

List of NLP Papers Using Bayesian Analysis

COMS E6998-11

Instructor: Shay B. Cohen

January 26, 2013

This list is a partial list of some of the natural language processing papers that make use of the Bayesian approach. The list currently is by no means exhaustive or comprehensive, but it gets updated over time.

- [1] A. Arun, C. Dyer, B. Haddow, P. Blunsom, A. Lopez, and P. Koehn. Monte carlo inference and maximization for phrase-based translation. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 102–110, Boulder, Colorado, June 2009. Association for Computational Linguistics.
- [2] S. B. Cohen, D. M. Blei, and N. A. Smith. Variational inference for adaptor grammars. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 564–572, Los Angeles, California, June 2010. Association for Computational Linguistics.
- [3] S. B. Cohen and N. A. Smith. Covariance in unsupervised learning of probabilistic grammars. *Journal of Machine Learning Research (JMLR)*, 11:3017–3051, 2010.
- [4] S. Goldwater and T. Griffiths. A fully Bayesian approach to unsupervised part-of-speech tagging. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 744–751, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [5] S. Goldwater and M. Johnson. Priors in Bayesian learning of phonological rules. In *Proceedings of the Seventh Meeting Meeting of the ACL Special Interest Group on Computational Phonology: SIGPHON 2004*, 2004.
- [6] A. Haghighi and D. Klein. Unsupervised coreference resolution in a nonparametric Bayesian model. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 848–855. Association for Computational Linguistics, 2007.

- [7] M. Johnson. PCFGs, topic models, Adaptor Grammars and learning topical collocations and the structure of proper names. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1148–1157, Uppsala, Sweden, July 2010. Association for Computational Linguistics.
- [8] M. Johnson and S. Goldwater. Improving nonparametric Bayesian inference: experiments on unsupervised word segmentation with adaptor grammars. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 317–325, Boulder, Colorado, June 2009. Association for Computational Linguistics.
- [9] M. Johnson, T. Griffiths, and S. Goldwater. Bayesian inference for PCFGs via Markov chain Monte Carlo. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 139–146, Rochester, New York, April 2007. Association for Computational Linguistics.
- [10] M. Johnson, T. L. Griffiths, and S. Goldwater. Adaptor Grammars: A framework for specifying compositional nonparametric Bayesian models. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 641–648. MIT Press, Cambridge, MA, 2007.
- [11] B. Jones, M. Johnson, and S. Goldwater. Semantic parsing with bayesian tree transducers. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 488–496, Jeju Island, Korea, July 2012. Association for Computational Linguistics.
- [12] K. Kurihara and T. Sato. Variational Bayesian grammar induction for natural language. In *8th International Colloquium on Grammatical Inference*, 2006.
- [13] P. Liang, M. Jordan, and D. Klein. Probabilistic grammars and hierarchical Dirichlet processes. In *The Oxford Handbook of Applied Bayesian Analysis*. Oxford University Press, 2009.
- [14] P. Liang, S. Petrov, M. Jordan, and D. Klein. The infinite PCFG using hierarchical Dirichlet processes. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 688–697, 2007.
- [15] L. Pearl, S. Goldwater, and M. Steyvers. Online learning mechanisms for Bayesian models of word segmentation. *Research on Language and Computation*, 8(2):107–132, 2011.
- [16] Y. W. Teh. A hierarchical Bayesian language model based on Pitman-Yor processes. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 985–992, 2006.

- [17] K. Toutanova and M. Johnson. A Bayesian LDA-based model for semi-supervised part-of-speech tagging. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*. MIT Press, Cambridge, MA, 2008.
- [18] E. Yamangil and S. M. Shieber. Bayesian synchronous tree-substitution grammar induction and its application to sentence compression. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 937–947, Uppsala, Sweden, July 2010. Association for Computational Linguistics.

The following is a list of papers that also use Bayesian statistics, but not necessarily for NLP. Still, they are quite useful to read as an NLP person.

- [1] Y. W. Teh, M. Jordan, M. Beal, and D. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101:1566–1581, 2006.
- [2] Y. W. Teh, K. Kurihara, and M. Welling. Collapsed variational inference for HDP. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*. MIT Press, Cambridge, MA, 2008.