

Esperienze di programmazione

Lezione 7

Gianna Del Corso <delcorso@di.unipi.it>

27 Marzo 2015

1 PageRank

Scopo dell'esercitazione è l'implementazione dell'algoritmo PageRank usando il metodo delle potenze o il metodo di Arnoldi (cioè utilizzando la funzione `eigs` di MatLab/OCTAVE). Si faranno in particolare esperimenti per valutare l'algoritmo per differenti valori del parametro α .

1. Si consideri un grafo diretto con n nodi. Si costruisca la matrice di adiacenza G , e la matrice P ottenuta dividendo gli elementi di ogni riga per l'outdegree del nodo corrispondente. Sia $\bar{P} = P + \mathbf{d}\mathbf{v}'$ con $d_i = 1$ se il nodo i non ha nessun arco uscente e $d_i = 0$ altrimenti e $v_i = 1/n$.

Sia $\hat{P} = (1 - \alpha)\bar{P} + \alpha\mathbf{e}\mathbf{v}'$.

2. Tenendo conto che per moltiplicare $\hat{P}^T \mathbf{x}$ possiamo procedere come segue nel caso il vettore \mathbf{x} sia tale che $\|\mathbf{x}\|_1 = 1$.
 - Si calcola $\mathbf{y} = (1 - \alpha)P^T \mathbf{x}$
 - $\gamma = \|\mathbf{x}\|_1 - \|\mathbf{y}\|_1$,
 - $\mathbf{y} = \mathbf{y} + \gamma \mathbf{v}$

Si scriva una funzione `function y=Px(P, x, alpha)` che implementi il prodotto $\mathbf{y} = \hat{P}^T \mathbf{x}$ secondo l'algoritmo precedente. Si verifichi su un esempio piccolo che il prodotto ottenuto è corretto.

3. Si implementi il metodo delle potenze per il calcolo del pagerank utilizzando la funzione scritta al punto precedente per eseguire ad ogni passo il prodotto $\hat{P}^T \mathbf{x}$. In particolare si scriva la funzione `function pr=Pr_method(G, alpha, eps)` procedendo come descritto qui di seguito
 - Si scelga un vettore di partenza casuale \mathbf{x}_0 di norma 1 unitaria.

- Fintanto che il residuo tra due iterazioni successive è maggiore di `eps` e se non sono state fatte più di 100 iterazioni si calcola il nuovo valore di $\mathbf{x} = \hat{P}\mathbf{x}$.
- Il vettore `pr` restituito è l'ultimo valore \mathbf{x} calcolato.

Si testi sulla matrice di Stanford e si faccia un plot del residuo.

4. Si implementi il calcolo di PageRank utilizzando il comando `eigs`. Per non dover memorizzare la matrice piena \hat{P} , si chiami la funzione `eigs` con una handle alla moltiplicazione matrice-vettore, ovvero si realizzi una funzione `function y=Px(x, P, alpha)` che restituisce $\mathbf{y} = \hat{P}^T\mathbf{x}$. Si noti che poichè in `eigs` non si può controllare la norma dei vettori nella moltiplicazione matrice-vettore, non si può utilizzare il “trucco” del punto precedente.
5. Si testino i due algoritmi usando i valori 0.15 e 0.01 per il parametro α . Si confronti la velocità di convergenza dei due algoritmi e il tempo impiegato (usando `tic` e `toc` di OCTAVE). Si confrontino i vettori ottenuti di pagerank per i due valori del parametro di teletrasporto.

2 DATI

A scopo di test si può utilizzare la matrice di adiacenza `stanford-web.dat`.

La matrice può essere letta con `load stanford-web.dat` e poi convertendo la lista con il comando `A=spconvert(stanford_web)`. Si modifichi `A` affinché la matrice risulti quadrata.