PageRank algoritmus

12. mája 2021

PageRank algoritmus

- Sergey Brin a Larry Page
- Stránky – na základe liniek chceme vybrať „najdôležitejšiu“.
- Hľadanie vlastného vektora.
- Markovovské reťazce.

PageRank algoritmus

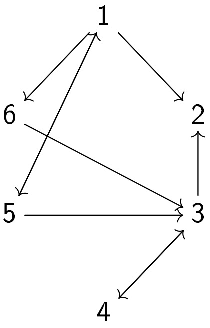
$$\vec{x} = \vec{e}A,$$

$\vec{x} = (x_1, \ldots, x_n)$ obsahuje ohodnotenia stránok, $\vec{e} = (1, 1, \ldots, 1)$ a matica A je

$$a_{ij} = \begin{cases} \frac{1}{n_i} & \text{ak } i\text{-ta stránka obsahuje odkaz na } j\text{-tu}, \\ 0 & \text{inak.} \end{cases}$$

pričom n_i označuje počet liniek na i-tej stránke.

PageRank algoritmus



$$A = \begin{pmatrix} 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

PageRank algoritmus

$$\vec{x}_0 = \vec{e}$$
$$\vec{x}_1 = \vec{x}_0 A$$
$$\vec{x}_2 = \vec{x}_1 A = \vec{x}_0 A^2$$
$$\dots$$
$$\vec{x}_n = \vec{x}_{n-1} A = \vec{x}_0 A^n$$

$$\vec{x} = \lim_{n\to\infty} \vec{x}_0 A^n$$

$$\vec{x} A = \lim_{n\to\infty} \vec{x}_0 A^{n+1} = \vec{x}$$

PageRank algoritmus

Definícia

Matica A sa nazýva *riadkovo stochastická*, ak súčet prvkov jej ľubovoľného riadku je 1.

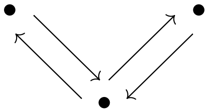
Tvrdenie

Ak matica A je riadkovo stochastická, tak číslo 1 je jej vlastnou hodnotou.

Tvrdenie

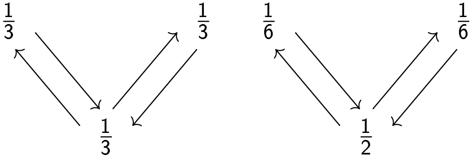
Nech A je riadkovo stochastická matica, $\vec{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ a $\vec{y} = (y_1, \ldots, y_n) = \vec{x}A$. Potom platí $\sum\limits_{i=1}^{n} y_i = \sum\limits_{i=1}^{n} x_i$.

Príklad – nekonverguje



x_1	x_2	x_3
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$
$\vdots$	$\vdots$	$\vdots$

Príklad – nekonverguje



PageRank algoritmus

Lema

Nech matica A je riadkovo stochastická a $a_{ij} > 0$ pre $i, j = 1, \dots, n$. Potom každý jej vlastný vektor prislúchajúci k vlastnej hodnote 1 má všetky súradnice rovnakého znamienka (všetky nezáporné alebo všetky nekladné).

Lema

Nech matica A je riadkovo stochastická a $a_{ij} \geq 0$ pre $i, j = 1, \dots, n$. Ak λ je jej vlastná hodnota, tak $|\lambda| \leq 1$.

PageRank algoritmus

Lema

Ak $\vec{\alpha}$, $\vec{\beta}$ sú lineárne nezávislé vektory, tak existujú koeficienty $c, d \in \mathbb{R}$ také, že $c\vec{\alpha} + d\vec{\beta}$ obsahuje prvky so zmiešanými znamienkami.

Dôsledok

Ak A je riadkovo stochastická matica, ktorej prvky sú kladné, tak podpriestor tvorený vlastnými vektormi k vlastnému číslu 1 je jednorozmerný.

$$G = \alpha A + (1-\alpha)\frac{1}{n}\vec{e}^T\vec{e}$$

Spektrálny rozklad

Veta (Spektrálny rozklad diagonalizovateľnej matice)

Ak je matica A typu $n \times n$ podobná s diagonálnou maticou $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, *tak existujú matice* $G_1, \ldots, G_n$ *také, že platí*

$$A = \lambda_1 G_1 + \lambda_2 G_2 + \cdots + \lambda_n G_n$$

a súčasne

$$G_1 + \cdots + G_n = I,$$

pre každé i platí $G_i^2 = G_i$ *a pre* $i \neq j$ *platí* $G_i G_j = 0$.

$$A^k = \lambda_1^k G_1 + \lambda_2^k G_2 + \cdots + \lambda_n^k G_n$$

Spektrálny rozklad

$$I = (\vec{\alpha}_1^T, \dots, \vec{\alpha}_n^T) \begin{pmatrix} \vec{\beta}_1 \\ \vdots \\ \vec{\beta}_n \end{pmatrix} = \vec{\alpha}_1^T \vec{\beta}_1 + \cdots + \vec{\alpha}_n^T \vec{\beta}_n = G_1 + \cdots + G_n.$$

$$I = \begin{pmatrix} \vec{\beta}_1 \\ \vdots \\ \vec{\beta}_n \end{pmatrix} (\vec{\alpha}_1^T, \dots, \vec{\alpha}_n^T) = \begin{pmatrix} \vec{\beta}_1 \vec{\alpha}_1^T & \dots & \vec{\beta}_1 \vec{\alpha}_n^T \\ \dots & \dots & \dots \\ \vec{\beta}_n \vec{\alpha}_1^T & \dots & \vec{\beta}_n \vec{\alpha}_n^T \end{pmatrix}$$

Dostaneme $\vec{\beta}_i \vec{\alpha}_i^T = 1$ a $\vec{\beta}_i \vec{\alpha}_j^T = 0$ pre $i \neq j$.

$$G_i^2 = \vec{\alpha}_i^T (\vec{\beta}_i \vec{\alpha}_i^T) \vec{\beta}_i = \vec{\alpha}_i^T \vec{\beta}_i = G_i$$

$$G_i G_j = \vec{\alpha}_i^T (\vec{\beta}_i \vec{\alpha}_j^T) \vec{\beta}_j = \vec{\alpha}_i^T .0. \vec{\beta}_j = 0.$$

PageRank algoritmus

$$A^k = G_1 + \lambda_2^k G_2 + \cdots + \lambda_n^k G_n$$

$$\vec{x_0} A^k = \vec{x_0} G_1 + \lambda_2^k \vec{x_0} G_2 + \cdots + \lambda_n^k \vec{x_0} G_n$$

Tvrdenie

Nech A je riadkovo stochastická matica a jej vlastné čísla sú $1, \lambda_2, \ldots, \lambda_n$. *Potom matica*

$$G = \alpha A + (1-\alpha)\frac{1}{n}\vec{e}^T\vec{e}$$

má vlastné hodnoty $1, \alpha\lambda_2, \ldots, \alpha\lambda_n$.