Modeling Cables Events

Allison J.B. Chaney

September 24, 2015

1 Introduction

Our objective is to detect historical events.

2 Related Work

- BBVI, DEFs; this deals with finding updates for non-conjuate models

- non uniform sampling (MMSB in PNAS); we want events to be represented appropriately. Also: how to update global stuff form non-uniform document sampling–it's a bit of challenge)

- spike and slab (e.g., Emperical study of SVI for Beta Bernoulli Process; also IBP papers) because we have variables that fit this pattern

- SSVI; this helps us with the event variables–they are bound together (is this the real purpose here?)

3 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

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

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

4 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: $\mathscr{L}(q) = \mathbb{E}_{q(\epsilon,\pi,\phi)}[\log p(\theta,\epsilon,\pi,\phi) - \log q(\epsilon,\pi,\phi)]$
- we define the approximating distribution q using the mean field assumption: $q(\epsilon,\pi,\phi) = \prod_i q(\epsilon_i) \prod_k \left[q(\phi_{0k}) \prod_i q(\pi_{ik})\right]$
- $q(\pi)$ and $q(\phi)$ are both gamma-distributed with variational parameters λ^π and λ^ϕ, respectively, where we use the softmax function $\mathscr{M}(x) = \log(1+\exp(x))$ 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

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

  $$\log p_{ik}^\pi(\theta,\epsilon,\pi,\phi) \triangleq \log p(\pi_{ik} \mid \alpha_0, \beta_0) + \mathbf{1}[\epsilon_i \neq 0] \sum_{j:f(a_i,c_j)\neq 0} \log p(\theta_{jk} \mid \phi_{0k}, c_j, a_i, d, \beta_c, \pi_{ik}, \epsilon_i),$$

  and

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

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

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

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

  and

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

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

For Reference The gamma distribution and derivatives:

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

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

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

The Poisson distribution and derivative:

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

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

The softmax function and derivative:

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

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

Algorithm 1: Inference for Cables Model

Input: document topics θ
Output: estimates of latent parameters entity topics ϕ, event topics π, and event occurances ϵ
Initialize λ^{π}, λ^{ϕ}, and λ^{ϵ} to respective priors
Initialize iteration count $t = 0$
while *change in validation likelihood* $< \delta$ **do**
    initialize $\sigma^{\pi} = 0$
    for *each sample* $s = 1, \ldots, S$ **do**
        for *each component* k **do**
            draw sample entity topics $\phi_{0k}[s] \sim \text{Gamma}(\mathscr{M}(\lambda_k^{\phi}))$
            set $p_k^{\phi}[s] = \log p(\phi_{0k}[s] \,|\, \alpha, \beta)$ `// see Eqn. 1`
            set $q_k^{\phi}[s] = \log q(\phi_{0k}[s] \,|\, \lambda_k^{\phi})$ `// Eqn. 1 with params` $\mathscr{M}(\lambda_k^{\phi})$
            set $g_k^{\phi}[s] = \nabla_{\lambda_k^{\phi}} \log q(\phi_{0k}[s] \,|\, \lambda_k^{\phi})$ `// see Eqns. 2, 3`
        end
        for *each timestep* i **do**
            draw sample event occurance $\epsilon_i[s] \sim \text{Poisson}(\mathscr{M}(\lambda_i^{\epsilon}))$
            set $p_i^{\epsilon}[s] = \log p(\epsilon_i[s] \,|\, \eta)$ `// see Eqn. 4`
            set $q_i^{\epsilon}[s] = \log q(\epsilon_i[s] \,|\, \lambda_i^{\epsilon})$ `// Eqn. 4 with param` $\mathscr{M}(\lambda_i^{\epsilon})$
            set $g_i^{\pi}[s] = \nabla_{\lambda_i^{\epsilon}} \log q(\epsilon_i[s] \,|\, \lambda_i^{\epsilon})$ `// see Eqn. 5`
            if $\epsilon_i[s] \neq 0$ **then**
                for *each component* k **do**
                    draw sample event topics $\pi_{ik}[s] \sim \text{Gamma}(\mathscr{M}(\lambda_{ik}^{\pi}))$
                    set $p_{ik}^{\pi}[s] = \log p(\pi_{ik}[s] \,|\, \alpha_0, \beta_0)$ `// see Eqn. 1`
                    set $q_{ik}^{\pi}[s] = \log q(\pi_{ik}[s] \,|\, \lambda_{ik}^{\pi})$ `// Eqn. 1 with params` $\mathscr{M}(\lambda_{ik}^{\pi})$
                    set $g_{ik}^{\pi}[s] = \nabla_{\lambda_{ik}^{\pi}} \log q(\pi_{ik}[s] \,|\, \lambda_{ik}^{\pi})$ `// see Eqns. 2, 3`
                end
            end
        end
    end
    for *each document* j *(sent on date* c_j *and has topics* θ_j*), sample* s *and component* k **do**
        set $\phi_{jk}[s] = \phi_j[s] + \sum_i f(a_i, c_j)\epsilon_i[s]\pi_{ik}[s]$
        set $p_{jk}^{\theta}[s] = \log p(\theta_{jk} \,|\, \phi_{jk}[s], \beta_c)$ `// see Eqn. 1`
        $p_k^{\phi}[s] \mathrel{+}= p_{jk}^{\theta}[s]$
        for *each timestep* i *where* $a_i \leq c_j < a_i + d$ **do**
            $p_i^{\epsilon}[s] \mathrel{+}= \sum_k p_{jk}^{\theta}[s]$
            if $\epsilon_i[s] \neq 0$ **then**
                $p_{ik}^{\pi}[s] \mathrel{+}= p_{jk}^{\theta}[s]$
                update $\sigma_i^{\pi} \mathrel{+}= 1$
            end
        end
    end
    set $\hat{\nabla}_{\lambda^{\phi}} \mathscr{L} \triangleq \frac{1}{S} \sum_s g^{\phi}[s](p^{\phi}[s] - q^{\phi}[s])$
    set $\hat{\nabla}_{\lambda^{\epsilon}} \mathscr{L} \triangleq \frac{1}{S} \sum_s g^{\epsilon}[s](p^{\epsilon}[s] - q^{\epsilon}[s])$
    set $\hat{\nabla}_{\lambda^{\pi}} \mathscr{L} \triangleq \frac{1}{\sigma_{\pi}} \sum_s g^{\pi}[s](p^{\pi}[s] - q^{\pi}[s])$
    set $\rho = (t + \tau)^{\kappa}$
    set $\lambda^{\pi} \mathrel{+}= \rho \hat{\nabla}_{\lambda^{\pi}} \mathscr{L}$
    set $\lambda^{\epsilon} \mathrel{+}= \rho \hat{\nabla}_{\lambda^{\epsilon}} \mathscr{L}$
    set $\lambda^{\phi} \mathrel{+}= \rho \hat{\nabla}_{\lambda^{\phi}} \mathscr{L}$
end
set $\mathbb{E}[\pi] = \lambda^{\pi,a}$
set $\mathbb{E}[\phi] = \lambda^{\phi,a}$
set $\mathbb{E}[\epsilon] = \lambda^{\epsilon}$
return $\mathbb{E}[\pi]$, $\mathbb{E}[\phi]$, $\mathbb{E}[\epsilon]$