Black Box Variational Inference on Latent Dirichlet Allocation

Aodong Li
al5350@nyu.edu

November 9, 2018

Abstract

Latend Dirichlet Allocation (LDA) [1] is an important member of topic models. It is used to model the underlying topics that a given set of documents represent. In order to infer the underlying topics, various inference algorithms can be applied. These algorithms are mainly divided into two categories, Markov Chain Monte Carlo (MCMC) and variational inference. Black Box Variational Inference (BBVI) [2] is an instance among them. Although BBVI is introduced as a generic algorithm to simplify inference problem, it has never been used for LDA. In this work, we try to apply BBVI to LDA and compare the performance with different inference algorithms, e.g., vanilla Gibbs sampling and variational inference. Their performances are evaluated on some dataset, 20 Newsgroups[1] or Reuters news dataset[2].

References

[1] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[2] Rajesh Ranganath, Sean Gerrish, and David Blei. Black box variational inference. In *Artificial Intelligence and Statistics*, pages 814–822, 2014.

[1] http://qwone.com/ jason/20Newsgroups/

[2] https://archive.ics.uci.edu/ml/datasets/reuters-21578+text+categorization+collection