Abstract

Currently used general purpose dictionaries are often too fine-grained, with narrow sense divisions that are not relevant for many Natural Language applications. WordNet, which is a widely used sense inventory for Word Sense Disambiguation, has the same problem. With different applications requiring different levels of sense granularities, producing sense clustered inventories of arbitrary sense granularity has evolved as a crucial task.

We try to exploit the resources available like human-labelled sense clusterings and semi-automatically generated domain labels of synsets, to estimate the similarity between synsets. Using supervision, we learn a model which predicts the probability of any two senses of a word to be merged. To learn a more generic model, we propose a graph based approach, which allows us to use the information learnt from supervision as well. Using this complete similarity measure, we propose a simple method for clustering synsets. We show that the coarse-grained sense inventory obtained significantly boosts the disambiguation of nouns on standard test sets.