Recognition Problem

2011.5.16

1.ME: an effective variational method

1.1) DPM-NIW

ME v.s. EE (Done!)

1.2) HDP-DM

Data: Naive Unequal Bars mentioned below

1. Hyper-parameters in Gibbs:

   (a) Tests:
      i. Test 1: Fix λ and the gamma prior of γ, change the gamma prior of $\alpha \in \{10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}\}$
      ii. Test 2: Fix λ and the gamma prior of α, change the gamma prior of $\gamma \in \{10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}\}$
      iii. Test 3: Fix and the gamma prior of γ and α, change $\lambda \in \{10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}\}$

   (b) Plots:
      i. Plot 1: Fixing λ,α and γ doesn't change performance too much
         x: parameter range;
         y:
         Likelihood on the data varying α,
         Likelihood on the data varying γ(fix+resample)
      ii. Plot 2: λ affects the shape of the configuration more than that by α and γ
         x: parameter range;
         y:
         number of topics varying γ,
         number of topics varying λ,
         average number of tables varying α,
         average number of tables varying λ

2. Initial number of clusters (K) in Gibbs:

   (a) Tests:(Fix a suitable set of hyper-parameters)
      i. Test 1: Run ME to stuck with K$in\{1, 10, 30, 50, 100\}$
      ii. Test 2: Run Gibbs to stuck with K$in\{1, 10, 30, 50, 100\}$
      iii. Test 3: Run Gibbs+ME+Gibbs to stuck with K$in\{1, 10, 30, 50, 100\}$

(b) Plots:

i. Plot 1: Better training likelihood
x: K range;
y:
Likelihood of ME on training data varying K
Likelihood of Gibbs on training data varying K
Likelihood of Gibbs+ME+Gibbs on training data varying K

ii. Plot 2: Better predictive likelihood
x: K range;
y:
Likelihood of ME on test data varying K
Likelihood of Gibbs on test data varying K
Likelihood of Gibbs+ME+Gibbs on test data varying K

iii. Plot 3: Better Configurations
x: K range;
y:
Configuration from ME
Configuration from Gibbs
Configuration from Gibbs+ME+Gibbs

iv. Plot 4: Evidence that Gibbs got stuck
x: iterations;
y:
Likelihood of Gibbs+ME+Gibbs

2.HDP topic Model: from synthetic bar data to real data

2.1) Comparison of statistics

Described in Griffiths et al.(2004), a synthetic bar data was generated to illustrate the power of Gibbs Sampler for LDA model. However, this bar data differs from real corpus in three important ways:

1. Unequal mass of the topics: Handled by LDA with asymmetric prior and HDP
2. Random noise in the document: HDP-UDM
3. Burstiness of topics in the document: HDP-BDM

Below is a detailed description to simulate NIPS corpus (Teh et al.) with bar data considering the above factors.

The similarness of the data statistics is evaluated in terms of sorted word distribution across the corpus (Fig 1) and averaged sorted word distribution in the document (Fig 2)
Settings:

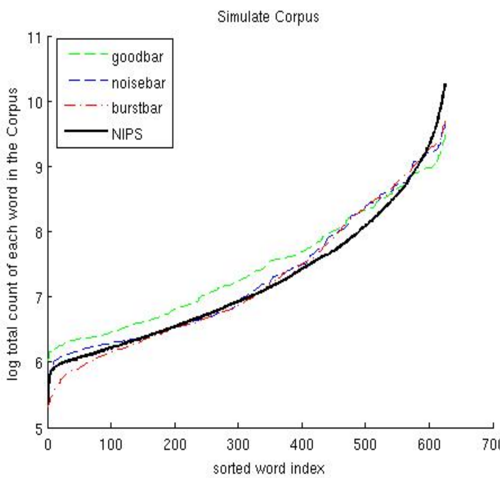
1. NIPS: following (Teh et al.), remove words with count less than 50 or more than 4,000
   Make the training data by randomly picking 90% of the documents from the corpus, which gives the following statistics:
   J=1566, $\overline{N}$=1001,W=5270
   In order to better visualize the comparison of the distribution, we evenly binned the 5270 words to 625

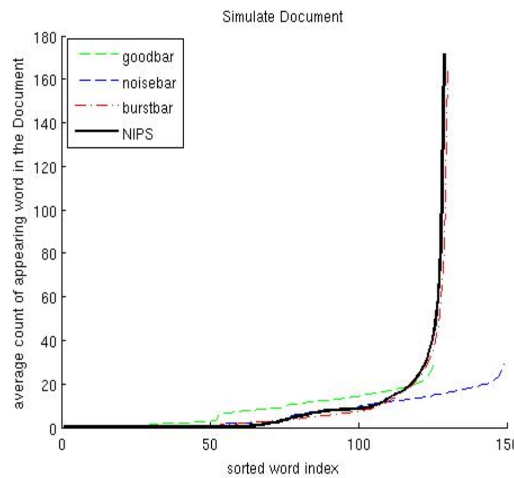
2. Bar Data:
   Basic statistics, J=1566, N=1001,W=625
   Heuristically approximate the distribution of the number of topics in the Document from the training data

   (a) Naive Unequal Bars:
       Tune the topic distribution to approximates the Corpus-Word Distribution of the training data

   (b) Unequal Bars+Noise:
       +Tune the noise level to approimate the first half of sorted Word-Document Distribution

   (c) Unequal Bars+Noise+Burstiness:
       +Tune the bursty word distribution in the biggest topic in the Document to approimate the second half of sorted Word-Document Distribution from the training data





2.2) Noisiness

1. Naive fix:
   Compare results from Gibbs with same settings on the NIPS training data and its denoised version.

2. HDP-UDM:
   Compare results from ME with same settings under HDP-UDM and HDP-DM.

2.3) Burstiness

1. No naive fix but we can show it is problematic for HDP-DM by encouraging small biased tables to be new topics.

2. Leave it for later development