

Grammar Induction: Beyond Local Search

Jason Eisner

Johns Hopkins University

JASON@CS.JHU.EDU

Editors: Jeffrey Heinz, Colin de la Higuera, and Tim Oates

Many approaches to probabilistic grammar induction operate by iteratively improving a single grammar, beginning with an initial guess. These local search paradigms include Expectation Maximization [1, 10] and its variants [13, 14, etc.]; hill-climbing [6] and Markov chain Monte Carlo [9]; and greedy merging [15] or splitting [16, 3, 12] of nonterminals.

Unfortunately, local search methods tend to get caught in local optima, even with random restarts [2, 13] or linguistically guided objective functions [4, 11]. I will suggest two techniques to try to avoid this problem.

One approach searches exhaustively within a region of parameter space, using branch-and-bound to eliminate subregions that cannot contain the optimum. This exploits relaxation and decomposition techniques from the operations research community [17]. The other approach is inspired by recent work on “deep learning” for vision [e.g., 8]. It uses spectral methods [7, 5] to build up featural representations of all substrings, without premature commitment to which substrings are constituents.

References

- [1] J. K. Baker. Trainable grammars for speech recognition. In Jared J. Wolf and Dennis H. Klatt, editors, *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, MIT, Cambridge, MA, June 1979.
- [2] Eugene Charniak. *Statistical Language Learning*. Language, Speech and Communication. MIT Press, Cambridge, Mass., 1993.
- [3] Stanley Chen. Bayesian grammar induction for language modeling. In *Proceedings of ACL*, 1995.
- [4] Shay Cohen and Noah A. Smith. Shared logistic normal distributions for soft parameter tying in unsupervised grammar induction. In *Proceedings of NAACL-HLT*, pages 74–82, 2009.
- [5] Shay B. Cohen, Karl Stratos, Michael Collins, Dean P. Foster, and Lyle Ungar. Spectral learning of latent-variable PCFGs. In *Proceedings of ACL*, pages 223–231, July 2012.
- [6] Carl de Marcken. On the unsupervised induction of phrase-structure grammars. In *Proceedings of the Third Workshop on Very Large Corpora*, 1995.

- [7] Paramveer S. Dhillon, Dean Foster, and Lyle Ungar. Multi-view learning of word embeddings via CCA. In *Advances in Neural Information Processing Systems (NIPS)*, volume 24, 2011.
- [8] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, July 2006.
- [9] Mark Johnson, Thomas L. Griffiths, and Sharon Goldwater. Bayesian inference for PCFGs via Markov chain Monte Carlo. In *Proceedings of NAACL-HLT*, 2007.
- [10] K. Lari and S. Young. The estimation of stochastic context-free grammars using the inside-outside algorithm. *Computer Speech and Language*, 4:35–56, 1990.
- [11] Tahira Naseem, Harr Chen, Regina Barzilay, and Mark Johnson. Using universal linguistic knowledge to guide grammar induction. In *Proceedings of EMNLP*, pages 1234–1244, 2010.
- [12] Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of COLING-ACL*, pages 433–440, July 2006.
- [13] Noah A. Smith. *Novel Estimation Methods for Unsupervised Discovery of Latent Structure in Natural Language Text*. PhD thesis, Johns Hopkins University, 2006.
- [14] Valentin I. Spitzkovsky, Hiyan Alshawi, and Daniel Jurafsky. From Baby Steps to Leapfrog: How “Less is More” in unsupervised dependency parsing. In *Proceedings of NAACL-HLT*, 2010.
- [15] Andreas Stolcke and Stephen M. Omohundro. Inducing probabilistic grammars by Bayesian model merging. In *Proceedings of ICGI*, Lecture Notes in Artificial Intelligence. Springer-Verlag, 1994.
- [16] J. Takami and S. Sagayama. A successive state splitting algorithm for efficient allophone modeling. In *Proceedings of ICASSP*, pages 573–576, 1992.
- [17] François Vanderbeck. On Dantzig-Wolfe decomposition in integer programming and ways to perform branching in a Branch-and-Price algorithm. *Operations Research*, 48 (1):111–128, January 2000.