
Memoized Online Variational Inference for Dirichlet Process Mixture Models

Michael C. Hughes and Erik B. Sudderth

Department of Computer Science, Brown University, Providence, RI 02912
mhughes@cs.brown.edu, sudderth@cs.brown.edu

Abstract

Variational inference algorithms provide the most effective framework for large-scale training of Bayesian nonparametric models. Stochastic online approaches are promising, but are sensitive to the chosen learning rate and often converge to poor local optima. We present a new algorithm, *memoized online variational inference*, which scales to very large (yet finite) datasets while avoiding the complexities of stochastic gradient. Our algorithm maintains finite-dimensional sufficient statistics from batches of the full dataset, requiring some additional memory but still scaling to millions of examples. Exploiting nested families of variational bounds for infinite nonparametric models, we develop principled birth and merge moves allowing non-local optimization. Births adaptively add components to the model to escape local optima, while merges remove redundancy and improve speed. Using Dirichlet process mixture models for image clustering and denoising, we demonstrate major improvements in robustness and accuracy.

1 Introduction

Bayesian nonparametric methods provide a flexible framework for unsupervised modeling of structured data like text documents, time series, and images. They are especially promising for large datasets, as their nonparametric priors should allow complexity to grow smoothly as more data is seen. Unfortunately, contemporary inference algorithms do not live up to this promise, scaling poorly and yielding solutions that represent poor local optima of the true posterior. In this paper, we propose new scalable algorithms capable of escaping local optima. Our focus is on clustering data via the Dirichlet process (DP) mixture model, but our methods are much more widely applicable.

Stochastic online variational inference is a promising general-purpose approach to Bayesian nonparametric learning from streaming data [1]. While individual steps of stochastic optimization algorithms are by design scalable, they are extremely vulnerable to local optima for non-convex unsupervised learning problems, frequently yielding poor solutions (see Fig. 2). While taking the best of multiple runs is possible, this is unreliable, expensive, and ineffective in more complex structured models. Furthermore, the noisy gradient step size (or learning rate) requires external parameters which must be fine-tuned for best performance, often requiring an expensive validation procedure. Recent work has proposed methods for automatically adapting learning rates [2], but these algorithms' progress on the overall variational objective remains local and non-monotonic.

In this paper, we present an alternative algorithm, *memoized online variational inference*, which avoids noisy gradient steps and learning rates altogether. Our method is useful when all data may not fit in memory, but we can afford multiple full passes through the data by processing successive batches. The algorithm visits each batch in turn and updates a cached set of sufficient statistics which accurately reflect the *entire dataset*. This allows rapid and noise-free updates to global parameters at every step, quickly propagating information and speeding convergence. Our memoized approach is generally applicable in any case batch or stochastic online methods are useful, including topic models [1] and relational models [3], though we do not explore these here.