Modeling Cables Events

Allison J.B. Chaney

August 18, 2015

1 Generative model

We start with a fitted LDA model where documents are represented in terms of topics (θ, a $D \times K$ matrix), and topics are represented as a distribution over words (β, a $K \times V$ matrix). This model fit, along with document metadata, are our observations. Alternatively, we can tack on the LDA generative process to the model below.

- for each day i with date a_i:
  - generate event occurrence/strength $\epsilon \sim \text{Poisson}(\eta_\epsilon)$, where η_ϵ is a fixed, non-negative hyperparameter for the mean event strength
  - generate the day/event's description in terms of each topic k: $\pi_{ik} \sim \text{Gamma}(\alpha_0, \beta_0)$, where α_0 and β_0 are fixed hyperparameters.
- draw the entity's base topics: $\phi_{0k} \sim \text{Gamma}(\alpha, \beta)$ (eventually for each entity, but for now, just limit data to only one entity)
- For each cable j on date c_j:
  - set cable topic parameter: $\phi_{jk} = \phi_{0k} + \sum_i f(a_i, c_j)\pi_{ik}\epsilon_i$, where f is defined below.
  - draw cable topic: $\theta_{jk} \sim \text{Gamma}(\beta_c \phi_{jk}, \beta_c)$

Note that

$$f(a, c) = \begin{cases} 1 - \frac{c-a}{d}, & \text{if } a \le c < a + d \\ 0, & \text{otherwise,} \end{cases}$$

where d is the time distance (in days) after event a at which point the event is no longer relevant.

2 Inference

For now, we assume that we know the LDA topics β and only observe the documents in terms of their topics θ; breaking this assumption makes inference a little more complicated as the updates for θ would have new dependencies. Pending the results, we should explore that vein.

Here, we follow the structure of the DEF paper to explain inference for this model.

- As usual, inference is the central computational problem.
- Variational inference minimizes the KL divergence from an approximating distribution q to the true posterior p.
- This is equivalent to maximizing the ELBO: $\mathcal{L}(q) = \text{E}_{q(\epsilon,\pi,\phi)}[\log p(\theta, \epsilon, \pi, \phi) - \log q(\epsilon, \pi, \phi)]$
- we define the approximating distribution q using the mean field assumption: $q(\epsilon, \pi, \phi) = \prod_i q(\epsilon_i) \prod_k [q(\phi_{0k}) \prod_i q(\pi_{ik})]$

- $q(\pi)$ and $q(\phi)$ are both gamma-distributed with variational parameters λ^{π} and λ^{ϕ}, respectively, where we use the softmax function $\mathcal{M}(x) = \log(1 + \exp(x))$ to constrain the free variational parameters. $q(\epsilon)$ is Poisson-distributed with variational parameter λ^{ϵ}; the free parameter is also constrained by the softmax function.
- the expectations under q (needed to maximize the ELBO) do not have a simple analytic form, so we use "black box" VI techniques
- for each variable, we can write the log probability of all terms containing that variable, giving us

$$\log p_i^{\epsilon}(\theta, \epsilon, \pi, \phi) = \log p(\epsilon_i \mid \eta_\epsilon) + \sum_{j: f(a_i, c_j) \neq 0} \sum_k \log p(\theta_{jk} \mid \phi_{0k}, c_j, a_i, d, \beta_c, \pi_{ik}, \epsilon_i),$$

$$\log p_{ik}^{\pi}(\theta, \epsilon, \pi, \phi) = \log p(\pi_{ik} \mid \alpha_0, \beta_0) + \sum_{j: f(a_i, c_j) \neq 0} \log p(\theta_{jk} \mid \phi_{0k}, c_j, a_i, d, \beta_c, \pi_{ik}, \epsilon_i),$$

and

$$p_k^{\phi}(\theta, \epsilon, \pi, \phi) = \log p(\phi_{0k} \mid \alpha, \beta) + \sum_i \sum_j \log p(\theta_{jk} \mid \phi_{0k}, c_j, a_i d, \beta_c, \pi_{ik}, \epsilon_i).$$

- Then we can write the gradients with respect to the variational parameters as:

$$\nabla_{\lambda_i^{\epsilon}} \mathcal{L} = \mathrm{E}_q \left[\nabla_{\lambda_i^{\epsilon}} \log q(\epsilon_i \mid \lambda_i^{\epsilon}) \left(\log p_i^{\epsilon}(\theta, \epsilon, \pi, \phi) - \log q(\epsilon_i \mid \lambda_i^{\epsilon}) \right) \right],$$

$$\nabla_{\lambda_{ik}^{\pi}} \mathcal{L} = \mathrm{E}_q \left[\nabla_{\lambda_{ik}^{\pi}} \log q(\pi_{ik} \mid \lambda_{ik}^{\pi}) \left(\log p_{ik}^{\pi}(\theta, \epsilon, \pi, \phi) - \log q(\pi_{ik} \mid \lambda_{ik}^{\pi}) \right) \right],$$

and

$$\nabla_{\lambda_k^{\phi}} \mathcal{L} = \mathrm{E}_q \left[\nabla_{\lambda_k^{\phi}} \log q(\phi_{0k} \mid \lambda_k^{\phi}) \left(\log p_k^{\phi}(\theta, \epsilon, \pi, \phi) - \log q(\phi_{0k} \mid \lambda_k^{\phi}) \right) \right].$$

Using this framework, we construct our black box algorithm below.

For Reference The gamma distribution and derivatives:

$$\log \mathrm{Gamma}(x \mid a, b) = ab \log b - \log \Gamma(ab) + (ab - 1) \log x - bx \tag{1}$$

$$\nabla_a \log \mathrm{Gamma}(x \mid \mathcal{M}(a), \mathcal{M}(b)) = \mathcal{M}'(a)\mathcal{M}(b) \left[\log \mathcal{M}(b) - \Psi(\mathcal{M}(a)\mathcal{M}(b)) + \log x \right] \tag{2}$$

$$\nabla_b \log \mathrm{Gamma}(x \mid \mathcal{M}(a), \mathcal{M}(b)) = \mathcal{M}'(b) \left[\mathcal{M}(a) \left((\log \mathcal{M}(b) + 1) - \Psi(\mathcal{M}(a)\mathcal{M}(b)) + \log x \right) - x \right] \tag{3}$$

The Poisson distribution and derivative:

$$\log \mathrm{Poisson}(x \mid \lambda) = x \log \lambda - \log(x!) - \lambda \tag{4}$$

$$\nabla_\lambda \log \mathrm{Poisson}(x \mid \mathcal{M}(\lambda)) = \mathcal{M}'(\lambda) \left[\frac{x}{\mathcal{M}(\lambda)} - 1 \right]. \tag{5}$$

The softmax function and derivative:

$$\mathcal{M}(x) = \log(1 + e^x)$$

$$\mathcal{M}'(x) = \frac{e^x}{1 + e^x}$$

Algorithm 1 Black Box Variational Inference for Cables Model

Input: document topics θ
Initialize λ^{π}, λ^{ϕ}, and λ^{ϵ} to respective priors
Initialize iteration count $t = 0$
repeat
  $t \mathrel{+}= 1$
  initialize $\hat{\lambda}^{\pi}$, $\hat{\lambda}^{\phi}$, and $\hat{\lambda}^{\epsilon}$ to 0 matrices
  for $j = 1, \ldots, B$ (batch size, or # of document samples) **do**
    Sample a document j that is sent on date c_j and has topics θ_j
    for $s = 1, \ldots, S$ **do**
      for $k = 1, \ldots, K$ **do**
        draw sample $\phi_{0k}[s] \sim \text{Gamma}(\mathcal{M}(\lambda_k^{\phi}))$
        set $p_k^{\phi}[s] = \log p(\phi_{0k}[s] \mid \alpha, \beta)$ ▷ see Eqn. 1
        set $q_k^{\phi}[s] = \log q(\phi_{0k} \mid \lambda_k^{\phi})$ ▷ Eqn. 1 with parameters $\mathcal{M}(\lambda_k^{\phi,a}), \mathcal{M}(\lambda_k^{\phi,b})$
        set $g_k^{\phi}[s] = \nabla_{\lambda_k^{\phi}} \log q(\phi_{0k}[s] \mid \lambda_k^{\phi})$ ▷ see Eqns. 2, 3
        for each event i on date $a_i \in (c_j - d, c_j]$ **do**
          draw sample $\pi_{ik}[s] \sim \text{Gamma}(\mathcal{M}(\lambda_{ik}^{\pi}))$
          draw sample $\epsilon_i[s] \sim \text{Poisson}(\mathcal{M}(\lambda_i^{\epsilon}))$
          set $p_{ik}^{\pi}[s] = \log p(\pi_{ik}[s] \mid \alpha_0, \beta_0)$ ▷ see Eqn. 1
          set $q_{ik}^{\pi}[s] = \log q(\pi_{ik}[s] \mid \lambda_{ik}^{\pi})$ ▷ Eqn. 1 with parameters $\mathcal{M}(\lambda_{ik}^{\pi,a}), \mathcal{M}(\lambda_{ik}^{\pi,b})$
          set $g_{ik}^{\pi}[s] = \nabla_{\lambda_{ik}^{\pi}} \log q(\pi_{ik}[s] \mid \lambda_{ik}^{\pi})$ ▷ see Eqns. 2, 3
      for each event i on date $a_i \in (c_j - d, c_j]$ **do**
        set $p_i^{\epsilon}[s] = \log p(\epsilon_i[s] \mid \eta)$ ▷ see Eqn. 4
        set $q_i^{\epsilon}[s] = \log q(\epsilon_i \mid \lambda_i^{\epsilon})$ ▷ Eqn. 4 with parameter $\mathcal{M}(\lambda_i^{\epsilon})$
        set $g_i^{\pi}[s] = \nabla_{\lambda_i^{\epsilon}} \log q(\epsilon_i[s] \mid \lambda_i^{\epsilon})$ ▷ see Eqn. 5
      for $k = 1, \ldots, K$ **do**
        set $\phi_{jk}[s] = \phi_{0k}[s] + \sum_{i:a_i \in (c_j - d, c_j]} \epsilon_i[s] \pi_{ik}[s] f(a_i, c_j)$
        set $p_k^{\theta}[s] = \log p(\theta_j \mid \phi_{jk}[s], \beta_c)$ ▷ see Eqn. 1
        for each event i **do**
          update $\hat{\lambda}_{ik}^{\pi} \mathrel{+}= \frac{1}{SB} g_{ik}^{\pi}[s](p_{ik}^{\pi}[s] + p_k^{\theta}[s] - q_{ik}^{\pi}[s])$
        update $\hat{\lambda}_k^{\phi} \mathrel{+}= \frac{1}{SB} g_{ik}^{\phi}[s](p_k^{\phi}[s] + p_k^{\theta}[s] - q_k^{\phi}[s])$
      for each event i **do**
        update $\hat{\lambda}_i^{\epsilon} \mathrel{+}= \frac{1}{SB} g_i^{\epsilon}[s](p_i^{\epsilon}[s] + \sum_k p_k^{\theta}[s] - q_i^{\epsilon}[s])$
  set $\rho = (t + \tau)^{\kappa}$
  update event content for each event i and topic k: $\lambda_{ik}^{\pi} = \lambda_{ik}^{\pi} + \rho \hat{\lambda}_{ik}^{\pi}$
  update general entity topics for each topic k: $\lambda_k^{\phi} = \lambda_k^{\phi} + \rho \hat{\lambda}_k^{\phi}$
  update event occurrences for each event i: $\lambda_k^{\epsilon} = \lambda_i^{\epsilon} + \rho \hat{\lambda}_i^{\epsilon}$
until change in validation likelihood $< \delta$