Deep Poisson Factor Modeling

Ricardo Henao, Zhe Gan, James Lu and Lawrence Carin
Department of Electrical and Computer Engineering
Duke University, Durham, NC 27708
\{r.henao,zhe.gan,james.lu,lcarin\}@duke.edu

1 Graphical Model

Figure 1: Graphical models. (a) Poisson Factor Analysis (PFA) module. Nodes (b_k, \hat{z}_n and y_n) and edges drawn with dashed lines correspond to the discriminative PFA. (b) Deep Poisson factor model. Filled and empty nodes represent observed and latent variables, respectively.

2 Inference Details

2.1 MCMC

Conditional posteriors (layer index omitted for clarity):

- \( \psi_k \sim \text{Dirichlet}(\eta + x_{1k}, \ldots, \eta + x_{Mk}) \),
- \( \theta_{kn} \sim \text{Gamma}(r_k h_{kn} + x_{kn}, b^{-1}) \),
- \( h_{kn} \sim \delta(x_{kn} = 0)\text{Bernoulli}(\tilde{\pi}_{kn}(\tilde{\pi}_{kn} + 1 - \pi_{kn})^{-1}) + \delta(x_{kn} \geq 1) \),
- \( r_k \sim \text{Gamma} \left( 1 + \sum_n u_{kn}, 1 - \sum_n h_{kn} \log(1 - b) \right) \),
- \( z_{kn} \sim \delta(h_{kn} = 1)\text{Poisson}_+(\hat{\lambda}_{kn}) \).
where Poisson$_+ (\cdot)$ is the zero-truncated Poisson distribution and

$$
x_{mk} = \sum_{n=1}^{N} x_{mkn},
\quad x_{.kn} = \sum_{m=1}^{M} x_{mkn},
\quad \tilde{\pi}_{kn} = \pi_{kn} (1 - b)^{r_k},
\quad u_{kn} = \sum_{j=1}^{x_{.kn}} u_{knj}, \quad u_{knj} \sim \text{Bernoulli} \left( \frac{r_k}{r_k + j - 1} \right),
$$

(1)

Note that for multilayer models, $\pi^{(l)}_{kn} = 1 - \exp(\lambda^{(l+1)}_{kn})$. The data augmentation scheme for $r_k$ via $u_{kn}$ is described in [1].

For the discriminative DPF A, let’s denote latent counts for $\hat{y}_n$ as $\hat{x}_{ckn}$, with summaries analogous to (1), as $\hat{x}_{ck}$ and $\hat{x}_{.ck}$. Then,

$$
b_k \sim \text{Dirichlet} (\zeta + \hat{x}_{1ck}, \ldots, \zeta + \hat{x}_{Cck}),
\quad \theta_{kn} \sim \text{Gamma} (E[r_k]E[h_{kn}] + \sum_{m=1}^{M} \phi_{mkn}, b^{-1}),
\quad h_{kn} \sim \delta(x_{.kn} = 0) \text{Bernoulli}(\tilde{\pi}_{kn} (\tilde{\pi}_{kn} + 1 - \pi_{kn})^{-1}) + \delta(x_{.kn} \geq 1),
\quad r_k \sim \text{Gamma} \left( 1 + \sum_{n} E[u_{kn}], 1 - \sum_{n} E[p(h_{kn} = 1)] \log(1 - b) \right),
\quad z_{kn} \sim \text{Poisson$_+ (\tilde{\lambda}_{kn})$},
$$

where $E[x_{mkn}] = \phi_{mkn}$, $E[\tilde{\pi}_{kn}] = \pi_{kn} (1 - b)^{E[r_k]}$ and $E[u_{kn}] = \sum_{j=1}^{x_{.kn}} E[r_k] (E[r_k] + j - 1)^{-1}$.

List of Figures

1. Graphical models .................................................. 1
2. Representative meta-topics obtained from 20 News .................................. 3
3. Graph representation obtained from 20 News ........................................ 4
4. Graph representation obtained from RCV1 ............................................ 5
5. Graph representation obtained from Wiki .............................................. 6
6. Representative meta-topics obtained from medical records data .................. 7
7. Graph representation obtained from medical records data .......................... 8
Figure 2: Representative meta-topics obtained from 20 News. Meta-topic weights $\psi_k^{(2)}$ vs. layer-1 topics indices, with word lists corresponding to the top four words in layer-1 topics, $\psi_k^{(1)}$. 
Figure 3: Graph representation obtained from 20 News. Meta-topics are denoted by circles and layer-1 topics as boxes, with word lists corresponding to the top four words in layer-1 topics, $\psi_k^{(1)}$. For clarity, we only show the top four connections between meta-topics and their associated topics.
Figure 4: Graph representation obtained from RCV1. Meta-topics are denoted by circles and layer-1 topics as boxes, with word lists corresponding to the top four words in layer-1 topics, $\psi_k^{(1)}$. For clarity, we only show the top four connections between meta-topics and their associated topics.
Figure 5: Graph representation obtained from Wiki. Meta-topics are denoted by circles and layer-1 topics as boxes, with word lists corresponding to the top four words in layer-1 topics, $\psi(1)$. For clarity, we only show the top four connections between meta-topics and their associated topics.
Figure 6: Representative meta-topics obtained from medical records data. Meta-topic weights $\psi_k^{(2)}$ vs. layer-1 topics indices, with word lists corresponding to the top four words in layer-1 topics, $\psi_k^{(1)}$. 
Figure 7: Graph representation obtained from medical records data. Meta-topics are denoted by circles and layer-1 topics as boxes, with word lists corresponding to the top four words in layer-1 topics, $\psi^{(1)}$. For clarity, we only show the top four connections between meta-topics and their associated topics.
References