A Probabilistic Model for Recursive Factorized Image Features: Supplemental Materials

Sergey Karayev Mario Fritz Sanja Fidler Trevor Darrell

April 4, 2011

1 Extended Derivation

1.1 Derivation of the Gibbs Sampling equations

We derive the Gibbs Sampling formulas for our recursive LDA model. The derivation is analogous to the original - but adds the recursion and spatial grouping in each layer.

$$p(w, z_{0,\dots,L} | x_{0,\dots,L}, \alpha, \beta_{0,\dots,L}) \tag{1}$$

$$= \underbrace{\int_{\theta} p(\theta|\alpha) p(z_L|\theta) d\theta}_{\text{top layer } L} \prod_{l=1}^{L-1} \underbrace{\int_{\phi_l} p(\phi_l|\beta_l) p(z_{l-1}|\phi_l, z_l, x_l) d\phi_l}_{\text{layer } l} \underbrace{\int_{\phi_0} p(w|\phi_0, z_0, x_0) p(\phi_0|\beta) d\phi_0}_{\text{evidence layer}} \tag{2}$$

For each of the three parts we integrate out the multinomial parameters. Given the conjugate Dirichlet prior we obtain a closed-form solution:

top layer:

$$p(z_L = t_L|\alpha) = \underbrace{\int_\theta p(\theta|\alpha)p(z_L|\theta)d\theta}_{\text{top layer}} \tag{3}$$

$$= \prod_{d=1}^{D} \int_{\theta^{(d)}} p(\theta^{(d)}|\alpha) \prod_{n=1}^{N(d)} p(z_L^{(d,n)}|\theta^{(d)})d\theta^{(d)} \tag{4}$$

$$= \prod_{d=1}^{D} \int_{\theta^{(d)}} \frac{\Gamma(T_L\alpha)}{\Gamma(\alpha)^{T_L}} \prod_{t_L=1}^{T_L} \left(\theta^{(d,t_L)}\right)^{\alpha-1} \prod_{t_L=1}^{T_L} \left(\theta^{(d,t_L)}\right)^{\#(z_L^{(d,\cdot)}=t_L)} d\theta^{(d)} \tag{5}$$

$$= \prod_{d=1}^{D} \int_{\theta^{(d)}} \frac{\Gamma(T_L\alpha)}{\Gamma(\alpha)^{T_L}} \prod_{t_L=1}^{T_L} \left(\theta^{(d,t_L)}\right)^{\#(z_L^{(d,\cdot)}=t_L)+\alpha-1} d\theta^{(d)} \tag{6}$$

$$= \prod_{d=1}^{D} \frac{\Gamma(T_L\alpha)}{\Gamma(\alpha)^{T_L}} \frac{\prod_{t_L=1}^{T_L} \Gamma(\alpha + \#(z_L^{(d,\cdot)} = t_L))}{\Gamma(T_L\alpha + \sum_{t_L=1}^{T_L} \#(z_L^{(d,\cdot)} = t_L))} \tag{7}$$

$$\underbrace{\int_{\theta^{(d)}} \frac{\Gamma(T_L\alpha + \sum_{t_L=1}^{T_L} \#(z_L^{(d,\cdot)} = t_L))}{\prod_{t_L=1}^{T_L} \Gamma(\alpha + \#(z_L^{(d,\cdot)} = t_L))} \prod_{t_L=1}^{T_L} \left(\theta^{(d,t_L)}\right)^{\#(z_L^{(d,\cdot)}=t_L)+\alpha-1} d\theta^{(d)}}_{\int_{\theta^{(d)}} \mathrm{Dir}(\#(z_L^{(d,\cdot)}=t_L)+\alpha)d\theta^{(d)}=1} \tag{8}$$

$$= \prod_{d=1}^{D} \frac{\Gamma(T_L\alpha)}{\Gamma(\alpha)^{T_L}} \frac{\prod_{t_L=1}^{T_L} \Gamma(\alpha + \#(z_L^{(d,\cdot)} = t_L))}{\Gamma(T_L\alpha + \sum_{t_L=1}^{T_L} \#(z_L^{(d,\cdot)} = t_L))} \tag{9}$$

$$= \prod_{d=1}^{D} \frac{\Gamma(T_L\alpha)}{\Gamma(\alpha)^{T_L}} \frac{\prod_{t_L=1}^{T_L} \Gamma(\alpha + \#(z_L^{(d,\cdot)} = t_L))}{\Gamma(T_L\alpha + \#(z_L^{(d,\cdot)}))} \tag{10}$$

intermediate layer:

$$S_l = X_l \times T_{l-1} \tag{11}$$

$$\text{e.g.} \quad \phi_l^{(t_l,\cdot,\cdot)} \in \mathbb{R}^{S_l} \tag{12}$$

$$p(z_{l-1} = t_{l-1}|z_l, x_l, \beta_l) = \underbrace{\int_{\phi_l} p(\phi_l|\beta_l) p(z_{l-1}|\phi_l, z_l, x_l) d\phi_l}_{\text{intermediate layer}} \tag{13}$$

$$= \prod_{t_l=1}^{T_l} \int_{\phi_l} p(\phi_l^{(t_l,\cdot,\cdot)}|\beta_l) \prod_{d=1}^{D} \prod_{n=1}^{N^{(d)}} p(z_{l-1}^{(d,n)}|\phi_l^{(t_l,x_l,\cdot)}) d\phi_l \tag{14}$$

$$= \prod_{t_l=1}^{T_l} \int_{\phi_l} \frac{\Gamma(|S_l|\ \beta_l)}{\Gamma(\beta_l)^{|S_l|}} \prod_{t_{l-1},x}^{S_l} \left(\phi_l^{(t_l,x,t_{l-1})}\right)^{\beta_l - 1} \prod_{t_{l-1},x}^{S_l} \left(\phi_l^{(t_l,x,t_{l-1})}\right)^{\overbrace{\#(z_{l-1} = t_{l-1} \wedge z_l = t_l \wedge x_l = x)}^{\text{expr}}} d\phi_l \tag{15}$$

$$= \prod_{t_l=1}^{T_l} \int_{\phi_l} \frac{\Gamma(|S_l|\beta_l)}{\Gamma(\beta_l)^{|S_l|}} \prod_{t_{l-1},x}^{S_l} \left(\phi_l^{(t_l,x,t_{l-1})}\right)^{\text{expr}+\beta_l-1} d\phi_l \tag{16}$$

$$= \prod_{t_l=1}^{T_l} \frac{\Gamma(|S_l|\ \beta_l)}{\Gamma(\beta_l)^{|S_l|}} \frac{\prod_{t_{l-1},x}^{S_l} \Gamma(\beta_l + \text{expr})}{\Gamma(|S_l|\ \beta_l + \sum_{t_{l-1},x}^{S_l} \text{expr})} \tag{17}$$

$$\int_{\phi_l} \frac{\Gamma(|S_l|\beta_l) + \sum_{t_{l-1},x}^{S_l} \text{expr})}{\prod_{t_{l-1},x}^{S_l} \Gamma(\beta_l + \text{expr})} \prod_{t_{l-1},x}^{S_l} \left(\phi_l^{(t_l,x,t_{l-1})}\right)^{\text{expr}+\beta_l-1} d\phi_l \tag{18}$$

$$= \prod_{t_l=1}^{T_l} \frac{\Gamma(|S_l|\ \beta_l)}{\Gamma(\beta_l)^{|S_l|}} \frac{\prod_{t_{l-1},x}^{S_l} \Gamma(\beta_l + \text{expr})}{\Gamma(|S_l|\ \beta_l + \sum_{t_{l-1},x}^{S_l} \text{expr})} \tag{19}$$

$$= \prod_{t_l=1}^{T_l} \frac{\Gamma(|S_l|\ \beta_l)}{\Gamma(\beta_l)^{|S_l|}} \frac{\prod_{t_{l-1},x}^{S_l} \Gamma(\beta_l + \#(z_{l-1} = t_{l-1} \wedge z_l = t_l \wedge x_l = x))}{\Gamma(|S_l|\ \beta_l + \sum_{t_{l-1},x}^{S_l} \#(z_{l-1} = t_{l-1} \wedge z_l = t_l \wedge x_l = x))} \tag{20}$$

$$= \prod_{t_l=1}^{T_l} \frac{\Gamma(|S_l|\ \beta_l)}{\Gamma(\beta_l)^{|S_l|}} \frac{\prod_{t_{l-1},x}^{S_l} \Gamma(\beta_l + \#(z_{l-1} = t_{l-1} \wedge z_l = t_l \wedge x_l = x))}{\Gamma(|S_l|\ \beta_l + \#(z_l = t_l))} \tag{21}$$

evidence layer: The derivation for the evidence layer is analogue to the intermediate layers:

$$S_0 = X_0 \times V \tag{22}$$

$$\text{e.g.}\quad \phi_0^{(t_0,\cdot,\cdot)} \in \mathbb{R}^{S_0} \tag{23}$$

$$p(w|z_0, x_0, \beta_0) = \underbrace{\int_{\phi_0} p(w|\phi_0, z_0, x_0) p(\phi_0|\beta) d\phi_0}_{\text{evidence layer}} \tag{24}$$

$$= \prod_{t_0=1}^{T_0} \int_{\phi_0} p(\phi_0^{(t_0,\cdot,\cdot)}|\beta_0) \prod_{d=1}^{D} \prod_{n=1}^{N^{(d)}} p(w^{(d,n)}|\phi_0^{(t_0,x_0,\cdot)}) d\phi_0 \tag{25}$$

$$= \prod_{t_0=1}^{T_0} \int_{\phi_0} \frac{\Gamma(|S_0|\,\beta_0)}{\Gamma(\beta_0)^{|S_0|}} \prod_{v,x}^{S_0} \left(\phi_0^{(t_0,x,v)}\right)^{\beta_0-1} \prod_{v,x}^{S_0} \left(\phi_0^{(t_0,x,v)}\right)^{\overbrace{\#(w=v \wedge z_0=t_0 \wedge x_0=x)}^{\text{expr}}} d\phi_0 \tag{26}$$

$$\tag{27}$$

$$= \prod_{t_0=1}^{T_0} \int_{\phi_0} \frac{\Gamma(|S_0|\,\beta_0)}{\Gamma(\beta_0)^{|S_0|}} \prod_{v,x}^{S_0} \left(\phi_0^{(t_0,x,v)}\right)^{\text{expr}+\beta_0-1} d\phi_0 \tag{28}$$

$$= \prod_{t_0=1}^{T_0} \frac{\Gamma(|S_0|\,\beta_0)}{\Gamma(\beta_0)^{|S_0|}} \frac{\prod_{v,x}^{S_0} \Gamma(\beta_0 + \text{expr})}{\Gamma(|S_0|\,\beta_0 + \sum_{v,x}^{S_0} \text{expr})} \int_{\phi_0} \frac{\Gamma(|S_0|\,\beta_0 + \sum_{v,x}^{S_0} \text{expr})}{\prod_{v,x}^{S_0} \Gamma(\beta_0 + \text{expr})} \prod_{v,x}^{S_0} \left(\phi_0^{(t_0,x,v)}\right)^{\text{expr}+\beta_0-1} d\phi_0 \tag{29}$$

$$= \prod_{t_0=1}^{T_0} \frac{\Gamma(|S_0|\,\beta_0)}{\Gamma(\beta_0)^{|S_0|}} \frac{\prod_{v,x}^{S_0} \Gamma(\beta_0 + \#(w=v \wedge z_0=t_0 \wedge x_0=x))}{\Gamma(|S_0|\,\beta_0 + \sum_{v,x}^{S_0} \#(w=v \wedge z_0=t_0 \wedge x_0=x))} \tag{30}$$

$$= \prod_{t_0=1}^{T_0} \frac{\Gamma(|S_0|\,\beta_0)}{\Gamma(\beta_0)^{|S_0|}} \frac{\prod_{v,x}^{S_0} \Gamma(\beta_0 + \#(w=v \wedge z_0=t_0 \wedge x_0=x))}{\Gamma(|S_0|\,\beta_0 + \#(z_0=t_0))} \tag{31}$$

The conditional probabilities for a single z can be expressed as follows:

top layer:

$$p(z_L^{(d,n)}|z_L^{\overline{(d,n)}}, z_{L-1}, x_L, \alpha, \beta_L) = \frac{p(z_L, z_{L-1}|\alpha, x_L, \beta_L)}{p(z_L^{\overline{(d,n)}}, z_{L-1}|\alpha, x_L, \beta_L)} \tag{32}$$

$$= \frac{p(z_L|\alpha)}{p(z_L^{\overline{(d,n)}}|\alpha)} \frac{p(z_{L-1}|z_L, x_L, \beta_L)}{p(z_{L-1}|z_L^{\overline{(d,n)}}, x_L, \beta_L)} \tag{33}$$

intermediate layer:

$$p(z_l^{(d,n)}|z_l^{\overline{(d,n)}}, z_{l-1}, z_{l+1}, x_l, x_{l+1}, \beta_l, \beta_{l+1}) \tag{34}$$
$$= \frac{p(z_l, z_{l-1}|z_{l+1}, x_l, x_{l+1}, \beta_l, \beta_{l+1})}{p(z_l^{\overline{(d,n)}}, z_{l-1}|z_{l+1}, x_l, x_{l+1}, \beta_l, \beta_{l+1})} \tag{35}$$
$$= \frac{p(z_{l-1}|z_l, x_l, \beta_l)}{p(z_{l-1}|z_l^{\overline{(d,n)}}, x_l, \beta_l)} \frac{p(z_l|z_{l+1}, x_{l+1}, \beta_{l+1})}{p(z_l^{\overline{(d,n)}}|z_{l+1}, x_{l+1}, \beta_{l+1})} \tag{36}$$

evidence layer:

$$p(z_0^{(d,n)}|z_0^{\overline{(d,n)}}, z_1, x_l, x_{l+1}, \beta_0, \beta_1) \tag{37}$$
$$= \frac{p(z_0, w|z_1, x_0, x_1, \beta_0, \beta_1)}{p(z_0^{\overline{(d,n)}}, w|z_1, x_0, x_1, \beta_0, \beta_1)} \tag{38}$$
$$= \frac{p(w|z_0, x_0, \beta_0)}{p(w|z_0^{\overline{(d,n)}}, x_0, \beta_0)} \frac{p(z_0|z_1, x_1, \beta_1)}{p(z_0^{\overline{(d,n)}}|z_1, x_1, \beta_1)} \tag{39}$$

This leaves us with 3 types of terms that we have to compute:

$$\frac{p(z_L|\alpha)}{p(z_L^{\overline{(d,n)}}|\alpha)} = \frac{\prod_{d=1}^{D} \frac{\Gamma(T_L\alpha)}{\Gamma(\alpha)^{T_L}} \frac{\prod_{t_L=1}^{T_L} \Gamma(\alpha+\#(z_L^{(d,\cdot)}=t_L))}{\Gamma(T_L\alpha+\#(z_L^{(d,\cdot)}))}}{\prod_{d=1}^{D} \frac{\Gamma(T_L\alpha)}{\Gamma(\alpha)^{T_L}} \frac{\prod_{t_L=1}^{T_L} \Gamma(\alpha+\#(z_L^{\overline{(d,n)}}=t_L))}{\Gamma(T_L\alpha+\#(z_L^{\overline{(d,n)}}))}} \tag{40}$$

$$= \frac{\alpha + \#(z_L^{\overline{(d,n)}} = t_L)}{T_L\alpha + \#(z_L^{\overline{(d,n)}})} \tag{41}$$

$$\frac{p(z_{l-1}|z_l, x_l, \beta_l)}{p(z_{l-1}|z_l^{\overline{(d,n)}}, x_l, \beta_l)} = \frac{\prod_{t_l=1}^{T_l} \frac{\Gamma(|S_l|\ \beta_l)}{\Gamma(\beta_l)^{|S_l|}} \frac{\prod_{t_{l-1},x}^{S_l} \Gamma(\beta_l+\#(z_{l-1}=t_{l-1}\wedge z_l=t_l\wedge x_l=x))}{\Gamma(|S_l|\ \beta_l+\#(z_l=t_l))}}{\prod_{t_l=1}^{T_l} \frac{\Gamma(|S_l|\ \beta_l)}{\Gamma(\beta_l)^{|S_l|}} \frac{\prod_{t_{l-1},x}^{S_l} \Gamma(\beta_l+\#(z_{l-1}=t_{l-1}\wedge z_l^{\overline{(d,n)}}=t_l\wedge x_l=x))}{\Gamma(|S_l|\ \beta_l+\#(z_l^{\overline{(d,n)}}=t_l))}} \tag{42}$$

$$= \frac{\beta_l + \#(z_{l-1} = t_{l-1} \wedge z_l^{\overline{(d,n)}} = t_l \wedge x_l = x)}{|S_l|\ \beta_l + \#(z_l^{\overline{(d,n)}} = t_l)} \tag{43}$$

$$\frac{p(w|z_0, x_0, \beta_0)}{p(w|z_0^{\overline{(d,n)}}, x_0, \beta_0)} = \frac{\prod_{t_0=1}^{T_0} \frac{\Gamma(|S_0|\ \beta_0)}{\Gamma(\beta_0)^{|S_0|}} \frac{\prod_{v,x}^{S_0} \Gamma(\beta_0+\#(w=v\wedge z_0=t_0\wedge x_0=x))}{\Gamma(|S_0|\ \beta_0+\#(z_0=t_0))}}{\prod_{t_0=1}^{T_0} \frac{\Gamma(|S_0|\ \beta_0)}{\Gamma(\beta_0)^{|S_0|}} \frac{\prod_{v,x}^{S_0} \Gamma(\beta_0+\#(w=v\wedge z_0^{\overline{(d,n)}}=t_0\wedge x_0=x))}{\Gamma(|S_0|\ \beta_0+\#(z_0^{\overline{(d,n)}}=t_0))}} \tag{44}$$

$$= \frac{\beta_0 + \#(w = v \wedge z_0^{\overline{(d,n)}} = t_0 \wedge x_0 = x)}{|S_0|\ \beta_0 + \#(z_0^{\overline{(d,n)}} = t_0)} \tag{45}$$

1.2 Formal Definition of χ

The spatial distribution $\chi_0 \in \mathbb{R}^{T_0 \times X_0}$ and $\chi_1 \in \mathbb{R}^{T_1 \times X_1}$ are directly computed from ϕ_0 and ϕ_1 respectively by summing the multinomial coefficients over the vocabulary:

$$p\left(\chi_0^{(t_0,\cdot)} | \phi_0^{(t_0,\cdot,\cdot)}\right) = \begin{cases} 1 \ \text{if} \ \chi_0^{(t_0,x_0)} = \sum_{v=1}^{V} \phi_0^{(t_0,x_0,v)} \\ 0 \ \text{else} \end{cases} \tag{46}$$

$$p\left(\chi_1^{(t_1,\cdot)} | \phi_1^{(t_1,\cdot,\cdot)}\right) = \begin{cases} 1 \ \text{if} \ \chi_1^{(t_1,x_1)} = \sum_{t_0=1}^{T_0} \phi_1^{(t_1,x_1,t_0)} \\ 0 \ \text{else} \end{cases} \tag{47}$$