The Algorithm of Google

A Brief Explanation of PageRank

Zhang Ruiyang

University College London

ruiyang.zhang.20@ucl.ac.uk

18 Oct 2021

Abstract

This talk will be on the PageRank algorithm, which is used to rank the relative importance of webpages. It is largely influential, and has a lot to do with the success of Google search engine. This talk will focus on the Mathematical aspect of the algorithm, which can be viewed as an application of Markov chain and eigendecomposition. It is self-contained and requires no prerequisite, but it would be helpful to have some knowledge of discrete-time Markov chain and eigendecomposition of a matrix.

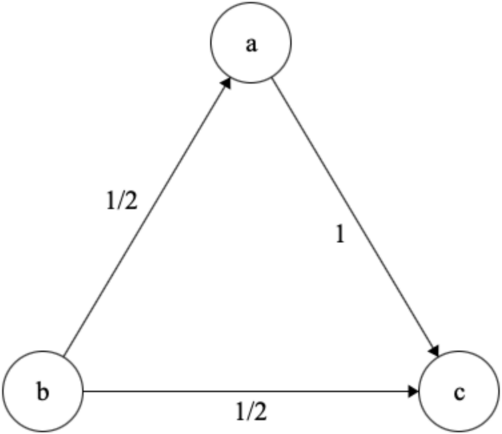
Goal

A computationally fast way to rank the relative importance of webpages, so we know what to return first after a search.

Observations of Webpages

- Webpages have links
- Links have directions
- Links on important pages are important
- Easy to manipulate webpages
- (Sink pages)

A Model of the Question



Definition of Markov Chain (1)

- **State space** I is a countable set with each $i \in I$ as a possible state. It is countable since we are working with discrete-time.
- **Distribution** on I is a collection $\lambda = (\lambda_i, i \in I)$ with λ_i for all i, and $\sum \lambda_i = 1$ due to normality.
- Working in the **probability space** $(\Omega, \mathcal{F}, \mathbb{P})$: sample space Ω, events $\mathcal{F}$ and probability measure $\mathbb{P}$.
- For a random variable $X : \Omega \to I$, we have $\lambda_i = \mathbb{P}(X = i)$.
- **Stochastic matrix** $P = (p_{ij} : i, j \in I)$ with every row $(p_{ij} : j \in I)$ being a distribution. p_{ij} here denotes the probability of going from state i to state j. (Not to be confused with the probability measure $\mathbb{P}$)

Definition of Markov Chain (2)

$(X_n)_{n\geq 0}$ is a Markov chain with **initial distribution** λ and **transition matrix** P if for $n \geq 0$ and $i_0, \ldots, i_{n+1} \in I$, we have

(1) $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$

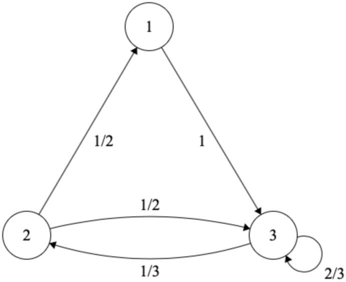
(2) $\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) = p_{i_n i_{n+1}}$

which we will then call it Markov(λ, P) in short.

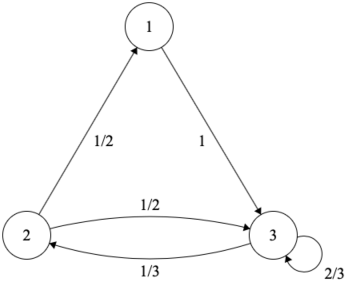
For any particular combination of positions of states X_0 to X_N for an integer N,

$$
\begin{aligned}
&\mathbb{P}(X_0 = i_0, \ldots, X_N = i_N) \\
&= \mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1 | X_0 = i_0) \cdots \mathbb{P}(X_N = i_N, | X_0 = i_0 \ldots, X_{N-1} = i_{N-1}) \\
&= \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{N-1} i_N}.
\end{aligned}
$$

Example of a Markov Chain



Example of a Markov Chain



The state space I is $\{1, 2, 3\}$, and let the initial distribution be $\lambda = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. For example, using the diagram, we can get

$$\begin{aligned}
&\mathbb{P}(X_0 = 1, X_1 = 3, X_2 = 3, X_3 = 2) \\
&\quad = \mathbb{P}(X_0 = 1)\mathbb{P}(X_1 = 3|X_0 = 1)\mathbb{P}(X_2 = 3|X_1 = 3)\mathbb{P}(X_3 = 2|X_2 = 3) \\
&\quad = \frac{1}{3} \times 1 \times \frac{2}{3} \times \frac{1}{3} = \frac{2}{27}.
\end{aligned}$$

Markov Property

Property of "memoryless" - the past does not depend on the future, only the current state does.

Let $(X_n)_{n\geq 0}$ be Markov(λ, P). Then, condition on $X_m = i$, $(X_{m+n})_{n\geq 0}$ is Markov(δ_i, P) and is independent of the random variables $X_0, \ldots, X_m$ where $\delta_i = (\delta_{ij} : j \in I)$ is the unit mass at i where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Equivalently,

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) = p_{i_n i_{n+1}}.$$

Markov Property

Property of “memoryless" - the past does not depend on the future, only the current state does.

Let $(X_n)_{n\geq 0}$ be Markov(λ, P). Then, condition on $X_m = i$, $(X_{m+n})_{n\geq 0}$ is Markov(δ_i, P) and is independent of the random variables $X_0, \ldots, X_m$ where $\delta_i = (\delta_{ij} : j \in I)$ is the unit mass at i where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Equivalently,

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) = p_{i_n i_{n+1}}.$$

“Life is like a Markov chain, your future only depends on what you are doing now, and independent of your past.”

Transition over Multiple Steps

The probability of going from state i and get to state j in two steps is

$$\sum_{k \in I} p_{ik} p_{kj}.$$

Using some knowledge from Linear Algebra, we can summarise all two-step transitions using P^2 of the stochastic matrix P. This can be extended to n-step transition of P^n.

Transition over Multiple Steps

The probability of going from state i and get to state j in two steps is

$$\sum_{k \in I} p_{ik} p_{kj}.$$

Using some knowledge from Linear Algebra, we can summarise all two-step transitions using P^2 of the stochastic matrix P. This can be extended to n-step transition of P^n.

Question: How would this matrix P^n look like, as $n \to \infty$?

Eigendecomposition of a Matrix

For $n \times n$ matrix A and some $n \times 1$ vector x, $Ax = \lambda x$ of some scalar λ. Such λ will be called an **eigenvalue**, and the vector x will be its corresponding **eigenvector**.

For eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ of A and the corresponding eigenvectors $v_1, v_2, \cdots, v_n$, we will have the eigenmatrix $Q = [v_1 \quad v_2 \quad \cdots \quad v_n]$ and a diagonal matrix Λ with the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ on its diagonal, we would have $A = Q\Lambda Q^{-1}$, the **eigendecomposition** of A.

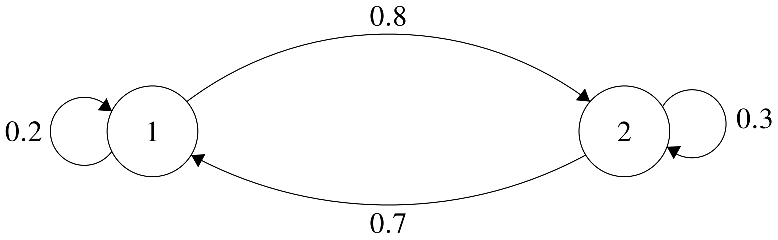
Eigendecomposition of a Matrix

For $n \times n$ matrix A and some $n \times 1$ vector x, $Ax = \lambda x$ of some scalar λ. Such λ will be called an **eigenvalue**, and the vector x will be its corresponding **eigenvector**.

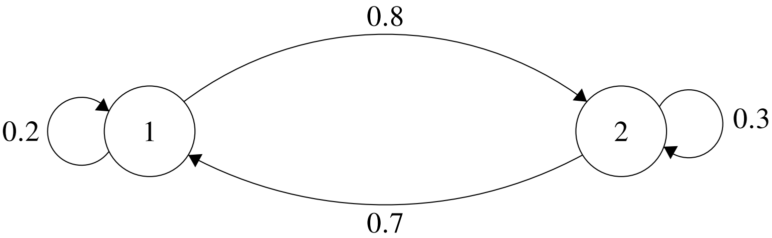
For eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ of A and the corresponding eigenvectors $v_1, v_2, \cdots, v_n$, we will have the eigenmatrix $Q = [v_1 \quad v_2 \quad \cdots \quad v_n]$ and a diagonal matrix Λ with the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ on its diagonal, we would have $A = Q\Lambda Q^{-1}$, the **eigendecomposition** of A.

One amazing property: $A^k = Q\Lambda^k Q^{-1}$, while Λ^k is a diagonal matrix with $\lambda_1^k, \lambda_2^k, \cdots, \lambda_n^k$ on its diagonal.

Example

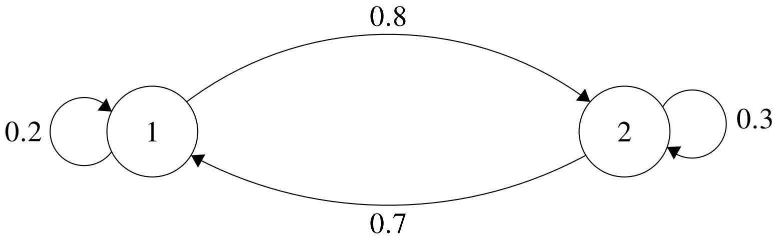


Example



Here, the transition matrix is $P = \begin{pmatrix} 0.2 & 0.8 \\ 0.7 & 0.3 \end{pmatrix}$. Its two eigenvalues are 1 and -0.5, so we have $P = Q \begin{pmatrix} 1 & 0 \\ 0 & -0.5 \end{pmatrix} Q^{-1}$, and $P^n = Q \begin{pmatrix} 1^n & 0 \\ 0 & (-0.5)^n \end{pmatrix} Q^{-1}$ of some eigenmatrix Q.

Example



Here, the transition matrix is $P = \begin{pmatrix} 0.2 & 0.8 \\ 0.7 & 0.3 \end{pmatrix}$. Its two eigenvalues are 1 and -0.5, so we have $P = Q \begin{pmatrix} 1 & 0 \\ 0 & -0.5 \end{pmatrix} Q^{-1}$, and $P^n = Q \begin{pmatrix} 1^n & 0 \\ 0 & (-0.5)^n \end{pmatrix} Q^{-1}$ of some eigenmatrix Q.

Thus, $\lim_{n\to\infty} P^n = Q \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} Q^{-1} = \begin{pmatrix} 7/15 & 8/15 \\ 7/15 & 8/15 \end{pmatrix}$, which indicates the equilibrium distribution of the chain.

PageRank

Assume we have n webpages $I_1, I_2, \cdots, I_n$. If I_1 has links directing to I_2 and I_3, a random surfer of the internet will have 50% chance of going to I_2, and 50% chance of going to I_3, just like a Markov chain. So, we can construct a transition matrix P of n webpages in that way, and the equilibrium distribution will offer us a ranking of the relative importance of pages.

In the previous example of two states, we have the equilibrium distribution $(7/15 \quad 8/15)^T$, meaning that we will be at State 1 for 7/15 of the time and at State 2 for the remaining 8/15. Then, we say that State 2 is a more important page than State 1.

PageRank

Assume we have n webpages $I_1, I_2, \cdots, I_n$. If I_1 has links directing to I_2 and I_3, a random surfer of the internet will have 50% chance of going to I_2, and 50% chance of going to I_3, just like a Markov chain. So, we can construct a transition matrix P of n webpages in that way, and the equilibrium distribution will offer us a ranking of the relative importance of pages.

In the previous example of two states, we have the equilibrium distribution $(7/15 \quad 8/15)^T$, meaning that we will be at State 1 for $7/15$ of the time and at State 2 for the remaining $8/15$. Then, we say that State 2 is a more important page than State 1.

Wait a second ... Will this always work?

Sink pages.

Sink pages.

To make sure an equilibrium distribution exists, we will introduce a damping factor, so there is some chance for the random surfer to leave the current page and go to any other pages (with equal probability). This way, we will always have an equilibrium distribution.

Extensions

- A more elaborated discussions on when a matrix can be eigendecomposed
- Other metrics to rank pages
- Other applications of eigendecomposition
- Can we ‘eigendecompose’ things other than matrix?

Reference

- Markov Chain - J. R. Norris
- The PageRank Citation Ranking: Bringing Order to the Web