

Zadania z PageRank

Joanna Kołodziejczyk

Celem zajęć jest przećwiczenie obliczeniowo algorytmu PageRank, co pozwoli na zrozumienie koncepcji i sprawdzenie praktyczne jego działania.

1 Przykład - zadanie za 1 pkt

Dana jest mała sieć stron z węzłami oznaczonymi: A, B, C, D:

- A ma połączenie/linki do B, C, D
- B ma połączenie/link do A
- C ma połączenie/link do A
- D ma połączenie/link do A

Przyjmijmy, że współczynnik tłumienia jest równy $d = \frac{7}{8}$

Aby obliczyć PageRank dla każdego węzła należy rozwiązać następujące równanie (wykład wzór (9)):

$$P = \left((1-d) \frac{1}{n} E + dA^T \right) P \quad (1)$$

$$\begin{pmatrix} p_A \\ p_B \\ p_C \\ p_D \end{pmatrix} = \left[\frac{1}{8} \cdot \frac{1}{4} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} + \frac{7}{8} \begin{pmatrix} 0 & 1 & 1 & 1 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \end{pmatrix} \right] \begin{pmatrix} p_A \\ p_B \\ p_C \\ p_D \end{pmatrix}$$

Należy rozwiązać (**1 pkt**) równanie przyjmując równość wszystkich p_x , gdzie $x \in \{A, B, C, D\}$, co oznacza takie samo prawdopodobieństwo wyboru każdego węzła (suma wszystkich p_x musi być równa 1). Można się wspierać przewodnikiem ze strony <http://pi.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html>.

2 Zadanie 1 - za 3 pkt

Dana jest mała sieć stron z węzłami oznaczonymi: X, Y, Z, V, U:

- X ma połączenie/linki do Y, Z, V
- Y ma połączenie/link do Z

- Z ma połączenie/link do Y i V
- V ma połączenie/link do Z
- U ma połączenie/link do Y, Z, V

Narysuj schemat i wykonaj następujące zadania:

1. Podać wartości PageRank dla węzła X i węzła U jako funkcję współczynnika tłumienia (*d* wykład wzór 11 - PR dla pojedynczego węzła). Wykorzystać wzór z wykładu. **(1 pkt)**
2. Należy wykazać, że węzły Y i V mają ten sam PageRank, bez względu na współczynnik tłumienia. **(1 pkt)**
3. Ustalić współczynnik tłumienia na 0.75. Rozpisać macierz A (rozkład prawdopodobieństw przejść) i powiązane z nią równania PageRank. Następnie obliczyć PageRank dla każdego węzła. **(1 pkt)**

3 Rozliczenie zadania

Wyniki, np skan obliczeń, mają mieć postaci dokumentu .pdf (lub kodu np z notatnika Jupiter). Nazwa pliku powinna zawierać nazwisko autora.