List of NLP Papers Using Bayesian Analysis COMS E6998-11

Instructor: Shay B. Cohen

May 8, 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. The list was compiled with a lot of help from the students attending COMS E6998-11.

[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. Chen and P. Gopalakrishnan. Clustering via the bayesian information criterion with applications in speech recognition. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 2, pages 645–648. IEEE, 1998.

[3] S. F. Chen. Bayesian grammar induction for language modeling. In *Proceedings of ACL*, 1995.

[4] D. Chiang, J. Graehl, K. Knight, A. Pauls, and S. Ravi. Bayesian inference for finite-state transducers. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 447–455, Los Angeles, California, June 2010. Association for Computational Linguistics.

[5] 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.

[6] S. B. Cohen and N. A. Smith. Covariance in unsupervised learning of probabilistic grammars. *Journal of Machine Learning Research (JMLR)*, 11:3017–3051, 2010.

[7] A. Davies and Z. Ghahramani. Language-independent bayesian sentiment mining of twitter. In *In The Fifth Workshop on Social Network Mining and Analysis (SNA-KDD 2011)*, August.

[8] M. Dreyer and J. Eisner. Discovering morphological paradigms from plain text using a Dirichlet process mixture model. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 616–627, Edinburgh, July 2011. Supplementary material (9 pages) also available.

[9] J. R. Finkel and C. D. Manning. Hierarchical bayesian domain adaptation. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 602–610, Boulder, Colorado, June 2009. Association for Computational Linguistics.

[10] J. Gao and M. Johnson. A comparison of Bayesian estimators for unsupervised Hidden Markov Model POS taggers. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 344–352, Honolulu, Hawaii, October 2008. Association for Computational Linguistics.

[11] 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.

[12] 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.

[13] 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.

[14] A. Haghighi and D. Klein. An entity-level approach to information extraction. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 291–295, Uppsala, Sweden, July 2010. Association for Computational Linguistics.

[15] Q. Huo and C. Lee. A bayesian predictive classification approach to robust speech recognition. *Speech and Audio Processing, IEEE Transactions on*, 8(2):200–204, 2000.

[16] H. D. III. Non-parametric bayesian areal linguistics. *CoRR*, abs/0906.5114, 2009.

[17] 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.

[18] M. Johnson and S. Goldwater. Improving nonparameteric 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.

[19] 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.

[20] 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.

[21] 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.

[22] K. Kurihara and T. Sato. Variational Bayesian grammar induction for natural language. In *8th International Colloquium on Grammatical Inference*, 2006.

[23] W. Li and A. McCallum. Pachinko allocation: Dag-structured mixture models of topic correlations. In *Proceedings of the 23rd international conference on Machine learning*, ICML '06, pages 577–584, 2006.

[24] 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.

[25] 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.

[26] C. Lin, Y. He, and R. Everson. A comparative study of bayesian models for unsupervised sentiment detection. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, CoNLL '10, pages 144–152, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

[27] C. Mermer and M. Saraclar. Bayesian word alignment for statistical machine translation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 182–187, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.

[28] D. Norris and J. McQueen. Shortlist b: a bayesian model of continuous speech recognition. *Psychological review*, 115(2):357, 2008.

[29] M. Paul, A. Finch, P. R. Dixon, and E. Sumita. Dialect translation: Integrating bayesian co-segmentation models with pivot-based smt. In *Proceedings of the First Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties*, pages 1–9, Edinburgh, Scotland, July 2011. Association for Computational Linguistics.

[30] 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.

[31] M. Post and D. Gildea. Bayesian learning of a tree substitution grammar. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACLShort '09, pages 45–48, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[32] S. Ravi and K. Knight. Deciphering foreign language. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 12–21, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.

[33] S. Roberts, D. Husmeier, I. Rezek, and W. Penny. Bayesian approaches to gaussian mixture modeling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(11):1133–1142, 1998.

[34] F. Sebastiani. Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1):1–47, 2002.

[35] H. Shindo, Y. Miyao, A. Fujino, and M. Nagata. Bayesian symbol-refined tree substitution grammars for syntactic parsing. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 440–448, Jeju Island, Korea, July 2012. Association for Computational Linguistics.

[36] S. Tan, X. Cheng, Y. Wang, and H. Xu. Adapting naive bayes to domain adaptation for sentiment analysis. In *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, ECIR '09, pages 337–349, Berlin, Heidelberg, 2009. Springer-Verlag.

[37] 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.

[38] I. Titov and A. Klementiev. A bayesian model for unsupervised semantic parsing. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 1445–1455, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[39] 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.

[40] 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.

[41] H. Yu and V. Hatzivassiloglou. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, EMNLP '03, pages 129–136, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics.

[42] H. Zhang, C. Quirk, R. C. Moore, and D. Gildea. Bayesian learning of non-compositional phrases with synchronous parsing. In *Proceedings of ACL-08: HLT*, pages 97–105, Columbus, Ohio, June 2008. 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] D. M. Blei, T. L. Griffiths, M. I. Jordan, and J. B. Tenenbaum. Hierarchical topic models and the nested chinese restaurant process. In *Advances in Neural Information Processing Systems*, page 2003. MIT Press, 2004.

[2] Y. W. Teh, M. Jordan, M. Beal, and D. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101:1566–1581, 2006.

[3] 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.