Point Estimation for Spatio-Temporal Topic Modeling

Nathan Wycoff

November 21, 2018

1 Introduction

This document outlines how inference will be undertaken on the SpatioTemporal Topic Model. As we intend to run this on a large dataset, parallelism is at the back of our minds at each step. Scalability dictates that an optimization based routine be used. The nonconjugacy of our model means variational methods may require approximations to the evidence lower bound (Murphy textbook), and while such an option may be feasible, it is more straightforward to carry out a Maximum of the Posterior (MAP) approach.

The overall structure of the algorithm is block coordinate ascent of the joint posterior via an Expectation Maximization (Dempsey et al) algorithm. Such a posterior is highly multimodal, even to the extent that a multiple restart approach is insufficient to guarantee reproducibility. As such, picking intelligent initializations is an important question which we will have to address.

2 E Step

3 M Step: Block Coordinate Ascent

We shall here go over the algo, step by step.

3.1 Inference for ϕ: A Simple Sum

The E step gives us $\hat{Z}$, where $\hat{Z}_{y,l,m,n,k}$ gives the probability that word n of document m at time y and location l belongs to document k. Here we seek an estimate of the topic word matrix ϕ. With some thought, we may convince ourselves that a reasonable thing to do is to tabulate how many of each token belongs to each topic, then suitably normalize (the reader more thoroughly convinced by mathematical derivation than thought experiment is referred to Appendix). To be precise:

The algo will go here.

3.2 Inference for α/Θ: A Generalized Linear Model

3.3 Inference for Ω: A Latent Kalman Filter