

# A latent model for inducing frame semantic roles

Yang Zhou

July 17, 2013

## 1 Introduction

## 2 Model Concepts

### 2.1 Graph

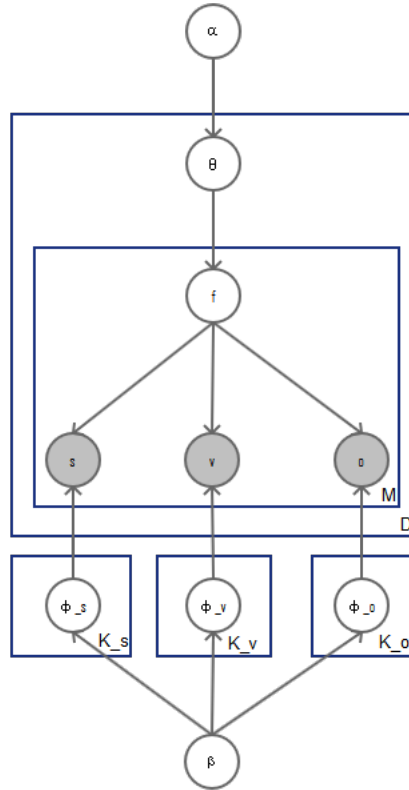


Figure 1: Plate graph

$$\theta \sim \text{Dirichlet}(\alpha)$$

$$\phi_s \sim \text{Dirichlet}(\beta)$$

$$\phi_v \sim \text{Dirichlet}(\beta)$$

$$\phi_o \sim \text{Dirichlet}(\beta)$$

$$\begin{aligned}
f_m &\sim \text{Multinomial}(\theta^{(m)}) \\
s \mid f_m, \phi_s^{(f_m)} &\sim \text{Multinomial}(f_m) \\
v \mid f_m, \phi_v^{(f_m)} &\sim \text{Multinomial}(f_m) \\
o \mid f_m, \phi_o^{(f_m)} &\sim \text{Multinomial}(f_m)
\end{aligned}$$

## 2.2 Generating process

1. Sample frame distribution  $\theta$  from  $\text{Dirichlet}(\alpha)$
2. Sample subject distribution  $\phi_s$  from  $\text{Dirichlet}(\beta)$
3. Sample predicate distribution  $\phi_v$  from  $\text{Dirichlet}(\beta)$
4. Sample object distribution  $\phi_o$  from  $\text{Dirichlet}(\beta)$

### Document-Frame

5. For each frame in  $d$ , sample  $f_m \sim \text{Multinomial}(\theta_m)$ :

### Frame-Roles

6. For each semantic roles in  $f_m$ :
  - (a) Sample  $s_m \sim \text{Multinomial}(\phi_s^{(f_m)})$
  - (b) Sample  $v_m \sim \text{Multinomial}(\phi_v^{(f_m)})$
  - (c) Sample  $o_m \sim \text{Multinomial}(\phi_o^{(f_m)})$

## 3 Model Detail

A deduction of the model.

### 3.1 Likelihood

The joint distribution of all known and hidden variables given the hyperparameters:

$$P(\vec{s}_m, \vec{v}_m, \vec{o}_m, \vec{f}_m, \vec{\theta}_m, \Phi^s, \Phi^v, \Phi^o \mid \vec{\alpha}, \vec{\beta}) \quad (1)$$

$$\begin{aligned}
&= \prod_{i=1}^{N_m} P(s_{m,n} \mid \vec{\phi}_{f_{m,n}}^s) P(v_{m,n} \mid \vec{\phi}_{f_{m,n}}^v) P(o_{m,n} \mid \vec{\phi}_{f_{m,n}}^o) P(f_{m,n} \mid \vec{\theta}_m) P(\vec{\theta}_m \mid \vec{\alpha}) \\
&\quad \cdot P(\vec{\Phi}^s \mid \vec{\beta}) P(\vec{\Phi}^v \mid \vec{\beta}) P(\vec{\Phi}^o \mid \vec{\beta})
\end{aligned} \quad (2)$$

Likelihood of a document:

$$P(\vec{d}_m \mid \vec{\alpha}, \vec{\beta}) \quad (3)$$

$$= P(\vec{s}_m, \vec{v}_m, \vec{o}_m \mid \vec{\alpha}, \vec{\beta}) \quad (4)$$

$$= \int P(\vec{s}_m, \vec{v}_m, \vec{o}_m \mid \vec{\theta}_m, \vec{\beta}) \cdot P(\vec{\theta}_m \mid \vec{\alpha}) d\vec{\theta}_m \quad (5)$$

$$= \int \prod_{n=1}^{N_m} \sum_{f_{m,n}} P(s_{m,n}, v_{m,n}, o_{m,n} \mid f_{m,n}, \vec{\beta}) P(f_{m,n} \mid \vec{\theta}_m) \cdot P(\vec{\theta}_m \mid \vec{\alpha}) d\vec{\theta}_m \quad (6)$$

$$= \iiint P(\vec{\theta}_m \mid \vec{\alpha}) P(\vec{\Phi}^s \mid \vec{\beta}) P(\vec{\Phi}^v \mid \vec{\beta}) P(\vec{\Phi}^o \mid \vec{\beta}) \\ \cdot \prod_{i=1}^{N_m} \sum_{f_{m,n}} P(s_{m,n} \mid \vec{\phi}_{f_{m,n}}^s) P(v_{m,n} \mid \vec{\phi}_{f_{m,n}}^v) P(o_{m,n} \mid \vec{\phi}_{f_{m,n}}^o) \cdot P(f_{m,n} \mid \vec{\theta}_m) d\vec{\Phi}^s d\vec{\Phi}^v d\vec{\Phi}^o d\vec{\theta}_m \quad (7)$$

Likelihood of a corpus:

$$P(W \mid \vec{\alpha}, \vec{\beta}) \quad (8)$$

$$= \prod_{m=1}^M P(\vec{s}_m, \vec{v}_m, \vec{o}_m \mid \vec{\alpha}, \vec{\beta}) \quad (9)$$

$$= \prod_{m=1}^M \iiint P(\vec{\theta}_m \mid \vec{\alpha}) P(\vec{\Phi}^s \mid \vec{\beta}) P(\vec{\Phi}^v \mid \vec{\beta}) P(\vec{\Phi}^o \mid \vec{\beta}) \\ \cdot \prod_{i=1}^{N_m} \sum_{f_{m,n}} P(s_{m,n} \mid \vec{\phi}_{f_{m,n}}^s) P(v_{m,n} \mid \vec{\phi}_{f_{m,n}}^v) P(o_{m,n} \mid \vec{\phi}_{f_{m,n}}^o) \cdot P(f_{m,n} \mid \vec{\theta}_m) d\vec{\Phi}^s d\vec{\Phi}^v d\vec{\Phi}^o d\vec{\theta}_m \quad (10)$$

### 3.2 Inference

The target of inference is the distribution:

$$P(\vec{f} \mid \vec{s}, \vec{v}, \vec{o}) \quad (11)$$

$$= \frac{P(\vec{s}, \vec{v}, \vec{o} \mid \vec{f})}{\sum_{f_m} P(\vec{s}, \vec{v}, \vec{o}, \vec{f})} \quad (12)$$

Because computing (12) directly costs high.

**Joint Distribution**

$$P(\vec{s}, \vec{v}, \vec{o}, \vec{f} \mid \vec{\alpha}, \vec{\beta}) \quad (13)$$

$$= P(\vec{s}, \vec{v}, \vec{o} \mid \vec{f}, \vec{\beta}) P(\vec{f} \mid \vec{\alpha}) \quad (14)$$

Deducing  $P(\vec{s}, \vec{v}, \vec{o} \mid \vec{f}, \vec{\beta})$  and  $P(\vec{f} \mid \vec{\alpha})$  separately

$$P(\vec{s}, \vec{v}, \vec{o} \mid \vec{f}, \vec{\beta}) \quad (15)$$

$$= \iiint P(\vec{s}, \vec{v}, \vec{o} \mid \vec{f}, \Phi^s, \Phi^v, \Phi^o) \cdot P(\Phi^s, \Phi^v, \Phi^o \mid \vec{\beta}) d\Phi^s d\Phi^v d\Phi^o \quad (16)$$

$$= \int P(\vec{s} \mid \vec{f}, \Phi^s) P(\Phi^s \mid \vec{\beta}) d\Phi^s \int P(\vec{v} \mid \vec{f}, \Phi^v) P(\Phi^v \mid \vec{\beta}) d\Phi^v \int P(\vec{o} \mid \vec{f}, \Phi^o) P(\Phi^o \mid \vec{\beta}) d\Phi^o \quad (17)$$

The three part of (17) are the same, so take  $\int P(\vec{s} \mid \vec{f}, \Phi^s) P(\Phi^s \mid \vec{\beta}) d\Phi^s$  as an example:

$$\int P(\vec{s} \mid \vec{f}, \Phi^s) P(\Phi^s \mid \vec{\beta}) d\Phi^s \quad (18)$$

$$= \int \prod_{k=1}^K \prod_{t=1}^{V_s} \phi_{k,t}^{n_{s,k}^{(t)}} \prod_{k=1}^K P(\vec{\phi}_k^s \mid \vec{\beta}) d\vec{\phi}_k^s \quad (19)$$

The  $n_{s,k}^{(t)}$  means the number of times a subject occurred under a frame k.

$$= \int \prod_{k=1}^K \prod_{t=1}^{V_s} \phi_{k,t}^{n_{s,k}^{(t)}} \prod_{k=1}^K \left( \frac{\Gamma(\sum_{t=1}^{V_s} \beta_t)}{\prod_{t=1}^{V_s} \Gamma(\beta_t)} \prod_{t=1}^{V_s} \phi_{k,t}^{(s)\beta_t-1} \right) d\phi_k^s \quad (20)$$

$$= \left( \frac{\Gamma(\sum_{t=1}^{V_s} \beta_t)}{\prod_{t=1}^{V_s} \Gamma(\beta_t)} \right)^K \prod_{k=1}^K \int \prod_{t=1}^{V_s} \phi_{k,t}^{(s)\beta_t+n_{s,k}^{(t)}-1} d\phi_k^s \quad (21)$$

$$= \left( \frac{\Gamma(\sum_{t=1}^{V_s} \beta_t)}{\prod_{t=1}^{V_s} \Gamma(\beta_t)} \right)^K \prod_{k=1}^K \frac{\prod_{t=1}^{V_s} \Gamma(\beta_t + n_{s,k}^{(t)})}{\Gamma(\sum_{t=1}^{V_s} (\beta_t + n_{s,k}^{(t)}))} \quad (22)$$

So,  $\int P(\vec{v} \mid \vec{f}, \Phi^v) P(\Phi^v \mid \vec{\beta}) d\Phi^v$  and  $\int P(\vec{o} \mid \vec{f}, \Phi^o) P(\Phi^o \mid \vec{\beta}) d\Phi^o$  can be deduced identically:

$$\int P(\vec{v} \mid \vec{f}, \Phi^v) P(\Phi^v \mid \vec{\beta}) d\Phi^v \quad (23)$$

$$= \left( \frac{\Gamma(\sum_{t=1}^{V_v} \beta_t)}{\prod_{t=1}^{V_v} \Gamma(\beta_t)} \right)^K \prod_{k=1}^K \frac{\prod_{t=1}^{V_v} \Gamma(\beta_t + n_{v,k}^{(t)})}{\Gamma(\sum_{t=1}^{V_v} (\beta_t + n_{v,k}^{(t)}))} \quad (24)$$

$$\int P(\vec{o} \mid \vec{f}, \Phi^o) P(\Phi^o \mid \vec{\beta}) d\Phi^o \quad (25)$$

$$= \left( \frac{\Gamma(\sum_{t=1}^{V_o} \beta_t)}{\prod_{t=1}^{V_o} \Gamma(\beta_t)} \right)^K \prod_{k=1}^K \frac{\prod_{t=1}^{V_o} \Gamma(\beta_t + n_{o,k}^{(t)})}{\Gamma(\sum_{t=1}^{V_o} (\beta_t + n_{o,k}^{(t)}))} \quad (26)$$

And

$$P(\vec{f} \mid \vec{\alpha}) \quad (27)$$

$$= \int P(\vec{f} \mid \vec{\theta}) P(\vec{\theta} \mid \vec{\alpha}) d\vec{\theta} \quad (28)$$

$$= \int \prod_{m=1}^M \prod_{k=1}^K \theta_{m,k}^{n_m^{(k)}} \prod_{m=1}^M P(\vec{\theta}_k \mid \vec{\alpha}) d\vec{\theta}_k \quad (29)$$

$$= \int \prod_{m=1}^M \prod_{k=1}^K \theta_{m,k}^{n_m^{(k)}} \prod_{m=1}^M \left( \frac{\Gamma(\sum_{t=1}^K \alpha_k)}{\prod_{t=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_{m,k}^{\alpha_k - 1} \right) d\vec{\theta}_k \quad (30)$$

$$= \left( \frac{\Gamma(\sum_{t=1}^K \alpha_k)}{\prod_{t=1}^K \Gamma(\alpha_k)} \right)^K \prod_{m=1}^M \frac{\prod_{k=1}^K \Gamma(n_m^{(k)} + \alpha_k)}{\Gamma(\sum_{k=1}^K n_m^{(k)} + \alpha_k)} \quad (31)$$

$n_m^{(k)}$  means number of tokens in m are assigned to topic k.  
Thus, the **full conditional probability** is:

$$P(f_i \mid \vec{f}_{-i}, \vec{s}, \vec{v}, \vec{o}, \vec{\alpha}, \vec{\beta}) \quad (32)$$

$$= \frac{P(\vec{f}, \vec{s}, \vec{v}, \vec{o} \mid \vec{\alpha}, \vec{\beta})}{P(\vec{f}_{-i}, \vec{s}, \vec{v}, \vec{o} \mid \vec{\alpha}, \vec{\beta})} \quad (33)$$

$$= \frac{P(\vec{s}, \vec{v}, \vec{o} \mid \vec{f}, \vec{\beta}) P(\vec{f} \mid \vec{\alpha})}{P(\vec{s}_{-i}, \vec{v}_{-i}, \vec{o}_{-i} \mid \vec{f}_{-i}, \vec{\beta}) P(s_i, v_i, o_i \mid \vec{\beta}) P(\vec{f}_{-i} \mid \vec{\alpha})} \quad (34)$$

$$\propto \frac{P(\vec{s} \mid \vec{f}, \vec{\beta}) P(\vec{v} \mid \vec{f}, \vec{\beta}) P(\vec{o} \mid \vec{f}, \vec{\beta}) P(\vec{f} \mid \vec{\alpha})}{P(\vec{s}_{-i} \mid \vec{f}_{-i}, \vec{\beta}) P(\vec{v}_{-i} \mid \vec{f}_{-i}, \vec{\beta}) P(\vec{o}_{-i} \mid \vec{f}_{-i}, \vec{\beta}) P(\vec{f}_{-i} \mid \vec{\alpha})} \quad (35)$$

$$\propto \frac{P(\vec{s} \mid \vec{f}, \vec{\beta})}{P(\vec{s}_{-i} \mid \vec{f}_{-i}, \vec{\beta})} \cdot \frac{P(\vec{v} \mid \vec{f}, \vec{\beta})}{P(\vec{v}_{-i} \mid \vec{f}_{-i}, \vec{\beta})} \cdot \frac{P(\vec{o} \mid \vec{f}, \vec{\beta})}{P(\vec{o}_{-i} \mid \vec{f}_{-i}, \vec{\beta})} \cdot \frac{P(\vec{f} \mid \vec{\alpha})}{P(\vec{f}_{-i} \mid \vec{\alpha})} \quad (36)$$

$$\propto P(\vec{s} \mid \vec{f}, \vec{\beta}) \cdot P(\vec{v} \mid \vec{f}, \vec{\beta}) \cdot P(\vec{o} \mid \vec{f}, \vec{\beta}) \cdot P(\vec{f} \mid \vec{\alpha}) \quad (37)$$

Also, Taking  $P(\vec{s} \mid \vec{f}, \vec{\beta})$  as an example:

$$P(\vec{s} \mid \vec{f}, \vec{\beta}) \quad (38)$$

$$= \left( \frac{\Gamma(\sum_{t=1}^{V_s} \beta_t)}{\prod_{t=1}^{V_s} \Gamma(\beta_t)} \right)^K \prod_{k=1}^K \frac{\prod_{t=1}^{V_s} \Gamma(\beta_t + n_{s,k}^{(t)})}{\Gamma(\sum_{t=1}^{V_s} (\beta_t + n_{s,k}^{(t)}))} \quad (39)$$

$$\propto \prod_{k=1}^K \frac{\prod_{t=1}^{V_s} \Gamma(n_{s,k}^{(t)} + \beta_t)}{\Gamma(\sum_{t=1}^{V_s} (n_{s,k}^{(t)} + \beta_t))} \quad (40)$$

Remove factor not related to current subject:

$$= \prod_{k=1}^K \frac{\prod_{j \neq t}^{V_s} \Gamma(n_{s,k}^{(j)} + \beta_j) \cdot \Gamma(n_{s,k}^{(t)} + \beta_t)}{\Gamma(\sum_{t=1}^{V_s} (n_{s,k}^{(t)} + \beta_t))} \quad (41)$$

$$\propto \prod_{k=1}^K \frac{\Gamma(n_{s,k}^{(t)} + \beta_t)}{\Gamma(\sum_{t=1}^{V_s} (n_{s,k}^{(t)} + \beta_t))} \quad (42)$$

Let  $n_{s,k,-i}^{(t)}$  represents times of subjects except for the  $i$ -th subject that assigned to topic  $k$ .

$$= \prod_{k=1}^K \frac{\Gamma(n_{s,k,-i}^{(t)} + \beta_t + 1)}{\Gamma(1 + \sum_{t=1}^{V_s} (n_{s,k,-i}^{(t)} + \beta_t))} \quad (43)$$

Remove factor not related to current topic:

$$= \prod_{k \neq z}^K \frac{\Gamma(n_{s,k,-i}^{(t)} + \beta_t + 1)}{\Gamma(1 + \sum_{t=1}^{V_s} (n_{s,k,-i}^{(t)} + \beta_t))} \cdot \frac{\Gamma(n_{s,z,-i}^{(t)} + \beta_t + 1)}{\Gamma(1 + \sum_{t=1}^{V_s} (n_{s,z,-i}^{(t)} + \beta_t))} \quad (44)$$

$$\propto \frac{\Gamma(n_{s,z,-i}^{(t)} + \beta_t + 1)}{\Gamma(1 + \sum_{t=1}^{V_s} (n_{s,z,-i}^{(t)} + \beta_t))} \quad (45)$$

Use  $\Gamma(x+1) = x\Gamma(x)$ , we get:

$$= \frac{n_{s,z,-i}^{(t)} + \beta_t}{\sum_{t=1}^{V_s} (n_{s,z,-i}^{(t)} + \beta_t)} \frac{\Gamma(n_{s,z,-i}^{(t)} + \beta_t)}{\Gamma(\sum_{t=1}^{V_s} (n_{s,z,-i}^{(t)} + \beta_t))} \quad (46)$$

$$\propto \frac{n_{s,z,-i}^{(t)} + \beta_t}{\sum_{t=1}^{V_s} (n_{s,z,-i}^{(t)} + \beta_t)} \quad (47)$$

So the first part of (32) can be computed:

$$\propto \frac{n_{s,z,-i}^{(t)} + \beta_t}{\sum_{t=1}^{V_s} (n_{s,z,-i}^{(t)} + \beta_t)} \cdot \frac{n_{v,z,-i}^{(t)} + \beta_t}{\sum_{t=1}^{V_v} (n_{v,z,-i}^{(t)} + \beta_t)} \cdot \frac{n_{o,z,-i}^{(t)} + \beta_t}{\sum_{t=1}^{V_o} (n_{o,z,-i}^{(t)} + \beta_t)} \quad (48)$$

And the second part of (32) can be computed in the same way:

$$P(\vec{f} | \vec{\alpha}) \tag{49}$$

$$\propto \frac{n_{m,-i}^{(t)} + \alpha_k}{(\sum_{k=1}^K (n_m^k + \alpha_k)) - 1} \tag{50}$$

so the full conditional probability distribution of (32) is proportional to:

$$\frac{n_{s,z,-i}^{(t)} + \beta_t}{\sum_{t=1}^{V_s} (n_{s,z,-i}^{(t)} + \beta_t)} \cdot \frac{n_{v,z,-i}^{(t)} + \beta_t}{\sum_{t=1}^{V_v} (n_{v,z,-i}^{(t)} + \beta_t)} \cdot \frac{n_{o,z,-i}^{(t)} + \beta_t}{\sum_{t=1}^{V_o} (n_{o,z,-i}^{(t)} + \beta_t)} \cdot \frac{n_{m,-i}^{(t)} + \alpha_k}{(\sum_{k=1}^K (n_m^k + \alpha_k)) - 1} \tag{51}$$

**Multimonial parameters.** The state of Markov chain is  $S = \{\vec{s}, \vec{v}, \vec{o}, \vec{f}\}$

$$P(\vec{\theta}_m \mid S, \vec{\alpha}) = Dir(\vec{\theta}_m \mid \vec{n}_m + \vec{\alpha}) \quad (52)$$

$$P(\vec{\phi}_k^s \mid S, \vec{\beta}) = Dir(\vec{\phi}_k^s \mid \vec{n}_{s,k} + \vec{\beta}) \quad (53)$$

$$P(\vec{\phi}_k^v \mid S, \vec{\beta}) = Dir(\vec{\phi}_k^v \mid \vec{n}_{v,k} + \vec{\beta}) \quad (54)$$

$$P(\vec{\phi}_k^o \mid S, \vec{\beta}) = Dir(\vec{\phi}_k^o \mid \vec{n}_{o,k} + \vec{\beta}) \quad (55)$$

By using  $E(Dir(\vec{\alpha})) = a_i / \sum_i a_i$ , we get:

$$\phi_{k,t}^s = \frac{n_{s,k}^{(t)} + \beta_t}{\sum_{t=1}^{V_s} (n_{s,k}^{(t)} + \beta_t)} \quad (56)$$

$$\phi_{k,t}^v = \frac{n_{v,k}^{(t)} + \beta_t}{\sum_{t=1}^{V_v} (n_{v,k}^{(t)} + \beta_t)} \quad (57)$$

$$\phi_{k,t}^o = \frac{n_{o,k}^{(t)} + \beta_t}{\sum_{t=1}^{V_o} (n_{o,k}^{(t)} + \beta_t)} \quad (58)$$

$$\theta_{m,k} = \frac{n_m^{(k)} + \alpha_k}{\sum_{k=1}^K (n_m^{(k)} + \alpha_k)} \quad (59)$$

Using (56), (57), (58), (59) and (51), the Gibbs sampling can runs.

### 3.3 Pseudocode of Gibbs sampling

## 4 Ideas for final project

Ideas