Large Sample Asymptotic for Nonparametric Mixture Model with Count Data

Vu Nguyen, Dinh Phung, Trung Le and Svetha Venkatesh
Centre for Pattern Recognition and Data Analytics, Deakin University, Australia

MOTIVATION

- Scalability in Bayesian nonparametric.
- Hard clustering for fast inference.
- Small variance asymptotic [1,2]
- Multinomial observations with large number of trials [3].

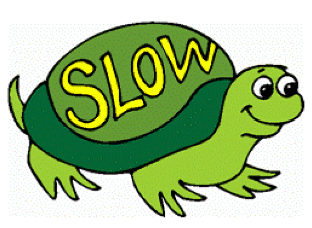
CONTRIBUTIONS

- A novel view of Large Sample Asymptotic (LSA).
- Theoretical derivation for the proposed LSA on Dirichlet Process Mixture model with count data.
- Comparative study with the existing clustering techniques.

Dirichlet Process Mixture - Gibbs Sampler

Gibbs sampler for posterior inference

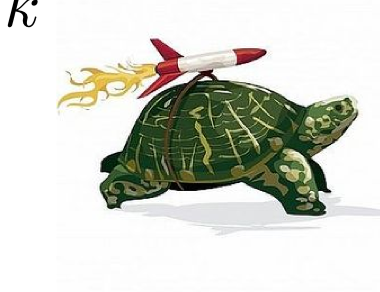
$$p(z_i = k \mid z_{-i}, x_i, \phi, \alpha, G) \propto \begin{cases} N_k \times p(x_i|\phi_{z_i}) & \text{used } k \\ \alpha \times \int_\phi p(x_i|\phi) dG(\phi) & \text{new } k \end{cases}$$



DPM - Small Variance Asymptotic [1,2]

Assume the data variance approaches to zero $\sigma \to 0$

$$\lim_{\sigma \to 0} \hat{\gamma}(z_i = k) = \begin{cases} D_{KL}(x_i, \phi_k) & \text{used } k \\ \lambda & \text{new } k \end{cases}$$



where $D_{KL}(x_i, \phi_k)$ is a KL divergence.

Large Sample Asymptotic (LSA)

Multinomial likelihood $p(x \mid \phi) = \frac{n!}{\prod_{d=1}^D x_d!} \prod_{d=1}^D \phi_d^{x_d}$

Average Log Likelihood $\bar{L} \equiv \log p(x \mid \phi)^{\frac{1}{n}}$

Equivalently $\bar{L} = \frac{1}{n} \log n! - \frac{1}{n} \sum_{d=1}^D \log x_d! + \sum_{d=1}^D \frac{x_d}{n} \log \phi_d$

Stirling Approximation $\log n! = n \log n - n + O(\log n)$

Alternative representation of Multinomial likelihood

$$P(x \mid \phi) = \exp\{-n D_{KL}(\hat{\boldsymbol{x}} || \phi) + O(\log n) + \sum_{d=1}^D O(\log x_d)\}$$

DPM - LSA

Sampling $z_i = k$

$$p(z_i = k) \propto N_k \times \exp[-n D_{KL}(\hat{\boldsymbol{x}}_i || \phi_{z_i}) + T]$$

Sampling $z_i = k^{\text{new}}$

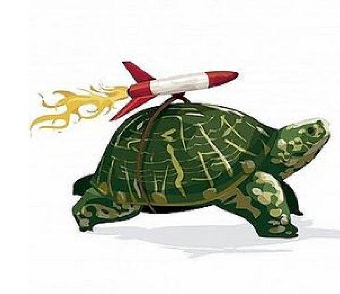
$$p(z_i = k^{\text{new}}) \propto p(z_i = k^{\text{new}} | z_{-i}, \alpha) p(x_i | z_i = k^{\text{new}}, G)$$

$$p(z_i = k^{\text{new}}) \propto \alpha \frac{n!}{\prod_{d=1}^D x_{id}!} \frac{\Gamma(\sum_{d=1}^D \gamma_d)}{\Gamma(\sum_{d=1}^D [\gamma_d + x_{id}])} \prod_{d=1}^D \frac{\Gamma(x_{id} + \gamma_d)}{\Gamma(\gamma_d)}$$

$$p(z_i = k^{\text{new}}) \propto C(x_i) \exp(-n\lambda) \exp[O(\log n)]$$

Let the number of trials in Multinomial distribution go to infinity

$$\lim_{n \to \infty} \hat{\gamma}(z_i = k) = \begin{cases} D_{KL}(\hat{x}_i || \phi_k) & \text{used } k \\ \lambda & \text{new } k. \end{cases}$$



Experiments

NUS WIDE dataset
- 3411 images
- SIFT features (500 dimensions)

Matlab environment

Approach	AP	K-means	GMM	DPM	DPmeans	DPM-LSA
Setting	Euclidian distance	Euclidian distance	Expectation Maximization	Collapsed Gibbs	$\lambda = 5960$	$\lambda = 1.73$
#Cluster	K=18	K=5-30	K=5-30	K=17	K=17	K=19
NMI	0.166	0.19(.01)	0.19(.01)	**0.188**	0.161	*0.174*
F1score	0.145	0.16(.01)	0.16(.01)	0.173	0.166	**0.184**
Time	167.8	14	15	1200	38	39.8

Reference

[1] Kulis, Brian, and Michael I. Jordan. "Revisiting k-means: New Algorithms via Bayesian Nonparametrics." ICML 2012.
[2] Jiang, Ke, Brian Kulis, and Michael I. Jordan. "Small-variance asymptotics for exponential family Dirichlet process mixture models." NIPS 2012.
[3] Shlens, Jonathon. "Notes on Kullback-Leibler Divergence and Likelihood." *arXiv preprint arXiv:1404.2000* (2014).