Pólya Urn Latent Dirichlet Allocation

using sparsity to reduce MCMC complexity in natural language processing

Imperial College London

Alexander Terenin (Imperial College London) · Måns Magnusson (Linköping University, SE),
Leif Jonsson (Linköping University, SE) · David Draper (University of California, Santa Cruz)

LINKÖPING UNIVERSITY

Abstract

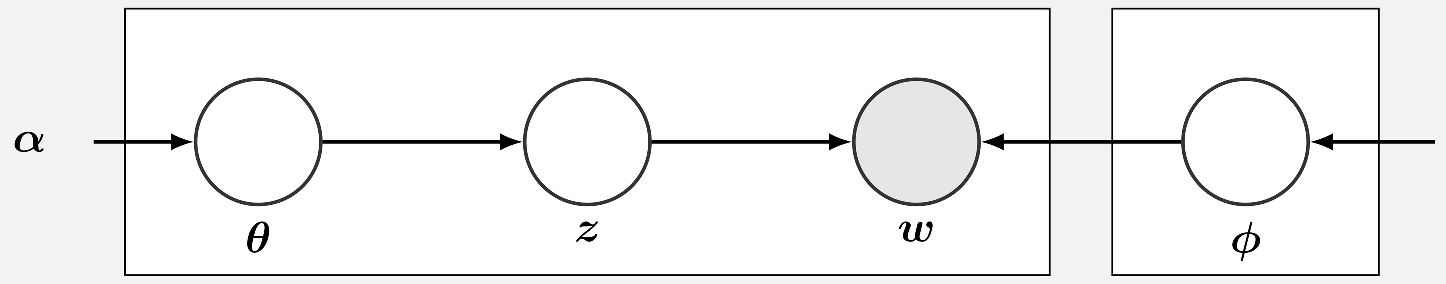
Dirichlet distributions are an essential building block in many Bayesian models, particularly those in Natural Language Processing. We propose a novel Pólya Urn based approximation to the Dirichlet distribution. We show that by using the Polya Urn approximation in a sparse partially collapsed Gibbs sampler for the popular Latent Dirichlet Allocation topic model, we can derive a sampler that is faster than the current state-of-the-art, both theoretically and empirically.

Latent Dirichlet Allocation

The canonical topic model [1].

– Used to learn about topics within a collection of documents.

– Very popular: 20,000+ citations on Google Scholar



Sparse MCMC

Sample topic indicators $z_1, .., z_{i,d}$:

$$z_{i,d} \mid \boldsymbol{z}_{-i,d} \propto \frac{n_{k,v(i)}^{-i} + \beta}{n_{k,\cdot}^{-i} + V\beta}(m_{d,k}^{-i} + \alpha).$$

Standard approach: split sum and precompute Alias table for α term [2].

Complexity: $O\left[\sum_{i=1}^{N} K_{d(i)}^{(\mathbf{m})}\right]$, Metropolis-Hastings based.

Completely sequential.

Poisson Pólya Urn approximation to the Dirichlet distribution

Definition. Let $\boldsymbol{x} \sim \mathrm{PPU}(\varpi, \boldsymbol{F})$ if we have for $\tilde{\gamma}_j \sim \mathrm{Pois}(\varpi F_j)$ that

$$\boldsymbol{x} = \left[\frac{\tilde{\gamma}_1}{\sum_{j=1}^{J}\tilde{\gamma}_j}, .., \frac{\tilde{\gamma}_J}{\sum_{j=1}^{J}\tilde{\gamma}_j}\right].$$

Theorem. Let $\boldsymbol{x} \sim \mathrm{PPU}(\varpi, \boldsymbol{F})$. Let $\boldsymbol{x}^* \sim \mathrm{Dir}(\varpi, \boldsymbol{F})$. Then we have $||\boldsymbol{x} - \boldsymbol{x}^*|| \to 0$ as $\varpi \to \infty$ for all $\boldsymbol{F}$ in the Levy-Prokhorov metric. [4]

Thus we may view the Poisson Pólya Urn distribution as an asymptotic approximation of the Dirichlet distribution using sparse vectors.

Parallel Doubly Sparse MCMC

Start with the partially collapsed Gibbs sampler (PCLDA) [3], replace the Dirichlet distribution with the Poisson Pólya Urn distribution.

Sample topic-word proportion matrix $\boldsymbol{\Phi}$, then topic indicators $\boldsymbol{z}$:

$$\boldsymbol{\phi}_k \sim \mathrm{PPU}(\boldsymbol{n}_k + \boldsymbol{\beta}) \qquad z_{i,d} \propto \phi_{k,v(i)}(m_{d,k}^{-i} + \alpha_k).$$

Advantages compared to PCLDA.

– $\boldsymbol{\Phi}$ becomes sparse.

– Bypasses memory bottleneck.

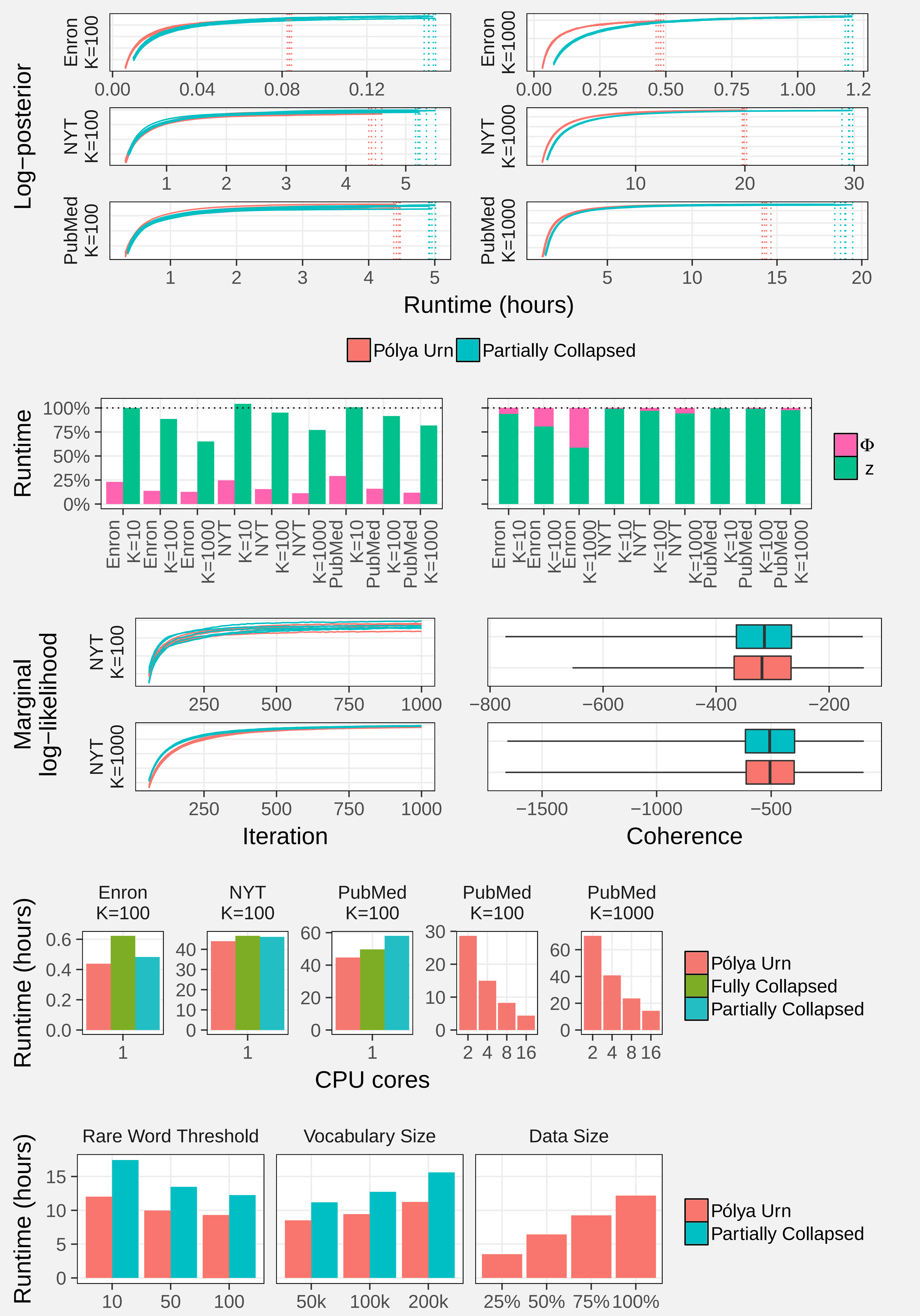
– Near-identical mixing properties.

– Same excellent parallelizability.

– Faster runtime for both $\boldsymbol{\Phi}$ and $\boldsymbol{z}$.

Complexity: $O\left[\sum_{i=1}^{N} \min\left\{K_{d(i)}^{(\mathbf{m})}, K_{v(i)}^{(\boldsymbol{\Phi})}\right\}\right]$.

Performance



References

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(1):993–1022, 2003.

[2] A. Q. Li, A. Ahmed, S. Ravi, and A. J. Smola. Reducing the sampling complexity of topic models. In *Proceedings of the 20th International Conference on Knowledge Discovery and Data Mining*, pages 891–900, 2014.

[3] M. Magnusson, L. Jonsson, M. Villani, and D. Broman. Sparse Partially Collapsed MCMC for Parallel Inference in Topic Models. *Journal of Computational and Graphical Statistics*, 26(4), 2017.

[4] A. Terenin, M. Magnusson, L. Jonsson, and D. Draper. Pólya Urn Latent Dirichlet Allocation: a doubly sparse massively parallel sampler. *arXiv:1704.03581*, 2017.

Future Work

– Understand approximation error from MCMC theory point of view.

– Poisson Pólya Urn approximation is generic and likely to be useful in many other Dirichlet-based hierarchical Bayesian models.