

Esame di Matematica Computazionale a.a. 2006-2007
Corso di Laurea in Informatica (triennale)
Esercizio N. 1

Costruire un mini motore di ricerca mediante Decomposizione a Valori Singolari (SVD).

1. Costruire la matrice X parole-documenti basata su 9 rapporti tecnici e considerando all'interno della matrice solo le parole riportate in corsivo per ciascun titolo dei rapporti tecnici elencati:
 - (a) *Human machine interface for computer applications*
 - (b) *A survey of user opinion of computer system response time*
 - (c) *The EPS user interface management system*
 - (d) *System and human system engineering testing of EPS*
 - (e) *Relation of user perceived response time to error measurement*
 - (f) *The generation of random, binary, ordered trees*
 - (g) *The intersection graph of paths in trees*
 - (h) *Graph minors IV: Widths of trees and well-quasi-ordering*
 - (i) *Graph minors: A survey*
2. Considerare le seguenti queries (q) al motore di ricerca:
 - (a) **human, interface, user**
 - (b) **survey, computer**
 - (c) **computer, trees, graph**
 - (d) **computer, time**
 - (e) **response, time**
3. Calcolare la decomposizione a valori singolari della matrice parole-documenti $X = U\Sigma V^T$
4. Scegliere un numero k di autovalori e autovettori (concetti) e ricavare le matrici ridotte U_k, Σ_k, V_k

5. Trasformare ciascun documento (colonna di X , x) nello spazio dei concetti come $\hat{x} = x^T U_k \Sigma_k^{-1}$. Quindi per ogni query q effettuare la trasformazione nello spazio dei concetti, \hat{q} , come $\hat{q} = q^T U_k \Sigma_k^{-1}$.
6. Valutare la somiglianza della query a ciascun documento calcolando il coseno, $\vartheta \in [-1, 1]$, tra il vettore della query e quello del documento (ambedue nello spazio dei concetti):

$$\vartheta = \frac{\hat{q}^T \hat{x}}{\|\hat{q}\| \|\hat{x}\|} = \frac{q^T U_k \Sigma_k^{-1} \Sigma_k^{-1} U_k^T x}{\|q^T U_k \Sigma_k^{-1}\| \|x^T U_k \Sigma_k^{-1}\|}$$

7. Ordinare i documenti in ordine decrescente per ϑ

Scegliere diversi valori del numero di concetti k . Effettuare gli esperimenti con la matrice X originaria e quella normalizzata su righe e colonne. Discutere i risultati ottenuti.

Ripetere l'esperimento su un'altra matrice parole-documenti creata a piacere (per esempio parole scelte da testi di canzoni, titoli di films, parole scelte da articoli di giornali o titoli di giornale stessi, ecc.)

Per lo svolgimento del progetto fare riferimento alle dispense sulla SVD disponibili alla pagina Web <http://www.na.iac.cnr.it/~bdv/mct0607.htm>, in particolare PDF1 e PDF3 (pagg. 9–14, in cui è riportato l'esempio proposto nel presente progetto).