Poisson Tucker Decomposition version of the Interaction-pattern Partitioned Topic Model

Bomin Kim

March 13, 2018

1 Generative Process

To maintain single interaction pattern assignments (instead of admixture form which adds huge complexity in network history calculations), we assume an interaction-pattern distribution over C unique interaction patterns

$$\boldsymbol{\psi} \sim \text{Dirichlet}\Big(\zeta, (\frac{1}{C}, \dots, \frac{1}{C})\Big), \tag{1}$$

where ζ is the concentration parameter, and then each document $d \in [D]$ draws an interaction pattern c_d as below:

$$c_d \sim \text{Multinomial}(\boldsymbol{\psi}). \tag{2}$$

Next, we model the contents using Poisson Tucker Decomposition of Schein et al. (2016). First, each interaction pattern $c \in [C]$ has the IP-specific topic distribution

$$\theta_{ck} \sim \text{Gamma}(a_c, b_c), \tag{3}$$

and each topic $k \in [K]$ has the topic-word distribution

$$\phi_{kv} \sim \text{Gamma}(\epsilon_0, \epsilon_0). \tag{4}$$

Then, the number of tokens of type v in document d is

$$w_{dv} \sim \text{Poisson}(\sum_{k=1}^{K} \theta_{c_d k}\phi_{kv}), \tag{5}$$

Therefore, $\boldsymbol{w}_d = (w_{d1}, \dots, w_{dV})$ is a very sparse vector with $\text{sum}(\boldsymbol{w}_d) = N_d$.

For tie generating process, we use the current version of the IPTM. For every possible author–recipient pair $(a, r)_{a \neq r}$, we define the "recipient intensity", which is the likelihood of document d being sent from a to r:

$$\lambda_{adr} = \boldsymbol{b}_{c_d}{}^{\top} \boldsymbol{x}_{adrc_d}, \tag{6}$$

where we place a Normal prior $\boldsymbol{b}_c \sim N(\boldsymbol{\mu}_b, \Sigma_b)$. Similarly, we hypothesize "If a were the author of document d, when would it be sent?" and define the "timing rate" for author i

$$\mu_{ad} = g^{-1}(\boldsymbol{\eta}_{c_d}^{\top} \boldsymbol{w}_{adc_d}), \tag{7}$$

with a Normal prior $\boldsymbol{\eta}_c \sim N(\boldsymbol{\mu}_\eta, \Sigma_\eta)$. We then follow the generalized linear model framework:

$$\begin{aligned} E(\tau_{ad}) &= \mu_{ad}, \\ V(\tau_{ad}) &= V(\mu_{ad}). \end{aligned} \tag{8}$$

2 Derivation

We first derive the sampling equation of θ and ϕ, respectively.

$$\theta_{ck}|\text{rest} \sim \text{Gamma}(a_c + \boldsymbol{w}_{ck}, b_c + \sum_{d:c_d=c} \sum_{v=1}^{V} \phi_{kv}), \tag{9}$$

where $\boldsymbol{w}_{ck} = \sum\limits_{d:c_d=c} \sum_{v=1}^{V} w_{dkv}$ with $w_{dkv} \sim \text{Multinomial}(w_{dv}, \theta_{c_d k}\phi_{kv})$.

$$\phi_{kv}|\text{rest} \sim \text{Gamma}(\epsilon_0 + \boldsymbol{w}_{kv}, \epsilon_0 + \sum_{d=1}^{D} \theta_{c_d k}), \tag{10}$$

where $\boldsymbol{w}_{kv} = \sum_{d=1}^{D} w_{dkv}$ with $w_{dkv} \sim \text{Multinomial}(w_{dv}, \theta_{c_d k}\phi_{kv})$.

Since u_{adr} is a binary random variable, new values may be sampled directly using

$$\begin{aligned} P(u_{adr} = 1 | \boldsymbol{u}_{ad\backslash r}, \boldsymbol{c}, \boldsymbol{b}, \delta, \boldsymbol{x}) &\propto \exp\{\delta + \lambda_{adr}\}; \\ P(u_{adr} = 0 | \boldsymbol{u}_{ad\backslash r}, \boldsymbol{c}, \boldsymbol{b}, \delta, \boldsymbol{x}) &\propto \text{I}(\|\boldsymbol{u}_{ad\backslash r}\|_1 > 0), \end{aligned} \tag{11}$$

where $I(\cdot)$ is the indicator function that is used to prevent from the instances where the author has no recipients to send the document.

New values for continuous variables $\delta, \boldsymbol{b}$, and $\boldsymbol{\eta}$ and σ_τ^2 (if applicable) cannot be sampled directly from their conditional posteriors, but may instead be obtained using the Metropolis–Hastings algorithm. With uninformative priors (i.e., $N(0, \infty)$), the conditional posterior over δ and $\boldsymbol{b}$ is

$$\prod_{d=1}^{D}\prod_{a=1}^{A}\frac{\exp\Big\{\log\big(\mathrm{I}(\|\boldsymbol{u}_{ad}\|_1 > 0)\big) + \sum_{r\neq a}(\delta + \lambda_{adr})u_{adr}\Big\}}{Z(\delta, \boldsymbol{\lambda}_{ad})}, \tag{12}$$

where the two variables share the conditional posterior and thus can be jointly sampled. Likewise, assuming uninformative priors on $\boldsymbol{\eta}$ (i.e., $N(0, \infty)$) and σ_τ^2 (i.e., half-Cauchy(∞)), the conditional posterior is

$$\prod_{d=1}^{D}\Big(\varphi_\tau(\tau_d; \mu_{a_d d}, \sigma_\tau^2) \times \prod_{a\neq a_d}\big(1 - \Phi_\tau(\tau_d; \mu_{ad}, \sigma_\tau^2)\big)\Big). \tag{13}$$

Finally, for each document $d \in [D]$, we sample interaction-pattern assignment from the discrete distribution over C interaction patterns using

$$\begin{aligned}
&P(c_d = c|\boldsymbol{\theta}, \zeta, \boldsymbol{u}, \boldsymbol{a}, \boldsymbol{t})\\
&\propto P(c_d = c|\boldsymbol{c}_{\backslash d}, \zeta)P(\boldsymbol{\theta}_c|a_c, b_c, \boldsymbol{w}, c_d = c, \boldsymbol{c}_{\backslash d}, \boldsymbol{\theta}_{\backslash c})\\
&\times P(\boldsymbol{w}_d|c_d = c, \boldsymbol{\theta}_c, \boldsymbol{\phi})\\
&\times P(a_d, t_d|c_d = c, \boldsymbol{\eta}, \sigma_\tau^2)P(\boldsymbol{u}|c_d = c, \boldsymbol{c}_{\backslash d}, \boldsymbol{b}, \delta)\\
&\propto (\hat{N}_{c,\backslash d} + \frac{\zeta}{C})\\
&\times \prod_{k=1}^{K}\mathrm{dGamma}(\theta_{ck}; a_c + \boldsymbol{w}_{ck}, b_c + \sum_{d:c_d=c}\sum_{v=1}^{V}\phi_{kv})\\
&\times \prod_{v=1}^{V}\mathrm{dPois}(w_{dv}; \sum_{k=1}^{K}\theta_{ck}\phi_{kv})\\
&\times \varphi_\tau(\tau_d; \mu_{a_d d}, \sigma_\tau^2) \times \prod_{a\neq a_d}\big(1 - \Phi_\tau(\tau_d; \mu_{ad}, \sigma_\tau^2)\big)\\
&\times \prod_{a=1}^{A}\frac{\exp\Big\{\log\big(\mathrm{I}(\|\boldsymbol{u}_{ad}\|_1 > 0)\big) + \sum_{r\neq a}(\delta + \lambda_{adr})u_{adr}\Big\}}{Z(\delta, \boldsymbol{\lambda}_{ad})}.
\end{aligned} \tag{14}$$