IR Tutorial

Nov 15, 2021

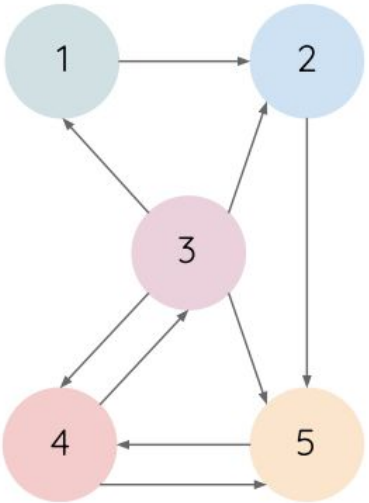
Topics to be covered

- PageRank
- Summarization

Evaluate the PageRank score of nodes in this graph

For simplicity of calculation treat \alpha = 1

I.e., random surfer will always follow the outward links.



PageRank computation

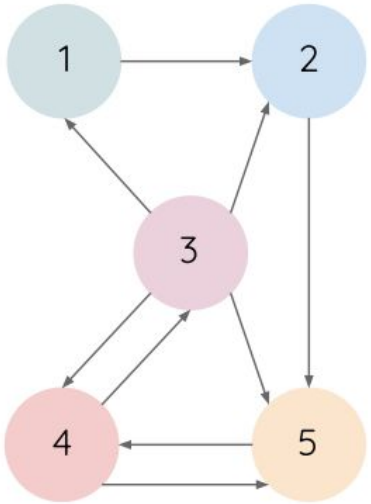
```
/* initialization */
for all nodes u in G: d(u) ← 1/N, where N = #nodes
for all nodes u in G: PR(u) ← d(u)
/* iteration */
do until PR vector converges
   for all nodes u in G
        for all nodes v that links to u
             t = Σ PR(v) / out-degree(v)
        PR(u) ← α  * t  + (1 – α ) * d(u)
   normalize scores
   check for convergence
end
```

Evaluate the PageRank score of nodes in this graph

Iteration 0

Initialise all pages to have PageRank score of 1/5

Nodes	Iteration 0
1	1/5
2	1/5
3	1/5
4	1/5
5	1/5



Evaluate the PageRank score of nodes in this graph

Iteration 1:

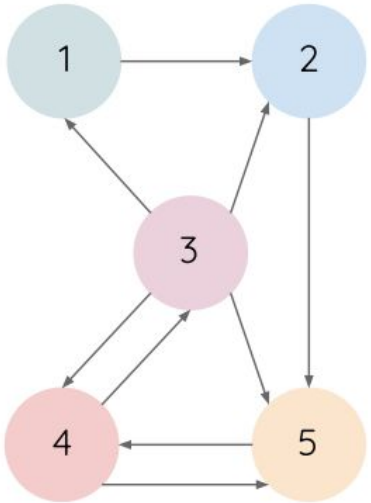
Page 1 has links from Page 3, and it has 4 out bound edges.

PR(1) = PR(3) / 4 = 1 / 20

Similarly for Page 2
PR (2) = PR(1) + PR(3) / 4

= 5/20

Nodes	Iteration 0	Iteration 1
1	1/5	1/20
2	1/5	5/20
3	1/5	1/10
4	1/5	5/20
5	1/5	7/20



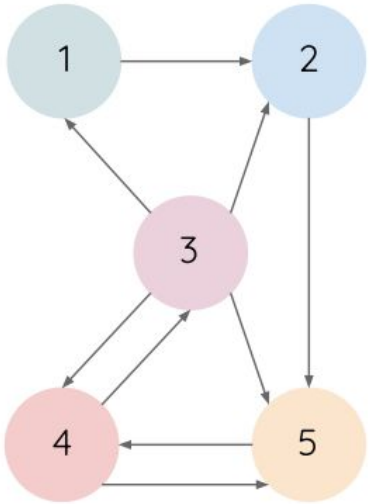
Similarly for other nodes.

Evaluate the PageRank score of nodes in this graph

Iteration 2:

Use the values in iteration 1 to compute that of iteration 2.

Nodes	Iteration 0	Iteration 1	Iteration 2
1	1/5	1/20	1/40
2	1/5	5/20	3/40
3	1/5	1/10	5/40
4	1/5	5/20	15/40
5	1/5	7/20	16/40



Evaluate the PageRank score of nodes in this graph

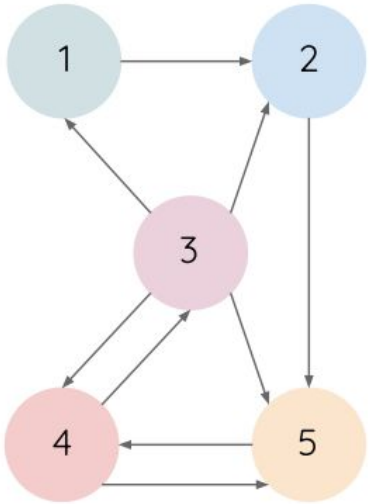
If the question asks, use \alpha = 0.8

Iteration 0 will remain unchanged

PageRank computation

/* initialization */
for all nodes u in G: $d(u) \leftarrow 1/N$, where N = #nodes
for all nodes u in G: $PR(u) \leftarrow d(u)$
/* iteration */
do until *PR* vector converges
  for all nodes *u* in G
    for all nodes *v* that links to *u*
      $t = \Sigma\ PR(v)$ / out-degree(*v*)
    $PR(u) \leftarrow \alpha * t + (1 - \alpha) * d(u)$
  normalize scores
  check for convergence
end

Nodes	d(u)	t
1	1/5	1/20
2	1/5	5/20
3	1/5	1/10
4	1/5	5/20
5	1/5	7/20



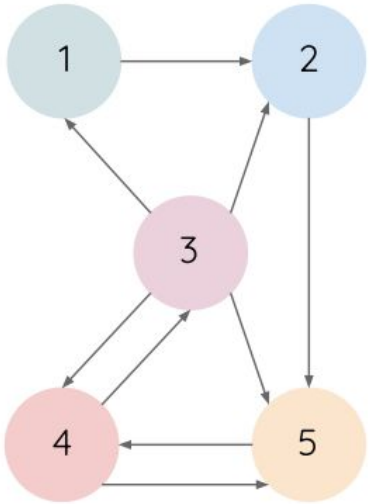
Evaluate the PageRank score of nodes in this graph

If the question asks, use \alpha = 0.8

Iteration 0 will remain unchanged

In iteration 1,

Nodes	d(u)	t
1	1/5	1/20
2	1/5	5/20
3	1/5	1/10
4	1/5	5/20
5	1/5	7/20



$$t = \Sigma\, PR(v) \,/\, \text{out-degree}(v)$$

$$PR(u) \leftarrow \alpha * t + (1 - \alpha) * d(u)$$

Probability of reaching to page `*u*' by following outward edges.

Probability of reaching to page `*u*' by teleportation..

Evaluate the PageRank score of nodes in this graph

If the question asks, use \alpha = 0.8

Iteration 0 will remain unchanged

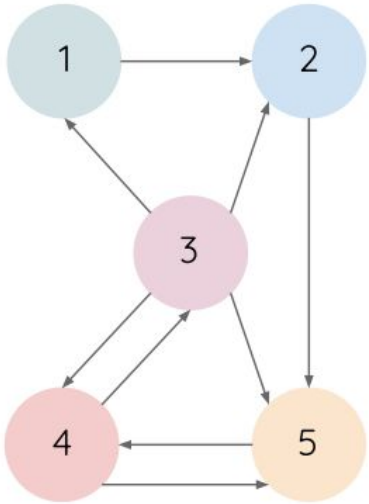
In iteration 1,

Pr(1) = 0.8 * 1/20 + 0.2 * 1/5

$$t = \Sigma\, PR(v) \,/\, \text{out-degree}(v)$$

$$PR(u) \leftarrow \alpha * t + (1 - \alpha) * d(u)$$

Nodes	d(u)	t
1	1/5	1/20
2	1/5	5/20
3	1/5	1/10
4	1/5	5/20
5	1/5	7/20



Probability of reaching to page `*u*' by following outward edges.

Probability of reaching to page `*u*' by teleportation..

Note that in the previous slides \alpha = 1, hence *Pr(u) = t*

LexRank

- LexRank → Applying the idea of PageRank to summarization

The document to summarize ; summary length = 50 words

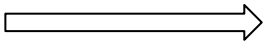
S1 : Junk foods taste good that's why it is mostly liked by everyone of any age group especially kids and children.
S2 : They generally ask for the junk food daily because they have been trained so by their parents from the childhood.
S3 : They never have been discussed by their parents about the harmful effects of junk foods over health.
S4 :Children will find one way or the other to have it.
S5 : Make sure you give them junk food in limited quantities and at healthy periods of time.

LexRank

- LexRank → Applying the idea of PageRank to summarization

The document to summarize ; length = 50 words

S1 : Junk foods taste good that's why it is mostly liked by everyone of any age group especially kids and children.
S2 : They generally ask for the junk food daily because they have been trained so by their parents from the childhood.
S3 : They never have been discussed by their parents about the harmful effects of junk foods over health.
S4 :Children will find one way or the other to have it.
S5 : Make sure you give them junk food in limited quantities and at healthy periods of time.



Step 1 : Sentence Representation

S1 : [0.5 0.0 0.14 0.009 …]

S2 : [0.01 0.12 0.9 0.54 ..]

S3 : […..]

S4 : […..]

S5 : […..]

LexRank

Step 1 : Sentence Representation

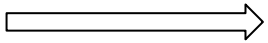
S1 : [0.5 0.0 0.14 0.009 …]

S2 : [0.01 0.12 0.9 0.54 ..]

S3 : […..]

S4 : […..]

S5 : […..]



Step 2 : Graph Formation

Compute similarity between the sentence pairs & form a graph (adjacency matrix)

	S1	S2	S3	S4	S5
S1	0	*0.9*	0.01	0.5	0.6
S2	0.9	*0*	0.7	0.85	0.9
S3	0.01	*0.7*	0	0.2	0.15
S4	0.5	*0.85*	0.2	0	0.05
S5	0.6	*0.9*	0.15	0.05	0

LexRank

Step 2 : Graph Formation

Compute similarity between the sentence pairs & form a graph (adjacency matrix)

	S1	S2	S3	S4	S5
S1	0	*0.9*	0.01	0.5	0.6
S2	0.9	*0*	0.7	0.85	0.9
S3	0.01	*0.7*	0	0.2	0.15
S4	0.5	*0.85*	0.2	0	0.05
S5	0.6	*0.9*	0.15	0.05	0

Step 3 : Row stochastic matrix

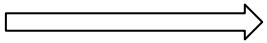
Entries in a row add up to 1

	S1	S2	S3	S4	S5
S1	0	*0.45*	0.01	0.25	0.29
S2	0.26	*0*	0.21	0.25	0.28
S3	0.09	*0.66*	0	0.19	0.06
S4	0.31	*0.53*	0.13	0	0.03
S5	0.35	*0.53*	0.08	0	0.04

LexRank

Step 3 : Row Stochastic matrix

	S1	S2	S3	S4	S5
S1	0	*0.45*	0.01	0.25	0.29
S2	0.26	*0*	0.21	0.25	0.28
S3	0.09	*0.66*	0	0.19	0.06
S4	0.31	*0.53*	0.13	0	0.03
S5	0.35	*0.53*	0.08	0	0.04



Step 4 : Sentence Importance Score

Compute PageRank algorithm

S1	S2	S3	S4	S5
[0.65	*0.8*	0.4	0.35	0.5]

LexRank

S1	S2	S3	S4	S5
[0.65	*0.8*	0.4	0.35	0.5]

Step 5 : Rank the sentence based on the scores

S2 : They generally ask for the junk food daily because they have been trained so by their parents from the childhood. [20 words]
S1 : Junk foods taste good that's why it is mostly liked by everyone of any age group especially kids and children. [20 words]
S5 : Make sure you give them junk food in limited quantities and at healthy periods of time. [16 words]
S3 : They never have been discussed by their parents about the harmful effects of junk foods over health. [17 words]
S4 :Children will find one way or the other to have it. [11 words]

length

50 words

Output

Junk foods taste good that's why it is mostly liked by everyone of any age group especially kids and school going children.

They generally ask for the junk food daily because they have been trained so by their parents from the childhood.

Rouge Computation

$$\frac{\text{number of n-grams found in model and reference}}{\text{number of n-grams in reference}}$$

$$\frac{\text{count}_{\text{match}}(\text{gram}_n)}{\text{count}(\text{gram}_n)}$$

Output

Junk foods taste good that's why it is mostly liked by everyone of any age group especially kids and children.

They generally ask for the junk food daily because they have been trained so by their parents from the childhood.

Reference

Junk foods are delicious. But they are unhealthy for kids and children. Parents should avoid giving their children junk food, even if they ask for it.

ROUGE 2 → Bi-grams
= 4 / 25