HawkesTopic: A Joint Model for Network Inference and Topic Modeling from Text-Based Cascades - Supplementary Material

A. Variational Inference

The joint likelihood of observing all the events E is:

$$
\begin{aligned}
&P(E, \boldsymbol{\eta}, \boldsymbol{z}, \boldsymbol{P} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{A}, \boldsymbol{\mu}) \\
&= \underbrace{\prod_{e \in E} \left\{ \left[\prod_{e': t_{e'} < t_e} P(\eta_e | \eta_{e'})^{P_{e,e'}} \right] \cdot P(\eta_e | \alpha_{v_e})^{P_{e,0}} \right\}}_{(1)} \\
&\times \underbrace{\prod_{e \in E} \left[\prod_{n=1}^{N_e} P(z_{e,n} | \eta_e) P(x_{e,n} | z_{e,n}, \beta) \right]}_{(2)} \\
&\times \underbrace{\prod_{v \in V} \frac{e^{-m_v} m_v^{n_v}}{n_v!} \times \prod_{e \in E} \left[\frac{e^{-G_e} G_e^{\ell_e}}{\ell_e!} \cdot \frac{\prod_{e' \in E} \kappa_{e,e'}^{P_{e,e'}}}{G_e!} \right]}_{(3)}
\end{aligned}
$$

Recall that for each event e, we have a hidden variable P_e as the parent indicator as introduced in section **??**. $P_{e,e'} = 1$ if event e' triggers event e, while $P_{e,0} = 1$ indicates that the event e is spontaneous. Moreover, $\kappa_{e,e'} = \kappa_{e'}(t_e, v_e)$ is the impulse response of event e' on event e. $n_u = \sum_{e \in E: v_e = u} P_{e,0}$ is the number of spontaneous events of node u, while $m_u = \mu_u \times T$ corresponds to the expected number of spontaneous events of node u in observation window $[0, T]$. Similarly, $\ell_e = \sum_{e' \in E} P_{e',e}$ is the number of events triggered by e, while $G_e = \sum_{u \in V} \int \kappa_e(t, u) dt = \sum_{u \in V} A_{e_v, u}$ corresponds to the expected number of events triggered by e.

The first part (1) in the likelihood corresponds to the diffusion of topics, while the second part (2) provides the likelihood of content given the topic of the document. The third part (3) is the complete likelihood of the MHP model with the parent relationship (Veen & Schoenberg, 2008).

We apply the full mean-field approximation for the posteriors distribution $P(\boldsymbol{\eta}, \boldsymbol{z}, \boldsymbol{P} | E, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{A}, \boldsymbol{\mu})$. Under the standard variational theory, the inference task becomes minimizing the KL divergence between $Q(\boldsymbol{\eta}, \boldsymbol{z}, \boldsymbol{P})$ and $P(\boldsymbol{\eta}, \boldsymbol{z}, \boldsymbol{P} | E, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{A}, \boldsymbol{\mu})$. This is equivalent to maximizing a lower bound on the log marginal likelihood obtained by Jensen's inequality:

$$
\begin{aligned}
\log P(E) &= \log \int P(\boldsymbol{\eta}, \boldsymbol{z}, \boldsymbol{P}, E) d\boldsymbol{\eta} d\boldsymbol{z} d\boldsymbol{P} \\
&\geq \mathbb{E}_Q \left[\log P(\boldsymbol{\eta}, \boldsymbol{z}, \boldsymbol{P}, E) \right] - \mathbb{E}_Q \left[Q(\boldsymbol{\eta}, \boldsymbol{z}, \boldsymbol{P}) \right] \\
&= \mathcal{L}(Q) \,.
\end{aligned}
$$

The explicit form of the variational objective $\mathcal{L}(Q)$ is as follows:

$$
\begin{aligned}
\mathcal{L}(Q) = &\sum_{e \in E} \left\{ -\frac{1}{2\sigma^2} r_{e,0} ||\hat{\eta}_e - \alpha_{v_e}||_2^2 \right. \\
&\left. + \sum_{e': t_{e'} < t_e} -\frac{1}{2\sigma^2} r_{e,e'} ||\hat{\eta}_e - \hat{\eta}_{e'}||_2^2 \right\} \\
&+ \sum_{e \in E} \sum_{n=1}^{N_e} \sum_{k=1}^{K} \phi_{d,n,k} \mathbb{E}_q(\eta_e) \left[\log \pi_{e,k} \right] \\
&+ \sum_{e \in E} \sum_{n=1}^{N_e} \sum_{k=1}^{K} \phi_{d,n,k} \log \beta_{k, x_{e,n}} \\
&+ \sum_{v \in V} \mathbb{E}_{q(\boldsymbol{P})} \left[-\gamma_v + n_v \log \gamma_v + \log(n_v!) \right] \\
&+ \sum_{e \in E} \mathbb{E}_{q(\boldsymbol{P})} \left[-G_e + \ell_e \log G_e + \log(\ell_e!) \right] \\
&+ \sum_{e \in E} \left(\sum_{e' \in E} r_{e,e'} \log(\kappa_{e,e'}) \right) \\
&- \sum_{e \in E} \left(r_{e,0} \log r_{e,0} + \sum_{e' \in E} r_{e,e'} \log r_{e,e'} \right) \\
&- \sum_{e \in E} \sum_{n=1}^{N_e} \phi_{d,n,k} \log \phi_{d,n,k} + \text{const}
\end{aligned}
$$

While the update for the variational distribution of parent relationship $q(\boldsymbol{P})$ is provided in section**??**, we provide the update equations for the rest variational distributions and model parameters.

A.1. Update of variational distribution

Update of $q(\eta)$ Following the formulation in (Wang & Blei, 2013), we can compute $f(\eta_e)$ by grouping the terms in $\mathcal{L}(Q)$ with $q(\eta_e)$ as follows;

$$f(\eta_e) = -\frac{1}{2\sigma^2} r_{e,0}||\eta_e - \alpha_{v_e}||_2^2 - \frac{1}{2\sigma^2}\sum_{e':t_{e'}<t_e} r_{e,e'}||\eta_e - \hat{\eta_{d'}}||_2^2 + \sum_{n=1}^{N_e}\sum_{k=1}^{K}\phi_{e,n,k}\log\pi_{e,k},$$

where $\pi_{e,k} = \frac{e^{\eta_{e,k}}}{\sum_j e^{\eta_{e,j}}}$ is the softmax of $\eta_{e,k}$. Then $\hat{\eta}_e = \arg\max_{\eta_e} f(\eta_e)$. Intuitively, the form of function $f(\eta_e)$ suggests that the topics of each document are determined by two factors: (i) the topics should explain the actual words in the document, (ii) the topics of the document should be similar to either the preference of the node if it is from a spontaneous event or to the document of its parent if it is triggers by another event. As the solution of $\hat{\eta}_e$ depends on the parameters η of other documents, we propose a coordinate ascend method that iterates over maximizing each $\hat{\eta}_e$. The optimization for each $\hat{\eta}_e$ is solved by the Newton Conjugate Gradient method where the first and second order derivatives are as follows:

$$\nabla f(\eta_e) = -\frac{1}{\sigma^2}(\eta_e - r_{e,0}\alpha_{v_e} - \sum_{e':t_{e'}<t_e} r_{e,e'}\hat{\eta}_{e'}) + \sum_{n=1}^{N_e}\phi_{e,n} - \pi_e\left[\sum_{n=1}^{N_e}\sum_{k=1}^{K}\phi_{e,n,k}\right]$$

$$\frac{\partial f(\eta_e)}{\partial\eta_{e,i}\partial\eta_{e,j}} = (-\pi_{e,i}\mathbf{1}[i=j] + \pi_{e,i}\pi_{e,j})\left[\sum_{n=1}^{N_e}\sum_{k=1}^{K}\phi_{d,n,k}\right] - \mathbf{1}[i=j]/\hat{\sigma}^2$$

Update of $q(\boldsymbol{P})$ Assume $f_\Delta(\Delta_t)$ is the pdf for the delay distribution and we use $f_N(x|\mu,\Sigma)$ as the pdf for the Gaussian distribution with mean μ and covariance matrix Σ. Then we the update by taking derivative of $\mathcal{L}(Q)$ with respect to $r_{e,0}$.

$$r_{e,0} \propto \mu_{v_e} f_N(\hat{\eta}_e|\alpha_{v_e}, \hat{\sigma}^2 I)$$
$$r_{e,e'} \propto A_{v_{e'},v_e} f_N(\hat{\eta}_e|\hat{\eta}_{e'}, \hat{\sigma}^2 I) f_\Delta(t_e - t_{e'}) .$$

Intuitively, we combine three aspects in our joint HawkesTopic model to decide the parent relationship for each event: $A_{v_{e'}}$ captures the influence between nodes; $f_N(\hat{\eta}_e|\hat{\eta}_{e'}, \hat{\sigma}^2 I)$ considers the similarity between the documents of the events and $f_\Delta(t_e - t_{e'})$ models the proximity of events in time. In contrast, the traditional MHP model uses only the time proximity and node influences to determine event's parent.

Update of $q(\boldsymbol{z})$ As in CTM model, taking the derivative of $\mathcal{L}(Q)$ with respect to $\phi_{e,n,k}$ and setting to zero lead to the update of $\phi_{e,n,k}$ as:

$$\phi_{e,n,k} \propto \frac{exp(\hat{\eta}_{e,k})}{\sum_j exp(\hat{\eta}_{e,j})}\beta_{k,x_{e,n}} .$$

A.2. Update of model parameters

Update for $\boldsymbol{\beta}$ As in CTM model, taking the derivative of $\mathcal{L}(Q)$ with respect to $\beta_{k,w}$ and setting it to zero gives us the following update for $\beta_{k,w}$:

$$\beta_{k,w} \propto \sum_{e\in E}\sum_{n=1}^{N_e}\sum_{k=1}^{K}\sum_{w=1}^{W}\phi_{e,n,k}x_{e,n,w} .$$

Update for $\boldsymbol{\alpha}$ Taking the derivative of $\mathcal{L}(Q)$ with respect to α_u and setting to zero lead to the update of α_u as:

$$\alpha_u = \sum_{e\in E} r_{e,0}\hat{\eta}_e\mathbf{1}[v_e = u] .$$

Treating $r_{e,0}$ as soft count, we update the topic prior α_u as the average over the topics of documents belonging to the spontaneous events of node u.

Update for $\boldsymbol{A}, \boldsymbol{\mu}$ Using the posterior distribution of the parent relationship as soft count, the update of the source influence matrix $\boldsymbol{A}$ and the base intensity $\boldsymbol{\mu}$ is similar to the traditional MHP model by taking derivative of $\mathcal{L}(Q)$ with respect to $A_{u,w}$ and μ_u:

$$A_{u,w} = \frac{\sum_{e\in E}\sum_{e':t_{e'}<t_e} r_{e,e'}\mathbf{1}[v_{e'} = u]\mathbf{1}[v_e = w]}{\sum_{e\in E}\mathbf{1}[v_e = u]}$$

$$\mu_u = \frac{\sum_{e\in E} r_{e,0}\mathbf{1}[v_e = u]}{T} .$$

B. Experiment setting

B.1. Synthetic data generation

We set the hyperparameters characterizing the node topic priors are set as $\alpha_1 = (2.0, 0.0, \ldots, 0.0)$, ..., $\alpha_5 = (0.0, \ldots, 0.0, 2.0)$. We set the node influence matrix corresponding to the circular diffusion network G as

$$A = \begin{pmatrix} 0.3 & 0.15 & 0 & 0 & 0 \\ 0 & 0.3 & 0.15 & 0 & 0 \\ 0 & 0 & 0.3 & 0.15 & 0 \\ 0 & 0 & 0 & 0.3 & 0.15 \\ 0.15 & 0 & 0 & 0 & 0.3 \end{pmatrix}$$

The base intensity μ for the processes of all nodes is set to 0.02. Regarding the parameters of the generative model, we set the dictionary size W to 500, $\lambda = 200$, $\alpha = 0.02$, $\sigma = 2.0$. The hyper-parameters characterizing the node topic priors are set as $\alpha_1 = (2.0, 0.0, \ldots, 0.0)$, ..., $\alpha_5 = (0.0, \ldots, 0.0, 2.0)$.

References

Veen, Alejandro and Schoenberg, Frederic P. Estimation of spacetime branching process models in seismology using an emtype algorithm. *Journal of the American Statistical Association*, 103:614–624, 2008.

Wang, Chong and Blei, David M. Variational inference in nonconjugate models. *J. Mach. Learn. Res.*, 14(1), April 2013.