Summary 11: Stochastic Variational Inference

Susan Cherry

This paper develops an algorithm for approximating posterior distributions that works well with massive datasets. Traditionally MCMC and variational inference have been used in statistics and machine learning for computing intractable posteriors, but neither scale easily to large and complex data sets. Instead, this paper introduces stochastic variational inference, which uses stochastic optimization to find the maximum of an objective function by following noisy estimates of its gradient. Stochastic variational inference has the form: 1) subsample data points from the data, 2) analyze the subsample using current variational parameters, 3) implemented closed form update of the parameters, 4) repeat. Unlike traditional algorithms which require analyzing the entire dataset repeatedly, this algorithm only requires that we analyze random subsamples.

Next, the paper derives stochastic variational inference. It defines local and global hidden variables and the requirements on the conditional distributions. Then it presents mean-field variational inference, which is an approximate inference technique that finds a tractable distribution over the hidden variables. The coordinate descent algorithm is derived. then the natural gradient of the variational objective is derived. Finally, stochastic optimization (which uses noisy estimates of a gradient to optimize an objective function) is reviewed and applied to variational inference. The authors use stochastic optimization with noisy estimates of the natural gradient of the variational object, where the estimates come from subsampling the data. Unlike traditional variational inference algorithms, stochastic variational inference can handle very large datasets.

Next, the authors describe stochastic variational inference for the latent dirichlet allocation and the hierarchical dirichlet process. The LDA is reviewed snd the algorithm for stochastic variational inference is specified on page 1327. A drawback of the LDA is that it requires the number of topics to be fixed in advanced, so the paper also discusses HDP and presents stochastic variational inference for HDP on page 1334. Next, the empirical performance and effectiveness of stochastic variational inference for LDA and HDP is investigated. The algorithms are tested on three collections of documents. The HDP consistently performed better than the LDA.

The paper concludes with a discussion of its results. Stochastic variational inference is useful because it allows us to analyze huge, complex datasets. The main idea is to use stochastic optimization to optimize the variational objective, after repeatedly subsampling the data to get noisy estimates of the natural gradient. Future areas of study include using stochastic optimization to move beyond closed form updates. For example, collapsed variational inference marginalizes out some of the hidden variables, which allows for a lower dimensional posterior. Structured variational distributions relax mean-field variation approximation and allow for better approximation of complex posteriors. The stochastic variational inference algorithms explained in this paper allow us to connect new developments in stochastic optimization to better methods of approximate posterior inference.