

Sistemi Informativi

Laboratorio 4

Catalin Copil

Mattia de Stefani

Giulio Lovisotto

April 29, 2015

1 Descrizione

Visto che abbiamo scelto di usare BM25 per il ranking, applicheremo la formula tenendo in considerazione i giudizi di rilevanza. Per il relevance feedback esplicito utilizzeremo il file `qrels-treceval.txt`. Ricordiamo che la formula di BM25 tiene già in considerazione i giudizi di rilevanza nella sua forma base.

$$\sum_{i \in Q} \log \left(\frac{(r_i + 0.5)/(R - r_i + 0.5)}{(n_i - r_i + 0.5)/(N - n_i - R + r_i + 0.5)} \right) \cdot \frac{(k_1 + 1)f_i}{k + f_i} \cdot \frac{(k_2 + 1)qf_i}{k_2 + qf_i}.$$

Ricordiamo che R è il numero di documenti rilevanti per la query in questione, mentre r_i è il numero di documenti rilevanti che contiene il termine i .

1.1 Relevance Feedback Esplicito

Il reperimento avverrà in 2 step. Nel primo verrà eseguito il ranking senza informazioni di rilevanza, e tra i primi N documenti verranno estratti quelli rilevanti usando il file `qrels-treceval.txt`. Poi verranno estratti i valori R, r_i tra i documenti rilevanti individuati e verranno usati per la seconda esecuzione dell'algoritmo.

1.2 Pseudo Relevance Feedback

Il reperimento avverrà in 2 step. Nel primo verrà eseguito il ranking senza informazioni di rilevanza, verranno considerati i primi N documenti come rilevanti. Poi verranno estratti i valori R, r_i tra i documenti rilevanti individuati e verranno usati per la seconda esecuzione dell'algoritmo.

2 Implementazione

Per il calcolo di R ed r_i , utilizzeremo la matrice che contiene la frequenza di occorrenza delle parole per ogni documento ($n_docs \times n_words$). Durante il reperimento, per ogni documento, per ogni termine i andremo a prendere il numero di documenti rilevanti che contiene il termine i , e lo salveremo in una mappa *map* (mappa $i \rightarrow r_i$). Useremo la mappa per calcolare il ranking rispetto ai valori R, r_i . I documenti rilevanti sono calcolati come descritto nella precedente sezione.

3 Risultati

Abbiamo notato che usando il relevance feedback esplicito la precisione *map* migliorava. Mentre utilizzando lo pseudo relevance feedback la precisione peggiorava leggermente. Figura 1 mostra i risultati utilizzando la versione base di BM25 (senza giudizi di rilevanza). Figura 2 mostra i risultati ottenuti utilizzando feedback esplicito. Figura 3 mostra i risultati ottenuti utilizzando pseudo relevance feedback.

runid	all	G12R7
num_q	all	43
num_ret	all	4300
num_rel	all	719
num_rel_ret	all	406
map	all	0.3156

Figure 1: Risultati *trec_eval* BM25 base.

runid	all	G12R8
num_q	all	43
num_ret	all	38957
num_rel	all	719
num_rel_ret	all	601
map	all	0.2966

Figure 2: Risultati *trec_eval* BM25 relevance feedback pseudo.

runid	all	G12R9
num_q	all	43
num_ret	all	38957
num_rel	all	719
num_rel_ret	all	599
map	all	0.3355

Figure 3: Risultati *trec_eval* BM25 relevance feedback esplicito.