

Información no estructurada

Práctica 4 – Pagerank y sistemas de recomendación

Javier Aróstegui Martín

Contenido

Parte 1: Pagerank	2
Fórmulas utilizadas:.....	2
Resultados:.....	2
R = 0.1 sin personalizar:.....	2
R = 0.85 sin personalizar:.....	2
R = 0.1 con personalización en los nodos A y D:	2
R = 0.85 con personalización en los nodos A y D:	2
Parte 2: Recomendación:	3
Fórmulas utilizadas:.....	3
Resultados:.....	4

Parte 1: Pagerank

Fórmulas utilizadas:

$$\vec{\pi}^{(k+1)t} = r\vec{\pi}^{(k)t}M + [r\vec{\pi}^{(k)t}\vec{a} + (1-r)]\frac{1}{n}\vec{e}^t$$

He utilizado esta fórmula para no tener que guardar la matriz G, que aunque en este ejemplo no es muy grande, puede llegar a ser muy ineficiente guardarla.

$$\vec{\pi}^{(k+1)t} = r\vec{\pi}^{(k)t}M + [r\vec{\pi}^{(k)t}\vec{a} + 1-r]\vec{v}^t$$

Para el pagerank personalizado he utilizado la misma fórmula, pero utilizando el vector v en vez del e

Resultados:

R = 0.1 sin personalizar:

En esta lista se muestran los resultados finales de pagerank sin personalizar desde el nodo A hasta el nodo E. En este caso se hacen 300 iteraciones, pero no son necesarias tantas para que el resultado converja.

```
[0.19105729 0.21389171 0.20338356 0.19105729 0.20061015]
```

R = 0.85 sin personalizar:

Cuando aumentamos r le estamos dando más valor a los nodos importantes y por tanto observamos que el nodo B acapara mayor pagerank ya que es el nodo más apuntado, por nodos con un pagerank relativamente alto y no apunta a nadie.

```
[0.14095691 0.31402498 0.20319762 0.14095691 0.20086359]
```

R = 0.1 con personalización en los nodos A y D:

En este caso al ser r muy pequeña le estamos dando mucha importancia al vector v (vector de personalización) y por tanto los nodos afines al usuario obtienen un pagerank muy elevado.

```
[0.45203538 0.02557702 0.04571508 0.45203538 0.02463715]
```

R = 0.85 con personalización en los nodos A y D:

Ahora que hemos aumentado r, ya no se le da tanta importancia al vector de personalización y por tanto el nodo B (al que le corresponde el pagerank más alto) recupera su valor alto. Los nodos A y D han aumentado su valor respecto al pagerank no personalizado.

```
[0.17699016 0.27210861 0.1967001 0.17699016 0.17721098]
```

Parte 2: Recomendación:

Fórmulas utilizadas:

Para la recomendación he utilizado **knn basado en usuarios** con las siguientes fórmulas para el rating esperado y la similitud:

$$\hat{r}(u, x) = c \sum_{v \in \eta_k(u)} \text{sim}(u, v) r(v, x)$$

$$c = \frac{1}{\sum_{v \in \eta_k(u)} |\text{sim}(u, v)|}$$

$$\eta_k(u) = \text{top}(k, \text{sim}(u, \cdot), U)$$

$$I(v) = \{y \in I \mid r(v, y) \neq \emptyset\}$$

$$\text{sim}(u, v) = \cos(r(u), r(v)) = \frac{\sum_{x \in I} r(u, x) r(v, x)}{\sqrt{\sum_{x \in I} r(u, x)^2 \sum_{x \in I} r(v, x)^2}}$$

Para la evaluación de recomendación he usado MAE, RMSE, P@k y R@k como se indica en el enunciado usando las siguientes formulas:

$$\text{MAE} = \frac{1}{|R_{\text{test}}|} \sum_{(u, x) \in R_{\text{test}}} |\hat{r}(u, x) - r_{\text{test}}(u, x)|$$

$$\text{RMSE} = \sqrt{\frac{1}{|R_{\text{test}}|} \sum_{(u, x) \in R_{\text{test}}} (\hat{r}(u, x) - r_{\text{test}}(u, x))^2}$$

$$P@k = \frac{1}{|U|} \sum_{u \in U} \frac{|\text{Rel}_u @ k|}{k}$$

$$R@k = \frac{1}{|U|} \sum_{u \in U} \frac{|\text{Rel}_u @ k|}{|\text{Rel}_u|}$$

Resultados:

K = 2:

```
K = 2
MAE: 2.08900999735994
RMSE: 2.5012653004401844
Precision@5: 0.38620689655172413
Recall@5: 0.16724142715522025
```

K = 5:

```
K = 5
MAE: 1.729852632334875
RMSE: 2.062445290898177
Precision@5: 0.4758620689655172
Recall@5: 0.20747954153126563
```

K = 10:

```
K = 10
MAE: 1.6939702849096283
RMSE: 1.9929473904990764
Precision@5: 0.503448275862069
Recall@5: 0.22814867125211952
```

Podemos observar como cuando aumentamos K, todas las métricas mejoran. Esto tiene sentido y nos indica que el algoritmo es capaz de aprender lo necesario cuando se le enseña suficiente información.