A latent model for inducing frame semantic roles

Yang Zhou

July 17, 2013

1 Introduction

2 Model Concepts

2.1 Graph

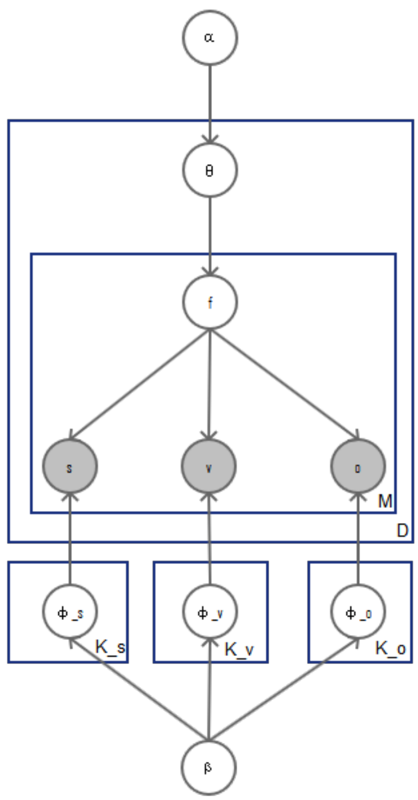

Figure 1: Plate graph

$$\theta \sim Dirichlet(\alpha)$$
$$\phi_s \sim Dirichlet(\beta)$$
$$\phi_v \sim Dirichlet(\beta)$$
$$\phi_o \sim Dirichlet(\beta)$$

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

2.2 Generating process

1. Sample frame distribution θ from $Dirichlet(\alpha)$
2. Sample subject distribution ϕ_s from $Dirichlet(\beta)$
3. Sample predicate distribution ϕ_v from $Dirichlet(\beta)$
4. Sample object distribution ϕ_o from $Dirichlet(\beta)$

   Document-Frame

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

   Frame-Roles

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

3 Model Detail

A deduction of the model.

3.1 Likelihood

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

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

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

Likelihood of a document:

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

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

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

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

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

Likelihood of a corpus:

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

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

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

3.2 Inference

The target of inference is the distribution:

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

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

Because computing (12) directly costs high.

Joint Distribution

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

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

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

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

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

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

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:

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

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

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

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

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

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

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:

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

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

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

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

And

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Remove factor not related to current subject:

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

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

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

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

Remove factor not related to current topic:

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

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

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

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

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

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

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

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

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

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

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 $S = \{\vec{s}, \vec{v}, \vec{o}, \vec{f}\}$

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

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

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

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

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

$$\phi^s_{k,t} = \frac{n^{(t)}_{s,k} + \beta_t}{\Sigma^{V_s}_{t=1}(n^{(t)}_{s,k} + \beta_t)} \tag{56}$$

$$\phi^v_{k,t} = \frac{n^{(t)}_{v,k} + \beta_t}{\Sigma^{V_v}_{t=1}(n^{(t)}_{v,k} + \beta_t)} \tag{57}$$

$$\phi^o_{k,t} = \frac{n^{(t)}_{o,k} + \beta_t}{\Sigma^{V_o}_{t=1}(n^{(t)}_{o,k} + \beta_t)} \tag{58}$$

$$\theta_{m,k} = \frac{n^{(k)}_m + \alpha_k}{\Sigma^{K}_{k=1}(n^{(k)}_m + \alpha_k)} \tag{59}$$

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

3.3 Pseudocode of Gibbs sampling

4 Ideas for final project

Ideas