Combining EM Training and the MDL Principle for an Automatic Verb Classification Incorporating Selectional Preferences

Sabine Schulte im Walde, Christian Scheible, Christian Hying, Helmut Schmid  
{schulte, scheiblec, hyingcn, schmid}@ims.uni-stuttgart.de  
SFB 732 “Incremental Specification in Context”  
Project D4 “Modular Lexicalisation of Probabilistic Context-Free Grammars”  
Institut für Maschinelbe Sprachverarbeitung, Universität Stuttgart

Main Goals
- A Statistical model for verb-argument tuples: \( p(\text{read, subj, obj, student, books}) \)
- Induction of a semantic verb classification with clustering techniques
- Learning of verbal selectional restrictions which are represented with WordNet concepts

Features of the Model
- Statistical soft clustering: verbs are assigned to one or more verb classes
- Representation of verbal polysemy by the assignment to multiple classes
- Training on verb-argument tuples with the Expectation-Maximization algorithm
- Generalization of selectional restrictions with Minimum Description Length principle
- The model is smooth because it generalizes over
  - the verbs of a cluster
  - the nouns instantiating the WordNet concepts representing the selectional restrictions

Probabilistic Verb Class Model
\[
p(\text{v, f, a}_1, \ldots, a_n) = \sum_p p(c)p(v|c)p(f|c)\prod_{i=1}^n p(r|c, f, i)p(a_i|r)
\]

- \( p(c) \) probability of verb class \( c \)
- \( p(v|c) \) probability of verb \( v \) in class \( c \)
- \( p(f|c) \) probability of frame \( f \) in class \( c \)
- \( p(r|c, f, i) \) probability that \( i \)th argument of frame \( f \) in class \( c \) is realised by WordNet concept \( r \) e.g., \( p(\text{person} | \text{cls}, \text{subj:pp}, 1) \)
- \( p(a|r) \) probability that WordNet concept \( r \) is realised by argument head \( a \) e.g., \( p(\text{professor} | \text{person}) \)

EM & MDL
The whole model is represented as a large graph
- Initialisation of selectional restrictions (SR) with top concept entity
- Random initial assignment of probabilities
- Expansion of SR by the next lower level
- Estimation of graph frequencies from training tuples using the Inside-Outside algorithm
- Re-estimation of the probabilities
- MDL pruning of the selectional restrictions

Experiments & Examples
- Tuples from BNC Viterbi parses (Carroll & Rooth, 1998)
- Only active clauses, no auxiliary, modal, or particle verbs, no pronouns
- 10/20 subcategorisation frame types
- Tuples with freq \( > 1 \) (51,569/55,980)
- 20/50/100 clusters and 50 iterations

Semantic Verb Classification
- Grouping of verbs according to semantic properties (Levin 1993)
  - Break a Solid Surface with an Instrument: break, crush, fracture, smash, etc.
- Goals:
  - Organisation of verbs wrt. shared properties
  - Generalization over verbs to counter sparse-data problems
- Applications:
  - Word sense disambiguation (Dorr & Jones 1996; Kohlbam & Lee 2005)
  - machine translation (Prescher et al., 2000; Koehn & Joang 2007)
  - document classification (Kleinaus & Kan 1998), etc.

Conversion of WordNet into a Markov Model (Abney/Light)
- Additional node for each word
- Additional hyponym links from each concept to the members of its synset
- A probability for each hyponym link

Path probabilities
- \( p(a|r) \) (and \( p(c,f,i) \)) is a sum of path probabilities
- Path probability = product of link probabilities

Evaluation
- Focus: statistical model of verb-argument tuples \( \rightarrow \) model predicts tuple probabilities
- Comparison of verb class model predictions with baseline model
- Baseline model without hidden variables
\[
p(v, f, a_1, \ldots, a_n) = p(v)p(f|v)\prod_i p(a_i|r_i)
\]
- Example:
  - \( p(\text{read, subj, pp-to, professor, audience}) \)
  - \( p(\text{read}) \) \( p(\text{subj:pp-to}) \) \( p(\text{professor}) \) \( p(\text{audience}) \)

Outlook
- Experiments with other languages and corpora
- Refinement of the model (representation of alternations and collocations, etc.)
- Refinement of the training (split and merge clusters, training on data slices)
- Applications
  - Induction of verb classes, subcategorisation, and selectional restrictions
  - Detection of verbal polysemy, verb alternation, and collocations
  - Automatic assignment of new nouns to WordNet synsets
  - Refinement of a PCFG parser with verb-argument association scores