Pólya Urn Latent Dirichlet Allocation

using sparsity to reduce MCMC complexity in natural language processing

Alexander Terenin
Imperial College London

Joint work with Måns Magnusson,
Leif Jonsson, and David Draper

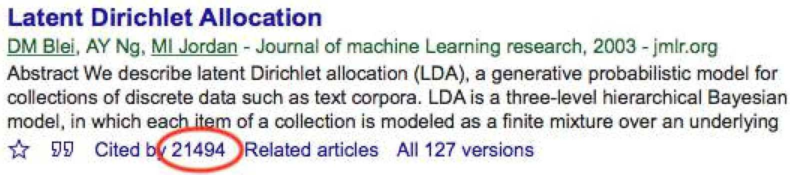
CRISM '18
July 10th, 2018

http://avt.im/

Imperial College London

LINKÖPING UNIVERSITY

Latent Dirichlet Allocation



Latent Dirichlet Allocation

DM Blei, AY Ng, MI Jordan - Journal of machine Learning research, 2003 - jmlr.org

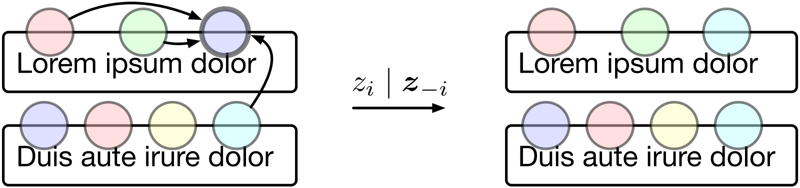
Abstract We describe latent Dirichlet allocation (LDA), a generative probabilistic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying

☆ ❞ Cited by 21494 Related articles All 127 versions

The canonical topic model – everybody uses it!

This work: compute as fast, parallel, and principled as possible

Sparse Fully Collapsed Gibbs Sampling (previous state-of-the-art)



Lorem ipsum dolor

Duis aute irure dolor

$\xrightarrow{z_i \mid \boldsymbol{z}_{-i}}$

Lorem ipsum dolor

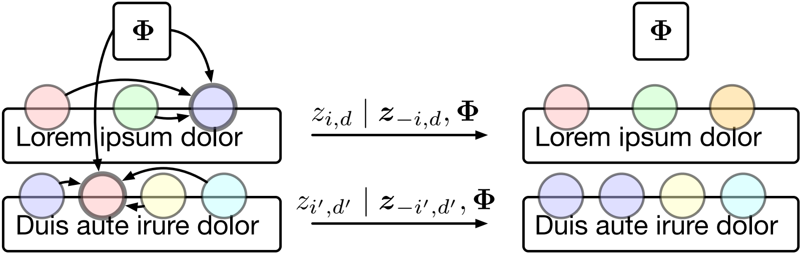
Duis aute irure dolor

✓ Relatively fast

☒ Sequential

☒ Not fully sparse

Sparse Partially Collapsed Gibbs Sampling


$\boldsymbol{\Phi}$
Lorem ipsum dolor
Duis aute irure dolor
$z_{i,d} \mid \boldsymbol{z}_{-i,d}, \boldsymbol{\Phi}$
$z_{i',d'} \mid \boldsymbol{z}_{-i',d'}, \boldsymbol{\Phi}$
$\boldsymbol{\Phi}$
Lorem ipsum dolor
Duis aute irure dolor

✓ Massively parallel

☒ Large dense matrix

☒ Still not fully sparse

Pólya Urn LDA

Dirichlet

$$\boldsymbol{x} = \left[\frac{\gamma_1}{\sum_{i=1}^{k} \gamma_i}, .., \frac{\gamma_k}{\sum_{i=1}^{k} \gamma_i}\right]$$

$$\gamma_i \sim \mathrm{Gamma}(\varpi F_i, 1)$$

Dense

$$\rightarrow$$

Poisson Pólya Urn

$$\boldsymbol{y} = \left[\frac{\tilde{\gamma}_1}{\sum_{i=1}^{k} \tilde{\gamma}_i}, .., \frac{\tilde{\gamma}_k}{\sum_{i=1}^{k} \tilde{\gamma}_i}\right]$$

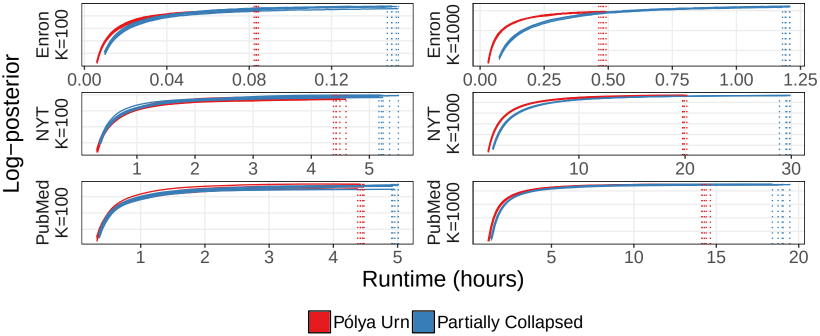
$$\tilde{\gamma}_i \sim \mathrm{Poisson}(\varpi F_i)$$

Sparse

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

✓ Dense matrix $\boldsymbol{\Phi}$ becomes sparse

Performance



✓ Faster runtime with no discernible loss in topic quality

A. Terenin, M. Magnusson, L. Jonsson, and D. Draper. Pólya urn latent Dirichlet allocation: a doubly sparse massively parallel sampler. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018. Accepted, to appear. Available at: arXiv:1704.03581.