An Architecture for Incorporating Semantics into Probabilistic Models

ulrike@coli.uni-sb.de

Ulrike Padó[1], Matthew Crocker[1], Frank Keller[2]
[1]Saarland University, [2]University of Edinburgh

Introduction

Probabilistic context-free models...

- ... assign probabilities to structural analyses of the input; Claim: *most probable = preferred*
- ... account for frequency effects (e.g., Jurafsky 1996)
- ... account for robustness of sentence processing (when using a wide-coverage grammar, e.g. Crocker & Brants 2000)

However, they have no notion of semantic processing!

- Thematic fit of verbs and prospective arguments influences initial parsing decisions in many constructions (e.g. NP/S)
- PCFGs can be lexicalised, but
  - This treats a semantic phenomenon on a collocational level
  - In practice, training data is very sparse

We propose a probabilistic wide-coverage modelling architecture that uses syntactic and semantic cues

- Cleanly extend existing models by a crucial and separate dimension
- Achieve broad coverage of corpus and experimental data

Architecture

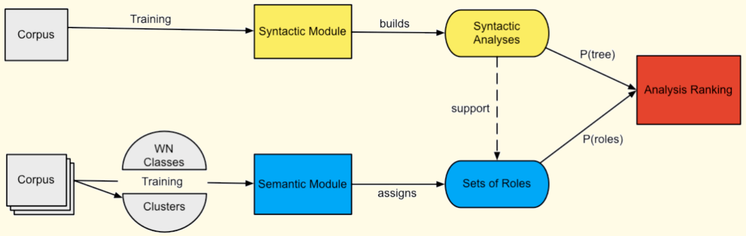
Standard: A probabilistic parser returns the most likely syntactic analyses at each word

- Syntactic probabilities are computed using a treebank grammar (induced from corpus)
  ⇒ Wide coverage on unseen text

Extension: Probabilistically assign thematic roles to each verb argument in the partial parse

- Plausibility of set of thematic roles is modelled by its probability
- Probability of individual role assignment is estimated from semantically annotated corpus
- Extract prospective argument heads from the partial parse
- Assign each verb-argument pair its most likely role (including adjunct roles)

The overall preferred analysis is determined by both constraints



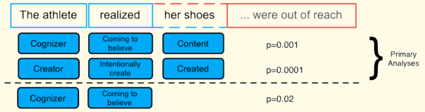
The model initially only considers those parses in which a new role can be assigned (cf. Pritchett (1992); if none of these parses is likely, it considers the remaining parses

- We predict disruption if the plausibility of the previously preferred analysis drops below that of another analysis or below threshold
- Future work: Predict (graded) effects quantitatively

Result: A probabilistic, incremental, wide-coverage model of sentence processing that accounts for semantic effects

A Test Case

NP/S ambiguity: The NP may belong to the verb as a direct object or to an embedded clause



Readers prefer the direct object reading regardless of subcat preference (Pickering et al. 2000)

- Unless contradicted by thematic fit or (later on) syntactic admissibility

Probabilistic models make an incorrect prediction

- Verb subcat preferences lead to an early, unchanged preference for the embedded clause reading

Our model will make the correct prediction

- Eagerness for role-assignment leads to initial preference for the object reading; preference is modified through thematic fit of arguments and syntactic probability

Testing the Semantic Module

Task: Model human judgment data with thematic role predictions

- Correlate judgments and model predictions

Approach: Estimate $P(role, verb_{frame}, arg\text{-}head)$ from corpus

- Compute as $P(verb_{frame})P(role|verb_{frame})P(arg\text{-}head|verb_{frame}, role)$

Problem: Semantically annotated corpora needed (PropBank/FrameNet);

large sparse data problem

Solution: Class-based smoothing (Instead of counting token frequencies, count class frequencies)

- Also model influence of infrequent words
- Verbs: Induce classes by clustering
- Nouns: Too sparse for clustering, use WordNet

Training and test data:

- Cluster and estimate probabilities from PropBank / FrameNet
- Test on 100 verb-argument-role triples with judgments on 1-7 scale from McRae et al. (1998)

Results:

Smoothing Scheme	Coverage	Correlation (ρ_S)
None	2 (2%)	ns
Clusters, FN	17 (17%)	ρ=0.515, p<0.05
Clusters, FN + WN noun synsets	18 (18%)	ρ=0.634, p<0.01

Conclusions:

- Semantic module reliably predicts human judgments
- Smoothing enlarges coverage, strengthens correlation
- Training data is still sparse. Current work: automatically annotate larger data set (parts of BNC) with role information to extend training set

References:
Crocker, M. and Brants, Th. (2000). Wide-coverage probabilistic sentence processing. JPR 29(6).
Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. Cognitive Science 20.
McRae, K., Spivey-Knowlton, M. and Tanenhaus, M. (1998). Modelling the influence of thematic fit (and other constraints) in on-line sentence comprehension. JML 38.
Pickering, M., Traxler, M., Crocker, M. (2000). Ambiguity resolution in sentence processing: Evidence against frequency based accounts. JML 43.
Pritchett, B. (1992). Grammatical competence and parsing performance. University of Chicago Press.