

Kapitola 5

O vyhledávači Google

Jak funguje Google ? - obsah

- *Jak funguje Google ?*
Počítá vlastní vektory matice webu
Perronova-Frobeniova teorie

Jak porovnat důležitost stránek

důležitost stránky je přímo úměrná součtu důležitostí stránek, které na ni odkazují

zdánlivě definice kruhem

jak zaznamenat, které stránky na které dokazují ?

matice webu $H = (h_{ij})$

Příklad grafu webu

Definice důležitosti stránek

neznámou důležitost i -té stránky označíme x_i , $i = 1, 2, \dots, N$

N je počet stránek, které Google indexuje

slovní definici *důležitost i -té stránky je přímo úměrná součtu důležitostí stránek, které na ni odkazují*

zapíšeme jako

Jaká vlastní čísla a vlastní vektory má matice webu H ?

H je reálná matice

má tedy N komplexních vlastních čísel včetně algebraických násobností

definice: *spektrální poloměr* komplexní čtvercové matice A je

$$\rho(A) = \max\{|\lambda| : \lambda \text{ je vlastní číslo matice } A\} \geq 0$$

$H = (h_{ij})$ je *nezáporná matice*, tj. každý prvek $h_{ij} \geq 0$

Perronova-Frobeniova teorie nezáporných matic

Perronova-Frobeniova věta

je-li $A = (a_{ij})$ reálná nezáporná matice řádu n , pro kterou je A^k kladná pro nějaké $k \in \mathbb{N}$, pak platí

1. spektrální poloměr $\rho(A) > 0$
2. spektrální poloměr $\rho(A)$ je (kladné) vlastní číslo matice A
3. algebraická (a tedy také geometrická) násobnost $\rho(A)$ je 1
4. existuje kladný vlastní vektor $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ příslušný $\rho(A)$
5. vlastní vektor \mathbf{x} příslušný $\rho(A)$ a splňující $\|\mathbf{x}\|_1 = x_1 + x_2 + \dots + x_n = 1$ je určený jednoznačně,
Perronův vektor

co znamená předpoklad, že $A^k > 0$ pro nějaké k ?

Široké možnosti použití

jak porovnávat výkonnost fotbalových týmů ?

jak porovnávat výkonnost tenisových hráčků ?

Jak vylepšit matici webu - obsah

- *Jak vylepšit matici webu*
 - Započtením počtu odkazů na stránce
 - Teleportace
 - Frobeniova teorie

Pravděpodobnost kliku

stránka, ze které vede spousta odkazů, přispěje důležitosti všech stránek, na které odkazuje, stejně jako stránka, která odkazuje pouze na jednu jinou

matici webu $H = (\mathbf{h}_1 | \mathbf{h}_2 | \dots | \mathbf{h}_N)$ upravíme tak, že sloupec \mathbf{h}_j nahradíme sloupcem $(1/k_j)\mathbf{h}_j$, kde k_j je počet odkazů na stránce j

nová matice $S = ((1/k_1)\mathbf{h}_1 | (1/k_2)\mathbf{h}_2 | \dots | (1/k_N)\mathbf{h}_N)$

náš malý příklad

Další úprava

S je stále nezáporná matice, lze použít Perronovu-Frobeniou teorii, pokud S^k je kladná pro nějaké k

matice $S = (s_{ij})$ říká, že ze stránky j přeskočím na stránku i s pravděpodobností s_{ij}

nulové sloupce v matici S jsou nevýhodou - nějakou důležitost mají, ale nedělí se o ni s ostatními stránkami

proto ještě v matici S nahradíme nulové sloupce sloupci $(1/N)\mathbf{1}$, kde $\mathbf{1} = (1, 1, \dots, 1)^N$

naš malý příklad

Stochastické matice

matice S je nezáporná, navíc má každý sloupec součet 1

takovým maticím se říká *stochastické*

lze je interpretovat jako *matici náhodné procházky po webu*

zvolím-li si počáteční stránku i s pravděpodobností p_{0i} , pak to, kde se nacházím na začátku klikání je popsáno počátečním vektorem

$$\mathbf{p}_0 = (p_{01}, p_{02}, \dots, p_{0N})^T$$

$\mathbf{p}_k = S^k \mathbf{p}_0$ pak udává

Teleportace

zůstává problém s předpokladem, že S^k je kladná pro nějaké $k \in \mathbb{N}$

zakladatelé Googlu jej vyřešili *teleportací*

řekli si, že náhodný chodec po webu kliká na odkazy v 90% případů a v 10% prostě přeskočí na jakoukoliv jinou stránku

to znamená, že matici S nahradili maticí

$$G = 0,9 S + 0,1 (1/N) \mathbf{1} \cdot \mathbf{1}^T$$

náš malý příklad

Frobeniova teorie kladných matic

matice G je stále stochastická, navíc má všechny prvky kladné

Frobeniova věta

je-li $A = (a_{ij})$ kladná matice řádu n , pak platí

1. spektrální poloměr $\rho(A) = 1$
2. spektrální poloměr 1 je vlastní číslo matice A
3. algebraická (a tedy také geometrická) násobnost $\rho(A)$ je 1
4. existuje kladný vlastní vektor $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ příslušný $\rho(A)$
5. vlastní vektor \mathbf{x} příslušný $\rho(A)$ a splňující $\|\mathbf{x}\|_1 = x_1 + x_2 + \dots + x_n = 1$ je určený jednoznačně, *Perronův vektor*
6. pro jakékoliv jiné vlastní číslo $\lambda \neq 1$ matice A platí $|\lambda| < 1$

Geometrická intuice za Frobeniovou větou

Mocninná metoda - obsah

- *Mocninná metoda*
 - Jak spočítat Perronův vektor
 - Výhody mocninné metody

Mocninná metoda 1

matici G chápeme jako matici nad komplexními čísly

je diagonalizovatelná jako komplexní matice

má vlastní čísla $\lambda_1 = 1, \lambda_2, \dots, \lambda_N$ (s algebraickými násobnostmi)

přepokládáme $\lambda_1 = 1 > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_N|$

v \mathbb{C}^N existuje báze $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_N$ z vlastních vektorů

počáteční stav \mathbf{p}_0 vyjádříme jako

$$\mathbf{p}_0 = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_N \mathbf{u}_N$$

Mocninná metoda 2

potom

$$\mathbf{p}_k = G^k \mathbf{p}_0 = a_1 \lambda_1^k \mathbf{u}_1 + a_2 \lambda_2^k \mathbf{u}_2 + a_3 \lambda_3^k \mathbf{u}_3 + \cdots + a_N \lambda_N^k \mathbf{u}_N$$

to znamená, že

\mathbf{p}_k konverguje pro $k \rightarrow \infty$ k

rychlost konvergence závisí na

Google uvádí, že mu stačí mezi 50 a 80 iteracemi

Proč mocninná metoda funguje pro tak velkou matici

1. matici G není nutné ukládat celou
2. stačí uložit pouze H
3. tu lze dobře komprimovat
4. při výpočtu se matice H (nebo G) nemění
5. jediný mezi výpočet, který je třeba dočasně uložit, je \mathbf{p}_i