

Metody Obliczeniowe w Nauce i Technice

Laboratorium 8

Page Rank

29 kwietnia 2020

Literatura

- <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>
- <https://snap.stanford.edu/data/>

Uwagi

Zadanie wyślij jako jupyter notebook. Jeśli korzystasz z zewnętrznych bibliotek podaj ich nazwy i wersje. Przy każdym z zadań opisz kolejne kroki, otrzymane wyniki, a także wnioski. Opisz jakie problemy można napotkać próbując wykorzystać takie metody dla bardzo dużych grafów (10^9 wierzchołków).

Zadanie 1 Prosty ranking wierzchołków

Zaimplementuj prosty model błędzenia przypadkowego po grafie skierowanym:

$$\mathbf{r}(u) = d \sum_{v \in B_u} \frac{\mathbf{r}(v)}{N_v}, \quad (1)$$

gdzie $\mathbf{r}(u)$ oznacza ranking wierzchołka u , parametr c jest używany w normalizacji, B_u jest zbiorem wierzchołków, z których wychodzą krawędzie do wierzchołka u , F_v oznacza zbiór wierzchołków, do których dochodzą krawędzie z wierzchołka v , a $N_v = |F_v|$. W zapisie macierzowym:

$$\mathbf{r} = d\mathbf{A}\mathbf{r}, \quad (2)$$

gdzie \mathbf{A} jest macierzą adiacencji grafu, w której każdy wiersz u jest przeskalowany wyjściowym stopniem wierzchołka u .

$$\mathbf{A}_{u,v} = \begin{cases} \frac{1}{N_u} & \text{jeśli krawędź } (u, v) \text{ istnieje} \\ 0 & \text{w przeciwnym wypadku} \end{cases} \quad (3)$$

Zauważ, że \mathbf{r} może zostać obliczony jako dominujący wektor własny macierzy \mathbf{A} za pomocą metody potęgowej (dominujący wektor własny \mathbf{q}_1 znormalizowany za pomocą normy L_1). Przetestuj poprawność obliczeń korzystając z 3 dowolnych silnie spójnych grafów skierowanych o liczbie wierzchołków większej niż 10.

Zadanie 2 Page Rank

Rozszerz model z poprzedniego zadania, dodając możliwość skoku do losowego wierzchołka grafu:

$$\mathbf{r}(u) = d \sum_{v \in B_u} \frac{\mathbf{r}(v)}{N_v} + (1 - d)\mathbf{e}(u), \quad (4)$$

W zapisie macierzowym:

$$\mathbf{r} = (d\mathbf{A} + (1 - d)\mathbf{e} \otimes \mathbf{1})\mathbf{r} \quad (5)$$

gdzie $\|\mathbf{r}\|_1 = 1$, a \mathbf{e} jest wektorem zawierającym prawdopodobieństwa odwiedzania wierzchołków przez losowy skok. Wykorzystaj metodę potęgową do obliczenia Page Rank jako dominującego wektora własnego macierzy $\mathbf{B} = d\mathbf{A} + (1 - d)\mathbf{e} \otimes \mathbf{1}$.

1. \mathbf{r}_0
2. **do**
3. $\mathbf{r}_{i+1} = \mathbf{B}\mathbf{r}_i$
4. $d = \|\mathbf{r}_i\|_1 - \|\mathbf{r}_{i+1}\|_1$
5. $\mathbf{r}_{i+1} = \mathbf{r}_{i+1} + d\mathbf{e}$
6. $\delta = \|\mathbf{r}_{i+1} - \mathbf{r}_i\|_1$
7. **while** $\delta > \epsilon$

Przetestuj działanie zaimplementowanego algorytmu Page Rank dla wybranych grafów z bazy SNAP. Przetestuj różne wartości parametru d (0.9, 0.85, 0.75, 0.6, 0.5) oraz różne postacie wektora \mathbf{e} , przykładowo $\mathbf{e} = \frac{1}{n}[1, 1, \dots, 1]$.