

Ohodnocovacie algoritmy v grafoch

Ján Suchal

Fakulta informatiky a informačných technológií
Slovenská technická univerzita v Bratislave

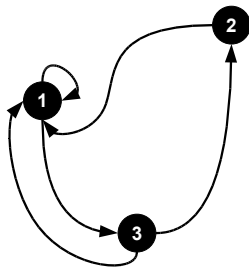
7. november 2007

Obsah

- 1 Grafy v maticovom počte
- 2 Ohodnocovacie algoritmy
 - Mocninová metóda
 - PageRank
 - HITS
 - NodeRanking
 - Model deravého kondenzátora
 - Šírenie aktivácie
- 3 Reálna aplikácia

Incidenčná matica grafu

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

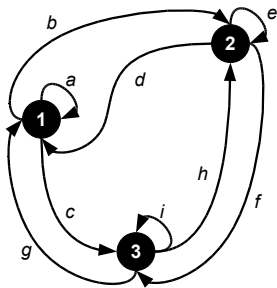


$$A_{ij} = \begin{cases} 1 & \text{ak existuje hrana z vrcholu } i \text{ do vrcholu } j \\ 0 & \text{inak} \end{cases}$$

Incidenčná matica ováňovaného grafu

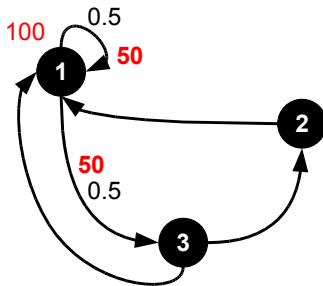
$$\mathbf{A} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

$$a, b, \dots, i \in \mathbb{R}$$



Násobenie matíc ako šírenie

$$\mathbf{r} = \begin{pmatrix} 100 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.5 & 0 & 0.5 \\ 1 & 0 & 0 \\ 0.5 & 0.5 & 0 \end{pmatrix} = \begin{pmatrix} 50 & 0 & 50 \end{pmatrix}$$



Mocninová metóda

- štartovací vektor \mathbf{r}_0
- $\mathbf{r}_k = \mathbf{r}_{k-1}\mathbf{A}$
- konverguje ku kladnému vlastnému vektoru (eigenvector)
 - stochastická – ide o uzavretý systém
 - ireducibilná – z každého vrcholu sa dá dostať do každého iného
 - aperiodická – periódy všetkých vrcholov sú jedna
- stabilný stav markovovského procesu

PageRank

- $\mathbf{M} = (1 - d)\mathbf{A} + d\frac{\mathbf{E}}{n}$
- \mathbf{E} jednotková matica, n počet vrcholov, d pravdepodobnosť skoku na náhodný vrchol
- model náhodného surfera
- vždy stochastická, ireducibilná, aperiodická matica
- $\mathbf{r}_k = \mathbf{r}_{k-1}\mathbf{M}$

HITS

- dve rôzne ohodnotenia každého vrcholu
 - autorita – $\mathbf{a}_k = \mathbf{h}_k \mathbf{A}$
 - rozcestnosť – $\mathbf{h}_k = \mathbf{a}_k \mathbf{A}^T$
- po úprave
 - autorita – $\mathbf{a}_k = \mathbf{a}_{k-1} \mathbf{A}^T \mathbf{A}$
 - rozcestnosť – $\mathbf{h}_k = \mathbf{h}_{k-1} \mathbf{A} \mathbf{A}^T$

NodeRanking

- pravdepodobnosť skoku na náhodný vrchol závisí od stupňa vrcholu
- $J_{ii} = \frac{1}{\sigma(i)+1}$
- $M = J \frac{E}{n} + (E - J)A$

Model deravého kondenzátora

- $\mathbf{M} = (1 - \gamma)\mathbf{I} + \alpha\mathbf{A}$
- γ pomer napätia, ktorý sa stratí α pomer napätia, ktorý sa prešíri na susedov
- napätie sa stráca \Rightarrow nutné dodávať dodatočne v každej iterácii
- $\mathbf{r}_k = \mathbf{r}_0 + \mathbf{r}_{k-1}\mathbf{M}$

Šírenie aktivácie

- $\mathbf{r}_k = \mathbf{r}_{k-1}\mathbf{A}$
- výsledné ohodnotenie $\mathbf{c}_k = \sum_{i=0}^k \mathbf{r}_i$
- s prahom $\mathbf{r}_k = \tau_\theta(\mathbf{r}_{k-1}\mathbf{A})$
- $(\tau_\theta(\mathbf{r}))_i = \begin{cases} \mathbf{r}_i & \text{ak } \mathbf{r}_i > \theta \\ 0 & \text{inak} \end{cases}$

Problémová doména

- www.hokejportal.sk
 - cca 400 registrovaných používateľov
- hodnotenie používateľov navzájom (stupnica -5 až 5)
 - cca 4000 hodnotení
- *“Kto je ako (ne)populárny?”*

Vlastnosti dôvery a nedôvery

- kladné hodnotenie - tranzitívnosť
 - *"Tvoj priateľ je aj môj priateľ."*
- záporné hodnotenie - tranzitívnosť?
 - *"Nepriateľ môjho nepriateľa je môj spojenec."*
 - *"Nepriateľ môjho nepriateľa je aj môj nepriateľ."*

Realizácia

- *“Nepriateľ môjho priateľa je aj môj nepriateľ.”*
- kladné hodnotenie \mathbf{T}
 - PageRank \Rightarrow vektor popularity používateľov \mathbf{t}
- záporné hodnotenie \mathbf{D}
 - vektor nepopularity $\mathbf{d} = \mathbf{tD}$
- výsledný vektor popularity
 - $\mathbf{r} = \mathbf{w}_t \mathbf{t} - \mathbf{w}_d \mathbf{d}$
 - $\mathbf{w}_t, \mathbf{w}_d$ sú počty kladných/záporných hodnotení

Vyhodnotenie

- “Väčšina používateľov s vysokým záporným hodnotením má zakázaný prístup alebo už neprispievajú dlhšie sami.”

Literatúra

- L. Page, S. Brin, R. Motwani, T. Winograd *The PageRank Citation Ranking: Bringing Order to the Web*
- J. M. Kleinberg *Authoritative Sources in a Hyperlinked Environment*
- J. Pujol, R. Sanguesa, J. Delgado *Extracting Reputation in Multiagent Systems by Means of Social Network Topology*
- Z. Huang, H. Chen, D. Zeng *Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering*
- M. Ceglowski, A. Coburn, J. Cuadrado *Semantic Search of Unstructured Data Using Contextual Network Graphs*
- R. Guha, R. Kumar, P. Raghavan, A. Tomkins *Propagation of Trust and Distrust*