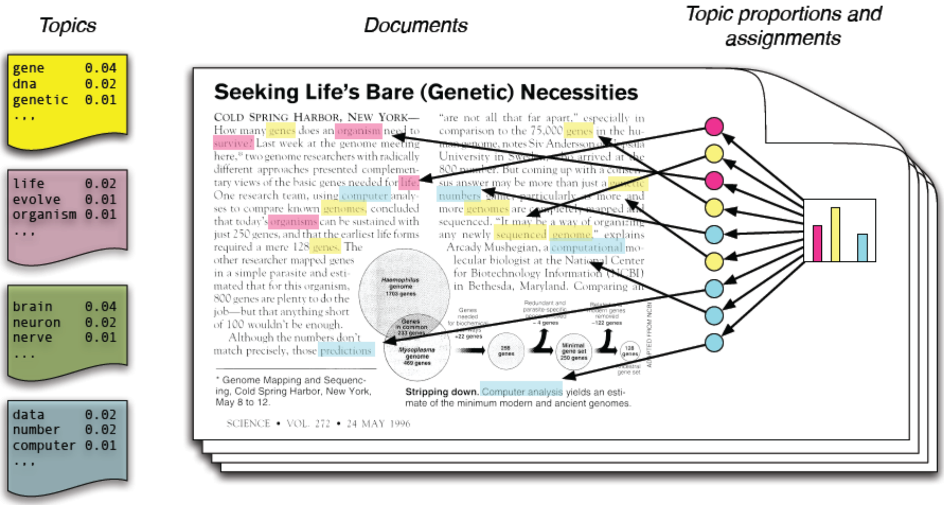
Latent Dirichlet Allocation

Patrice Béchard Théophile Gervet

Université de Montréal

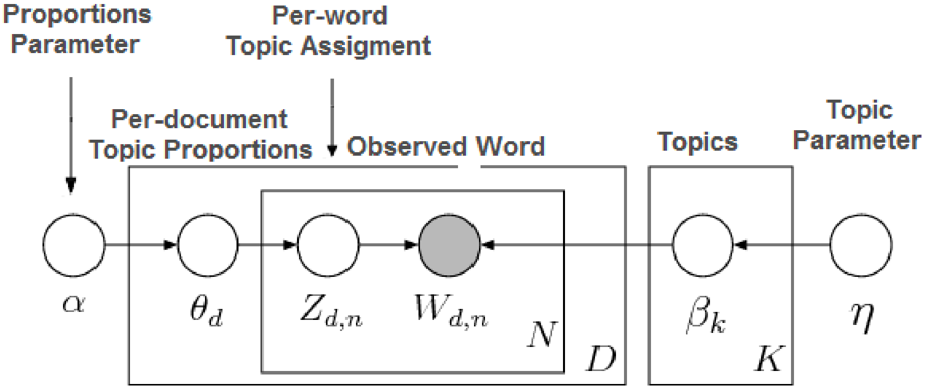
December 12, 2017

Context



Topics
gene 0.04
dna 0.02
genetic 0.01
...
life 0.02
evolve 0.01
organism 0.01
...
brain 0.04
neuron 0.02
nerve 0.01
...
data 0.02
number 0.02
computer 0.01
...
Documents
Topic proportions and assignments
Seeking Life's Bare (Genetic) Necessities
* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.
Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.
SCIENCE • VOL. 272 • 24 MAY 1996

Graphical Model



$$\prod_{k=1}^{K} p(\boldsymbol{\beta}_k|\boldsymbol{\eta}) \prod_{d=1}^{D} p(\boldsymbol{\theta}_d|\boldsymbol{\alpha}) \prod_{n=1}^{N_d} p(z_{d,n}|\boldsymbol{\theta}_d) p(w_{d,n}|z_{d,n}, \boldsymbol{\beta}_{1:K})$$

Generative Process

Formally, LDA assumes the following generative process for each document $\mathbf{w}_d$:

1. Sample a distribution over topics $\boldsymbol{\theta}_d \sim \mathrm{Dir}(\boldsymbol{\alpha})$
2. For each of the N_d words $w_{d,n}$ independently:
   - 2.1 Draw a topic from the distribution over topics : $z_{d,n} \sim \mathrm{Mult}(\boldsymbol{\theta}_d)$
   - 2.2 Draw a word from this topic : $w_{d,n} \sim \mathrm{Mult}(\boldsymbol{\beta}_{z,n})$

Intractable posterior mean :

$$p(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z} | \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = \frac{p(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z}, \mathbf{w} | \boldsymbol{\alpha}, \boldsymbol{\eta})}{p(\mathbf{w} | \boldsymbol{\alpha}, \boldsymbol{\eta})}$$

Variational EM

Intractable posterior : $$p(\boldsymbol{\theta}, \mathbf{z} | \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w} | \boldsymbol{\alpha}, \boldsymbol{\beta})}{p(\mathbf{w} | \boldsymbol{\alpha}, \boldsymbol{\beta})}$$

Log-likelihood : $$\log p(\mathbf{w} | \boldsymbol{\alpha}, \boldsymbol{\beta}) = \log \int \sum_{\mathbf{z}} p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w} | \boldsymbol{\alpha}, \boldsymbol{\beta}) d\boldsymbol{\theta}$$

Usual trick : $$\begin{aligned} \log p(\mathbf{w} | \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \log \int \sum_{\mathbf{z}} \frac{p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w} | \boldsymbol{\alpha}, \boldsymbol{\beta}) q(\boldsymbol{\theta}, \mathbf{z})}{q(\boldsymbol{\theta}, \mathbf{z})} d\boldsymbol{\theta} \\ &= \mathbb{E}_q \left[\log \int \sum_{\mathbf{z}} \frac{p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w} | \boldsymbol{\alpha}, \boldsymbol{\beta})}{q(\boldsymbol{\theta}, \mathbf{z})} d\boldsymbol{\theta} \right] \\ &\geq \mathbb{E}_q \left[\log p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w} | \boldsymbol{\alpha}, \boldsymbol{\beta}) \right] - \mathbb{E}_q \left[q(\boldsymbol{\theta}, \mathbf{z}) \right] \\ &= \mathcal{L}(q, \boldsymbol{\alpha}, \boldsymbol{\beta}) \end{aligned}$$

Variational EM

$$\text{E-step :} \qquad \max_q \mathcal{L} \Leftrightarrow \min_q \mathrm{KL}\left(q(\boldsymbol{\theta}, \mathbf{z}||p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta})\right)$$

$$\text{Variational distribution :} \qquad q(\boldsymbol{\theta}, \mathbf{z}|\gamma, \phi) = q(\boldsymbol{\theta}|\gamma) \prod_{n=1}^{N} q(z_n|\phi_n)$$

$$\text{New E-step :} \qquad (\gamma^*, \phi^*) = \arg\min_{\gamma,\phi} \mathrm{KL}(q(\boldsymbol{\theta}, \mathbf{z}|\gamma, \phi) \mid\mid p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}))$$

- Overall algorithm :

$$\text{E-step :} \qquad \phi_{n,k} \propto \boldsymbol{\beta}_{k,w_n} \exp\{\Psi(\gamma_k)\}$$

$$\boldsymbol{\gamma}_k = \boldsymbol{\alpha}_k + \sum_{n=1}^{N} \phi_{n,k}$$

$$\text{M-step :} \qquad \boldsymbol{\beta}_{k,j} \propto \sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{d,n,k} w_{d,n}^{(j)}$$

Gibbs Sampling

Collapsed (or *Rao-Blackwellized*) Gibbs sampling :

- $\boldsymbol{\theta}_d$ and $\boldsymbol{\beta}_k$ can be computed using only $z_{d,n}$.
- We sample $z_{d,n}$ from posterior $p(z_n|\mathbf{z}_{\neg n}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \mathbf{w})$.

Developed posterior in terms of counts :

$$p(z_n = k|\mathbf{z}_{\neg n}, \mathbf{w}) \propto \frac{n_{\neg n,k}^{(w_n)} + \boldsymbol{\eta}_n}{\sum_{n'=1}^{V}\left(n_{\neg n,k}^{(w_{n'})} + \boldsymbol{\eta}_{n'}\right)} \frac{n_{\neg n,k}^{(\mathbf{w}_n)} + \boldsymbol{\alpha}_k}{\sum_{k'=1}^{(K)}\left(n_{\neg n,k'}^{(\mathbf{w}_n)} + \boldsymbol{\alpha}_{k'}\right)}$$

Computation of parameters :

$$\boldsymbol{\beta}_{k,n} = \frac{n_k^{(w_n)} + \boldsymbol{\eta}_t}{\sum_{t=1}^{V} n_k^{(t)} + \boldsymbol{\eta}_t} \qquad \boldsymbol{\theta}_{d,k} = \frac{n_m^{(k)} + \boldsymbol{\alpha}_k}{\sum_{k=1}^{K} n_m^{(k)} + \boldsymbol{\alpha}_k}$$

References

Blei, David M, Ng, Andrew Y, and Jordan, Michael I. *Latent dirichlet allocation*. The Journal of Machine Learning Research, 3:993-1022, 2003.

Heinrich, Gregor. *Parameter estimation for text analysis*. Technical report, vsonix GmbH + University of Leipzig, Leipzig, Germany, 2005.

Darling, William M. *A theoretical and practical implementation tutorial on topic modeling and gibbs sampling*. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies, 2011.