# Modeling Dialogue Acts with Content Word Filtering and Speaker Preferences

## (Supplementary Material)

**Yohan Jo, Michael Miller Yoder, Hyeju Jang, Carolyn P. Rosé**
Language Technologies Institute
Carnegie Mellon University
{yohanj,yoder,hyejuj,cprose}@cs.cmu.edu

## 1 Derivation for Gibbs Sampling

**Sampling** $s$ For the current utterance $u$, refer to its parent utterance as $p$, its child utterances as $\mathbf{c} = \{c_1, \cdots, c_m\}$, and its sentences as $\mathbf{e} = \{e_1, \cdots, e_n\}$. The speaker of an utterance is denoted by $a_?$ with its subscript being the utterance, e.g., $a_u$. For brevity, $\mathbf{c}_i$ denotes $\{c_i, \cdots, c_m\}$, and $\mathbf{e}_i$ denotes $\{e_i, \cdots, e_n\}$. $\mathbf{a}$ denotes the author indices of all utterances. $\mathbf{s}$ denotes the utterance-level DA indices of all utterances, and $\mathbf{s_c}$ denotes the utterance-level DA indices of $\mathbf{c}$. $\mathbf{z}^F$ denotes the sentence-level DA indices of all sentences in the data, and $\mathbf{z}^F_\mathbf{e}$ denotes the sentence-level DA indices of $\mathbf{e}$. A subscript starting with $\setminus$ means exclusion, e.g., $\mathbf{s}_{\setminus u,\mathbf{c}}$ means the utterance-level DA indices of all utterances but $u$ and $\mathbf{c}$.

As we sample, $N^{SS}_{ij}$ is the frequency of the utterances whose parent's DA is $i$ and whose own DA is $j$. $N^{AS}_{ij}$ is the frequency of utterances whose speaker is $i$ and DA is $j$. $N^{SF}_{ij}$ is the frequency of sentences whose utterance has utterance-level DA $i$ and whose own sentence-level DA is $j$. A superscript with "\utterances/sentences" (e.g., $N^{SS\setminus u}_{ij}$) means to exclude those utterances or sentences from counting.

$$
\begin{aligned}
&P(s_u = j|\mathbf{a}, \mathbf{s}_{\setminus u}, \mathbf{z}^F, \nu, \gamma^S, \gamma^A, \alpha^F) && (1)\\
&\quad \propto P(s_u = j, \mathbf{s_c}, \mathbf{z}^F_{\mathbf{e}}|\mathbf{a}, \mathbf{s}_{\setminus u,\mathbf{c}}, \mathbf{z}^F_{\setminus \mathbf{e}}, \nu, \gamma^S, \gamma^A, \alpha^F)\\
&\quad = P(s_u = j|\mathbf{a}, \mathbf{s}_{\setminus u,\mathbf{c}}, \mathbf{z}^F_{\setminus \mathbf{e}}, \nu, \gamma^S, \gamma^A) && (2)\\
&\qquad \prod_{i=1}^{m} P(s_{c_i}|\mathbf{a}, \mathbf{s}_{\setminus \mathbf{c}_i}, \mathbf{z}^F_{\setminus \mathbf{e}}, \nu, \gamma^S, \gamma^A) && (3)\\
&\qquad \prod_{i=1}^{n} P(z^F_{e_i}|\mathbf{a}, \mathbf{z}^F_{\setminus \mathbf{e}_i}, \alpha^F). && (4)
\end{aligned}
$$

(2) is approximated as

$$
\nu \frac{N^{SS\setminus u,\mathbf{c}}_{s_p j} + \gamma^S}{\sum_k \left(N^{SS\setminus u,\mathbf{c}}_{s_p k} + \gamma^S\right)} + (1-\nu)\frac{N^{AS\setminus u}_{a_u j} + \gamma^A}{\sum_k N^{AS\setminus u}_{a_u k} + \gamma^A}.
$$

(3) is approximated as

$$
\prod_{i=1}^{m} \left( \nu \frac{N^{SS\setminus \mathbf{c}_i}_{j s_{c_i}} + \gamma^S}{\sum_k \left(N^{SS\setminus \mathbf{c}_i}_{jk} + \gamma^S\right)} + (1-\nu)\frac{N^{AS}_{a_{c_i} s_{c_i}} + \gamma^A}{\sum_k N^{AS}_{a_{c_i} k} + \gamma^A} \right).
$$

(4) is approximated as

$$
\prod_{i=1}^{n} \frac{N^{SF\setminus \mathbf{e}_i}_{j z^F_{e_i}} + \alpha^F}{\sum_k \left(N^{SF\setminus \mathbf{e}_i}_{jk} + \alpha^F\right)}.
$$

**Sampling the sentence-level DA $z^F_e$ of sentence $e$:** Let $\boldsymbol{\omega}$ denote all words in the data and $\boldsymbol{\omega}_e$ the words in sentence $e$. $\mathbf{w}$ and $\mathbf{w}_e$ are the word indices of $\boldsymbol{\omega}$ and $\boldsymbol{\omega}_e$, respectively, where $\mathbf{w}_{e>j} = \{w_{ej}, \cdots, w_{e|\boldsymbol{\omega}_e|}\}$. $\mathbf{l}$ is the DA/content indicators of all words. $\mathbf{z}^F$ is the sentence-level DAs of all sentences in the same utterance.

As we sample, $N^{FW}_{ij}$ is the frequency of the words whose $l$ is "DA" and word index is $j$. As before, a superscript with "\ words" means to exclude those words from counting.

$$
\begin{aligned}
&P(z_e^F = k|\mathbf{s}, \mathbf{z}_{\backslash e}^F, \mathbf{w}, \mathbf{l}, \mathbf{b}, \alpha^F, \beta) \\
&\quad \propto P(z_e^F = k, \mathbf{w}_e|\mathbf{s}, \mathbf{z}_{\backslash e}^F, \mathbf{w}_{\backslash e}, \mathbf{l}, \mathbf{b}, \alpha^F, \beta) \\
&\quad \propto P(z_e^F = k|\mathbf{s}, \mathbf{z}_{\backslash e}^F, \alpha^F) && (5) \\
&\qquad \prod_{\substack{j=1 \\ l_j=\text{DA}}}^{|\boldsymbol{\omega}_e|} P(w_{ej}|\mathbf{z}^F, \mathbf{l}, \mathbf{w}\backslash\mathbf{w}_{e>j}, \beta). && (6)
\end{aligned}
$$

(5) is approximated as

$$
\frac{N_{s_u k}^{SF\backslash e} + \alpha^F}{\sum_{k'} N_{s_u k'}^{SF\backslash e} + \alpha^F}.
$$

(6) is approximated as

$$
\prod_{\substack{j=1 \\ l_j\text{DA}}}^{|\boldsymbol{\omega}_e|} \frac{N_{k w_{ej}}^{FW\backslash\boldsymbol{\omega}_{e>j}} + \beta}{\sum_{w'} N_{k w'}^{FW\backslash\boldsymbol{\omega}_{e>j}} + \beta}.
$$

**Sampling the background topic $z_d^B$ of discussion $d$:** $\boldsymbol{\omega}_d$ is all words in $d$ and $\mathbf{w}_d$ is their word indices, where $\mathbf{w}_{d>j} = \{w_{dj}, \cdots, w_{d|\boldsymbol{\omega}_d|}\}$. Now $\mathbf{l}$ denotes all indicators in $d$.

As we sample, $N_j^B$ is the frequency of the discussions whose background topic is $j$. As before, a superscript with "$\backslash d$" means to exclude $d$ from counting.

$$
\begin{aligned}
&P(z_d^B = k|z_{\backslash d}^B, \mathbf{w}, \mathbf{l}, \alpha^B, \beta) \\
&\quad \propto P(z_d^B = k, \mathbf{w}_d|z_{\backslash d}^B, \mathbf{w}_{\backslash d}, \mathbf{l}, \alpha^B, \beta) \\
&\quad \propto P(z_d^B = k|z_{\backslash d}^B, \alpha^B) && (7) \\
&\qquad \prod_{\substack{j=1 \\ l_j=\text{content}}}^{|\boldsymbol{\omega}_d|} P(w_{dj}|\mathbf{z}^B, \mathbf{l}, \mathbf{w}\backslash\mathbf{w}_{d>j}, \beta). && (8)
\end{aligned}
$$

(7) is approximated as

$$
\frac{N_k^{B\backslash d} + \alpha^B}{\sum_{k'} N_k^{B\backslash d} + \alpha^B}.
$$

(8) is approximated as

$$\prod_{\substack{j=1 \\ l_j=\text{content}}}^{|\boldsymbol{\omega}_d|} \frac{N_{kw_{dj}}^{BW\setminus\boldsymbol{\omega}_{d>j}} + \beta}{\sum_{w'} N_{kw'}^{BW\setminus\boldsymbol{\omega}_{d>j}} + \beta}.$$

**Sampling $l$ of word $\omega$:** Let $\omega$ be in sentence $e$ and have word index $w$.

$$\begin{aligned} P(l = \text{DA}|\mathbf{l}_{\setminus\omega}, \mathbf{w}, \mathbf{z}^F, \eta, \beta) &\propto P(l = \text{DA}, w|\mathbf{l}_{\setminus\omega}, \mathbf{w}_{\setminus\omega}, \mathbf{z}^F, \eta, \beta) \\ &= P(l = \text{DA}|\eta) P(w|\mathbf{w}_{\setminus\omega}, \mathbf{z}^F, \mathbf{l}, \beta) \\ &\approx \eta \frac{N_{z_e^F w}^{FW\setminus\omega} + \beta}{\sum_{w'} N_{z_e^F w'}^{FW\setminus\omega} + \beta}. \end{aligned}$$

Similarly,

$$\begin{aligned} P(l = \text{content}|\mathbf{l}_{\setminus\omega}, \mathbf{w}, \mathbf{z}^B, \eta, \beta) &\propto P(l = \text{content}, w|\mathbf{l}_{\setminus\omega}, \mathbf{w}_{\setminus\omega}, \mathbf{z}^B, \eta, \beta) \\ &= P(l = \text{content}|\eta) P(w|\mathbf{w}_{\setminus\omega}, \mathbf{z}^B, \mathbf{l}, \beta) \\ &\approx (1 - \eta) \frac{N_{z_d^B w}^{BW\setminus\omega} + \beta}{\sum_{w'} N_{z_d^B w'}^{BW\setminus\omega} + \beta}. \end{aligned}$$