Modeling Cables Events

Allison J.B. Chaney

July 13, 2015

1 Model v1

1.1 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 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: $\phi_{jk} = \phi_{0k} + \sum_i f(a_i, c_j)\pi_{ik}$, where f is defined below.
  - draw cable topic: $\theta_{jk} \sim \text{Gamma}(\beta_c \phi_{jk}, \beta_c)$

Note that

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

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

1.2 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: $\mathcal{L}(q) = \text{E}_{q(\pi,\phi)}[\log p(\theta, \pi, \phi) - \log q(\pi, \phi)]$
- we define the approximating distribution q using the mean field assumption: $q(\pi, \phi) = \prod_k [q(\phi_{0k}) \prod_i q(\pi_{ik})]$
- $q(\pi)$ and $q(\phi)$ are both gamma-distributed, with variational parameters λ^π and λ^ϕ, respectively

- 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 probability of all terms containing that variable, giving us

$$p_{ik}^{\pi}(\theta, \pi, \phi) = p(\pi_{ik} \mid \alpha_0, \beta_0) \prod_j p(\theta_{jk} \mid \phi_{0k}, c, a, d, \beta_c, \pi)$$

and

$$p_{k}^{\phi}(\theta, \pi, \phi) = p(\phi_{0k} \mid \alpha, \beta) \prod_j p(\theta_{jk} \mid \phi_{0k}, c, a, d, \beta_c, \pi).$$

We can also write these in their log forms:

$$\log p_{ik}^{\pi}(\theta, \pi, \phi) = \log p(\pi_{ik} \mid \alpha_0, \beta_0) + \sum_j \log p(\theta_{jk} \mid \phi_{0k}, c, a, d, \beta_c, \pi)$$

and

$$\log p_{k}^{\phi}(\theta, \pi, \phi) = \log p(\phi_{0k} \mid \alpha, \beta) + \sum_j \log p(\theta_{jk} \mid \phi_{0k}, c, a, d, \beta_c, \pi).$$

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

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

and

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

Using this framework, we construct our black box algorithm below. The messiness comes from the full expansions of p, q, and $\nabla_{\lambda} q$.

Algorithm 1 Black Box variational inference for Cables Model v0

Input: document topics θ
Initialize λ^{π} and λ^{ϕ} randomly
Initialize $t_i^{\pi} = 0$ for all days i and $t^{\phi} = 0$
repeat
  Sample a document j that is sent on date c_j and has topics θ_j
  $t^{\phi} \mathrel{+}= 1$
  for $s = 1, \ldots, S$ **do**
    for $k = 1, \ldots, K$ **do**
      draw sample $\phi_{0k}[s] \sim \text{Gamma}(\lambda_k^{\phi})$
      for each event i on date $a_i \in (c_j - d, c_j]$ **do**
        draw sample $\pi_{ik}[s] \sim \text{Gamma}(\lambda_{ik}^{\pi})$
        $t_i^{\pi} \mathrel{+}= 1$
      set $\phi_{jk}[s] = \phi_{0k}[s] + \sum_{i:a_i \in \left(c_j - d, c_j\right]} \pi_{ik}[s] f(a_i, c_j)$
      set $p_k^{\theta}[s] = \log p(\theta_j \mid \phi_{jk}[s], \beta_c) = \beta_c \phi_{jk}[s] \log \beta_c - \log \Gamma(\beta_c \phi_{jk}[s]) + (\beta_c \phi_{jk}[s] - 1) \log \theta_k - \beta_c \theta_k$
      set $p_k^{\phi}[s] = \log p(\phi_{0k} \mid \alpha, \beta) = \alpha \log \beta - \log \Gamma(\alpha) + (\alpha - 1) \log \phi_{0k}[s] - \beta \phi_{0k}[s]$
      set (for each relevant i) $p_{ik}^{\pi}[s] = \log p(\pi_{ik} \mid \alpha_0, \beta_0) = \alpha_0 \log \beta_0 - \log \Gamma(\alpha_0) + (\alpha_0 - 1) \log \pi_k[s] - \beta_0 \pi_{ik}[s]$
      set $q_k^{\phi}[s] = \log q(\phi_{0k} \mid \lambda_k^{\phi}) = \lambda_k^{\phi,\alpha} \log \lambda_k^{\phi,\beta} - \log \Gamma(\lambda_k^{\phi,\alpha}) + (\lambda_k^{\phi,\alpha} - 1) \log \phi_{0k}[s] - \lambda_k^{\phi,\beta} \phi_{0k}[s]$
      set (for each relevant i) $q_{ik}^{\pi}[s] = \log q(\pi_{ik} \mid \lambda_{ik}^{\pi}) = \lambda_{ik}^{\pi,\alpha} \log \lambda_{ik}^{\pi,\beta} - \log \Gamma(\lambda_{ik}^{\pi,\alpha}) + (\lambda_{ik}^{\pi,\alpha} - 1) \log \pi_{ik}[s] - \lambda_{ik}^{\pi,\beta} \pi_{ik}[s]$
      set $g_k^{\phi}[s] = \nabla_{\lambda_k^{\phi}} \log q(\phi_{0k} \mid \lambda_k^{\phi}) = \langle \log \lambda_k^{\phi,\beta} - \Psi(\lambda_k^{\phi,\alpha}) + \log \phi_{0k}[s], \lambda_k^{\phi,\alpha} / \lambda_k^{\phi,\beta} - \phi_{0k}[s] \rangle$
      set (for each relevant i) $g_{ik}^{\pi}[s] = \nabla_{\lambda_{ik}^{\pi}} \log q(\pi_{ik} \mid \lambda_{ik}^{\pi}) = \langle \log \lambda_{ik}^{\pi,\beta} - \Psi(\lambda_{ik}^{\pi,\alpha}) + \log \pi_{ik}[s], \lambda_{ik}^{\pi,\alpha} / \lambda_{ik}^{\pi,\beta} - \pi_{ik}[s] \rangle$
  for each event i on date $a_i \in (c_j - d, c_j]$ **do**
    set $\hat{\lambda}_{ik}^{\pi} = \frac{1}{S} \sum_s g_{ik}^{\pi}[s](p_{ik}^{\pi}[s] + J p_k^{\theta}[s] - q_{ik}^{\pi}[s])$
    set $\rho_i^{\pi} = (t_i^{\pi} + \tau)^{\kappa}$
    update $\lambda_{ik}^{\pi} = \lambda_{ik}^{\pi} + \rho_i^{\pi} \hat{\lambda}_{ik}^{\pi}$
  set $\hat{\lambda}_k^{\phi} = \frac{1}{S} \sum_s g_{ik}^{\phi}[s](p_k^{\phi}[s] + J p_k^{\theta}[s] - q_k^{\phi}[s])$
  set $\rho^{\phi} = (t^{\phi} + \tau)^{\kappa}$
  update $\lambda_k^{\phi} = \lambda_k^{\phi} + \rho^{\phi} \hat{\lambda}_k^{\phi}$
until change in validation likelihood $< \delta$