



Reversible stochastic attribute-value grammars

Daniël de Kok, Barbara Plank, Gertjan van Noord

ACL 2011, Portland, Oregon

Motivation for reversible systems

- State of the art: attribute-value grammar with a maximum entropy model for parse disambiguation or fluency ranking.
- Why do we use separate models for parse disambiguation and fluency ranking?
- Counter-intuitive if we assume that some preferences are shared between parsing and generation.
- Historical accident or necessity?

Subject/object fronting in Dutch

Consider the two possible readings of the sentence *Jan zag de man* (*Jan saw the man*):

- $[Jan]_{su} zag [de man]_{obj}$
- $[Jan]_{obj} zag [de man]_{su}$

Subject fronting is preferred in Dutch, consequently:

- In parse disambiguation we prefer reading of a fronted NP as a subject.
- In fluency ranking we prefer realizations that have a fronted subject NP.

Stochastic AVGs

- Statistical modelling of ‘good’ derivations.
- Due to the use of constraints, relative frequencies as in PCFG are not applicable to AVGs.
- Probability of a derivation using a maximum entropy model (Abney, 1996):

$$p(d) = \frac{1}{Z} \exp \sum_i \lambda_i f_i(d)$$
$$Z = \sum_{d' \in \Omega} \exp \sum_i \lambda_i f_i(d') \quad (1)$$

- Reversible, but not practical: normalized over all possible derivations.

Conditional models

- When using a conditional model, we can normalize over the yield of parsing or generation (Johnson, 1999):

$$p(d|x) = \frac{1}{Z(x)} \exp \sum_i \lambda_i f_i(x, d)$$
$$Z(x) = \sum_{d' \in \Omega(x)} \exp \sum_i \lambda_i f_i(x, d') \quad (2)$$

- Where x is a sentence (parsing) or a logical form (generation).

Directionality

- Such models are not reversible: conditioned on input for one direction.
- Directionality is embodied in the constraint that is applied to each feature:

$$E_p(f_i) - E_{\tilde{p}}(f_i) = 0 \equiv \sum_{x \in X} \sum_{d \in \Omega(x)} \tilde{p}(x)p(d|x)f_i(x, d) - \tilde{p}(x, d)f_i(x, d) = 0 \quad (3)$$

- Where X consists of training instances for **either** parse disambiguation **or** fluency ranking.
- Requires training of two models.

Reversible SAVGs

- Observation: sentences and logical forms are both sets of constraints c that limit the set of all possible derivations.
- If preferences are indeed shared in parsing and generation, one model can be used for both tasks:

$$p(d|c) = \frac{1}{Z(c)} \exp \sum_i \lambda_i f_i(c, d) \quad (4)$$

- How to enforce feature constraints?

Constraints in reversible models

Two constraints per feature, one with respect to parse disambiguation training data S , the other with respect to fluency ranking training data L :

$$\sum_{s \in S} \sum_{d \in \Omega(s)} \tilde{p}(s) p(d|c = s) f_i(s, d) - \tilde{p}(c = s, d) f_i(s, d) = 0$$
$$\sum_{l \in L} \sum_{d \in \Omega(l)} \tilde{p}(l) p(d|c = l) f_i(l, d) - \tilde{p}(c = l, d) f_i(l, d) = 0 \quad (5)$$

Training leads to one weight per feature.

Does it work?

- Applied to the Alpino parser/generator for Dutch.
- No significant loss of accuracy in parsing and generation compared to directional models.
- When constructing new models, adding data from the other direction improves performance of models significantly.
- Read our paper for a description of our experiments.

Follow-up work

What features are discriminative in reversible models, and how does this compare to directional models?

Discriminative features in reversible stochastic attribute-value grammars, Daniël de Kok, Proceedings of the EMNLP 2011 Workshop on Language Generation and Evaluation

Thank you!