DGfS-CL

Sabine Schulte im Walde
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgat

Comparing Computational Models of Selectional Preferences – Second-order Co-Occurrence vs. Latent Semantic Clusters

Selectional preferences (i.e., semantic restrictions on the realisation of predicate complements) are of great interest to research in Computational Linguistics, both from a lexicographic and from an applied (wrt data sparseness) perspective. This poster presents a comparison of three computational approaches to selectional preferences: (i) an intuitive distributional approach that uses second-order co-occurrence of predicates and complement properties; (ii) an EM-based clustering approach that models the strengths of predicate–noun relationships by latent semantic clusters (Rooth *et al.*, 1999); and (iii) an extension of the latent semantic clusters by incorporating the MDL principle into the EM training, thus explicitly modelling the predicate–noun selectional preferences by WordNet classes (Schulte im Walde *et al.*, 2008).

The motivation of our work was driven by two main question: Concerning the distributional approach, we were interested not only in how well the model describes selectional preferences, but moreover which second-order properties were most salient. For example, a typical direct object of the verb *drink* is usually fluid, might be hot or cold, can be bought, might be bottled, etc. So are adjectives that modify nouns, or verbs that subcategorise nouns salient second-order properties to describe the selectional preferences of direct objects? Our second interest was in the actual comparison of the models: How does a very simple distributional model compare to much more complex approaches, especially with respect to model (iii) that explicitly incorporates selectional preferences?

References:

Mats Rooth, Stefan Riezler, Detlef Prescher, Glenn Carroll, and Franz Beil (1999). Inducing a Semantically-Annotated Lexicon via EM-Based Clustering. In *Proceedings ot the 37th Annual Meeting of the Association for Computational Linguistics.*

Sabine Schulte im Walde, Christian Hying, Christian Scheible, and Helmut Schmid (2008). Combining EM Training and the MDL Principle for an Automatic Verb Classification incorporating Selectional Preferences. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics.*