Calcolo del PageRank™ di Google per $c \longrightarrow 1$

"Una semplice alternativa"

Antonio Cicone
Università degli Studi di L'Aquila

in collaborazione con

Stefano Serra Capizzano
Università dell'Insubria

16 febbraio 2009

Indice I

G matrice stocastica per righe che contiene informazioni riguardanti gli iperlink presenti nelle pagine Web

PageRank™

Vettore y tale che: $y^T G = y^T$

Esistono in generale infinite soluzioni a causa della riducibilità della matrice G.

Modello di Cammino Libero

Si definisce c probabilità con cui l'utente segue gli iperlink del Web

$$G \rightsquigarrow G(c) = cG + (1-c)ev^T$$

e vettore di tutti 1, $v \geq 0$ vettore di personalizzazione, $\|v\|_1 = 1$

Adesso la soluzione di $y(c)^T G(c) = y(c)^T$ esiste unica in particolare modo per valori di $c \in \mathbb{C}$ in un intorno di 1.

Dalle proprietà del vettore di PageRank^{TM} sappiamo che esiste unico il limite

$$\lim_{c \to 1} y(c) = Nv = \tilde{y}$$

$N \in M_n(\mathbb{R})$ è detto *Proiettore Ergodico* ed è univocamente definito come

$$N = [\eta_1 \cdots \eta_n]$$

con η_i, $i = 1, \ldots, n$, autovettori sinistri non negativi di G relativi all'autovalore 1.
Perciò il PageRank^{TM} è dato da una combinazione convessa di questi autovettori

$$\tilde{y} = v_1\eta_1 + \ldots + v_n\eta_n$$

i quali, però, non sono noti a priori.

Idea per il calcolo del PageRank^{TM} $\tilde{y}$

Modifica algebrica della matrice G tale che i suoi autovettori rimangono costanti ed il metodo delle potenze diventa convergente in modo incondizionato.

$$G \rightsquigarrow G_\delta = \delta G + (1-\delta)I, \text{ con } \delta \in (0,1)$$

Si dimostra che

$$\lim_{n\to\infty} {G_\delta^T}^n v = Nv = \tilde{y}$$

La convergenza del metodo è si incondizionata, ma la sua velocità è funzione di δ.

Si vuole accelerare la velocità di convergenza agendo sul parametro

$$\min_{\delta\in(0,1)} \max_{\lambda_j(\delta)\in\sum(G_\delta)-\{1\}} |\lambda_j(\delta)| < 1$$

Generalizzare $G_\delta = \delta G + (1-\delta)I$, polinomio lineare in G a coefficienti non negativi e con l'autovalore 1 punto fisso della trasformazione, al caso di un generico polinomio $p(G)$.

Studiare in modo più approfondito il problema dell'ottimizzazione della velocità di convergenza del metodo con particolare enfasi a casi concreti.