

Algoritmi di feature selection

Simone Rutigliano

Corso di Laurea in Informatica Magistrale

11 dicembre 2014



Outline



PageRank

Implementazione del *Wrapper Model*:

- Utilizzare lo stesso algoritmo sia per la feature selection sia per la fase di raccomandazione
- Subset ottimizzato per la raccomandazione



HITS

Creazione del subset attraverso l'utilizzo dell'algoritmo di *Hyperlink-Induced Topic Search* basato sul *ranking* di risorse in base a due metriche:

- Hub
- Authority

Implementazioni trovate:

- <http://goo.gl/4pWAq4>
- <http://goo.gl/qSDXru>



SALSA

- Combinazione di HITS e PageRank
- Usa i punteggi di Hub e Authority
- Crea un grafo bipartito $G = (V_1 \cup V_2, E)$ dove
 - V_1 rappresenta il set degli Hub
 - V_2 rappresenta il set degli Authority
 - Una risorsa può essere contenuta sia in un set che nell'altro

Implementazione trovata:

- <http://goo.gl/DtHa4K>



ReConRank

Tratto dal paper [HHD06]

- Basato su due ranking:
 - **ResourceRank** : Associa uno score basato sul PageRank alle risorse del grafo RDF
 - **ContextRank** : Permette di inglobare la provenienza del contenuto semantico nel calcolo del ranking
- Computazione molto onerosa
- Implementazioni trovate
 - <http://goo.gl/PnZfNc>
 - <http://goo.gl/oCwQWe>



SimRank

Tratto dal paper [JW02] Algoritmo per il calcolo di similarità tra due nodi all'interno di un grafo G

- Esegue un random walk con ripartenza da un nodo fissato u su un grafo k -partito
- Gli score risultati misureranno la similarità tra il nodo u e tutti gli altri nodi del grafo

Implementazione trovata:

- <http://goo.gl/9cLDda>



TripleRank

Tratto dal paper [FSSS09]

- Consiste in una generalizzazione di HITS nel contesto dei Linked Data
- The algorithm was evaluated for faceted browsing and filtering semantic relations for a better Linked Data exploration experience

Implementazione trovata:

- <http://goo.gl/Pb3vEr> (Richiede l'utilizzo di Matlab)



mRMR

Tratto dall'articolo [PLD05] e approfondito in [Rut14]

- Consiste nel calcolo della
 - **minima ridondanza** tra le features
 - **massima rilevanza** delle features con la classe target

Implementazione trovata:

- <http://goo.gl/YQUx1s>



PICSS

Ranker by Information Content (PICSS) We propose our partitioned information content (PIC)-based semantic similarity measure, called PICSS, which is a combination of feature- and information content-based approaches. Feature in comune per la similarità
Pesa gli archi con information content e similarità tra feature



SHSEL - Information Gain

University of Mannheim



TSEL - Information Gain

University of Mannheim



Greedy Top Down

University of Mannheim



References



Thomas Franz, Antje Schultz, Sergej Sizov, and Steffen Staab.

Triplerank: Ranking semantic web data by tensor decomposition.

In Proceedings of the 8th International Semantic Web Conference, ISWC '09, pages 213–228, Berlin, Heidelberg, 2009. Springer-Verlag.



Aidan Hogan, Andreas Harth, and Stefan Decker.

Reconrank: A scalable ranking method for semantic web data with context.

In In 2nd Workshop on Scalable Semantic Web Knowledge Base Systems, 2006.

