Latent Dirichlet Allocation

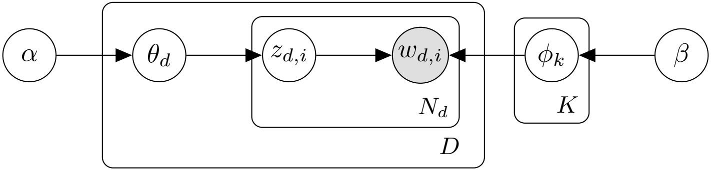
Kangcheng Hou

Zhejiang University

April 29, 2018

Introduction

Here is a PGM of LDA.



We want to sample the posterior distribution of the latent variable $z_{d,i}$ given $w_{d,i}$. One method is gibbs sampling, in every iteration, we sample z_i from distribution $p(z_i|z_{\neg i}, w)$. The joint distribution is

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

Posterior predictive of Dirichlet-Multinomial

Suppose we have dirichlet prior distribution,

$$p(\theta|\alpha) = \mathsf{Dir}(\theta|\alpha) \propto \prod_{k=1}^{K} \theta_i^{\alpha_i - 1}$$

Joint distribution I

The joint distribution is $p(w, z|\alpha, \beta) = p(w|z, \beta)p(z|\alpha)$. The first part is

$$p(w|z, \beta) = \int_{\phi_{1:K}} p(w|z, \phi_{1:K})p(\phi_{1:K}|\beta)d\phi_{1:K}$$

Let's see $p(w|z, \phi_{1:K})$ first,

$$p(w|z, \phi_{1:K}) = \prod_{i=1}^{W} \phi_{z_i, w_i}$$

Or we can rephrase it in another way, where we classify it by topic.

$$p(w|z, \phi_{1:K}) = \prod_{i=1}^{W} \phi_{z_i, w_i} = \prod_{k=1}^{K} \prod_{v=1}^{V} \phi_{k,v}^{n_k^{(v)}}$$

Joint distribution II

And

$$p(\phi_{1:K}|\beta) = \prod_{k=1}^{K} \prod_{v=1}^{V} \phi_{k,v}^{\beta_v - 1}$$

Adding these two together, we have

$$p(w|z,\beta) = \prod_{k=1}^{K} \frac{B(n_k + \beta)}{B(\beta)}$$

n_k represents the word apperance frequencies in topic k. The topic distribution $p(z|\alpha)$ can be derived similarly,

$$p(z|\alpha) = \prod_{d=1}^{D} \frac{B(n_d + \alpha)}{B(\alpha)}$$

Joint distribution III

So the joint distribution is

$$p(z, w|\alpha, \beta) = \prod_{k=1}^{K} \frac{B(n_k + \beta)}{B(\beta)} \prod_{d=1}^{D} \frac{B(n_d + \alpha)}{B(\alpha)}$$

$$p(z_i = k|z_{\neg i}, w, \alpha, \beta) = \frac{p(w, z)|\alpha, \beta}{p(w, z_{\neg i}|\alpha, \beta)} \propto p(w, z|\alpha, \beta)$$

And the multinomial parameters can be derived as follows

$$p(\theta_d|w, z, \alpha) \sim \mathrm{Dir}(\theta_m|n_m + \alpha)$$

$$p(\phi_d|w, z, \beta) \sim \mathrm{Dir}(\phi_d|n_d + \beta)$$