

Graeme Hirst
with Saif Mohammad and Meghana Marathe

Semantic distance measures with distributional profiles of coarse-grained concepts

Although semantic distance measures are applied to words in textual tasks such as building lexical chains, semantic distance is really a property of concepts, not words. We present a hybrid measure of semantic distance based on **distributional profiles of concepts** that we infer from text corpora. We use only a very coarse-grained inventory of concepts — each category of a published thesaurus is taken as a single concept — and yet obtain results on basic semantic distance tasks that are generally as good as methods that use fine-grained word-based measures. Because the measure is based on naturally occurring text, it is able to find word pairs that stand in non-classical relationships not found in WordNet. It can be applied cross-lingually, using a thesaurus in one language to measure semantic distance between words in another. We report the results of our experiments using this measure in text segmentation. In addition, we will discuss some recent challenges to the utility of lexical chains.