Reversible stochastic attribute-value grammars

Daniël de Kok, Barbara Plank, Gertjan van Noord

> *It is plausible that we have something in our heads that fills the function I am ascribing to grammar, though I am not insensitive to the claims of those who deny this. But it is altogether implausible that we have two such things, one for parsing and one for generation, essentially unrelated to one another.*

Martin Kay, 1975

Motivation

- Reversible grammars have been realized.
- Modern systems add an non-reversible statistical component for parse disambiguation and fluency ranking.
- Parse disambiguation and fluency ranking seem to be governed partially by the same preferences, for example: subject/object fronting in Dutch.
- Simplicity.
- What went wrong?

Attribute-value grammars

- Representation of lexical items as attribute-value structures.
- Analysis by means of unification of attribute-value structures.
- Unification is associative and commutative, consequently AVGs are non-directional.
- Since algorithms that perform parsing and generation exist, systems using AVGs can be made reversible.

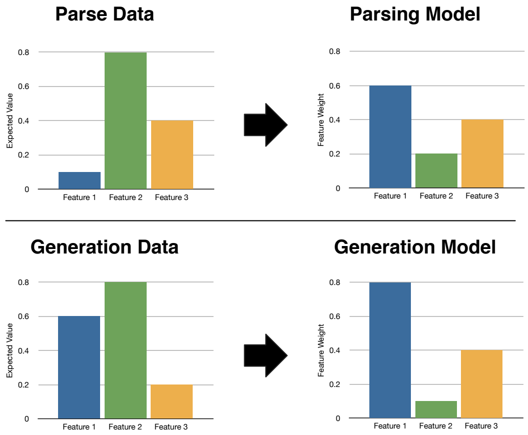
Stochastic AVGs

- What is the best parse?
- Due to the use of constraints, relative frequencies as in PCFG are not applicable to AVGs.
- Abney (1996): probability of an analysis using a maximum entropy model, based on features of the analysis...
- ... normalized over all possible derivations.
- Impractical.
- Reversible.

Stochastic AVGs (2)

- Johnson (1999): conditions the probability of a derivation on the (yield of) the sentence $p(d|s)$ using a conditional maximum entropy model.
- Maximum entropy models minimize assumptions not explained by the data, while obeying a constraint for each feature in the model.
- Practical, given a finite number of parses of a sentence.
- Same approach followed in generation.

Non-reversible models



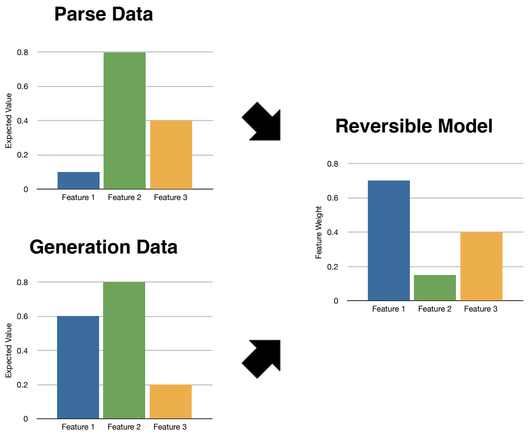
Stochastic AVGs

- Sacrifices reversibility: the probability of a derivation is conditioned on a sentence or logical form.

Reversible SAVGs

- Observation: sentences and logicals forms are both sets of constraints that limit the set of all possible derivations.
- If preferences are shared in parsing and generation, one model can be used.
- How to enforce feature constraints?

Reversible models



Training/Evaluation

- Training and evaluation as in previous work (van Noord, 2006; de Kok and van Noord, 2010).
- Training: Syntactic annotations from the Alpino treebank.
- Evaluation: Part of the Trouw 2001 newspaper, syntactic annotations from LASSY.

Features

- **Word adjacency:** trigram models of words and part-of-speech tags.
- **Dependency relations:** dependency relations with characteristics of the head and dependent.
- **Syntactic features**: features that record the application of grammar rules and features that describe more complex syntactic patterns.
- **Frame selection**: Features that assist in the selection of proper subcategorization frames for words.

Results (parse disambiguation)

Model	CA (%)	f-score (%)
Baseline	75.88	76.28
Oracle	94.86	95.09
Parse model	90.93	91.28
Reversible	90.87	91.21

Table: Concept Accuracy scores and f-scores in terms of named dependency relations for the parsing-specific model versus the reversible model.

Results (fluency ranking)

Model	GTM
Random	55.72
Oracle	86.63
Fluency	71.82
Reversible	71.69

Table: General Text Matcher scores for fluency ranking using various models.

Baseline scenarios

- Do parse disambiguation features contribute to a baseline fluency ranking model and vise versa?
- Scenarios:
  1. Parse disambiguation model with rule identifiers.
  2. Fluency ranking model with auxiliary trigram models.

Parse disambiguation baseline model

Model	**CA (%)**	**f-score (%)**
Rule id	86.10	86.52
Rule id + fluency	86.95	87.37

Table: Parse disambiguation bootstrapping scenarios: building a disambiguator based solely on rule identifiers, informed by fluency ranking training data.

Fluency ranking baseline model

Model	GTM
N-gram	68.91
Parse + N-gram	71.02

Table: Fluency ranking bootstrapping scenarios: building a ranker based solely on N-gram, informed by parse disambiguation training data.

Conclusions

- The reversible model does not perform significantly different from directional parse disambiguation and fluency ranking models.
- Feature weights from one direction can inform the other direction.
- Or rephrasing Martin Kay:

  > *It is plausible that we have something in our heads that prefers certain sentence meaning pairs over others. But it is altogether implausible that we have two such things, one for parsing and one for generation, essentially unrelated to one another.*

Question

- Do reversible models rely on features used in both directions, or do they rely on direction-specific features?
- See my presentation of *Discriminative features in reversible stochastic attribute-value grammars*.

Thank you!