding outdegrees nodes $D_{ij} = \begin{cases} i\text{'s outlink number} & i = j \\ 0 & i \neq j \end{cases}$

adjancy matrix $G_{ij} = \begin{cases} 1 & P_{ij} > 0 \\ 0 & P_{ij} = 0 \end{cases}$

Normally $x^T P$ vector matrix multiplication need $nnz(P)$ multiplication and $nnz(P)$ addition.
($nnz(P)$: Number of nonzeros of P)
but with decomposition $x^T . P = x^T D^{-1} G = (x^T) * diag(D^{-1}) G$
$*$ is componentwise multiplication so n multiplication.
and $[x^T * (diag(D^{-1}))] . G$ so for second operation
G only contains 0's and 1's hence there is $nnz(P)$ additions.

$nnz(P)$ multiplication + $nnz(P)$ addition
$\rightarrow$ n multiplication + $nnz(P)$ addition

storing G as inlink information.

Example : Node Inlink from
100 112 113 116 117 as 112 0 2 0

If we have similar adjancy list x y we can store as in terms of one another.

Example:

x	5	7	12	89	101	190	390
y	5	6	12	50	101	190	

y in terms of x $\rightarrow$ 1010110 $\boxed{6 \mid 50}$

first 1's, 0's give information about which columns are same and second table give what is extra in y.

Other stroge solution for stochasticity fixation problem is dangling nodes. (Dangling nodes have no outlink)

We got rid of by $\bar{P} = p_i^T \rightarrow e^T \frac{1}{n}$ or v^T
but this cause more storage issues so we can implicitly store information about dangling nodes by vector a. $a_i = 1$ if $p_i^T = [0, 0, \dots 0]$ and 0 otherwise
Hence $\bar{P} = P + a v^T$

5. Solution Methods for solving the PageRank Problem

either we solve $\pi^T \bar{\bar{P}} = \pi^T$ or $\pi^T (I - \bar{\bar{P}}) = 0^T$

$\pi^T e = 1$ is a normalization equation, this ensures that π^T is a probability vector. 5

π_i is the pagerank of page i.

We order pages by rank of π^T components.

5.1 Power Method

$$x^{(k)T} = x^{(k-1)T} \bar{\bar{P}} \qquad (x^{(0)T} \text{ is starting vector.})$$

$$= \alpha\, x^{(k-1)T} \bar{P} + (1-\alpha)\, x^{(k-1)T} e v^T = \alpha\, x^{(k-1)T} \bar{P} + (1-\alpha) v^T$$

$$= \alpha\, x^{(k-1)T} P + (\alpha\, x^{(k-1)T} a + (1-\alpha)) v^T \qquad (x^{(k-1)T} . e = 1)$$

6.3 Forcing Irreducibility

Irreducibility condition was provided by connecting every vertice to each other but this method might lead deviation from orginal matrice P.

Therefore, it can be shown that there are other methods for example minimal irreducibility by adding fake new vertice and this fake vertice connected all vertices.

Hence $\hat{P}$ is irreducible.

$$\hat{P} = \left(\begin{array}{c|c} \alpha \bar{P} & (1-\alpha) e \\ \hline v^T & 0 \end{array}\right)$$

We acquire $\hat{P}$ and stationary distribution of $\hat{P}$ let say $\hat{\pi}^T$ then we should look for relation between π^T and $\hat{\pi}^T$

by system of equations we have (6)

$$\hat{\pi}^T = \alpha \hat{\pi}^T \bar{P} + \hat{\pi}_{n+1} \cdot V^T$$

$$\hat{\pi}_{n+1} = (1-\alpha) \hat{\pi}^T \cdot e \Rightarrow \hat{\pi}_{n+1} = \frac{1-\alpha}{2-\alpha}$$

and relation between $\hat{\pi}^T$ and π^T is clearly removing $\hat{\pi}^T_{n+1}$ from $\hat{\pi}^T$ and normalizing, to do this $\hat{\pi}^T$ is multiplied by $\dfrac{1}{1-\hat{\pi}_{n+1}} = 2-\alpha$

with $2-\alpha$ gives $\hat{\pi}^T = \alpha \hat{\pi}^T \bar{P} + (1-\alpha) v^T$.

This is power method's equation hence we have stationary distribution similar to maximal irreducibility method but more precise.