Supplemental Material of “BTM: Topic Modeling over Short Texts”

1 Derivation of \( P(z_i|z_{-i}, B) \) in Gibbs sampling

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

\[
P(z_i|z_{-i}, B) = \frac{P(z, B)}{P(z_{-i}, B)} \propto \frac{P(B|z)P(z)}{P(B_{-i}|z_{-i})P(z_{-i})}.
\]

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

\[
P(B|z) = \int P(B, \Phi|z)P(\Phi)d\Phi
\]

\[
= \int \left( \prod_{i=1}^{N_B} P(b_i|z_i, \phi_{z_i}) \right) P(\Phi)d\Phi
\]

\[
= \int \prod_{k=1}^{K} \frac{\Gamma(W\beta)}{\Gamma(\beta)^W} \frac{1}{\prod_{w=1}^{W} \phi_{k,w}}^{n_{w|k}+\beta-1}d\phi_k
\]

\[
= \frac{\left( \frac{\Gamma(W\beta)}{\Gamma(\beta)^W} \right)^K}{K} \prod_{k=1}^{K} \frac{\prod_{w=1}^{W} \Gamma(n_{w|k} + \beta)}{\Gamma(n_{w|k} + W\beta)},
\]

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

\[
P(z) = \int P(z|\Theta)P(\Theta)d\Theta
\]

\[
= \int \left( \prod_{i=1}^{N_B} P(z_i|\Theta) \right) P(\Theta)d\Theta
\]

\[
= \int \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \prod_{k=1}^{K} \phi_{k}^{n_{.|k} + a - 1}d\Theta
\]

\[
= \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \prod_{k=1}^{K} \frac{\Gamma(n_{.|k} + \alpha)}{\Gamma(n_{.|k} + W\beta)},
\]

and \( P(B_{-i}|z_{-i}), P(z_{-i}) \) can be worked out in the same way:

\[
P(B_{-i}|z_{-i}) = \left( \frac{\Gamma(W\beta)}{\Gamma(\beta)^W} \right)^K \prod_{k=1}^{K} \frac{\prod_{w=1}^{W} \Gamma(n_{-i,w|k} + \beta)}{\Gamma(n_{-i,.|k} + W\beta)},
\]

\[
P(z_{-i}) = \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \prod_{k=1}^{K} \frac{\Gamma(n_{-i,.|k} + \alpha)}{\Gamma(N_B - 1 + K\alpha)}.
\]

\(^{1}\)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 $n_{-i,}$ is the count that does exclude biterm $b_i$. Considering that the Gamma function satisfies
\[
\Gamma(n_{i,k} + W\beta) = (n_{i,-1} + W\beta + 1)(n_{-1,i,k} + W\beta)\Gamma(n_{i,k} + W\beta).
\]
(6)

By replacing terms in Eq.(1) with those in Eqs.(2-6), we obtain the final conditional distribution:
\[
P(z_i = k | z_{-i}, B) \propto (n_{-i,k} + \alpha) \prod_{i=1}^{N_B} P(z_i | \theta) P(\theta | \alpha) = \text{Dir}(\theta | \alpha + n),
\]
(7)
\[
P(\phi_k | z, B, \beta) = \frac{1}{Z_{\phi_k}} \prod_{i=1}^{N_B} P(z_i | \theta) P(\theta | \alpha) = \text{Dir}(\phi_k | \beta + n_k),
\]
(8)

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

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

Given the hyperparameters $\alpha$ and $\beta$, biterm set $B$ and their topic assignments $z$, we can derive the probability of the parameters $\Phi$ and $\Theta$ by utilizing the Bayes’ rule and Dirichlet-multinomial conjugate property:
\[
P(\theta | z, \alpha) = \frac{1}{Z_{\theta}} \prod_{i=1}^{N_B} P(z_i | \theta) P(\theta | \alpha) = \text{Dir}(\theta | \alpha + n),
\]
(7)
\[
P(\phi_k | z, B, \beta) = \frac{1}{Z_{\phi_k}} \prod_{i=1}^{N_B} P(z_i | \theta) P(\theta | \alpha) = \text{Dir}(\phi_k | \beta + n_k),
\]
(8)

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

Note that the expectation of the Dirichlet distribution Dir($x|\alpha$) is $E(x_i) = \frac{\alpha_i}{\sum_{k=1}^K \alpha_k}$. Based on Eqs.(7-8), we estimate $\phi_{k,w}$ and $\theta_k$ using their expectations:
\[
\phi_{k,w} = \frac{n_{w,k} + \beta}{n_{-,k} + W\beta},
\]
\[
\theta_k = \frac{n_k + \alpha}{N_B + \alpha K}.
\]