Supplementary: Scalable Bilinear Non-negative Latent Factor Models for Multi-Relational Data

1 Gibbs Sampling for BPBFM-2

Just as we did for BPBFM-1, we can express each latent count $\mathcal{X}^r_{ij}$ in BPBFM-2 (which models each Λ^r as $\Lambda^r = \sum_{m=1}^M \eta_{mr} G^m$) as a sum of the following form: $\mathcal{X}^r_{ij} = \sum_{k_1}^K \sum_{k_2}^K \sum_m^M X^r_{ik_1k_2mj}$ where $\mathcal{X}^r_{ik_1k_2mj} \sim \text{Poisson}(u_{ik_1}\eta_{mr}G^m_{k_1k_2}u_{jk_2})$. We further define $\mathcal{X}^{\cdot}_{\cdot k_1k_2m\cdot} = \sum_i^N \sum_j^N \sum_{r=1}^R \mathcal{X}^r_{ik_1k_2mj}$ and $\mathcal{X}^r_{\cdots m\cdot} = \sum_i^N \sum_j^N \sum_{k_1=1}^K \sum_{k_2=1}^K \mathcal{X}^r_{ik_1k_2mj}$, using additive property of Poisson distribution

$$\mathcal{X}^{\cdot}_{\cdot k_1k_2m\cdot} \sim \text{Poisson}(\theta_{k_1k_2}G^m_{k_1k_2}\sum_{r=1}^R \eta_{mr}) \quad (1)$$

$$\mathcal{X}^r_{\cdots m\cdot} \sim \text{Poisson}(\eta_{mr}\sum_{k_1=1}^K\sum_{k_2=1}^K \theta_{k_1k_2}G^m_{k_1k_2}) \quad (2)$$

With these defined, we proceed to give the update equations for the Gibbs sampler for BPBFM-2.

Sampling $\mathcal{X}^r_{ij}$: X^r_{ij} is sampled just as in model-1.

Sampling $\mathcal{X}^r_{ik_1k_2mj}$: $\mathcal{X}^r_{ik_1k_2mj}$ can be sampled as

$$\mathcal{X}^r_{ik_1k_2j} \sim \text{Mult}(\mathcal{X}^r_{ij}; \frac{u_{ik_1}\eta_{mr}G^m_{k_1k_2}u_{jk_2}}{\sum_{k_1=1}^K\sum_{k_2=1}^K u_{ik_1}\Lambda^r_{k_1k_2}u_{jk_2})} \quad (3)$$

Sampling $\mathbf{U}_{:,k}$: Using Dirichlet-multinomial conjugacy, each column of $\mathbf{U}$ can be sampled as

$$\mathbf{U}_{:,k} \sim \text{Dir}(a + \mathcal{X}^{\cdot}_{1k\cdots}, a + \mathcal{X}^{\cdot}_{2k\cdots}, \ldots, a + \mathcal{X}^{\cdot}_{Nk\cdots}) \quad (4)$$

where $\mathcal{X}^{\cdot}_{ik\cdots} = \sum_{k_2=1}^K \sum_{j=1}^N \sum_{m=1}^M \sum_{r=1}^R \mathcal{X}^r_{ikk_2mj}$.

Sampling d^m_m:. Marginalizing out $G^m_{k_1k_2}$ from Eq.(1), we have

$$\mathcal{X}^{\cdot}_{\cdot k_1k_2m\cdot} \sim \text{NegBin}((\epsilon^m)^{\delta_{k_1k_2}}d^m_{k_1}(d^m_{k_2})^{1-\delta_{k_1k_2}}, p_{k_1k_2}) \quad (5)$$

where $p_{k_1k_2} = \frac{\theta_{k_1k_2}\sum_{r=1}^R \eta_{mr}}{\theta_{k_1k_2}\sum_{r=1}^R \eta_{mr}+\beta}$. Using the data augmentation scheme proposed we used for BPBFM-1, d^m_k can be sampled by first sampling

$$\ell^m_{kk_2} \sim \sum_{t=1}^{X^{\cdot}_{\cdot k_1k_2m\cdot}} \text{Bern}(\frac{(\epsilon^m)^{\delta_{k_1k_2}}d^m_{k_1}(d^m_{k_2})^{1-\delta_{k_1k_2}}}{(\epsilon^m)^{\delta_{k_1k_2}}d^m_{k_1}(d^m_{k_2})^{1-\delta_{k_1k_2}}+t-1}) \quad (6)$$

and then sampling

$$d^m \sim \text{Ga}(\frac{\gamma_0}{K} + \sum_{k_2}^K \ell^m_{kk_2}, \frac{1}{c_0 - \sum_{k_2}^K (\epsilon^m)^{\delta_{kk_2}}(d^m_{k_2})^{1-\delta_{kk_2}}\ln(1-p_{kk_2})}) \quad (7)$$

Sampling ϵ^m: ϵ^m can be sampled as

$$\epsilon^r \sim \text{Ga}(e_0 + \sum_k^K \ell^r_{kk}, \frac{1}{f_0 - \sum_k^K d^m_k \ln(1-p_{kk})}) \quad (8)$$

Sampling $G^m_{k_1k_2}$: Using Gamma-Poisson conjugacy, $G^m_{k_1k_2}$ can be sampled by

$$G^m_{k_1k_2} \sim \text{Ga}((\epsilon^m)^{\delta_{k_1k_2}}d^r_{k_1}(d^r_{k_2})^{1-\delta_{k_1k_2}} + X^{\cdot}_{\cdot k_1k_2m\cdot}, \frac{1}{\beta + \theta_{k_1k_2}\sum_{r=1}^R \eta_{mr}}) \quad (9)$$

Sampling η_{mr}: Using equation (2) and Gamma-Poisson conjugacy, η_{mr} can be sampled by

$$\eta_{mr} \sim \text{Ga}(h_0 + \mathcal{X}^r_{\cdots m\cdot}, \frac{1}{q_0 + \sum_{k_1=1}^K\sum_{k_2=1}^K \theta_{k_1k_2}G^m_{k_1k_2}}) \quad (10)$$

2 Online Gibbs Sampling

In this section, we provide the details of the online Gibbs sampling algorithms for both of our models. Our online Gibbs sampling algorithms are based on the idea of the recently developed Bayesian Conditional Density Filtering (BCDF) framework (Guhaniyogi et al., 2014). They key idea in BCDF is to process data in small minibatches, and maintain and update sufficient statistics of the model parameters with each new minibatch of the data. In our models, these sufficient statistics are the latent counts.

2.1 Online Gibbs Sampling for BPBFM-1

Denoting I_t as indices of valid triplets in minibatch selected at iteration t, and I as the indices of all the valid triplets in training data. Define $\mathcal{X}^{r,t}_{ik\cdot\cdot} = \frac{|I|}{|I_t|}\sum_{k_2=1}^K\sum_{j=1,ij\in I_t}^N \mathcal{X}^r_{ikk_2j}$, $\mathcal{X}^t_{ik\cdot\cdot} = \frac{|I|}{|I_t|}\sum_{r=1}^R \mathcal{X}^{r,t}_{ik\cdot\cdot}$, and $\mathcal{X}^{r,t}_{\cdot k_1k_2\cdot} = \frac{|I|}{|I_t|}\sum_{i,j\in I_t}^N \mathcal{X}^{r,t}_{ik_1k_2j}$, where $|I|$ and I_t are cardinalities of the two sets. Then similar to batch Gibbs Sampling, define following quantities for $t \leq 2$: $\mathcal{X}^{r,t}_{ik\cdot\cdot} = (1-\rho)\mathcal{X}^{r,t-1}_{ik\cdot\cdot} + \rho\frac{|I|}{|I_t|}\sum_{k_2=1}^K\sum_{j=1,ij\in I_t}^N \mathcal{X}^r_{ikk_2j}$, $\mathcal{X}^t_{ik\cdot\cdot} = (1-\rho)\mathcal{X}^{t-1}_{ik\cdot\cdot} + \rho\frac{|I|}{|I_t|}\sum_{r=1}^R \mathcal{X}^{rt}_{ik\cdot\cdot}$, and $\mathcal{X}^{r,t}_{\cdot k_1k_2\cdot} = (1-\rho)\mathcal{X}^{r,t-1}_{\cdot k_1k_2\cdot} + \rho\frac{|I|}{|I_t|}\sum_{ij,ij\in I_t}^N \mathcal{X}^{r,t}_{ik_1k_2j}$. Here $\rho = (t+t_0)^{-w}$ is a decaying learning rate, as used in other online inference algorithms, such as stochastic variational inference (Hoffman et al., 2013). Here, $t_0 > 0$ and $w \in (0.5, 1]$ are required to guarantee convergence. With these defined, online Gibbs sampling at iteration t proceeds as:

Sampling $\mathbf{U}_{:,k}$: Each column of $\mathbf{U}$ can be sampled as

$$\mathbf{U}_{:,k} \sim \text{Dir}(a + \mathcal{X}^t_{1k\cdot\cdot}, a + \mathcal{X}^t_{2k\cdot\cdot}, \ldots, a + \mathcal{X}^t_{Nk\cdot\cdot}) \quad (11)$$

Sampling d^r_k: d^r_k can be sampled by first sampling

$$\ell^r_{kk_2} \sim \sum_{t=1}^{X^{r,t}_{\cdot k_1 k_2 \cdot}} \text{Bern}(\frac{(\epsilon^r)^{\delta_{k_1k_2}} d^r_{k_1} (d^r_{k_2})^{1-\delta_{k_1k_2}}}{(\epsilon^r)^{\delta_{k_1k_2}} d^r_{k_1} (d^r_{k_2})^{1-\delta_{k_1k_2}} + t - 1}) \quad (12)$$

and then sampling

$$d^r_k \sim \quad (13)$$
$$\text{Ga}(\frac{\gamma_0}{K} + \sum_{k_2}^K \ell^r_{kk_2}, \frac{1}{c_0 - \sum_{k_2}^K (\epsilon^r)^{\delta_{kk_2}} (d^r_{k_2})^{1-\delta_{kk_2}} \ln(1 - p_{kk_2})})$$

Sampling ϵ^r: ϵ^r can be sampled as

$$\epsilon^r \sim \text{Ga}(e_0 + \sum_k^K \ell^r_{kk}, \frac{1}{f_0 - \sum_k^K d^r_k \ln(1 - p_{kk})}) \quad (14)$$

Sampling $\Lambda^r_{k_1k_2}$: $\Lambda^r_{k_1k_2}$ can be sampled by

$$\Lambda^r_{k_1k_2} \sim \quad (15)$$
$$\text{Ga}((\epsilon^r)^{\delta_{k_1k_2}} d^r_{k_1} (d^r_{k_2})^{1-\delta_{k_1k_2}} + \mathcal{X}^{r,t}_{\cdot k_1k_2\cdot}, \frac{1}{\beta + \theta_{k_1k_2}})$$

$\mathcal{X}^r_{ij}$, ϵ^r and $\mathcal{X}^r_{ik_1k_2j}$ are sampled the same way as the batch Gibbs sampling.

2.2 Online Gibbs Sampling for BPBFM-2

Similar to online BPBFM-1, we define $\mathcal{X}^{\cdot,t}_{\cdot k_1k_2m\cdot} = (1-\rho)\mathcal{X}^{\cdot,t-1}_{\cdot k_1k_2m\cdot} + \rho\frac{|I|}{|I_t|}\sum_{ij,ij\in I_t}\sum_{r=1}^R \mathcal{X}^r_{ik_1k_2mj}$, $\mathcal{X}^{r,t}_{\cdot\cdot\cdot m\cdot} = (1-\rho)\mathcal{X}^{r,t-1}_{\cdot\cdot\cdot m\cdot} + \rho\frac{|I|}{|I_t|}\sum_{ij,ij\in I_t}\sum_{k_1=1}^K\sum_{k_2=1}^K \mathcal{X}^r_{ik_1k_2mj}$, and $\mathcal{X}^{\cdot,t}_{ik\cdots} = (1 - \rho)\mathcal{X}^{\cdot,t-1}ik\cdots + \rho\frac{|I|}{|I_t|}\sum_{k_2=1}^K\sum_{j=1,ij\in I_t}^N\sum_{m=1}^M\sum_{r=1}^R \mathcal{X}^r_{ikk_2mj}$. With these defined, we proceed to give the update equations for the online Gibbs sampler for model-2:

Sampling $\mathbf{U}_{:,k}$: Each column of $\mathbf{U}$ can be sampled as

$$\mathbf{U}_{:,k} \sim \text{Dir}(a + \mathcal{X}^{\cdot,t}_{1k\cdots}, a + \mathcal{X}^{\cdot,t}_{2k\cdots}, \ldots, a + \mathcal{X}^{\cdot,t}_{Nk\cdots}) \quad (16)$$

Sampling d^m_m: d^m_k can be sampled by first sampling

$$\ell^m_{kk_2} \sim \sum_{t=1}^{X^{\cdot,t}_{\cdot k_1k_2m\cdot}} \text{Bern}(\frac{(\epsilon^m)^{\delta_{k_1k_2}} d^m_{k_1} (d^m_{k_2})^{1-\delta_{k_1k_2}}}{(\epsilon^m)^{\delta_{k_1k_2}} d^m_{k_1} (d^m_{k_2})^{1-\delta_{k_1k_2}} + t - 1}) \quad (17)$$

and then sampling

$$d^m_k \sim \quad (18)$$
$$\text{Ga}(\frac{\gamma_0}{K} + \sum_{k_2}^K \ell^m_{kk_2}, \frac{1}{c_0 - \sum_{k_2}^K (\epsilon^m)^{\delta_{kk_2}} (d^m_{k_2})^{1-\delta_{kk_2}} \ln(1 - p_{kk_2})})$$

Sampling $G^m_{k_1k_2}$: Using Gamma-Poisson conjugacy, $G^m_{k_1k_2}$ can be sampled by

$$G^m_{k_1k_2} \sim \quad (19)$$
$$\text{Ga}((\epsilon^m)^{\delta_{k_1k_2}} d^r_{k_1} (d^r_{k_2})^{1-\delta_{k_1k_2}} + X^{\cdot,t}_{\cdot k_1k_2m\cdot}, \frac{1}{\beta + \theta_{k_1k_2}\sum_{r=1}^R \eta_{mr}})$$

Sampling η_{mr}: η_{mr} can be sampled by

$$\eta_{mr} \sim \text{Ga}(h_0 + \mathcal{X}^{r,t}_{\cdot\cdot\cdot m\cdot}, \frac{1}{q_0 + \sum_{k_1=1}^K\sum_{k_2=1}^K \theta_{k_1k_2} G^m_{k_1k_2}}) \quad (20)$$

$\mathcal{X}^r_{ij}$, $\mathcal{X}^r_{ik_1k_2mj}$, ϵ^m ϵ^m can be sampled the same way as the batch Gibbs sampling.

References

R. Guhaniyogi, S. Qamar, and D. B. Dunson. Bayesian conditional density filtering. *arXiv preprint arXiv:1401.3632*, 2014.

M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley. Stochastic variational inference. *JMLR*, 2013.