Rethinking collocation: Directional association measure and its potential role in semantic similarity computation

Ying Liu
University of California, Davis
yiliu@ucdavis.edu

As one of the most important notions in the field of lexical semantics, collocation / association measure has attracted nothing but countless attention in the past 50 years. Though previous studies have accomplished fruitful results coming up with scientific methods for association measure, all of them tend to be bidirectional, or symmetric (Gries, 2013). What they have ignored is the fact that associations are not necessarily reciprocal in length (Ellis, 2007), since within a collocation, one word could be more predictive of the other than the other way around. It is not unlikely that the probability of (word1|word2) doesn’t equal to that of (word2|word1). In other words, ρ(regardless|of) may not be the same as ρ(of|regardless).

Based on former studies, Gries(2013) proposed to use a directional approach for association measure, ΔP, which arose out of associative learning theory:

\[ ΔP = ρ(\text{outcome}|\text{cue}=\text{present}) - ρ(\text{outcome}|\text{cue}=\text{absent}) \]

As can be told from the formula, ΔP is more sensitive than traditional measures as it could select out the collocates in a collation which are more predictive of the others, and thus may set the tone for the collocation. Other than that, ΔP is also comparatively much easier to compute, even in a relatively large corpus. What’s more, since it ΔP is based on associative learning theory, a (psycho-)linguistic foundation for this measure has already been assumed.

Given the advantages of directional association measure, in this paper I try to propose as well as explore the possibility of applying ΔP in semantic similarity computation. Result of previous studies (Gries, 2001) has shown that feasibility of this thought is quite promising, which is the similarity of the meanings of two adjectives could be measured on the basis of how many collocates of one adjective are also collocates of other. Here I present a corpus-based study of the –ic and –ical adjectives from the Brown Corpus, then compare the results with those measured by the widely-used Cosine similarity measure, which hopefully would shed new lights also on the modeling of word meanings.

Reference:
Gries, St. Th. 2013. “50-something years of work on collocations: what is or should be next ...”. International Journal of Corpus Linguistics 18(1). 137-165.