

Generative Models for SpatioTemporal Topic Modeling

Newt Gingrich

September 11, 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 $w_{y,l,m,n}$.

Algorithm 1 StLda1

```
1:                                     ▷ Generate Spatiotemporal priors
2:  $\Omega_1 \sim N(\mathbf{0}, \sigma_\omega \mathbf{I})$ 
3: for  $y \in \{2, \dots, Y\}$  do
4:    $\epsilon_t \sim \text{iid} N(0, \sigma_\omega)$ 
5:    $\Omega_y \leftarrow \mathbf{B}\Omega_{y-1} + \epsilon_t$ 
6: for  $y \in \{1, \dots, Y\}$  do
7:    $\mathbf{A}_y \leftarrow \exp \odot \Omega_{y-1}$ 
8:                                     ▷ Generate Topics (standard LDA)
9: for  $k \in \{1, \dots, K\}$  do
10:   $\phi_k \sim \text{Dir}(\eta)$ 
11:                                     ▷ Generate Document-Topic Prevalences
12: for  $y \in \{1, \dots, Y\}$  do
13:   for  $l \in \{1, \dots, L\}$  do
14:     for  $m \in \{1, \dots, M\}$  do
15:        $\theta_{y,l,m} \sim \text{Dir}(\mathbf{A}_{y,l})$ 
16:                                     ▷ Generate Document-Topic Prevalences
17: for  $y \in \{1, \dots, Y\}$  do
18:   for  $l \in \{1, \dots, L\}$  do
19:     for  $m \in \{1, \dots, M\}$  do
20:       for  $n \in \{1, \dots, N_m\}$  do
21:          $z_{y,l,m,n} \sim \text{Mult}(\theta_{y,l,m})$ 
22:          $w_{y,l,m,n} \sim \text{Mult}(\phi_{z_{y,l,m,n}})$ 
```
