Supplemental Material of "BTM: Topic Modeling over Short Texts"

1 Derivation of $P(z_i|\mathbf{z}_{-i},\mathbf{B})$ in Gibbs sampling

Using the chain rule, the conditional distribution can be rewritten as:

$$P(z_i|\mathbf{z}_{-i},\mathbf{B}) = \frac{P(\mathbf{z},\mathbf{B})}{P(\mathbf{z}_{-i},\mathbf{B})} \propto \frac{P(\mathbf{B}|\mathbf{z})P(\mathbf{z})}{P(\mathbf{B}_{-i}|\mathbf{z}_{-i})P(\mathbf{z}_{-i})}. \tag{1}$$

In Eq.(1), $P(\mathbf{B}|\mathbf{z})$ can be obtained by integrating out $\mathbf{\Phi}$:

$$\begin{aligned} P(\mathbf{B}|\mathbf{z}) &= \int P(\mathbf{B}|\mathbf{z},\mathbf{\Phi})P(\mathbf{\Phi})\mathrm{d}\mathbf{\Phi} \\ &= \int \left(\prod_{i=1}^{N_B} P(b_i|z_i,\boldsymbol{\phi}_{z_i})\right) P(\mathbf{\Phi})\mathrm{d}\mathbf{\Phi} \\ &= \int \prod_{k=1}^{K} \left(\frac{\Gamma(W\beta)}{\Gamma(\beta)^W} \prod_{w=1}^{W} \phi_{k,w}^{n_{w|k}+\beta-1} \mathrm{d}\boldsymbol{\phi_k}\right) \\ &= \left(\frac{\Gamma(W\beta)}{\Gamma(\beta)^W}\right)^K \prod_{k=1}^{K} \frac{\prod_{w=1}^{W}\Gamma(n_{w|k}+\beta)}{\Gamma(n_{\cdot|k}+W\beta)}, \end{aligned} \tag{2}$$

where $\Gamma(\cdot)$ is the standard Gamma function[1], $n_{w|k}$ is the number of times word w assigned to topic k, and $n_{\cdot|k} = \sum_{w=1}^{W} n_{w|k}$. $P(\mathbf{z})$ can be obtained by integrating out $\boldsymbol{\theta}$:

$$\begin{aligned} P(\mathbf{z}) &= \int P(\mathbf{z}|\boldsymbol{\theta})P(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \\ &= \int \left(\prod_{i=1}^{N_B} P(z_i|\boldsymbol{\theta})\right) P(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \\ &= \int \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \prod_{k=1}^{K} \theta_k^{n_k+\alpha-1}\mathrm{d}\boldsymbol{\theta} \\ &= \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \frac{\prod_k \Gamma(n_k+\alpha)}{\Gamma(N_B+K\alpha)}. \end{aligned} \tag{3}$$

and $P(\mathbf{B}_{-i}|\mathbf{z}_{-i})$, $P(\mathbf{z}_{-i})$ can be worked out in the same way:

$$P(\mathbf{B}_{-i}|\mathbf{z}_{-i}) = \left(\frac{\Gamma(W\beta)}{\Gamma(\beta)^W}\right)^K \prod_{k=1}^{K} \frac{\prod_{w=1}^{W}\Gamma(n_{-i,w|k}+\beta)}{\Gamma(n_{-i,\cdot|k}+W\beta)}, \tag{4}$$

$$P(\mathbf{z}_{-i}) = \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \frac{\prod_{k=1}^{K}\Gamma(n_{-i,k}+\alpha)}{\Gamma(N_B-1+K\alpha)}, \tag{5}$$

[1] Please refer to http://en.wikipedia.org/wiki/Gamma_function. Particularly, when x is a positive integer, the Gamma function is defined as $\Gamma(x) = (x-1)!$

where $n_{-i,\cdot}$ is the count that does exclude biterm b_i. Considering that the Gamma function satisfies $\Gamma(x+1) = x\Gamma(x)$, and $n_{\cdot|k} = n_{-i,\cdot|k} + 2$, thus we have

$$\Gamma(n_{\cdot|k} + W\beta) = (n_{-i,\cdot|k} + W\beta + 1)(n_{-i,\cdot|k} + W\beta)\Gamma(n_{-i,\cdot|k} + W\beta). \tag{6}$$

By replacing terms in Eq.(1) with those in Eqs.(2-6), we obtain the final conditional distribution:

$$P(z_i = k|\mathbf{z}_{-i}, \mathbf{B}) \propto (n_{-i,k} + \alpha)\frac{(n_{-i,w_{i,1}|k} + \beta)(n_{-i,w_{i,2}|k} + \beta)}{(n_{-i,\cdot|k} + W\beta + 1)(n_{-i,\cdot|k} + W\beta)}.$$

Acknowledgement: In the published version of both our TKDE'14 and WWW'13 papers, $(n_{-i,\cdot|k} + W\beta + 1)$ in the denominator in right hand was mistakenly written as $(n_{-i,\cdot|k} + W\beta)$. Thanks to Konishi Takuya point out it.

2 Derivation of the estimation of $\phi_{k,w}$ and θ_k in Gibbs sampling

Given the hyperparamters α and β, biterm set $\mathbf{B}$ and their topic assignments $\mathbf{z}$, we can derive the probability of the parameters $\mathbf{\Phi}$ and $\boldsymbol{\theta}$ by utilizing the Bayes' rule and Dirichlet-multinomial conjugate property:

$$P(\boldsymbol{\theta}|\mathbf{z}, \alpha) = \frac{1}{Z_\theta}\prod_{i=1}^{N_B} P(z_i|\boldsymbol{\theta})P(\boldsymbol{\theta}|\alpha) = \mathrm{Dir}(\boldsymbol{\theta}|\alpha + \mathbf{n}), \tag{7}$$

$$P(\boldsymbol{\phi}_k|\mathbf{z}, \mathbf{B}, \beta) = \frac{1}{Z_{\boldsymbol{\phi}_k}}\prod_{i=1}^{N_B} P(z_i|\boldsymbol{\theta})P(\boldsymbol{\theta}|\alpha) = \mathrm{Dir}(\boldsymbol{\phi}_k|\beta + \mathbf{n_k}), \tag{8}$$

where vector $\mathbf{n} = \{n_k\}_{k=1}^K$, vector $\mathbf{n_k} = \{n_{w|k}\}_{w=1}^W$, $Z_{\boldsymbol{\theta}}$ and $Z_{\boldsymbol{\phi}_k}$ are normalization factors, $\mathrm{Dir}(\cdot)$ denotes the probability density function of a Dirichlet distribution.

Note that the expectation of the Dirichlet distribution $\mathrm{Dir}(\mathbf{x}|\boldsymbol{\alpha})$ is $\mathrm{E}(x_i) = \frac{\alpha_i}{\sum_{k=1}^K \alpha_k}$. Based on Eqs.(7-8), we estimate $\phi_{k,w}$ and θ_k using their expectations:

$$\phi_{k,w} = \frac{n_{w|k} + \beta}{n_{\cdot|k} + W\beta},$$

$$\theta_k = \frac{n_k + \alpha}{N_B + K\alpha}.$$