Latent Dirichlet Allocation

Jakub Glinka

March, 2018

Model definition

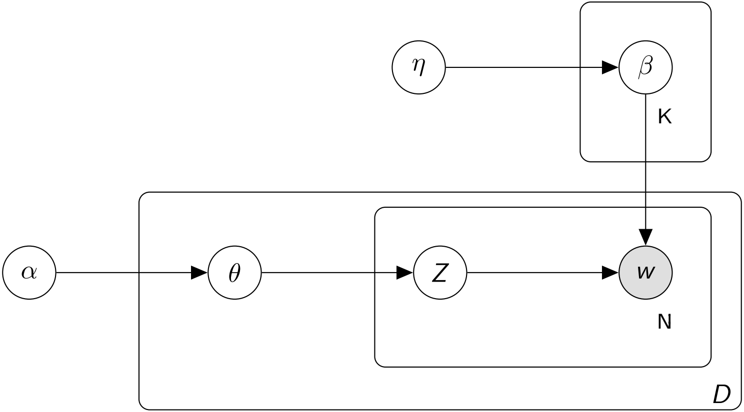
Model definition

- $\theta_{d=1,\dots,M} \sim \mathrm{Dir}_K(\alpha)$
- $\beta_{k=1,\dots,K} \sim \mathrm{Dir}_V(\eta)$
- $z_{d=1,\dots,M,i=1,\dots,N_d} \sim \mathrm{Multinomial}\ _K(\theta_d)$
- $w_{d=1,\dots,M,i=1,\dots,N_d} \sim \mathrm{Multinomial}\ _V(\beta_{z_{di}})$

where

- η is the parameter of the Dirichlet prior on the per-topic word distribution (known and symetric)
- α is the parameter of the Dirichlet prior on the per-document topic distributions (known and symetric)
- θ_d is the topic distribution for document d
- β_k is the word distribution for topic k
- z_{di} is the topic for the i-th word in document d
- w_{di} is a specific word from d-th document belonging to V

Model in plate notation



Model likelihood

$$\log p(z,\theta,\beta,w|\alpha,\eta) = \log p(w|z,\beta)p(z|\theta)p(\theta|\alpha)p(\beta|\eta) =$$

$$= \log\left(\prod_{k=1}^{K} p(\beta_k|\eta) \times \prod_{d=1}^{D} p(\theta_d|\alpha) \times \prod_{d=1}^{D}\prod_{i=1}^{N_d} p(w_{di}|\beta_{z_{di}})p(z_{di}|\theta_d)\right) =$$

$$= \sum_{k=1}^{K} \log p(\beta_k|\eta) + \sum_{d=1}^{D} \log p(\theta_d|\alpha) + \sum_{d=1}^{D}\sum_{i=1}^{N_d} \log p(w_{di}|\beta_{z_{di}})p(z_{di}|\theta_d) =$$

$$= \sum_{k=1}^{K} \log p(\beta_k|\eta) + \sum_{d=1}^{D}\left(\log p(\theta_d|\alpha) + \sum_{i=1}^{N_d} \log p(w_{di}|\beta_{z_{di}})p(z_{di}|\theta_d)\right) =$$

$$= \sum_{d=1}^{D}\left(\log p(\theta_d|\alpha) + \sum_{i=1}^{N_d} \log\left(p(w_{di}|\beta_{z_{di}})p(z_{di}|\theta_d)\right) + \frac{1}{D}\sum_{k=1}^{K} \log p(\beta_k|\eta)\right)$$

Classicial Bayesian Inference

$$p(\theta|x) \propto p(x,\theta) = p(x|\theta)p(\theta)$$

Variational Bayes Inference

Evidence Lower Bound (ELBO)

$$\mathcal{L}(x, \phi) := \mathbb{E}_q \left[\log p(x, \theta)\right] - \mathbb{E}_q \left[\log q(\theta|\phi)\right]$$

where $q(\theta|\phi)$ is a parametric approximation of true posterior [1].

General recipe:

- define flexible (and reasonable) family of posterior distributions
- derive analytical form of $\mathcal{L}(x, \phi)$
- group independent parameters
- find coordinate ascent updates

[1] https://en.wikipedia.org/wiki/Variational_Bayesian_methods

LDA ELBO

From the definition of the **E**vidence **L**ower **B**ound:

$$\mathcal{L}(w, \phi, \gamma, \lambda) := \mathbb{E}_{z,\theta,\beta}\left[\log p(w, z, \theta, \beta | \alpha, \eta)\right] - \mathbb{E}_{z,\theta,\beta}\left[\log q(z, \theta, \beta)\right]$$

with posterior of the form

$$q(z, \beta, \theta) = \prod_d \prod_i q(z_{di}) \times \prod_d q(\theta_d) \times \prod_k q(\beta_k)$$

where [2]

- $q(z_{di}) = p(z_{di}|\phi) = \text{Multinomial}_K(\phi_{dw_{di}})$
- $q(\theta_d) = p(\theta_d|\gamma) = \text{Dir}_K(\gamma_d)$
- $q(\beta_k) = p(\beta_k|\lambda) = \text{Dir}_V(\lambda_k)$

[2] i.e. full factorial design.

20 painfull hours later...*

ELBO coordiate ascent algorithm

Algorithm 1 coordinate ascent VB algorithm for LDA

while improvement $\mathcal{L}(w, \phi, \gamma, \lambda) > \epsilon$ **do**
    while change $\gamma > \epsilon$ **do**
        E-Step:
        for $d \in 1, ..., D$ **do**
            $\phi_{dvk} \propto \exp\left\{\mathbb{E}_\beta\left[\log \beta_{kv}\right] + \mathbb{E}_\theta\left[\log \theta_{dk}\right]\right\}$
            $\gamma_{dk} = \alpha_k + \sum_{v=1}^{V} n_{dv}\phi_{dvk}$
        end for
    end while
    M-Step:
    $\lambda_{kv} = \eta_v + \sum_{d=1}^{D} n_{dv}\phi_{dvk}$
end while

Implementation

Jupyter notebook

References

Andrew Ng et al. (2003) Latent Dirichlet Allocation *Journal of Machine Learning Research* 3 (2003) 993-1022

Matthew Hoffman (2010) Online Learning for Latent Dirichlet Allocation *Advances in Neural Information Processing Systems 23*