Introduction

Problem: Discovering latent semantic structures from data efficiently, particularly inferences on non-conjugate logistic-normal topic models.

Topic Model Background:
- Latent Dirichlet Allocation (LDA)
- gCTM (M=1, P=12)
- Y!LDA (M=40, P=480)
- gCTM (M=40, P=480)

Limitations:
- Inapplicability to real world for large scale of data
- Training time is almost kept constant. Parallel gCTM enjoys nice scalability.

Fast Approximate Sampling

Fast approximate sampling method to draw \( P(W, \eta) \) samples, reducing the time complexity from \( O(n) \) to \( O(1) \).

Parallel Implementation

- Delays of over-4-64-46 shared distributed sampler [1]
- No communication is needed inferring \( \eta_j \) and \( \lambda_j \)
- Broadcasting the global variables \( \mu \) and \( \Sigma \) to every machine after each iteration.

For each iteration:
- For each document \( d \):
  - Draw topic mixing proportions \( \theta_d \sim \text{Dir}(\alpha) \)
  - For each word \( w \): (a) draw topic \( z_w \sim \text{Mult}(\theta_d) \), \( \Phi_z \sim \text{Dir}(\beta) \)
    - Sample Logistic-Normal parameters with data augmentation technique
    - For \( \phi_d \) per (b) \( \phi_d \sim \text{Mult}(\theta_d, \beta) \)
      - For each word \( w \)
        - \( \lambda_j \sim \text{Poisson}(\lambda_j, \eta) \)
        - \( \mu \sim \text{Normal}(\mu, \Sigma) \)

Gibbs Sampling with Data Augmentation

Integrate-out \( \theta \). Gibbs sampling for marginalized distribution:

\[
p(\eta, Z|W) \propto p(Z|W) \prod_{d=1}^D p(z_d|z_{d-1})
\]

For each document \( d \): (a) draw topic \( z_D \sim \text{Mult}(\theta_d) \), \( \Phi_z \sim \text{Dir}(\beta) \)

- Marginalizing out topic proportions \( \theta_d \)
- Sampling Logistic-Normal parameters with data augmentation technique

- For each word \( w \)
  - \( \phi_d \sim \text{Mult}(\theta_d, \beta) \)
  - \( \lambda_j \sim \text{Poisson}(\lambda_j, \eta) \)
  - \( \mu \sim \text{Normal}(\mu, \Sigma) \)

Conclusion

- We present a scalable Gibbs sampling algorithm for logistic-normal topic models.
- Experimental results show significant improvement in time efficiency over existing variational methods, with slightly better perplexity.
- The algorithm enjoys excellent scalability, suggesting the ability to extract large structures from massive data.

Selected References