Adaptive Markov Chain Monte Carlo Methods for Bayesian Conditional Random Fields

Charles Sutton
School of Informatics
The University of Edinburgh
10 Crichton Street
Edinburgh EH8 9AB
csutton@inf.ed.ac.uk

Yichuan Zhang
School of Informatics
The University of Edinburgh
10 Crichton Street
Edinburgh EH8 9AB
Y.Zhang-60@sms.ed.ac.uk

Abstract

Conditional random fields (CRFs) are a family of powerful graphical models that have been applied successfully in many areas, such as natural language processing, image modelling and information extraction. The maximum likelihood training approach on CRFs suffers from overfitting. Overfitting can be reduced by using the maximum a posteriori (MAP) approach, but only the single most probable model parameters is used in prediction. To overcome this shortcoming, Bayesian Conditional Random Fields (BCRFs) are trained by estimating the posterior distribution and the prediction is made by marginalising over model parameters. Because in general the posterior distribution cannot be computed exactly, the approximation techniques are crucial for effective training. Approximations to BCRFs based on variational methods and Laplace methods have been studied by Qi et al. [3] and Welling and Parise [4]. One important assumption of these methods is that the posterior distribution can be well approximated by a Gaussian-like distribution. In contrast, Monte Carlo methods can provide an asymptotically accurate approximation to the target distribution, but sampling from a high dimensional distribution can be very inefficient.

In this work, we aim at an efficient Markov Chain Monte Carlo (MCMC) method for BCRFs. Like many optimisation techniques used in MAP training, the Hamiltonian Monte Carlo (HMC) method uses gradient information to reduce the random walk behaviour, so HMC method is able to sample more efficiently than many other MCMC methods. However, to our knowledge, there is no previous works on applying HMC method to BCRFs, so we use HMC method to train Linear-chain BCRFs. Evaluating the performance of BCRFs using HMC on a small natural language segmentation dataset, we find significantly improved results over standard MAP estimation. However, we find that, in general, a better result can only be attained when the HMC method takes a relative large number of leapfrog steps, 50 leaps in our experiments, and a carefully chosen step size. This observation agrees with the experience in previous works [1]. More importantly, because computing the gradient is the bottleneck of performance, for the same number of samples, the computational time grows linearly with the number of leapfrog steps.

Because of these drawbacks of HMC, we try to find a more efficient sampling method. Therefore, we propose an adaptive MCMC scheme that modifies a Metropolis-Hastings (MH) sampler like proposal based on the previous samples. Importantly, our scheme avoids the well-known problems with the validity of adaptive schemes by limiting the adaption to depend on at most the previous k

samples. Our scheme has two important differences from other MCMC methods that maintain a population of particles, namely, non-reversibility, that may be advantageous [2], and the form of adaptation can be quite general without sacrificing validity. A simple example of this framework is a random walk MH sampler with a Gaussian proposal whose variance is chosen as the empirical variance of the previous k samples. We are currently exploring more complex adaptation schemes that use gradient information to search the state space more efficiently.

References

[1] Neal, R.M. (2010) MCMC using Hamiltonian dynamics in S. Brooks, A. Gelman, G. Jones, and X. Meng (Ed.), *Handbook of Markov Chain Monte Carlo* Chapman & Hall / CRC Press.

[2] Neal, R.M. (2004) Improving Asymptotic Variance of MCMC Estimators: Non-reversible Chains are Better. Technical Report No. 0406, Dept. of Statistics, University of Toronto.

[3] Qi, Y., Szummer, M., & Minka, T. P. (2005) Bayesian conditional random fields. *Proc. AI & Statistics*.

[4] Welling, M., & Parise, S. (2006) Bayesian random fields: The Bethe-Laplace approximation. *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, Arlington, VA: AUAI Press.