Technical Report for
A Hierarchical Aspect-Sentiment Model for Online Reviews

Suin Kim[1], **Jianwen Zhang**[2], **Zheng Chen**[2], **Alice Oh**[1], **Shixia Liu**[2]
[1]Department of Computer Science, KAIST, Republic of Korea
[2]Microsoft Research Asia, Beijing, P.R. China
[1]{suin.kim@kaist.ac.kr, alice.oh@kaist.edu}, [2]{jiazhan,zhengc,shliu}@microsoft.com

This full-length technical report describes the detailed derivation of model inference process, which was omitted due to the space limit on the published paper.

In this report we describe the detailed derivation for the inference of posterior distribution of HASM using Gibbs sampling. Specifically, the task of inference is to learn the posterior $p(T|\boldsymbol{w})$ of the tree from data. Using collapsed Gibbs sampling, we infer the posterior $p(\boldsymbol{c}, \boldsymbol{s}, \boldsymbol{p}|\boldsymbol{w}, \alpha, \eta, \boldsymbol{\beta}, \gamma)$ of three groups of variables, 1) $\boldsymbol{c}$, the aspect-sentiment node of each sentence 2) $\boldsymbol{s}$, the sentiment of each sentence 3) $\boldsymbol{p}$, the subjectivity of each word in a sentence. Other variables are integrated out and do not need sampling.

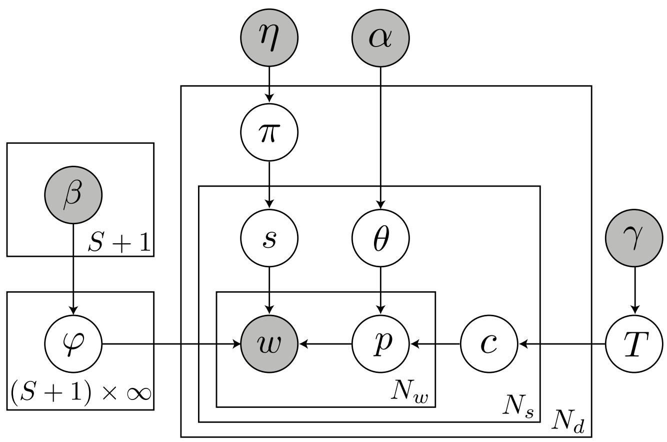

Figure 1: Graphical representation of HASM.

We start from the total probability of HASM:

$$\mathcal{L} = P(T|\gamma) \prod_{k=1}^{\infty} \prod_{s=1}^{S+1} P(\varphi_{s,k}|\beta_s) \prod_{i}^{N_d} P(\pi_i|\eta) \prod_{j}^{N_s} P(s_{i,j}|\pi) P(\theta_{i,j}|\alpha) P(c_{i,j}|T) \prod_{k}^{N_w} P(p_{i,j,k}|\theta, c) P(w_{i,j,k}|s, p, c, \varphi),$$

and integrate out four variables T, φ, π, and θ:

$$\int_T \int_\varphi \int_\pi \int_\theta \mathcal{L}\, \mathrm{d}\theta\, \mathrm{d}\pi\, \mathrm{d}\varphi\, \mathrm{d}T = \int_T P(T|\gamma) \prod_i^{N_d} \prod_j^{N_s} P(c_{i,j}|T)\, \mathrm{d}T \tag{1}$$

$$\times \int_\varphi \prod_{k=1}^{\infty} \prod_{s=0}^{S} P(\varphi_{s,k}|\beta_s) \prod_i^{N_d} \prod_j^{N_s} \prod_k^{N_w} P(w_{i,j,k}|s,p,c,\varphi)\, \mathrm{d}\varphi \tag{2}$$

$$\times \int_\pi \prod_i^{N_d} P(\pi_i|\eta) \prod_j^{N_s} P(s_{i,j}|\pi)\, \mathrm{d}\pi \tag{3}$$

$$\times \int_\theta \prod_i^{N_d} \prod_j^{N_s} P(\theta_{i,j}|\alpha) \prod_k^{N_w} P(p_{i,j,k}|\theta, c)\, \mathrm{d}\theta, \tag{4}$$

and

$$(2) = \int_\varphi \prod_{k=1}^{\infty} \prod_{s=0}^{S} \frac{\Gamma(\sum_{r=1}^{V} \beta_{s,r})}{\prod_{r=1}^{V} \Gamma(\beta_{s,r})} \prod_{r=1}^{V} \varphi_{s,k,r}^{\beta_{s,r}-1} \prod_i^{N_d} \prod_j^{N_s} \prod_k^{N_w} P(w_{i,j,k}|s,p,c,\varphi)\, \mathrm{d}\varphi$$

$$= \int_\varphi \prod_{k=1}^{\infty} \prod_{s=0}^{S} \frac{\Gamma(\sum_{r=1}^{V} \beta_{s,r})}{\prod_{r=1}^{V} \Gamma(\beta_{s,r})} \prod_{r=1}^{V} \varphi_{s,k,r}^{\beta_{s,r}+n_{k,s}^{w,(r)}-1}\, \mathrm{d}\varphi \tag{5}$$

$$= \prod_{k=1}^{\infty} \prod_{s=0}^{S} \frac{\Gamma(\sum_{r=1}^{V} \beta_{s,r})}{\prod_{r=1}^{V} \Gamma(\beta_{s,r})} \frac{\prod_{r=1}^{V} \Gamma(\beta_{s,r} + n_{k,s}^{w,(r)})}{\Gamma(\sum_{r=1}^{V} \beta_{s,r} + n_{k,s}^{w,(r)})}. \tag{6}$$

Note that equation (6) is derivable from (5) using the property of Dirichlet distribution. In the same way,

$$(3) = \int_\pi \prod_i^{N_d} P(\pi_i|\eta) \prod_j^{N_s} P(s_{i,j}|\pi)\, \mathrm{d}\pi$$

$$= \int_\pi \prod_i^{N_d} \frac{\Gamma(\sum_{r=1}^{S} \eta_r)}{\prod_{r=1}^{S} \Gamma(\eta_r)} \prod_{r=1}^{S} \eta_r^{\pi_{i,r}-1} \prod_j^{N_s} P(s_{i,j}|\pi)\, \mathrm{d}\pi$$

$$= \int_\pi \prod_i^{N_d} \frac{\Gamma(\sum_{r=1}^{S} \eta_r)}{\prod_{r=1}^{S} \Gamma(\eta_r)} \prod_{r=1}^{S} \eta_r^{\pi_{i,r}+n_i^{s,(r)}-1}\, \mathrm{d}\pi = \prod_i^{N_d} \frac{\Gamma(\sum_{r=1}^{S} \eta_r)}{\prod_{r=1}^{S} \Gamma(\eta_r)} \frac{\prod_{r=1}^{S} \Gamma(\eta_r + n_i^{s,(r)})}{\Gamma(\sum_{r=1}^{S} \eta_r + n_i^{s,(r)})}$$

$$(4) = \int_\theta \prod_i^{N_d} \prod_j^{N_s} P(\theta_{i,j}|\alpha) \prod_k^{N_w} P(p_{i,j,k}|\theta, c)\, \mathrm{d}\theta$$

$$= \int_\theta \prod_i^{N_d} \prod_j^{N_s} \frac{\Gamma(\sum_{r=0}^{1} \alpha_r)}{\prod_{r=0}^{1} \Gamma(\alpha_r)} \prod_{r=0}^{1} \alpha_r^{\theta_{i,j,r}+n_{i,j}^{p,(r)}-1}\, \mathrm{d}\theta = \prod_i^{N_d} \prod_j^{N_s} \frac{\Gamma(\sum_{r=0}^{1} \alpha_r)}{\prod_{r=0}^{1} \Gamma(\alpha_r)} \frac{\prod_{r=0}^{1} \Gamma(\alpha_r + n_{i,j}^{p,(r)})}{\Gamma(\sum_{r=0}^{1} \alpha_r + n_{i,j}^{p,(r)})},$$

while (1) simply follows the inference of rCRP prior.

Aspect Sampling (Sampling c).

We follow the sampling procedure of rCRP, beginning from the root dish and moving down along the path. Each sentence contains statistics about sentiment of the sentence and subjectivity of words. Such information is being kept while the sentence is moving along the path or assigned to a new node. There are three possibilities for aspect-sentiment node Φ_k of each sentence i in document d:

1. $P(\text{Select the node } \Phi_k) \propto m_k \times P(\boldsymbol{w}_{di}|s, \boldsymbol{p}, \Phi_k, \boldsymbol{\beta})$
2. $P(\text{Select a child } c \text{ of } \Phi_k) \propto M_c \times P(\boldsymbol{w}_{di}|s, \boldsymbol{p}, c, \boldsymbol{\beta})$
3. $P(\text{Create a new child}) \propto \gamma \times P(\boldsymbol{w}_{di}|s, \boldsymbol{p}, \phi, \boldsymbol{\beta})$,

and the recursive assign process for each sentence stops if a node is selected by the first or the third assignment.

The probability of generating words in sentence i, $w \in i$ in document d from sentiment s, subjectivity $\boldsymbol{p}$, and node Φ_k is

$$
\begin{aligned}
P(\boldsymbol{w}_{di}|s, \boldsymbol{p}, \Phi_k, \boldsymbol{\beta}) \propto (2)& \\
&= \prod_{k=1}^{\infty} \prod_{s=0}^{S} \frac{\Gamma(\sum_{r=1}^{V} \beta_{s,r})}{\prod_{r=1}^{V} \Gamma(\beta_{s,r})} \frac{\prod_{r=1}^{V} \Gamma(\beta_{s,r} + n_{k,s}^{w,(r)})}{\Gamma(\sum_{r=1}^{V} \beta_{s,r} + n_{k,s}^{w,(r)})} \\
&\propto \prod_{l=0}^{1} \frac{\prod_{r=1}^{V} \Gamma(\beta_{s_i \times l,r} + n_{k,s_i \times l}^{w,(r)})}{\Gamma(\sum_{r=1}^{V} \beta_{s_i \times l,r} + n_{k,s_i \times l}^{w,(r)})} \\
&\propto \prod_{l=0}^{1} \frac{\Gamma(n_{k,s_i \times l,-i}^{w,(\cdot)} + \hat{\beta}_{s_i \times l})}{\prod_{w \in i} \Gamma(n_{k,s_i \times l,-i}^{w,(w)} + \beta_{s_i \times l,w})} \frac{\prod_{w \in i} \Gamma(n_{k,s_i \times l}^{w,(w)} + \beta_{s_i \times l,w})}{\Gamma(n_{k,s_i \times l}^{w,(\cdot)} + \hat{\beta}_{s_i \times l})}.
\end{aligned}
$$

Sentiment Sampling (Sampling s).

We then sample the sentiment polarity s_{di} of each sentence. Beta hyperparameter η controls the sentiment distribution for each document. Higher η implies more prior confidence that distribution of sentiment polarities are likely to be even. The probability of assigning k^{th} sentiment for sentence i in document d is

$$
\begin{aligned}
P(s_{di} = k|\boldsymbol{w}, \boldsymbol{s}, \boldsymbol{p}, \boldsymbol{c}, \boldsymbol{\beta}) \propto (2) \times (3)& \\
&\propto \prod_{i}^{N_d} \frac{\Gamma(\sum_{r=1}^{S} \eta_r)}{\prod_{r=1}^{S} \Gamma(\eta_r)} \frac{\prod_{r=1}^{S} \Gamma(\eta_r + n_i^{s,(r)})}{\Gamma(\sum_{r=1}^{S} \eta_r + n_i^{s,(r)})} \times P(\boldsymbol{w}_{di}|k, \boldsymbol{p}, \boldsymbol{c}_{di}, \boldsymbol{\beta}) \\
&\propto \frac{\prod_{r=1}^{S} \Gamma(\eta_r + n_i^{s,(r)})}{\Gamma(\sum_{r=1}^{S} \eta_r + n_i^{s,(r)})} \times P(\boldsymbol{w}_{di}|k, \boldsymbol{p}, \boldsymbol{c}_{di}, \boldsymbol{\beta}) \\
&\propto \prod_{r=1}^{S} \Gamma(\eta_r + n_i^{s,(r)}) P(\boldsymbol{w}_{di}|k, \boldsymbol{p}, \boldsymbol{c}_{di}, \boldsymbol{\beta}) \\
&\propto (n_{d,-i}^{s,(k)} + \eta) P(\boldsymbol{w}_{di}|k, \boldsymbol{p}, \boldsymbol{c}_{di}, \boldsymbol{\beta}).
\end{aligned}
$$

Subjectivity Sampling (Sampling p).

Finally we sample the subjectivity of each word. Subjectivity sampling is similar to the Gibbs sampling process in a basic LDA model with two topics: $\{$0-*Non-subjective*, 1-*Subjective*$\}$. The probability of assigning subjectivity p for p_{dij}, which is generated from aspect-sentiment node Φ_k is

$$
\begin{aligned}
P(p_{dij} = p|\boldsymbol{w}, \boldsymbol{s}, \boldsymbol{p}, \boldsymbol{c}, \boldsymbol{\beta}) \propto (2) \times (4)& \\
&= \prod_{k=1}^{\infty} \prod_{s=0}^{S} \frac{\Gamma(\sum_{r=1}^{V} \beta_{s,r})}{\prod_{r=1}^{V} \Gamma(\beta_{s,r})} \frac{\prod_{r=1}^{V} \Gamma(\beta_{s,r} + n_{k,s}^{w,(r)})}{\Gamma(\sum_{r=1}^{V} \beta_{s,r} + n_{k,s}^{w,(r)})} \times \prod_{i}^{N_d} \prod_{j}^{N_s} \frac{\Gamma(\sum_{r=0}^{1} \alpha_r)}{\prod_{r=0}^{1} \Gamma(\alpha_r)} \frac{\prod_{r=0}^{1} \Gamma(\alpha_r + n_{i,j}^{p,(r)})}{\Gamma(\sum_{r=0}^{1} \alpha_r + n_{i,j}^{p,(r)})} \\
&\propto \prod_{l=0}^{1} \frac{\prod_{r=1}^{V} \Gamma(\beta_{s_{di} \times l,r} + n_{k,s_{di} \times l}^{w,(r)})}{\Gamma(\sum_{r=1}^{V} \beta_{s_{di} \times l,r} + n_{k,s_{di} \times l}^{w,(r)})} \times \frac{\prod_{r=0}^{1} \Gamma(\alpha_r + n_{i,j}^{p,(r)})}{\Gamma(\sum_{r=0}^{1} \alpha_r + n_{i,j}^{p,(r)})} \\
&\propto \prod_{l=0}^{1} \frac{\prod_{r=1}^{V} \Gamma(\beta_{s_{di} \times l,r} + n_{k,s_{di} \times l}^{w,(r)})}{\Gamma(\sum_{r=1}^{V} \beta_{s_{di} \times l,r} + n_{k,s_{di} \times l}^{w,(r)})} \times \prod_{r=0}^{1} \Gamma(\alpha_r + n_{i,j}^{p,(r)}) \\
&\propto (n_{d,i,-j}^{p,(p)} + \alpha) \frac{n_{k,s_{di} \times p,-j}^{w,(v)} + \beta_{s_{di} \times p,v}}{\sum_{r=1}^{V} n_{k,s_{di} \times p,-j}^{w,(r)} + \hat{\beta}_{s_{di} \times p}}.
\end{aligned}
$$

Estimating $\boldsymbol{\varphi}$.

In our collapsed Gibbs sampling algorithm, some latent variables such as $\boldsymbol{\varphi}$ and $\boldsymbol{\theta}$, are integrated out. After the Gibbs sampling process, a topic $\boldsymbol{\varphi}$ can be obtained by Bayesian estimation:

$$
\hat{\boldsymbol{\varphi}} = \int \boldsymbol{\varphi} \cdot p(\boldsymbol{\varphi}|\boldsymbol{w}, \boldsymbol{\beta}, \boldsymbol{c}, \boldsymbol{p}, \boldsymbol{s}) d\,\boldsymbol{\varphi}.
$$