Vine Parsing and Minimum Risk Reranking for Speed and Precision

Markus Dreyer
David A. Smith
Noah A. Smith

Johns Hopkins University
Design Goals

- Speed
- Precision
**Unlabeled Parser**

Smoothed MLE

**Labeler**

Smoothed MLE

Log-linear

**Reranker**

$U$-best unlabeled parses

$U \times L$-best labeled parses

CoNLL-2006 • M. Dreyer, D. A. Smith, N. A. Smith • *Vine Parsing and Minimum Risk Reranking*

Collins (2000)

Charniak and Johnson (2005)
Split-Head assumption
Projectivity
Vine Grammar

Eisner and N. Smith (2005)
According to estimates, would changes cut filings by more than a third. (from the Penn Treebank)
According to estimates, some would changes the rule cut filings by more than a third.

(b = 4)

(from the Penn Treebank)
According to some estimates, would changes the rule cut filings by more than a third.

$ b = 3 $
According to estimates, some changes would cut filings by more than a third.

(from the Penn Treebank)
According to estimates, changes to the rule would cut filings by insider more than a third. (from the Penn Treebank)
Different bounds for left and right children

Speed - Accuracy Tradeoff

Choose bounds:
90% of original dependencies untouched

\( O(nB_{left}^2 + nB_{right}^2) \)

(from the Penn Treebank)
Minimum Risk Training

Deterministic annealing

D. Smith and Eisner (2006)
CoNLL-2006 • M. Dreyer, D. A. Smith, N. A. Smith • Vine Parsing and Minimum Risk Reranking

**Labeled**

Median 67.6

Japanese (82.9), Portuguese (75.3), Bulgarian (74.8), Chinese (71.6), German (71.0)

**Unlabeled**

Median 77.5

Japanese (86.0), Portuguese (82.4), Bulgarian (82.0), Swedish (79.5), Chinese (77.6)
Summary

- Parsing constraints
- Linear-time inference and decoding
- Minimum Risk reranking
- High precision, mediocre recall

Future Work

- Better estimation
- Better labeler (label bigrams)
- More fine-grained parsing constraints (length bounds given head)