

B. REFERENCES

- [1] Apache hadoop. <http://hadoop.apache.org/>.
- [2] C. E. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, pages 1152–1174, 1974.
- [3] A. Beutel, W. Xu, V. Guruswami, C. Palow, and C. Faloutsos. CopyCatch: stopping group attacks by spotting lockstep behavior in social networks. In *WWW*, 2013.
- [4] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):15, 2009.
- [5] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Trans. on PAMI*, 17(8):790–799, 1995.
- [6] J. Dean and S. Ghemawat. MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113, 2008.
- [7] A. Doan, R. Ramakrishnan, and A. Y. Halevy. Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4):86–96, 2011.
- [8] J. Howe. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4, 2006.
- [9] K. Jiang, B. Kulis, and M. I. Jordan. Small-variance asymptotics for exponential family Dirichlet process mixture models. In *NIPS*, 2012.
- [10] E. Kamar, S. Hacker, and E. Horvitz. Combining human and machine intelligence in large-scale crowdsourcing. In *AAMAS*, 2012.
- [11] L. Kim. The most expensive keywords in Google adwords. <http://www.wordstream.com/blog/ws/2011/07/18/most-expensive-google-adwords-keywords>, 2011.
- [12] B. Kulis and M. I. Jordan. Revisiting k-means: New algorithms via Bayesian nonparametrics. In *ICML*, 2012.
- [13] X. Li, Y. Liu, M. Zhang, and S. Ma. Fraudulent support telephone number identification based on co-occurrence information on the web. In *AAAI*, 2014.
- [14] X. Li, M. Zhang, Y. Liu, S. Ma, Y. Jin, and L. Ru. Search engine click spam detection based on bipartite graph propagation. In *WSDM*, 2014.
- [15] A. Metwally, D. Agrawal, and A. El Abbadi. Detectives: detecting coalition hit inflation attacks in advertising networks streams. In *WWW*, 2007.
- [16] A. Metwally, D. A. A. El Abbadi, and Q. Zheng. Hide and seek: Detecting hit inflation fraud in streams of web advertising networks. *Technical Report, University of California at Santa Barbara*, 2006.
- [17] R. Oentaryo, E.-P. Lim, M. Finegold, D. Lo, F. Zhu, C. Phua, E.-Y. Cheu, G.-E. Yap, K. Sim, M. N. Nguyen, et al. Detecting click fraud in online advertising: a data mining approach. *JMLR*, 15(1):99–140, 2014.
- [18] X. Pan, J. E. Gonzalez, S. Jegelka, T. Broderick, and M. Jordan. Optimistic concurrency control for distributed unsupervised learning. In *NIPS*, 2013.
- [19] A. Roychowdhury, K. Jiang, and B. Kulis. Small-variance asymptotics for hidden Markov models. In *NIPS*, 2013.
- [20] R. Snow, B. O’Connor, D. Jurafsky, and A. Y. Ng. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *EMNLP*, 2008.
- [21] Y. Tian and J. Zhu. Learning from crowds in the presence of schools of thought. In *SIGKDD*, 2012.
- [22] G. Wang, C. Wilson, X. Zhao, Y. Zhu, M. Mohanlal, H. Zheng, and B. Y. Zhao. Serf and turf: crowdurfing for fun and profit. In *WWW*, 2012.
- [23] Y. Wang and J. Zhu. Small-variance asymptotics for Dirichlet process mixtures of SVMs. In *AAAI*, 2014.