

RELAZIONE ESERCITAZIONE 4

Roger Ferrod, Simone Cullino, Davide Giosa

SOMMARIO

La seguente esercitazione ha lo scopo di calcolare la similarità semantica tra due termini usando *Nasari* e confrontarla con un'annotazione manuale. L'annotazione delle 100 coppie di termini è stata svolta, in modo indipendente, da due componenti del gruppo utilizzando una scala da 0 a 4; su di essa vengono calcolati gli indici di correlazione per verificare la conformità delle opinioni. La consegna richiede, quindi, di selezionare i sensi tali da massimizzare la similarità tra le parole della coppia. Una volta individuati tali sensi, si ricavano i *BabelID* corrispondenti e si effettua una chiamata a *BabelNet* volta a recuperare la relativa definizione. Segue un confronto tra i sensi individuati dall'algoritmo e quelli riconosciuti durante l'annotazione.

IMPLEMENTAZIONE

In input è stato fornito il file contenenti i vettori *Nasari* (rappresentazione *Embedded*) sulla base del quale vengono effettuate le analisi per la *concept similarity*.

Viene fornito in input anche un file contenente tutte le parole delle coppie (in italiano) con i relativi *BabelID*, in modo tale da non dover accedere, con eccessive richieste, a *BabelNet* per tradurre in inglese le parole dei file di annotazione. All'annotazione manuale segue il calcolo della media delle annotazione e gli indici di correlazione (*Pearson* e *Spearman*).

In seguito viene applicato il calcolo della similarità, basata sulla similarità coseno, estraendo la coppia di sensi (i.e. *BabelNet synset*) che ne massimizza il valore e la relativa definizione ottenuta da *BabelNet*.

Il risultato finale viene stampato su un file di output.

RISULTATI

Abbiamo innanzitutto confrontato i due file di annotazioni ottenendo le seguenti correlazioni: *Person* 0.934, *Spearman* 0.942.

Inseguito, abbiamo visionato il file contenente i risultati dell'algoritmo notando come in alcuni casi quest'ultimo non abbia individuato gli stessi sensi pensati in fase di annotazione, ad esempio la coppia ('campagna', 'scimmia') è stata interpretata come:

- Algoritmo
 - Campagna: comune di Torre d'Isola (provincia di Pavia).
 - Scimmia: segno dell'astrologia cinese.
- Annotatore
 - Campagna: paesaggio rurale
 - Scimmia: animale

Questa discrepanza si presenta nel 28% dei casi, mentre il 19% delle coppie sono state scartate dal calcolo perché prive di glossa o di *BabelNet Synset*. Ne risulta un'accuratezza del 65% (53 casi su 81).