Graphical Model

Politeness group

December 5, 2014

Let N sentences be given containing words $w_{n1}, \ldots, w_{nM_n}$ for $n = 1, \ldots, N$. Each word w_{nm} is generated from a distribution of words ϕ_z, with the possibilities background (neutral), polite or impolite. These ϕ_z are Dirichlet distributions with hyperparameter β. In order to see which distribution a word is generated from, we attach a (latent) variable z_{nm} to each word. Furthermore, a sentence as a whole can be either polite or impolite, which we store in the (observed) variable y_n.

We make the assumption that polite sentences only contain words that were generated from the polite distribution or the background distribution, where the distribution over both distributions is specified by μ_{polite}. Similarly, $\mu_{impolite}$ is used for impolite sentences to give the distribution over impolite words and backgrounds words. Both μ_i are Beta distributions with hyperparameter α_i. The generation of a topic z_{nm} for word w_{nm} is sampled from some μ_{y_n}, where y_n picks either the polite or the impolite distribution.

In order to visualize which words are polite, which are impolite and which are background, we want to sample the topics z_i for the words w_i of a new sentence. For this purpose, the sentence first needs to be classified as either polite or impolite (which gives the value for y) and then we need to sample from $p(z_i|W, y, \Phi, z_{-i}, \mu)$, where W consists of all words in the sentence, z_{-i} are the other topics and $\Phi = \phi_1, \phi_2, \phi_3 = \phi_{background}, \phi_{polite}, \phi_{impolite}$.

We see that

$$p(Z, W, Y, \mu, \Phi|\alpha, \beta) = p(\mu|\alpha) \prod_{i=1}^{3} p(\phi_i|\beta) \prod_{n=1}^{N} \prod_{m=1}^{M} p(w_{nm}|\phi_{z_{nm}}) p(z_{nm}|\mu, y_n),$$

hence

$$p(Z, W|\alpha, \beta) = \int p(Z, W, Y, \mu, \Phi|\alpha, \beta) d\Phi d\mu =$$

$$\left(\int p(\mu|\alpha) \prod_{n=1}^{N} \prod_{m=1}^{M} p(z_{nm}|\mu, y_n) d\mu \right) \left(\int \prod_{i=1}^{3} p(\phi_i|\beta) \prod_{n=1}^{N} \prod_{m=1}^{M} p(w_{nm}|\phi_{z_{nm}}) d\Phi \right).$$

Both parts will be worked out separately.

$$\int \prod_{i=1}^{2} p(\mu_i|\alpha) \prod_{n=1}^{N} \prod_{m=1}^{M} p(z_{nm}|\mu_{y_n}) d\mu =$$

$$\int \prod_{i=1}^{2} \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \mu_i^{\alpha_1 - 1} (1 - \mu_i)^{\alpha_2 - 1} \prod_{i=1}^{2} \prod_{\{n: y_n = i\}} \prod_{m=1}^{M} p(z_{nm}|\mu_i) d\mu =$$

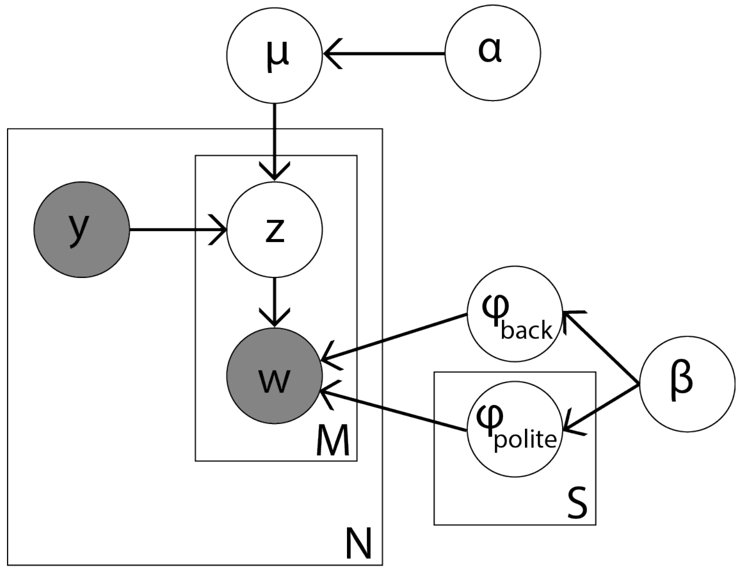

Figure 1: Graphical model for our topic model.

$$\prod_{i=1}^{2} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int \mu_i^{\alpha_1-1}(1-\mu_i)^{\alpha_2-1} \prod_{\{n:y_n=i\}} \mu_i^{C(0,n)}(1-\mu_i)^{C(1,n)} d\mu_i$$

where $C(0,n), C(1,n)$ denotes the number of background or (im)polite words in sentence n respectively. Let $C(1), C(2)$ denote the number of sentences that are polite or impolite respectively, then we can rewrite the expression above as:

$$\prod_{i=1}^{2} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int \mu_i^{\alpha_1-1}(1-\mu_i)^{\alpha_2-1} \mu_i^{C(0,n)C(i)}(1-\mu_i)^{C(1,n)C(i)} d\mu_i =$$

$$\prod_{i=1}^{2} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int \mu_i^{\alpha_1-1+C(0,n)C(i)}(1-\mu_i)^{\alpha_2-1+C(1,n)C(i)} d\mu_i =$$

$$\prod_{i=1}^{2} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{\Gamma(\alpha_1+C(0,n)C(i))\Gamma(\alpha_2+C(1,n)C(i))}{\Gamma(\alpha_1+\alpha_2+(C(0,n)+C(1,n))C(i))} \propto$$

$$\prod_{i=1}^{2} \frac{\Gamma(\alpha_1+C(0,n)C(i))\Gamma(\alpha_2+C(1,n)C(i))}{\Gamma(\alpha_1+\alpha_2+M_nC(i))}.$$

Before rewriting the second integral, note that we assume a fixed vocabulary of some size V_i of possible words that the ϕ_i can generate.

$$\int \prod_{i=1}^{3} p(\phi_i|\beta) \prod_{n=1}^{N} \prod_{m=1}^{M} p(w_{nm}|\phi_{z_{nm}}) d\Phi \propto$$

$$\int \prod_{i=1}^{3} \prod_{w=1}^{V_i} \phi_i(w)^{\beta-1} \prod_{i=1}^{3} \prod_{w=1}^{V} \phi_i(w)^{C_i(w)} d\Phi =$$

$$\prod_{i=1}^{3}\int\prod_{w=1}^{V_i}\phi_i(w)^{\beta-1+C_i(w)}d\phi_i=\prod_{i=1}^{3}\frac{\prod_{w=1}^{V_i}\Gamma(\beta+C_i(w))}{\Gamma(\sum_{w=1}^{V_i}\beta+C_i(w))}$$

for $C_i(w)$ the number of (n,m) such that $w_{nm}=w$ and $z_{nm}=i$.

Combining the equations above gives

$$p(Z,W,Y,\mu,\Phi|\alpha,\beta)=\prod_{i=1}^{2}\frac{\Gamma(\alpha_1+C(0,n)C(i))\Gamma(\alpha_2+C(1,n)C(i))}{\Gamma(\alpha_1+\alpha_2+M_nC(i))}\prod_{i=1}^{3}\frac{\prod_{w=1}^{V_i}\Gamma(\beta+C_i(w))}{\Gamma(\sum_{w=1}^{V_i}\beta+C_i(w))}$$

such that

$$p(z_{nm}=i|Z_{-nm},W_{-nm},\Phi,\mu)=\frac{p(z_{nm}=i,Z_{-nm},W,Y,\mu,\Phi|\alpha,\beta)}{p(Z_{-nm},W_{-nm},Y,\mu,\Phi|\alpha,\beta)}\propto$$

$$\prod_{j=1}^{2}\frac{[\Gamma(\alpha_1+(\tilde{C}(0,n)+\delta_{i,\text{back}})\tilde{C}(j))\Gamma(\alpha_2+(\tilde{C}(1,n)+\delta_{i,\text{(im)polite}})\tilde{C}(j))][\Gamma(\alpha_1+\alpha_2+M_n\tilde{C}(j))]}{[\Gamma(\alpha_1+\alpha_2+M_n\tilde{C}(j))][\Gamma(\alpha_1+\tilde{C}(0,n)\tilde{C}(j))\Gamma(\alpha_2+\tilde{C}(1,n)\tilde{C}(j))]}\times$$

$$\frac{\Gamma(\beta+\tilde{C}_i(w_{nm})+1)][\Gamma(\sum_{w=1}^{V_i}\beta+\tilde{C}_i(w)]}{[\Gamma(\sum_{w=1}^{V_i}\beta+\tilde{C}_i(w)+\delta_{w,w_{nm}})][\Gamma(\beta+\tilde{C}_i(w_{nm}))]}$$

for $\tilde{C}$ the counts on (W_{-nm},Z_{-nm}), the dataset with (w_{nm},z_{nm}) removed. The equality $(x+n-1)\cdots(x+1)x=\frac{\Gamma(x+n)}{\Gamma(x)}$ and some simplifications give

$$\prod_{j\in\{\text{polite,impolite}\}}\prod_{k=0}^{\tilde{C}(j)-1}[\alpha_{\delta_{i,\text{(im)polite}},n)}+\tilde{C}(\delta_{i,\text{(im)polite}},n)\tilde{C}(j))+k)]\times\frac{\beta+\tilde{C}_i(w_{nm})}{V_i\beta+\sum_{w=1}^{V_i}\tilde{C}_i(w)}=$$

$$\prod_{j\in\{\text{polite,impolite}\}}\prod_{k=0}^{C(j)-1}[\alpha_{\delta_{i,\text{(im)polite}},n)}+(C(\delta_{i,\text{(im)polite}},n)-1)C(j))+k)]\times\frac{\beta+C_i(w_{nm})-1}{-1+V_i\beta+\sum_{w=1}^{V_i}C_i(w)}.$$

Recall that $C_i(w)$ is the number of (a,b) such that $w_{ab}=w$ and $z_{ab}=i$, $C(0,n),C(1,n)$ denotes the number of background respectively (im)polite words in sentence n and $C(1),C(2)$ denote the number of sentences that are polite respectively impolite. Furthermore, V_i is the total number of unique words with class (polite, impolite, background) i, i.e. the vocabulary for the distribution ϕ_i.