Poisson-Minibatching for Gibbs Sampling with Convergence Rate Guarantees

Ruqi Zhang and Christopher De Sa

Cornell University

Scale Gibbs Sampling by Subsampling

Gibbs sampling is one of the most popular Markov chain Monte Carlo (MCMC) methods

+ Converge asymptotically to the desired distribution
+ Work very well in practice
– Prohibitive cost on large-scale datasets or models

Scale Gibbs Sampling by Subsampling

Gibbs sampling is one of the most popular Markov chain Monte Carlo (MCMC) methods

+ Converge asymptotically to the desired distribution

+ Work very well in practice

– Prohibitive cost on large-scale datasets or models

Subsampling methods to scale MCMC

+ Reduce computational cost significantly

– No guarantees on the accuracy and the efficiency

Scale Gibbs Sampling by Subsampling

Gibbs sampling is one of the most popular Markov chain Monte Carlo (MCMC) methods

+ Converge asymptotically to the desired distribution
+ Work very well in practice
– Prohibitive cost on large-scale datasets or models

Subsampling methods to scale MCMC

+ Reduce computational cost significantly
– No guarantees on the accuracy and the efficiency

We show how to scale Gibbs sampling by subsampling with guarantees on the accuracy, convergence rate, and computational efficiency

Inference on Graphical Models

Consider factor graphs

$$\pi(x_{1:n}) = \frac{1}{Z} \cdot \prod_{\phi \in \Phi} \exp\left(\phi(x_{1:n})\right)$$

Sample from π by Gibbs sampling

Loop

Select a variable x_i to sample at random

Compute the conditional distribution of x_i based on all factors ϕ that depend on x_i

Resample variable x_i from the conditional distribution

End Loop

Inference on Graphical Models

Consider factor graphs

$$\pi(x_{1:n}) = \frac{1}{Z} \cdot \prod_{\phi \in \Phi} \exp\left(\phi(x_{1:n})\right)$$

Sample from π by Gibbs sampling

Loop

Select a variable x_i to sample at random

Compute the conditional distribution of x_i based on all factors ϕ that depend on x_i

Resample variable x_i from the conditional distribution

End Loop

Very expensive when the factor set is large!

Can we subsample factors to compute conditional distributions?

Previous Work

Scale MCMC with subsampling methods: [Welling and Teh, 2011], [Maclaurin and Adams, 2014], [Bardenet et.al., 2017] ...

Christopher De Sa, Vincent Chen and Wing Wong. *Minibatch Gibbs Sampling on Large Graphical Models*. ICML 2018

Main idea:

- Use conditional distributions based on subsampled factors as proposal distributions
- Add the Metropolis-Hastings (M-H) step to correct the bias

Previous Work

Scale MCMC with subsampling methods: [Welling and Teh, 2011], [Maclaurin and Adams, 2014], [Bardenet et.al., 2017] ...

Christopher De Sa, Vincent Chen and Wing Wong. *Minibatch Gibbs Sampling on Large Graphical Models*. ICML 2018

Main idea:

- Use conditional distributions based on subsampled factors as proposal distributions
- Add the Metropolis-Hastings (M-H) step to correct the bias

Limitations:

- The Metropolis-Hastings step is expensive
- Only support sampling from discrete distributions

Poisson-Minibatching

Introduce an auxiliary Poisson variable for each factor to control whether a factor is used or not

$$s_\phi | x_{1:n} \sim \text{Poisson}\left(\frac{\lambda M_\phi}{L} + \phi(x_{1:n})\right)$$

Poisson-Minibatching

Introduce an auxiliary Poisson variable for each factor to control whether a factor is used or not

$$s_\phi | x_{1:n} \sim \text{Poisson}\left(\frac{\lambda M_\phi}{L} + \phi(x_{1:n})\right)$$

The joint distribution

$$\pi(x_{1:n}, s_{\phi\in\Phi}) \propto \exp\left(\sum_{\phi\in\Phi}\left(s_\phi \log\left(1 + \frac{L}{\lambda M_\phi}\phi(x_{1:n})\right) + s_\phi \log\left(\frac{\lambda M_\phi}{L}\right) - \log\left(s_\phi!\right)\right)\right)$$

A factor ϕ contributes to the energy only when $s_\phi > 0$, thus the algorithm computes conditional distributions with only a subset of factors

Poisson-Minibatching

Introduce an auxiliary Poisson variable for each factor to control whether a factor is used or not

$$s_\phi | x_{1:n} \sim \text{Poisson}\left(\frac{\lambda M_\phi}{L} + \phi(x_{1:n})\right)$$

The joint distribution

$$\pi(x_{1:n}, s_{\phi\in\Phi}) \propto \exp\left(\sum_{\phi\in\Phi}\left(s_\phi \log\left(1 + \frac{L}{\lambda M_\phi}\phi(x_{1:n})\right) + s_\phi \log\left(\frac{\lambda M_\phi}{L}\right) - \log\left(s_\phi!\right)\right)\right)$$

A factor ϕ contributes to the energy only when $s_\phi > 0$, thus the algorithm computes conditional distributions with only a subset of factors

- Expected number of factors being used $\ll$ the factor set size
- Stationary distribution of $x_{1:n}$ does not change even without the M-H step
- Sampling a set of Poisson variables is cheap

Algorithm of Poisson-Minibatching Gibbs Sampling (Poisson-Gibbs)

Loop

Select a variable x_i to sample at random

Resample s_ϕ from its conditional distribution given $x_{1:n}$

Compute the conditional distribution based on the chosen factors ϕ such that $s_\phi > 0$

Resample variable x_i from the conditional distribution

End Loop

- Simple to implement
- No Metropolis-Hastings step

Theoretical Guarantees on Convergence Rate

The convergence rate of our method can be slowed down by at most a constant compared to that of Gibbs sampling

- Provide recipe of setting the hyperparameter minibatch size to make this constant $O(1)$

Sample from Continuous Distributions

Difficulty: non-trivial to sample from continuous conditional distributions

Our Solution: Double Chebyshev Approximation method

- Get polynomial approximation of the PDF by using Chebyshev approximation twice
- Generate a sample by inverse transform sampling

Sample from Continuous Distributions

Difficulty: non-trivial to sample from continuous conditional distributions

Our Solution: Double Chebyshev Approximation method

- Get polynomial approximation of the PDF by using Chebyshev approximation twice
- Generate a sample by inverse transform sampling

Theoretical Guarantees on the accuracy and the efficiency

- Stationary distribution of $x_{1:n}$ does not change
- The convergence rate of our method can be slowed down by at most a constant compared to that of Gibbs sampling

Summary

- Scaling MCMC methods while maintaining theoretical guarantees is hard
- We propose *Poisson-minibatching Gibbs sampling* which solves this problem using the auxiliary variable method
- We provide theoretical guarantees on the accuracy, convergence rate and computational efficiency
- For more details—including experiments—come see our poster!

Thank you!

Poster #158, 5:30 – 7:30 today