Spectral Learning Algorithms for Natural Language Processing (NAACL 2013 Tutorial)

Shay B. Cohen, Michael Collins, Dean P. Foster, Karl Stratos and Lyle Ungar

June 7, 2013

References

[1] A. Anandkumar, D. Foster, D. Hsu, S. M. Kakade, and Y. Liu. A spectral algorithm for latent dirichlet allocation. arXiv:1204.6703, 2012.

[2] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky. Tensor decompositions for learning latent-variable models. arXiv:1210.7559, 2012.

[3] R. Bailly, A. Habrar, and F. Denis. A spectral approach for probabilistic grammatical inference on trees. In *Proceedings of ALT*, 2010.

[4] B. Balle and M. Mohri. Spectral learning of general weighted automata via constrained matrix completion. In P. Bartlett, F.C.N. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 2168–2176. 2012.

[5] B. Balle, A. Quattoni, and X. Carreras. A spectral learning algorithm for finite state transducers. In *Proceedings of ECML*, 2011.

[6] A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of COLT*, 1998.

[7] P. F. Brown, P. V. deSouza, R. L. Mercer, V. J. Della Pietra, and J. C. Lai. Class-based n-gram models of natural language. *Computational Linguistics*, 18:467–479, 1992.

[8] S. B. Cohen, K. Stratos, M. Collins, D. F. Foster, and L. Ungar. Spectral learning of latent-variable PCFGs. In *Proceedings of ACL*, 2012.

[9] S. B. Cohen, K. Stratos, M. Collins, D. P. Foster, and L. Ungar. Experiments with spectral learning of latent-variable PCFGs. In *Proceedings of NAACL*, 2013.

[10] M. Collins and Y. Singer. Unsupervised models for named entity classification. In *In Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 100–110, 1999.

[11] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39:1–38, 1977.

[12] P. Dhillon, J. Rodu, M. Collins, D. P. Foster, and L. H. Ungar. Spectral dependency parsing with latent variables. In *Proceedings of EMNLP*, 2012.

[13] J. Goodman. Parsing algorithms and metrics. In *Proceedings of ACL*, 1996.

[14] D. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, 16(12):2639–2664, 2004.

[15] H. Hotelling. Relations between two sets of variants. *Biometrika*, 28:321–377, 1936.

[16] D. Hsu, S. M. Kakade, and T. Zhang. A spectral algorithm for learning hidden Markov models. In *Proceedings of COLT*, 2009.

[17] H. Jaeger. Observable operator models for discrete stochastic time series. *Neural Computation*, 12(6), 2000.

[18] T. K. Landauer, P. W. Foltz, and D. Laham. An introduction to latent semantic analysis. *Discourse Processes*, (25):259–284, 1998.

[19] F. M. Luque, A. Quattoni, B. Balle, and X. Carreras. Spectral learning for non-deterministic dependency parsing. In *Proceedings of EACL*, 2012.

[20] T. Matsuzaki, Y. Miyao, and J. Tsujii. Probabilistic CFG with latent annotations. In *Proceedings of ACL*, 2005.

[21] A. Parikh, L. Song, and E. P. Xing. A spectral algorithm for latent tree graphical models. In *Proceedings of The 28th International Conference on Machine Learning (ICML 2011)*, 2011.

[22] S. Petrov, L. Barrett, R. Thibaux, and D. Klein. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of COLING-ACL*, 2006.

[23] L. Saul, F. Pereira, and O. Pereira. Aggregate and mixed-order markov models for statistical language processing. In *In Proceedings of the Second Conference on Empirical Methods in Natural Language Processing*, pages 81–89, 1997.

[24] A. Tropp, N. Halko, and P. G. Martinsson. Finding structure with randomness: Stochastic algorithms for constructing approximate matrix decompositions. In *Technical Report No. 2009-05*, 2009.

[25] S. Vempala and G. Wang. A spectral algorithm for learning mixtures of distributions. *Journal of Computer and System Sciences*, 68(4):841–860, 2004.