Learning Context Selection Models for Knowledge-based WSD

Evgenia Wasserman Pritsker
Computer Science dep.
University of Haifa

Einat Minkov
Information Systems dep.
University of Haifa

Word sense disambiguation (WSD) is the task of selecting the most appropriate sense for an ambiguous word mention given its context. Approaches to WSD are often classified according to the main source of knowledge used in sense differentiation (Agirre and Edmonds, 2007): annotated corpora (supervised), row unannotated corpora (unsupervised), and lexical knowledge databases ('knowledge-based'). All these methods assume that the relevant context is predetermined and provided as input.

We focus on knowledge-based WSD methods, which use available lexical resources (dictionaries, thesauri and lexical knowledge bases like Wordnet and Wikipedia) to find maximal agreement, or *similarity*, between the target word meaning and the given context. In most WSD systems the disambiguation process can be divided into two subtasks: context selection and sense prediction. As of a today many efforts have been invested into the sense prediction part, and there are smart and comprehensive algorithms in this area, but context selection methods are standard and relatively scarce. A typical approach for the context selection is a window-based method, which considers the words surrounding the ambiguous word mention within a pre-defined window size, where all words are assumed equal importance. The context, picked up in this way, may include unrelated and non-discriminative words, which should have negative effect on WSD performance.

In this work, we propose to automatically select informative context words. We outline a learning framework that aims at identifying useful contextual cues for knowledge-based algorithms. To that end, various types of features are proposed and evaluated that represent context-target word pairs using syntactic and lexico-statistical information, in addition to simple word distance. Importantly, the constructed model is independent from the lexical identity of the word whose sense is being predicted. This means that the learned prediction models are general rather than word specific, and therefore fit within unsupervised WSD settings.

The main contributions of this work are the following:

- We believe that it is first to outline a learning framework that performs context selection for unsupervised knowledge-based WSD. The framework is general, and can be applied towards learning of specialized contextual models.
- We test the conjecture that learning context models is favorable to using uninformed general context models. To that end, we experimented with a number of benchmark WSD datasets and two different knowledge-based WSD methods - Personalized PageRank (Agirre and Soroa, 2009) and Gloss Vectors (Patwardhan and Pedersen, 2006). The results demonstrate consistent improvement in WSD performance. On a large noun lexical sample dataset the improvements were found to be statistically significant, and in some settings our results exceeded state-of-the-art performance.

- We analyze the contribution of the diverse feature types and show that word lexico-statistical information (PMI, IDF and a number of senses) is the most influential in the considered setting, rather than simple word distance.

References

Eneko Agirre and Philip Edmonds, "*Word Sense Disambiguation: Algorithms and Applications"*, Springer, 2007.

Eneko Agirre and Aitor Soroa, "Personalizing PageRank for word sense disambiguation", in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, 2009.

Siddharth Patwardhan and Ted Pedersen, "Using WordNet-based Context Vectors to estimate the semantic relatedness of concepts", in *Proceedings of the EACL 2006 Workshop on Making Sense of Sense*, 2006.