

CAIM Lab, Session 5:

Pagerank

David Sanchez

Josep Cordón

18/11/2019

1 - Introducció

Aquesta sessió de laboratori ha consistit en implementar una de les versions de l'algoritme Page Rank. Aquest algoritme va néixer amb l'objectiu de mesurar la importància de les pàgines web per poder assignar un valor a cada una d'elles útil a l'hora d'oferir els resultats d'una cerca segons la seva rellevància. Tot i això aquest algoritme també s'ha aplicat a camps totalment diferents, i en el nostre cas l'utilitzarem en una xarxa d'aeroports i rutes.

2 - Implementació

Primerament hem definit les diferents estructures de dades que ens permetran implementar el mètode alternatiu proposat per calcular el Page rank:

- **Airport:** classe amb els atributs de cada aeroport (codi, nom, rutes, diccionari ruta-pes, out-weight i PageIndex)
- **Edge:** classe amb els atributs de cada ruta (origen, destinació, nombre de rutes)
- **edgeList** : Llista de rutes
- **edgeHash:** hash de les rutes i pes
- **airportList:** llista d'aeroports
- **airportHash:** hash key IATA code -> Airport
- **OutAir** : llista dels valors *out* de cada aeroport
- **PageRank:** page ranks

Pel que fa a al nombre d'iteracions, hem definit una *stopping condition* que té en compte la convergència de les iteracions consecutives, concretament l'hem fixat tinguent en compte sis decimals.

Un altre aspecte que hem hagut de tenir en compte és la condició de que després de cada iteració la suma dels elements de P sigui 1. Per això hem realitzat una normalització d'aquests elements al final de cada iteració.

3 - Experimentació

Un cop hem realitzat la implementació de l'algoritme hem realitzat diferents experiments per tal de veure com la variació dels paràmetres afecta el resultat.

Stopping Condition

En aquest cas hem variat els decimals a tenir en compte a l'hora de realitzar més o menys iteracions. Amb aquestes variacions hem vist que com menys decimals tenim en compte l'algoritme acaba més ràpid realitzant menys iteracions.

A mesura que augmenta el nombre de iteracions els resultats són més precisos fins a arribar al punt de convergència on hem vist que encara que es realitzin més iteracions no té efecte en la precisió dels resultats.

Damping Factor

Normalment els valor d'aquest factor varien entre 0.8 i 0.9, en el nostre cas:

- **Damping factor = 0.8** -> 8 iteracions per convergir
- **Damping factor = 0.85** -> 8 iteracions per convergir
- **Damping factor = 0.9** -> 9 iteracions per convergir

- **Damping factor = 0.2** -> 4 iteracions per convergir

Variant aquest valor hem vist que a mesura que augmenta aquest valor, l'algoritme tarda més a convergir i realitza un major nombre de iteracions.