Natural Language Understanding using Knowledge Bases and Random Walks

One of the key challenges for creating the semantic representation of a text is mapping words found in a natural language text to their meanings. This task, Word Sense Disambiguation (WSD), is confounded by the fact that words have multiple meanings, or senses, dictated by their use in a sentence and the domain. We present an algorithm that employs random walks over the graph structure of knowledge bases, yielding state-of-the-art results for WSD on both general and biomedical texts, as well as in Named-Entity Disambiguation. We also show that the same algorithm can be successfully applied to Word Similarity and to enrich texts with related concepts, yielding improvements in Information Retrieval. Finally, we argue that knowledge-based approaches are complementary to other approaches, like supervised machine learning or unsupervised distributional embeddings.