LJST: Appendix

Anonymous CoNLL submission

0.1 Model Inference for LJST

The total probability of the model based words, topics and sentiment labels can be decomposed as follows:

$$p(\mathbf{w},\mathbf{z},\mathbf{l},\varphi,\pi,\theta,\boldsymbol{\alpha},\beta,\boldsymbol{\gamma}) = \prod_{j=1}^{T}\prod_{k=1}^{S} p(\varphi_{j,k};\beta)\prod_{d=1}^{D} p(\pi_{d,j};\boldsymbol{\gamma})p(\theta_d;\boldsymbol{\alpha}) \prod_{i=1}^{N_d} p(z_{d,i}|\theta_d)p(l_{d,j,i}|\pi_{d,j})p(w_{d,j,k,i}|\varphi_{z_{d,i},l_{d,j,i}}) \tag{1}$$

In order to use *collapsed Gibbs sampling*, we integrate out φ, π and θ.

$$\begin{aligned}p(\mathbf{w},\mathbf{z},\mathbf{l},\boldsymbol{\alpha},\beta,\boldsymbol{\gamma}) &= \int_{\varphi}\int_{\pi}\int_{\theta} p(\mathbf{w},\mathbf{z},\mathbf{l},\varphi,\pi,\theta,\boldsymbol{\alpha},\beta,\boldsymbol{\gamma})\,d\varphi\,d\pi\,d\theta = \\ &\int_{\varphi}\prod_{j}\prod_{k} p(\varphi_{j,k};\beta)\prod_{d}\prod_{i} p(w_{d,j,k,i}|\varphi_{z_{d,i},l_{d,j,i}})\,d\varphi \\ &\int_{\pi}\prod_{d}\prod_{j} p(\pi_{d,j};\boldsymbol{\gamma})\prod_{i} p(l_{d,j,i}|\pi_{d,j})\,d\pi \\ &\int_{\theta}\prod_{d}\prod_{i} p(z_{d,i}|\theta_d)p(\theta_d;\boldsymbol{\alpha})\,d\theta\end{aligned} \tag{2}$$

As, φ, π and θ are independent variables, we can split the 3 terms in Eq. 2 and calculate them separately. Further, each individual terms $\varphi_{j,k}$, $\pi_{d,j}$ and θ_d are independent. Thus, we can interchange the product and integration in each of the 3 terms.
For each document d, θ_d, we replace the term $p(\theta_d;\boldsymbol{\alpha})$ with corresponding Dirichlet distribution and $p(z_{d,i}|\theta_d)$ with Multinomial distribution to get:

$$\int_{\theta_d} p(\theta_d;\boldsymbol{\alpha})\prod_{i} p(z_{d,i}|\theta_d)\,d\theta_d = \int_{\theta_d}\left(\frac{\Gamma(\sum_{j=1}^{T}\alpha_j)}{\prod_{j=1}^{T}\Gamma(\alpha_j)}\right)\theta_{d,j}^{N_{d,j}}\prod_{j}\theta_{d,j}^{\alpha_j-1}\,d\theta_d. \tag{3}$$

Using the property of Dirichlet distribution we get:

$$\int_{\theta_d}\left(\frac{\Gamma(\sum_{j=1}^{T}\alpha_j+\sum_{j=1}^{T}N_{d,j})}{\prod_{j=1}^{T}\Gamma(\alpha_j+N_{d,j})}\right)\prod_{j}\theta_{d,j}^{N_{d,j}+\alpha_j-1}\,d\theta_d = 1.$$

This leads us to :

$$p(z) = \prod_{d} p(z_d) = \prod_{d}\int_{\theta_d} p(\theta_d;\boldsymbol{\alpha})\prod_{i} p(z_{d,i}|\theta_d)\,d\theta_d = \left(\frac{\Gamma(\sum_{j=1}^{T}\alpha_j)}{\prod_{j=1}^{T}\Gamma(\alpha_j)}\right)^{D}\cdot\prod_{d}\frac{\prod_{j}\Gamma(N_{d,j}+\alpha_j)}{\Gamma(N_d+\sum_{j}\alpha_j)}. \tag{4}$$

Similarly, we calculate $p(l)=\prod_{d}\prod_{j} p(l_{d,j})$ using the formula -

$$\begin{aligned}&\prod_{d}\prod_{j}\int_{\pi_{d,j}} p(\pi_{d,j};\boldsymbol{\gamma_d})\prod_{i} p(l_{d,j,i}|\pi_{d,j})\,d\pi_{d,j} = \\ &\prod_{d}\prod_{j}\int_{\pi_{d,j}}\prod_{k}\pi_{d,j,k}^{\gamma_{d,k}-1}\left(\frac{\Gamma(\sum_{k=1}^{S}\gamma_{d,k})}{\prod_{k=1}^{S}\Gamma(\gamma_{d,k})}\right)\pi_{d,j,k}^{N_{d,j,k}}\,d\pi_{d,j} = \\ &\left(\frac{\Gamma(\sum_{k=1}^{S}\gamma_{d,k})}{\prod_{k=1}^{S}\Gamma(\gamma_{d,k})}\right)^{D\times T}\cdot\prod_{d}\prod_{j}\frac{\prod_{k}\Gamma(N_{d,j,k}+\gamma_{d,k})}{\Gamma(N_{d,j}+\sum_{k}\gamma_{d,k})}\end{aligned} \tag{5}$$

Variable	Type	Description
D	integer	Number of documents
T	integer	Number of topics
S	integer	Number of sentiment class
V	integer	Number of words in vocabulary
B	integer	Number of bi-terms in bi-term vocab
N_d	integer	Number of words in document d
$\boldsymbol{\alpha}$	vector	document-topic prior of size T
α_j	positive real	document-topic prior for topic j
β	positive real	topic-sentiment-word prior
$\boldsymbol{\gamma}$	matrix	document-topic-sentiment prior of size $D \times S$
$\boldsymbol{\gamma_d}$	vector	document-topic-sentiment prior for document d
$\gamma_{d,j}$	positive real	document-topic-sentiment prior for document d and topic j
θ	vector	document-topic probability distribution of size $D \times T$
θ_d	vector	topic proportions of document d
pi	tensor	document-topic-sentiment distribution of size $D \times T \times S$
$pi_{d,j}$	vector	sentiment proportions of document d and topic j
φ	tensor	topic-sentiment-word distribution of size $T \times S \times V$
$\varphi_{j,k}$	vector	word proportions of topic j and sentiment k
$N_{d,j,k}$	non-negative integer	Number of words in document d for which sentiment k is assigned to topic j
$N_{j,k}$	non-negative integer	Number of words for which sentiment k is assigned to topic j
$N_{d,j}$	non-negative integer	Number of times a word from document d is assigned to topic j
$N_{j,k,v}$	non-negative integer	Number of times topic j and sentiment k is assigned to word v

Table 1: Notation

Finally, we obtain $p(w) = \prod_j \prod_k p(w_{j,k})$ by replacing $\prod_{d=1}^{D} \prod_{i=1}^{N_d}$ with $\prod_{i=1}^{V}$ in the formula

$$\prod_j \prod_k \int_{\varphi_{j,k}} p(\varphi_{j,k};\beta) \prod_d \prod_i p(w_{d,j,k,i}|\varphi_{j,k})\, d\varphi_{j,k} =$$
$$\prod_j \prod_k \int_{\varphi_{j,k}} p(\varphi_{j,k};\beta) \prod_i p(w_{j,k,i}|\varphi_{j,k})\, d\varphi_{j,k} =$$
$$\prod_j \prod_k \int_{\varphi_{j,k}} \prod_i \varphi_{j,k,i}^{\beta-1} \left(\frac{\Gamma(\sum_{i=1}^{V} \beta)}{\prod_{i=1}^{V} \Gamma(\beta)} \right) \varphi_{j,k,i}^{N_{j,k,i}}\, d\varphi_{j,k} =$$
$$\left(\frac{\Gamma(V\beta)}{\Gamma(\beta)^V} \right)^{T\times S} \cdot \prod_j \prod_k \frac{\prod_i \Gamma(N_{j,k,i}+\beta)}{\Gamma(N_{j,k}+V\beta)} \tag{6}$$

The goal of Gibbs sampling is to calculate $p(\mathbf{z}, \mathbf{l} \,|\, w_t, \boldsymbol{\alpha}, \beta, \boldsymbol{\gamma})$ instead of Eq. 2, by taking approximation $p(z_t, l_t \,|\, w_t, \mathbf{z^{-t}}, \mathbf{l^{-t}}, \boldsymbol{\alpha}, \beta, \boldsymbol{\gamma})$, for each word w_t, in document d. Here t is the index of the word in document d.

$$p(z_t, l_t \,|\, w_t, \mathbf{z^{-t}}, \mathbf{l^{-t}}, \boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}) \propto$$
$$p(z_t, l_t, \mathbf{z^{-t}}, \mathbf{l^{-t}} \,|\, w_t, \boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}) =$$
$$p(z_t, l_t \,|\, w_t, \boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}) \cdot p(, \mathbf{z^{-t}}, \mathbf{l^{-t}} \,|\, w_t, \boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}) \tag{7}$$

If we put the values from Eq. 4, 5 and 6 in each of the terms in 7, we get

$$p(z_t, l_t \,|\, w_t, \mathbf{z^{-t}}, \mathbf{l^{-t}}, \boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}) \propto$$
$$\prod_j \prod_k \frac{\Gamma(N_{j,k,w_t}+\beta)}{\Gamma(N_{j,k}+V\beta)} \cdot \prod_j \frac{\prod_k \Gamma(N_{d,j,k}+\gamma_{d,k})}{\Gamma(N_{d,j}+\sum_k \gamma_{d,k})} \cdot$$
$$\frac{\prod_j \Gamma(N_{d,j}+\alpha_j)}{\Gamma(N_d+\sum_j \alpha_j)} \tag{8}$$

and

$$p(z_t = j, l_t = k \,|\, w_t, \mathbf{z^{-t}}, \mathbf{l^{-t}}, \boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}) \propto$$
$$\frac{\Gamma(N_{j,k,w_t}+\beta)}{\Gamma(N_{j,k}+V\beta)} \cdot \frac{\Gamma(N_{d,j,k}+\gamma_{d,k})}{\Gamma(N_{d,j}+\sum_k \gamma_{d,k})} \cdot \frac{\Gamma(N_{d,j}+\alpha_j)}{\Gamma(N_d+\sum_j \alpha_j)} \tag{9}$$

For further simplification, we use formula $\Gamma(x+1) = x \cdot \Gamma(x)$ to get

$$\Gamma(N_{j,k,w_t}+\beta) = \Gamma(N_{j,k,w_t}^{-t}+\beta+1) \propto N_{j,k,w_t}^{-t}+\beta$$

Following this on other terms - $N_{j,k}, N_{d,j,k}, N_{d,j}$ and N_d we get

$$p(z_t = j, l_t = k \,|\, \mathbf{w}, \mathbf{z^{-t}}, \mathbf{l^{-t}}, \boldsymbol{\alpha}, \beta, \gamma) \propto$$
$$\frac{N_{j,k,w_t}^{-t}+\beta}{N_{j,k}^{-t}+V\beta} \cdot \frac{N_{d,j,k}^{-t}+\gamma_{d,k}}{N_{d,j}^{-t}+\sum_k \gamma_{d,k}} \cdot \frac{N_{d,j}^{-t}+\alpha_j}{N_d^{-t}+\sum_j \alpha_j} \tag{10}$$

0.2 Model Inference for Bi-LJST

There are some changes in the probability distribution for Bi-LJST. Let $\mathcal{B}$ snd B_d denote the vocabulary of bi-terms and the number of bi-terms in the document d. Further assume that a bi-term is represented as $b_t = (w_p, w_q)$. The joint probability of bi-terms, topics and sentiment labels can be captured by,

$$p(b_t, \mathbf{z}, \mathbf{l}) = p(b_t \,|\, \mathbf{l}, \mathbf{z}) p(\mathbf{l}|\mathbf{z}) p(\mathbf{z}) \tag{11}$$

As the terms, $p(\mathbf{l}|\mathbf{z})$ and $p(\mathbf{z})$ do not depend on bi-term b_t, we can use directly Eq. 4, 5. Further, the generative process of Bi-LJST, word pairs within a bi-term are conditionally independent i.e. $p(b_t \,|\, \mathbf{l}, \mathbf{z}) = p(w_p \,|\, \mathbf{l}, \mathbf{z}) \cdot p(w_q \,|\, \mathbf{l}, \mathbf{z})$. This leads us to topic sentiment assignment corresponding to bi-term b_t as

$$p(z_t = j, l_t = k \,|\, b_t, \mathbf{z}^{-\mathbf{t}}, \mathbf{l}^{-\mathbf{t}}, \boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}) \propto \frac{(N_{j,k,w_p}^{-t} + \beta) \cdot (N_{j,k,w_q}^{-t} + \beta)}{(N_{j,k}^{-t} + V\beta)^2} \cdot \frac{N_{d,j,k}^{-t} + \gamma_{d,k}}{N_{d,j}^{-t} + \sum_k \gamma_{d,k}} \cdot \frac{N_{d,j}^{-t} + \alpha_j}{N_d^{-t} + \sum_j \alpha_j} \tag{12}$$

where, p_1 and p_2 are the index of w_p and w_q respectively in vocabulary.