Generative Models for SpatioTemporal Topic Modeling

Newt Gingrich

October 4, 2018

1 AR Process on log Dirichlet Concentration

Notation:

Indices:

- K number of topics
- V size of vocabulary
- L number of locations
- Y number of time points
- P number of spatial dimensions

Hyperparameters:

- η dirichlet hyperparameter for word-topic concentrations

Parameters:

- $\Omega \in \mathbb{R}^{Y \times L \times K}$ log concentration parameters for document-topic prevalence.
- σ_ω isotropic error for matrix AR process on Ω.
- $\phi \in \mathbb{R}^{K \times V}$ topic - word probability matrix.
- $w_{y,l,m,n}$ the word in position n of the m'th document at location l at time y.
- $z_{y,l,m,n}$ the topic associated with $w_{y,l,m,n}$.
- $\boldsymbol{B}^{L \times L}$ matrix based on distances between the spatial locations divided by a scalar length scale
- $\boldsymbol{A}^{Y \times L \times K}$ matrix based on an AR(1) process

Algorithm 1 StLda1

▷ Generate Spatiotemporal priors

$\Omega_1 \sim N(\mathbf{0}, \sigma_\omega \mathbf{I})$

for $y \in \{2, \ldots, Y\}$ **do**

  $\epsilon_t \sim \text{iid}N(0, \sigma_\omega)$

  $\Omega_y \leftarrow \mathbf{B}\Omega_{y-1} + \epsilon_t$

▷ Generate Topics (standard LDA)

for $k \in \{1, \ldots, K\}$ **do**

  $\phi_k \sim \text{Dir}(\eta)$

▷ Generate Document-Topic Prevalences

for $y \in \{1, \ldots, Y\}$ **do**

  for $l \in \{1, \ldots, L\}$ **do**

    for $m \in \{1, \ldots, M\}$ **do**

      $\theta_{y,l,m} \sim \text{N}(\Omega_{y,l}, \sigma_\theta)$

▷ Generate Document-Topic Prevalences

for $y \in \{1, \ldots, Y\}$ **do**

  for $l \in \{1, \ldots, L\}$ **do**

    for $m \in \{1, \ldots, M\}$ **do**

      for $n \in \{1, \ldots, N_m\}$ **do**

        $z_{y,l,m,n} \sim \text{Mult}(\theta_{y,l,m})$

        $w_{y,l,m,n} \sim \text{Mult}(\phi_{z_{y,l,m,n}})$