Depth-bounding is effective: Improvements and evaluation of unsupervised PCFG induction

Lifeng Jin¹, Finale Doshi-Velez², Timothy Miller³, William Schuler¹ and Lane Schwartz⁴

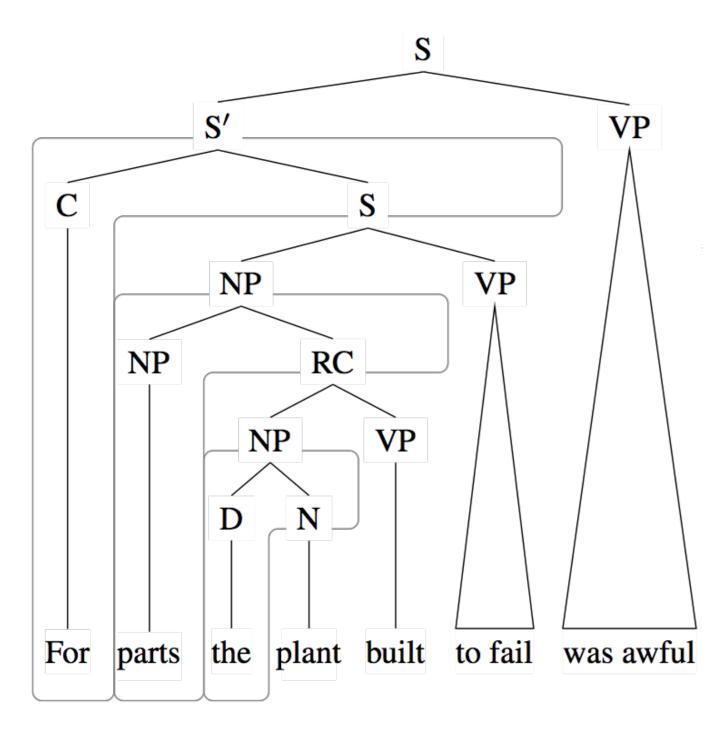
¹The Ohio State University, ²Harvard University, ³Harvard Medical School, ⁴University of Illinois jin@ling.osu.edu



Introduction

- Unsupervised PCFG grammar induction from raw text is hard (Johnson et al., 2007; Liang et al., 2009).
- But unlike a PCFG, natural languages do not have infinite nested center embeddings (Chomsky and Miller, 1963), because they are hard.
- We use this constraint for grammar induction, and directly compare induction models with or without it.
- We carry out model behavior analysis and show that constraining such recursion depths helps induce better grammars significantly.

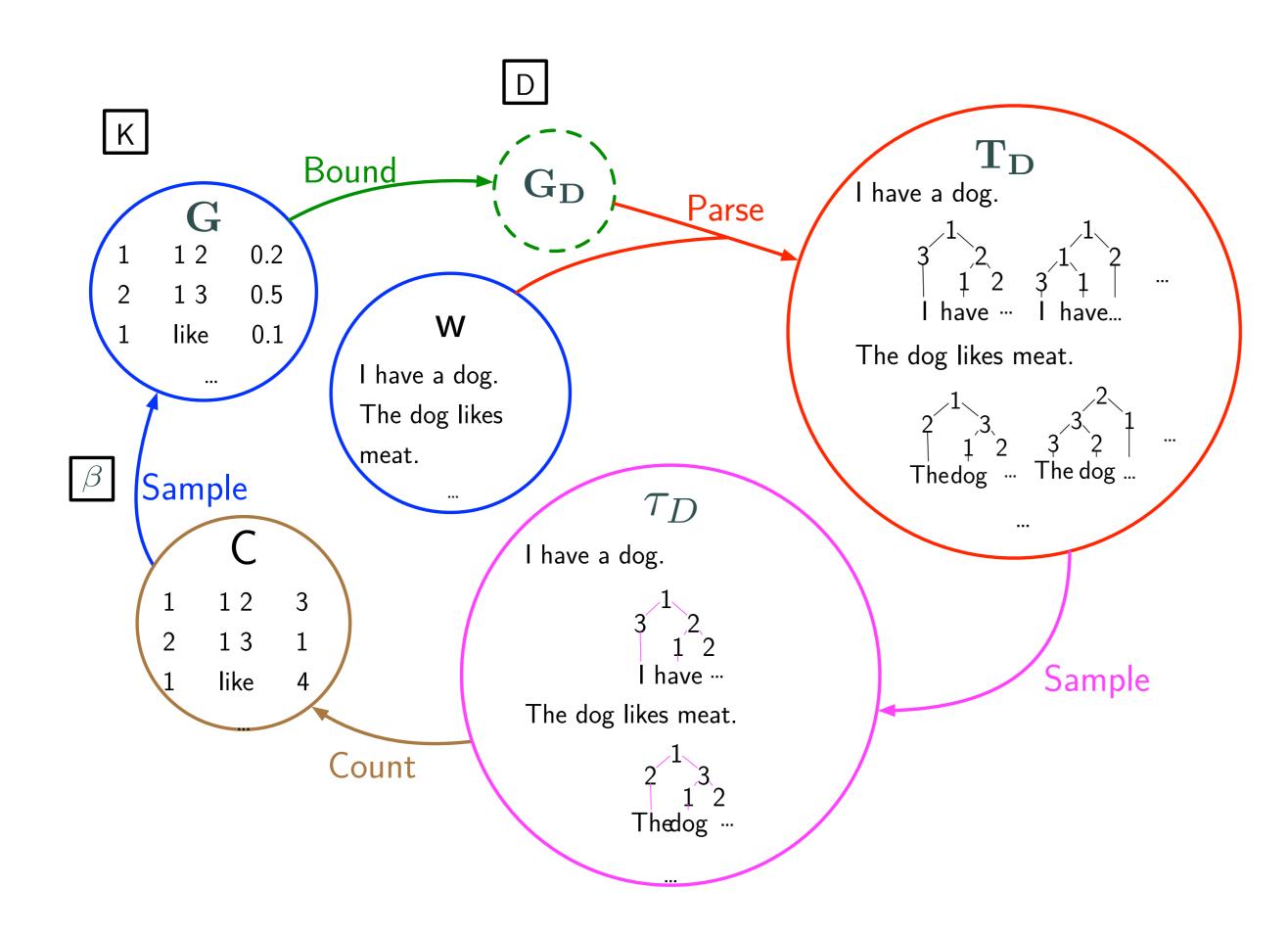
Defining depth with left corner parsing



Stack elements after the word *the* in a left-corner parse of the sentence *For parts the* plant built to fail was awful.

- There are 3 nested center embeddings in the sentence.
- This requires using 3 stack elements to parse, therefore the sentence has maximum depth of 3.
- And it is hard to understand because of the nesting.

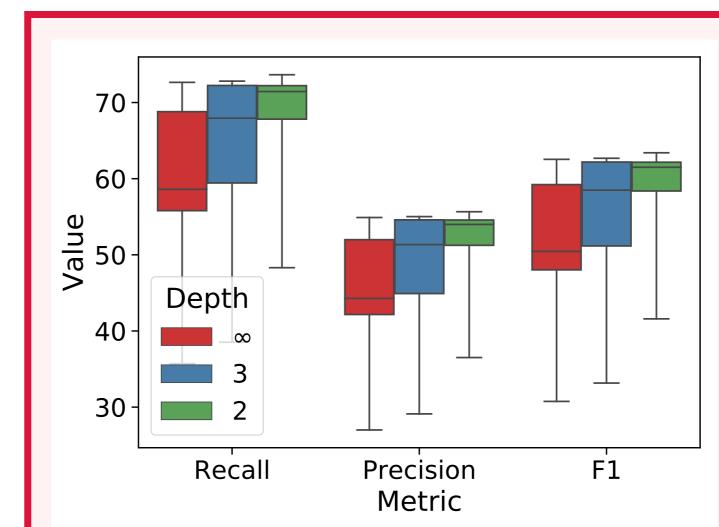
The (un)bounded grammar inducer



- The inducer is a depth-boundable chart-based Dirichlet-multinomial inducer for efficient parsing and sampling (cf. Jin et al., 2018).
- The bounding process can be switched on and off, facilitating comparison between bounded and unbounded models.
- Hyperparameters include K, number of non-terminal categories; D, maximum depth, and β , the Dirichlet symmetric concentration parameter.

Analysis: Depth-bounding is effective

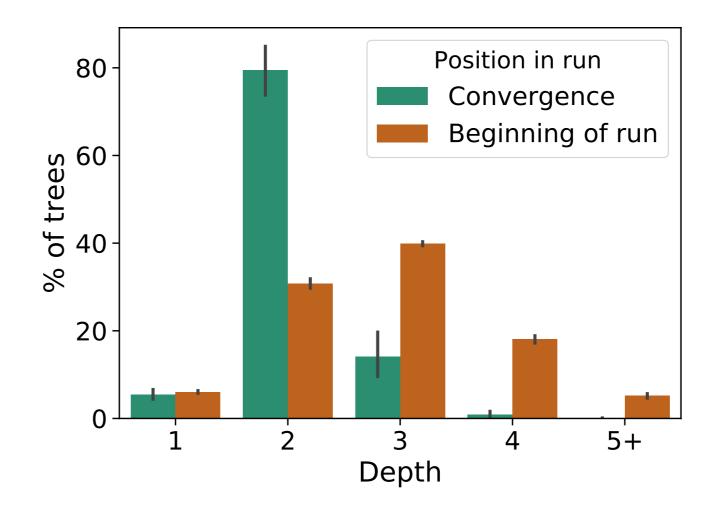
We use 3 different depth settings (2, 3 and ∞) for 20 runs, collecting 60 different models at convergence on the WSJ20dev corpus for model behavior analysis.



PARSEVAL scores for runs with different depth limits.

- Bounding helps the inducer find better grammars by weeding out a lot of bad grammars.
- The difference of all PARSEVAL scores between depth ∞ and depth 2 is significant.
- Most of the models are better than the right-branching baseline.

Low depth is preferred



Distribution of trees with a certain depth at initialization and convergence with unbounded models.

- Random grammars produce deeper trees.
- But the data support only shallow trees.

Multilingual parsing evaluation

We did 3 runs on English (WSJ20test), Chinese (CTB20) and German (NEGRA20) corpora. We bounded the models at depth 2, used 15 non-terminal categories, and 0.2 as the Dirichlet hyperparameter.

Posterior inference on constituents (PIoC)

But how do we know which run is the best run?

• PIoC: (1) find the MAP split points for a constituent; (2) flatten constituents with high posterior uncertainty for their split points. No need to do model selection:

System	Rec	Prec	F1
Best	73.65	55.66	63.40
Best w/ PIoC	73.59	56.41	63.87
All w/ PIoC	72.99	59.21	65.38
All w/ PIoC w/o best	73.00	59.06	65.29

• The constituents with high uncertainty are NPs like *the long history*, coinciding with the fact that they are not split into smaller constituents by annotators.

Multilingual parsing results

System	English		Chinese			German			
	Rec	Prec	F1	Rec	Prec	F1	Rec	Prec	F1
CCL	61.7	60.1	60.9	35.3	39.2	37.1	44.4	27.2	33.7
UPPARSE	40.5	47.8	43.9	33.8	44.0	38.2	55.5	41.9	47.7
DB-PCFG	70.5	53.0	60.5	_	_	_	_	_	-
this work	73.1	55.6	63.1	43.8	35.1	38.9	59.1	31.2	40.8

- Our system has highest recall across the board, but lower precision due to binary branching trees in some of these corpora
- We achieve new state-of-the-art results on two languages, English and Chinese

Conclusions

- Depth-bounding is indeed effective in helping the grammar inducer to find better grammars by limiting search space of possible grammars.
- PIoC is effective in finding better trees without doing meticulous model selection.
- The proposed system is able to achieve state-of-the-art or competitive results across different languages.
- It is possible to induce high quality PCFGs with raw text only with or without bounding, but depth-bounding is preferred.

References

Noam Chomsky and George A Miller. 1963. Introduction to the formal analysis of natural languages. In *Handbook of Mathematical Psychology*, pages 269–321. Wiley, New York, NY.

Lifeng Jin, William Schuler, Finale Doshi-Velez, Timothy A Miller, and Lane Schwartz. 2018. Unsupervised Grammar Induction with Depth-bounded PCFG. *Transactions of the Association for Computational Linguistics*.

Mark Johnson, Thomas L. Griffiths, and Sharon Goldwater. 2007. Bayesian Inference for PCFGs via Markov chain Monte Carlo. *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference (ACL)*, pages 139–146.

Percy Liang, Michael I. Jordan, and Dan Klein. 2009. Probabilistic Grammars and Hierarchical Dirichlet Processes. The Handbook of Applied Bayesian Analysis.