Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 2, No. 1, February : 2023

A STUDY OF MATHEMATICAL ASPECT OF GOOGLE PAGE RANK ALGORITHM

Divya Acharya Department of Humanities and Applied Sciences, Atharva College of Engineering, Mumbai University,Malad-95
Ancy Dsouza Department of Humanities and Applied Sciences, Atharva College of Engineering, Mumbai University,Malad-95
Priyanka Malgaonkar Department of Humanities and Applied Sciences, Atharva College of Engineering, Mumbai University,Malad-95
Deepika Panchal Department of Humanities and Applied Sciences, Atharva College of Engineering, Mumbai University,Malad-95

Abstract: Page Rank algorithm is the core concept used in Google Search Engine . It is a widely known algorithm which helps in giving rank to the web pages and hence, it determines the order in which they are to be displayed during a search .It was developed in 1998.Almost all web search tools still use the Page Rank algorithm as their foundation. In this paper, we will get an overview of application of linear algebra and graphs in Google's Page Rank algorithm.

Keywords: *Stochastic matrix, Reducible matrix, damping factor.*

I. INTRODUCTION :

Eigen values and Eigen vectors is a well known concept in Linear Algebra associated with linear Transformations and hence matrices. Following is the definition :

Eigen Value: A be an order n matrix . A real or a complex number λ is an Eigen value of A if there exists an $n \times 1$ matrix $X \neq 0$ satisfying, AX= λ X. [1]

II. APPLICATION OF LINEAR ALGEBRA:

Let us see how eigen values and eigen vectors concept is used in page rank algorithm to get the order of preference of web pages.

We can use graph to represent the linking of web pages and further define

"Hyperlink Matrix" , H =(H_{ij}) to describe the linking as follows,

Suppose, a page i has p_i outlinks. Let $H_{ij} = 1/p_i$, if page i has link to page $j \neq i$,

$$= 0, otherwise.$$

H has dominant eigen value λ = 1and we find eigen vector X corresponding to it.

The page corresponding to largest entry i.e, pagerank in X will get the preference in the search.

.To compute the Eigen vector, different iterative methods can be used. One of these method is **Power Method**:

In this , we start by picking a vector J^0 and then define a sequence of vectors J^k by,

$$\mathbf{J^k = H. J^{k+1}}$$

The method is based on the following principle:

" The sequence J^k will converge to the Eigen vector J." [2]

The above method has some restrictions if

(i) there is a dangling node.

(ii) a group of pages forming a sink.

Hence the matrix H may require some modifications. It is achieved as follows

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 2, No. 1, February : 2023

(i) For a dangling node

The ith row of H, however, has only zero entries if page i has no outlinks.. It is called as a dangling node. Due to this nodes, pagerank of some or all pages may become 0.
In order to fill these type of rows we modify H to S by defining,

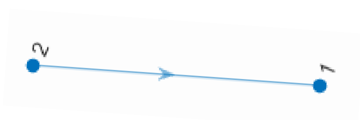
$$\boldsymbol{S = H + eu^T} \ ,$$

where e = (e_i) is a column matrix of order $n \times 1$ and $e_i = 1$,if page i has no outlinks
$= 0$ otherwise;
and u =(u_i) is a column matrix with $u_i \geq 0$ and $\|u\|_1 = 1$.

Choose u = 1/n **B**, where **B** = $\begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{bmatrix}_{n \times 1}$, i.e, all entries are 1 in **B** and n is the number of nodes in the graph. [2]

Above definition of matrix S ensures it does not have a zero column in it and hence iterative methods will give eigen vector with positive entries.
(Such a matrix S is stochastic matrix.)

For example, consider



H = $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, using power method we get J0 = $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, J1 = $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ which fails to be an eigen vector

(ii) A group of pages forming a sink.

Sometimes we can also come across a group of pages which may form a **sink** (ex3 in result)and due to this sink the iterative method to find eigen vector may not converge.
And due to this sink other pages may not get page ranks even if they may be having link.

Thus, if there is a sink present then S is a block matrix given by,

S = $\begin{pmatrix} * & 0 \\ * & * \end{pmatrix}$ which is a reducible matrix and as in above example PageRanks of some pages can be zero inspite of each of them having links. [2]
To overcome this, S is modified to an irreducible matrix G as follows,
G is defined as,

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 2, No. 1, February : 2023

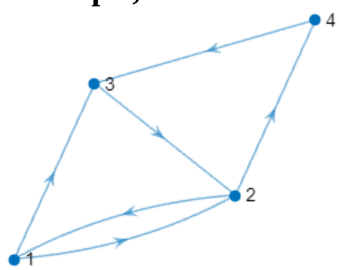
$$\mathbf{G} = \boldsymbol{\mu}\mathbf{S} + (\mathbf{1} - \boldsymbol{\mu})\mathbf{u}\mathbf{B}^{\mathbf{T}}, \qquad 0 \le \mu < 1,\ u_i \ge 0, \qquad \|u\|_1 = 1$$

G is called “Google matrix.”
A damping factor μ of 0.85 is used.
The iterative methods used to get eigen vector converge faster for $\mu = 0.85$.
G is stochastic and irreducible having a clear dominating eigenvalue that is 1.
If the eigenvalues of S are, 1 and $\lambda_2, \lambda_{3,\ldots\ldots,} \lambda_n$ where $|\lambda_j| \le 1$ for j = 2 to n, then the eigenvalues of G are 1 and $\mu\lambda_2, \mu\lambda_{3,\ldots\ldots,} \mu\lambda_n$.
Thus, PageRank vector is achieved iteratively from G using power method. [3]

III. RESULT:

“Hyperlink matrix” gives following eigen vector:

For example, consider ex1. where the webpage linkage is represented using Matlab.



The Hyperlink Matrix is given by

$$H = \begin{bmatrix} 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1 & 0 \\ 1/2 & 0 & 0 & 1 \\ 0 & 1/2 & 0 & 0 \end{bmatrix}^T \quad \text{and eigen vector } X = \begin{bmatrix} 0.250000 \\ 0.375000 \\ 0.250000 \\ 0.125000 \end{bmatrix}$$

As score of page 2 is 0.375 which is more than others, it will get more preference than other pages

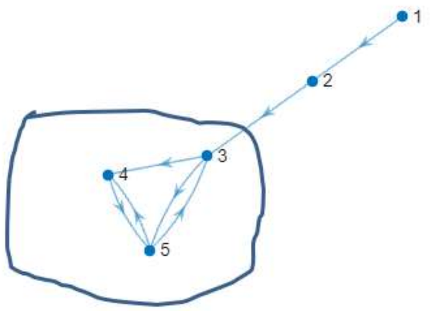
Dangling node give following result:

Power method fails to give an eigen vector.

$$\text{Eigen vector} = \begin{bmatrix} \text{NaN} \\ \text{NaN} \\ \text{NaN} \\ \text{NaN} \end{bmatrix}$$

Sink gives following eigen vector:

For example, consider ex3.

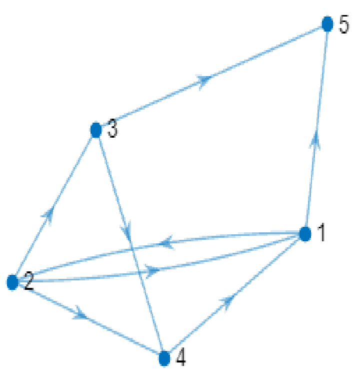


X = 0 0 0.2222 0.3333 0.4444 is the eigen vector

$$X = \begin{matrix} 0 \\ 0 \\ 0.2222 \\ 0.3333 \\ 0.4444 \end{matrix}$$

"Google Matrix" gives following eigen vector:

Using matlab we get following page rank for corresponding link of pages.



$$\text{Eigen vector } X = \begin{bmatrix} 0.288461538461538 \\ 0.230769230769231 \\ 0.115384615384615 \\ 0.173076923076923 \\ 0.192307692307692 \end{bmatrix}$$

$$\text{rank} = \begin{bmatrix} 1 \\ 2 \\ 5 \\ 4 \\ 3 \end{bmatrix}$$

Page 1 will get preference than other pages.

[4]

III. CONCLUSION:

In this paper, we have seen a review of how the concept of eigen values of Linear Algebra has an application in the Page Rank algorithm. It uses power method which is an iterative method to calculate the eigen vector. The hyperlink matrix H is corrected to a stochastic matrix S because of dangling nodes and then to Google matrix G which is irreducible and row stochastic. It uses the value of damping factor μ = 0.85. The PageRank vector is prone to changes in the damping factor μ.

IV. REFERENCES:

[1]. María José Beltrán Meneu1 , Marina Murillo Arcila2 , Enrique Jordá Mora, A Teaching Proposal For The Study Of Eigenvectors And Eigenvalues, Journal Of Technology And Science Education Jotse,

2017 – 7(1): 100-113 – Online Issn: 2013-6374 – Print Issn: 2014-5349 Https://Doi.Org/10.3926/Jotse.260

[2]. David Austin, Grand Valley State University,How Google Finds Your Needle in the Web's Haystack in Feature Column from the AMS.
[3]. Amy N Langville and Carl D Meyer. Google's PageRank and beyond: The science of search engine rankings. Princeton University Press, 2011
[4]. https://personalpages.manchester.ac.uk/staff/yanghong.huang/teaching/MATH36032/html/labdemo16.html