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,scheibcn,hyingcn,schmid}@ims.uni-stuttgart.de

SFB 732 "Incremental Specification in Context" Project D4 "Modular Lexicalisation of Probabilistic Context-Free Grammars"

Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

Main Goals

- > A Statistical model for verb-argument tuples: p(<read, subj:obj, student, book>)
- > Induction of a semantic verb classification with clustering techniques
- > Learning of verbal selectional restrictions which are represented with WordNet concepts

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:

Features of the Model

- > Statistical **soft clustering** approach: 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
- Organisation of verbs wrt. shared properties
- Generalization over verbs to counter sparse-data problems
- Applications
 - word sense disambiguation (Dorr & Jones 1996; Kohomban & Lee 2005)
 - o machine translation (Prescher et al. 2000; Koehn & Joang 2007)
 - o document classification (Klavans & Kan 1998), etc.

Probabilistic Verb Class Model p(speak, subj-pp.to, professor, audience) $p(v, f, a_1, ..., a_{n_f}) = \sum p(c) p(v | c) p(f | c) \prod \sum p(r | c, f, i) p(a_i | r)$

- probability of verb class c p(c)
- probability of verb v in class c p(v|c)
- probability of frame *f* in class *c p(f|c)*
- probability that *i*th argument of frame *f* in class *c* is realised by WordNet concept *r* p(r|c,f,i)e.g., *p*(person | *c*3, *subj-pp.to*, 1)
- probability that WordNet concept *r* is realised by argument head *a p*(*a*|*r*) e.g., p(professor | person)

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

Two types of WordNet-HMMs

- Many partial WordNet HMMs for p(r|c,f,i)
- One a priori model for *p(a|r)*

Path probabilities

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



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

| | ClusterViewer - cluster: c3 | | | | ClusterViewer - frame: subj, slot: subj | |
|---|--|---|--|--|---|---|
| <- previou | s c3 (p=0.0460 | 512) | ; | next | Slot: subj | |
| Dimension: (Vc3) | | | Dimension: (Ac3) | | word study 7048627 | selectional preference |
| show suggest indicate reveal find imply demonstrate conclude confirm say report state claim note illustrate collect estimate prove | probability 0.405593 0.261479 0.0858937 0.0452668 0.0365477 0.0194393 0.0188025 0.0141904 0.00905879 0.00730457 0.00608054 0.00386967 0.00368099 0.00367489 0.00337297 | O sul tha ob sul ap obj o su to su | it subj j subj oj adv 2 obj subj oj subj pp-in bj pp-in | 0.535841 0.452097 0.00649998 0.00389853 0.000548383 0.000404125 0.000297971 0.000281138 0.000131246 1.32126e-56 | study 7048027 survey research name, epithet result study, work evidence datum analysis 7067876 report figure experiment, experimentation 5798043 example written communication, written language, black and white representation, mental representation, internal representation statistic prose experience chapter poll, opinion poll, public opinion poll, canvass record | 0.0630798 0.0615228 0.0479986 0.0466557 0.0439606 0.0419544 0.0376109 0.0346575 0.0304419 0.0292167 0.0280233 0.0275861 0.022131 0.0182848 0.0162484 0.0129201 0.0125366 0.0123137 0.0109277 0.0107317 |

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 $p(v, f, a_1, \dots, a_{n_f}) = p(v)p(f | v)\prod_{i=1}^{f} p(a_i | a_1^{i-1}, v, f, f_i)$
- Example:
- *p*(*speak*, *subj*:*pp-to*, *professor*, *audience*) = p(speak) p(subj:pp-to | speak) p(professor | speak, subj:pp-to, subj)





p(audience | professor, speak, subj:pp-to, pp-to)

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