Matrix Computations
Chapter 4: Eigenvalues, Eigenvectors, and Eigendecomposition

Section 4.5 Power method for PageRank

Jie Lu
ShanghaiTech University

Case Study: PageRank

An algorithm used by Google to rank the pages of a search result [1]

More important webpages are likely to receive more links from other websites

Determine the importance of each webpage based on the quality and quantity of links pointing to it

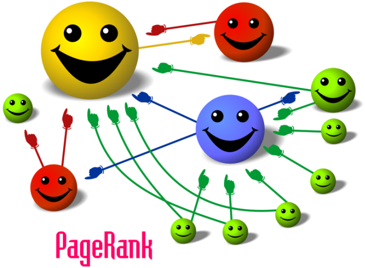


Figure: PageRank. Source: Wikipedia

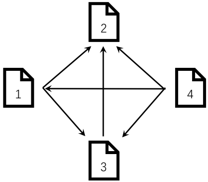
[1] K. Bryan and L. Tanya, "The 25, 000, 000, 000 eigenvector: The linear algebra behind Google," *SIAM Review*, vol. 48, no. 3, pp. 569–581, 2006.

PageRank Model

Let v_i be the importance score of page $i = 1, \ldots, n$, $\mathcal{L}_i$ be the set of pages containing a link to page i, and c_j be the number of outgoing links from page j

$$\sum_{j \in \mathcal{L}_i} \frac{v_j}{c_j} = v_i, \quad \forall i = 1, \ldots, n$$

Example:



$$\begin{aligned} \mathcal{L}_1 &= \{4\} & c_1 &= 2 \\ \mathcal{L}_2 &= \{1, 3, 4\} & c_2 &= 0 \\ \mathcal{L}_3 &= \{1, 4\} & c_3 &= 1 \\ \mathcal{L}_4 &= \emptyset & c_4 &= 3 \end{aligned}$$

$$\begin{bmatrix} 0 & 0 & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & 1 & \frac{1}{3} \\ \frac{1}{2} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

Notation and Definitions

Notation: For any $x, y \in \mathbb{R}^n$,

- $\mathbf{x} \geq \mathbf{y}$ means that $x_i \geq y_i$ for all i
- $\mathbf{x} > \mathbf{y}$ means that $x_i > y_i$ for all i
- $\mathbf{x} \ngeq \mathbf{y}$ means that $\mathbf{x} \geq \mathbf{y}$ does not hold
- The same notation applies to matrices

Definitions:

- $\mathbf{x}$ is said to be non-negative if $\mathbf{x} \geq \mathbf{0}$, and non-positive if $-\mathbf{x} \geq \mathbf{0}$
- $\mathbf{x}$ is said to be positive if $\mathbf{x} > \mathbf{0}$, and negative if $-\mathbf{x} > \mathbf{0}$
- The same definitions apply to matrices
- A square matrix $\mathbf{A}$ is said to be column-stochastic if $\mathbf{A} \geq \mathbf{0}$ and $\mathbf{A}^T\mathbf{1} = \mathbf{1}$
  - Each column $\mathbf{a}_i$ of column-stochastic $\mathbf{A}$ satisfies $\mathbf{a}_i^T\mathbf{1} = \sum_{j=1}^n a_{ji} = 1$

PageRank Problem

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a matrix s.t. $a_{ij} = 1/c_j$ if $j \in \mathcal{L}_i$ and $a_{ij} = 0$ if $j \notin \mathcal{L}_i$

Problem: Find a non-negative $\mathbf{v}$ s.t. $\mathbf{A}\mathbf{v} = \mathbf{v}$

- $\mathbf{A}$ is extremely large and sparse, so we choose the power method

Questions:

- Does a solution to $\mathbf{A}\mathbf{v} = \mathbf{v}$ always exist? Or, is $\lambda = 1$ always an eigenvalue of $\mathbf{A}$?
- Does $\mathbf{A}\mathbf{v} = \mathbf{v}$ have a non-negative solution? Or, is there a non-negative eigenvector associated with $\lambda = 1$?
- Is the solution to $\mathbf{A}\mathbf{v} = \mathbf{v}$ unique? Or, would there exist more than one eigenvector associated with $\lambda = 1$?
  - A unique solution is desired for PageRank
- Is $\lambda = 1$ the only eigenvalue that is the largest in modulus?
  - Required by the power method

PageRank Matrix Properties

Observation: In PageRank, $\mathbf{A}$ is column-stochastic if all pages have outgoing links

Properties: Let $\mathbf{A}$ be column-stochastic. Then,

- $\lambda = 1$ is an eigenvalue of $\mathbf{A}$
- $|\lambda| \leq 1$ for any eigenvalue λ of $\mathbf{A}$

Implications: There exists a solution to $\mathbf{Av} = \mathbf{v}$ and $\lambda = 1$ is an eigenvalue with the largest modulus

Remaining questions: We still don't know

- whether $\mathbf{v} \geq \mathbf{0}$ or not
- whether $\lambda = 1$ is the *only* eigenvalue that has the largest modulus (i.e., whether its algebraic multiplicity is 1 and no other distinct eigenvalues have modulus 1)

We resort to *non-negative matrix theory* to find the answers

Non-Negative Matrix Theory

Theorem (Perron-Frobenius)

Let $\mathbf{A}$ *be a positive square matrix. There exists an eigenvalue* ρ *of* $\mathbf{A}$ *s.t.*

- ρ *is real and* $\rho > 0$
- $|\lambda| < \rho$ *for any eigenvalue* λ *of* $\mathbf{A}$ *with* $\lambda \neq \rho$
- *There exists a positive eigenvector associated with* ρ
- *The algebraic multiplicity of* ρ *is* 1 *(so the geometric multiplicity of* ρ *is also* 1*)*

Theorem (more general matrix, weaker result)

Let $\mathbf{A}$ *be a non-negative square matrix. There exists an eigenvalue* ρ *of* $\mathbf{A}$ *s.t.*

- ρ *is real and* $\rho \geq 0$
- $|\lambda| \leq \rho$ *for any eigenvalue* λ *of* $\mathbf{A}$
- *There exists a non-negative eigenvector associated with* ρ

Modified PageRank Model

From the theorem for non-negative matrices, there exists a non-negative solution to $\mathbf{Av} = \mathbf{v}$, but we don't know whether there exists another solution $\mathbf{v}'$ and whether $\mathbf{v}' \ngeq \mathbf{0}$

For PageRank, we actually consider a modified version of $\mathbf{A}$

$$\tilde{\mathbf{A}} = (1 - \beta)\mathbf{A} + \beta \begin{bmatrix} 1/n & \dots & 1/n \\ \vdots & & \vdots \\ 1/n & \dots & 1/n \end{bmatrix}$$

where $0 < \beta < 1$ (typical value is $\beta = 0.15$), so that $\tilde{\mathbf{A}}$ is positive

From the Perron-Frobenius Theorem,

- $\lambda = 1$ is the only eigenvalue that has the largest modulus
- There exists only one eigenvector associated with $\lambda = 1$, either positive or negative
- Therefore, the power method can work