

MODELING AND INFERENCE WITH THE DIRICHLET PROCESS

Michael I. Jordan

Department of Electrical Engineering and Computer Science

Department of Statistics

University of California, Berkeley

(jordan@cs.berkeley.edu; <http://www.cs.berkeley.edu/~jordan>)

Abstract

The Dirichlet process (DP) is a random measure that is useful in nonparametric Bayesian modeling, specifically in model-based clustering. In this presentation, I consider several extensions of the DP. Motivated by problems in bioinformatics and information retrieval, I first consider involving groups of clustering problems, in which it is desirable to share clusters among groups. While a simple Bayesian hierarchy in which multiple DP mixtures share a fixed base measure fail to yield the desired sharing, a more fully Bayesian notion of a "hierarchical Dirichlet process" (HDP) resolves the sharing problem [1]. I discuss representations of the HDP based on extensions of stick-breaking and Chinese restaurant processes, and show how these representations yield MCMC algorithms for posterior inference. I also discuss alternative variational methods for posterior inference, showing that these methods can yield comparable accuracy at less computational cost [2]. Finally, I discuss empirical Bayesian approaches to Dirichlet process mixtures which involve inference with respect to the base measure [3]. (Joint work with David Blei, Jon McAuliffe, Yee Whye Teh and Matt Beal).

References:

- [1] Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (to appear). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*.
- [2] Blei, D. M., & Jordan, M. I. (2005). Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1, 121-144.
- [3] McAuliffe, J. D., Blei, D. M., & Jordan, M. I. (2006). Nonparametric empirical Bayes for the Dirichlet process mixture model. *Statistics and Computing*, 16, 5-14.

Key Words: Nonparametric Bayes, Dirichlet process, variational inference