Expectation-Propagation for Latent Dirichlet Allocation

Tomonari MASADA @ Nagasaki University

December 10, 2012

1

This manuscript contains a derivation of the update formulae of expectation-propagation (EP) presented in the following paper:

Thomas Minka and John Lafferty.
Expectation-Propagation for the Generative Aspect Model.
in Proc. of the 18th Conference on Uncertainty in Artificial Intelligence, pp. 352-359, 2002.¹

2

A joint distribution for document \(d\) in LDA can be written as below.

\[
p(x, \theta_d|\alpha) = p(\theta_d|\alpha)p(x|\theta_d)
\]

\[
= p(\theta_d|\alpha) \prod_{i=1}^{n_d} p(x_i|\theta_d) = p(\theta_d|\alpha) \prod_{i=1}^{n_d} \left( \sum_k \theta_{dk} \phi_{kx_i} \right)
\]

\[
= p(\theta_d|\alpha) \prod_w \left( \sum_k \theta_{dk} \phi_{kw} \right)^{n_{dw}},
\]

(1)

where \(p(\theta_d|\alpha)\) is a Dirichlet prior distribution.

We approximate \(p(w|\theta_d) = \sum_k \theta_{dk} \phi_{kw}\) by \(t_w(\theta_d) = s_w \prod_k \theta_{dk}^{\beta_{wk}}\).

Then we obtain an approximated joint distribution as follows:

\[
p(x, \theta_d|\alpha) \approx p(\theta_d|\alpha) \prod_w t_w(\theta_d)^{n_{dw}} = p(\theta_d|\alpha) \prod_w \left( s_w \prod_k \theta_{dk}^{\beta_{wk}} \right)^{n_{dw}}
\]

\[
= \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_w s_w \prod_k \theta_{dk}^{\gamma_{dk}-1+\sum_w \beta_{wk} n_{dw}}
\]

(2)

Let \(\gamma_{dk} = \alpha_k + \sum_w \beta_{wk} n_{dw}\). That is,

\[
p(x_d, \theta_d|\alpha) \approx \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_w s_w \prod_k \theta_{dk}^{\gamma_{dk}-1}.
\]

(3)

An approximated posterior \(q(\theta_d)\) can be obtained as follows:

\[
q(\theta_d) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_w s_w \prod_k \theta_{dk}^{\gamma_{dk}-1}
\]

\[
= \frac{\Gamma(\sum_k \gamma_{dk})}{\prod_k \Gamma(\gamma_{dk})} \prod_k \theta_{dk}^{\gamma_{dk}-1}.
\]

(4)

For each word token in document $d$, we remove its contribution to $\prod_w t_w(\theta_d)^{n_{dw}} = \prod_w (s_w \prod_k \theta^\beta_{dk})^{n_{dw}}$.

When the token is a token of word $w$, this corresponds to a division by $s_w \prod_k \theta^\beta_{dk}$.

Note that we here consider a removal of word token, not a removal of word type.

Then we obtain an ‘old’ posterior after this division as follows:

$$q_{\gamma_{cl_w}}(\theta_d) = \frac{\Gamma(\sum_k (\gamma_{dk} - \beta_{wk}))}{\prod_k \Gamma(\gamma_{dk} - \beta_{wk})} \prod_k \theta^\gamma_{dk} - \beta_{wk} - 1 = \frac{\Gamma(\sum_k \gamma_{cl_w})}{\prod_k \Gamma(\gamma_{cl_w})} \prod_k \theta^\gamma_{cl_w} - 1,$$

where $\gamma_{cl_w} = \gamma_{dk} - \beta_{wk}$.

By combining $p(w|\theta_d) = \sum_k \theta_{dk} \Phi_{kw}$ with $q_{\gamma_{cl_w}}(\theta_d)$, we obtain something similar to joint distribution as follows:

$$\langle \sum_k \theta_{dk} \Phi_{kw} \rangle \cdot \frac{\Gamma(\sum_k \gamma_{cl_w})}{\prod_k \Gamma(\gamma_{cl_w})} \prod_k \theta^\gamma_{cl_w} - 1.$$

Therefore, by integrating $\theta_d$ out, we obtain something similar to evidence as follows:

$$\int \sum_k \theta_{dk} \Phi_{kw} \frac{\Gamma(\sum_k \gamma_{cl_w})}{\prod_k \Gamma(\gamma_{cl_w})} \prod_k \theta^\gamma_{cl_w} - 1 d\theta_d = \frac{\sum_k \phi_{kw} \gamma_{cl_w}}{\sum_k \Phi_{kw} \gamma_{cl_w}} \cdot \frac{\Gamma(\sum_k \gamma_{cl_w})}{\prod_k \Gamma(\gamma_{cl_w})} \prod_k \theta^\gamma_{cl_w} - 1.$$

Consequently, we obtain a new posterior as follows:

$$\hat{q}(\theta_d) = \frac{\sum_k \phi_{kw} \gamma_{cl_w}}{\sum_k \phi_{kw} \gamma_{cl_w}} \cdot \frac{\Gamma(\sum_k \gamma_{cl_w})}{\prod_k \Gamma(\gamma_{cl_w})} \prod_k \theta^\gamma_{cl_w} - 1.$$

### 4

Based on Eq. (6), we calculate $E_q[\theta_{dk}]$ and $E_q^2[\theta_{dk}]$.

$$E_q[\theta_{dk}] = \int \theta_{dk} \hat{q}(\theta_d) d\theta_d$$

$$= \frac{\sum_k \phi_{kw} \gamma_{cl_w}}{\sum_k \phi_{kw} \gamma_{cl_w}} \cdot \frac{\Gamma(\sum_k \gamma_{cl_w})}{\prod_k \Gamma(\gamma_{cl_w})} \prod_k \theta^\gamma_{cl_w} - 1 d\theta_d$$

$$= \frac{\gamma_{cl_w}}{\sum_k \phi_{kw} \gamma_{cl_w}} \cdot \frac{\gamma_{cl_w}}{\gamma_{cl_w} + 1} \cdot \frac{\gamma_{cl_w}}{\gamma_{cl_w} + 1}$$

$$= \frac{\gamma_{cl_w}}{\sum_k \phi_{kw} \gamma_{cl_w}} \cdot \frac{\gamma_{cl_w}}{\gamma_{cl_w} + 1} \cdot \frac{\gamma_{cl_w}}{\gamma_{cl_w} + 1}$$

$$E_q^2[\theta_{dk}] = \int \theta_{dk} \hat{q}(\theta_d) d\theta_d$$

$$= \frac{\sum_k \phi_{kw} \gamma_{cl_w}}{\sum_k \phi_{kw} \gamma_{cl_w}} \cdot \frac{\Gamma(\sum_k \gamma_{cl_w})}{\prod_k \Gamma(\gamma_{cl_w})} \prod_k \theta^\gamma_{cl_w} - 1 d\theta_d$$

$$= \frac{\gamma_{cl_w}}{\sum_k \phi_{kw} \gamma_{cl_w}} \cdot \frac{\gamma_{cl_w}}{\gamma_{cl_w} + 1} \cdot \frac{\gamma_{cl_w}}{\gamma_{cl_w} + 1}$$

$$= \frac{\gamma_{cl_w}}{\sum_k \phi_{kw} \gamma_{cl_w}} \cdot \frac{\gamma_{cl_w} + 1}{\gamma_{cl_w} + 1} \cdot \frac{\gamma_{cl_w} + 1}{\gamma_{cl_w} + 1}$$

where $\gamma_{cl_w} = \sum_k \gamma_{cl_w}$.

We determine a Dirichlet distribution $\text{Dir}(\gamma_d)$ so that the mean $\gamma_d$ and the average second moment$^2$

$$\frac{1}{k} \sum_k \frac{\gamma_{cl_w}(\gamma_{cl_w} + 1)}{\gamma_{cl_w}},$$

are equal to those of $\hat{q}(\theta_d)$, where $\gamma_d = \sum_k \gamma_d$.  

---

$^2$http://www.ucl.ac.uk/statistics/research/pdfs/135.zip
To be precise, we obtain $\gamma'_d$ by solving the following equations:

$$\frac{\gamma'_d}{\gamma'_d} = E_q[\theta_{dk}] \text{ for each } k, \text{ and}$$

$$\frac{1}{K} \sum_k \frac{\gamma'_d}{\gamma'_d} = \frac{1}{K} \sum_k E_q[\theta_{dk}^2].$$

We can obtain $\gamma'_d$ as below.

$$\sum_k E_q[\theta_{dk}]\frac{\gamma'_d}{\gamma'_d} = \sum_k E_q[\theta_{dk}^2]$$

$$\sum_k E_q[\theta_{dk}] (E_q[\theta_{dk}] \frac{\gamma'_d}{\gamma'_d} + 1) = \sum_k E_q[\theta_{dk}^2] (\gamma'_d + 1)$$

$$\sum_k E_q[\theta_{dk}]^2 \gamma'_d + \sum_k E_q[\theta_{dk}] = \sum_k E_q[\theta_{dk}^2] \gamma'_d + \sum_k E_q[\theta_{dk}^2]$$

$$\sum_k (E_q[\theta_{dk}]^2 - E_q[\theta_{dk}^2]) \gamma'_d = \sum_k (E_q[\theta_{dk}^2] - E_q[\theta_{dk}^2])$$

$$\gamma'_d = \frac{\sum_k (E_q[\theta_{dk}^2] - E_q[\theta_{dk}^2])}{\sum_k (E_q[\theta_{dk}^2] - E_q[\theta_{dk}^2])}, \text{ and}$$

$$\gamma'_d = \frac{\sum_k (E_q[\theta_{dk}^2] - E_q[\theta_{dk}^2])}{\sum_k (E_q[\theta_{dk}^2] - E_q[\theta_{dk}^2])} \cdot E_q[\theta_{dk}^2].$$

5

We now have a Dirichlet distribution that approximates the posterior $\tilde{q}(\theta_d)$ in Eq. (6), i.e.,

$$\tilde{q}(\theta_d) = \frac{\sum_k \gamma_{dk}^w (\sum_k \theta_{dk} \phi_{kw})}{\sum_k \phi_{kw} \gamma_{dk}^w} \left( \frac{\Gamma(\sum_k \gamma_{dk}^w)}{\prod_k \Gamma(\gamma_{dk}^w)} \prod_k \theta_{dk}^{\gamma_{dk}^w - 1} \right).$$

To obtain a density function of Dirichlet distribution from Eq. (6), we replace the term $\left( \sum_k \theta_{dk} \phi_{kw} \gamma_{dk}^w \right)$ by $\theta_w' = \theta_{w'} \prod_k \theta_{dk}^{\gamma_{dk}^w}$. We set the resulting function equal to the density function of $\text{Dir}(\gamma'_d)$ and obtain the following equation:

$$\frac{\sum_k \gamma_{dk}^w}{\sum_k \phi_{kw} \gamma_{dk}^w} \theta_w' \prod_k \theta_{dk}^{\gamma_{dk}^w} = \Gamma(\sum_k \gamma_{dk}^w) \prod_k \theta_{dk}^{\gamma_{dk}^w}.$$ 

Consequently, we can estimate $\theta'_w$ and $\beta'_{w,k}$ as follows:

$$\beta'_{w,k} = \gamma_{dk} - \gamma_{dk}^w,$$

$$\gamma_{dk}^w = \frac{\sum_k \phi_{kw} \gamma_{dk}^w}{\sum_k \gamma_{dk}^w}.$$