Coding Plan

Donglai Wei

2010.10.27

1.Attribute and indexing

1. First-Level DP (table assign):$log\{\frac{\Gamma(\gamma)}{\Gamma(m_{..}+\gamma)}\Pi_{k=1}^{K}[\Gamma(m_{\cdot k})]\gamma^{K}\}$
2. Second-Level DP (customer assign):$\sum_{j=1}^{J} log\{[\frac{\Gamma(\alpha)}{\Gamma(n_{j..}+\alpha)}\Pi_{k=1}^{K}(\Gamma(n_{j\cdot k}))]\alpha^{m_{\cdot\cdot}}\}$
3. Likelihood: $\sum_{k=1}^{K}\{log(\frac{\Gamma(W\lambda_0)}{\Gamma(n_{\cdot\cdot k}+W\lambda_0)})+\sum_{w=1}^{W} log(\frac{\Gamma(\lambda_0+n_{\cdot\cdot k}^{w})}{\Gamma(\lambda_0)})\}$

Though Teh's 2nd version code only has beta sampler which samples z_{ji} instead of t_{ji}, k_{jt}, we can still reuse his data structure by assuming that dishes can have at most one table in each restaurant.

2.Method

Structure	**variable**	**ME_Package**	**Teh_Package**
First-Level DP	1) concentration parameter γ 2) number of data $m_{..}$ 3) number of cluster K 4) number of data in each cluster $m_{.k}$	prior.γ sum(mj) length(classes) classes.nt	hdp.dp{1}.α sum(hdp.dp{1}.classnd) sum(hdp.dp{1}.classnt) hdp.dp{1}.classnd
Second-Level DP	1) concentration parameter α 2) number of data $n_{j..}$ 3) number of cluster $m_{j.}$ 4) number of data in each cluster $n_{jt.}$	prior.α N(j) mj(j) tables{j}{t}.Nc	hdp.dp{j}.α hdp.dp{j}.numdata sum(hdp.dp{j}.classnt) hdp.dp{j}.classnd(k_{jt})
Likelihood	1) hyperparameter $lambda_0, W$ 2) Multinomial Distribution $n^w_{\cdot\cdot k}$	prior.phi,prior.W classesk.suff(w)	hdp.base.hh,length(hdp.base.hh) hdp.base.classqq(w,j)

Functions	**MEPackage**	**TehPackage**
Sampling	1) sample t_{ji} for reinitialization of restaurant	sample z_{ji}, β, m
Search Moves	1) Local Moves 2) Merge Moves 3) Decompose Moves	... ...
Utilities	1) data manipulation (add/delete structure) 2) change of likelihood 3) Graphic	... adddatalik.marglikelihoods ...