New Representation to Unify Ideas

Donglai Wei

2010.8.3

0) Notations:

Hierarchical Dirichlet Process Model with Dirichlet-Multinomial:

Hyper-parameter:
α, γ: HDP concentration parameter
$\vec{\lambda}$: Prior for Dirichlet Distribution(W=dim($\vec{\lambda}$):number of different words;uniform prior:$\lambda_1, ..., \lambda_W = \lambda_0$)

Hidden Variable:
(M-step)z: Discrete assignment(t_{ji}, k_{jt} correspond to customer,table assignment in Chinese Restaurant Franchise)
(E-step)θ: Multinomial parameter

Observation:
x:$\in$(1,...W)

Auxilary Counting Parameter:
n_{jtk}number of customers in table t Restaurant j serving dish k
$n_{jt.}$number of customers in table t in Restaurant j
$n_{j..}$number of customers in in Restaurant j
$n_{..k}$number of customers serving dish k
$n^w_{..k}$number of occurence of word w in dish k
$n_{j.k}$number of customers in the table in Restaurant J that serves dish k
m_{jk}number of tables in Restaurant j serving dish k
$m_{..}$number of tables in total
$m_{.k}$number of tables in dish k

1)Representation:

J Restaurants(Documents),K dishes(Topics) and W words

Goal:(Marginalize θ and Search over z:)
Maximize log Probability $L = log\ p(x,z|\lambda)$

Crazy Idea: Represent the data in J-W-K 3D coordinates

0.1) With reasonable γ,the object function prefer not to have two tables sharing the same dish in the same restaurant. So tables are just the projection of the Dishes into the Restaurants.

0.2) Imagine, in the beginning, every customer is served with a junk dish k=0
For the corresponding J-W plane, we can see it as a grid of integer points with coordinates (j,w).
For each point (j,w), it is associated with a number $n^w_{j..}$ (number of word w in Restaurant j).
The goal is to distribute these $n^w_{j..}$ into different ks to maximize L, under the constraints that $n^w_{j..} = \sum_{k=1}^K n^w_{j.k}$

1) Decompose Restaurant j: Fix J=j, Reconfig K-W plane
1.0) Sample new tables: drag all the points back to W axis(k=0)+sample promising Ks to form new table
1.1) Local-Table:exchanging datas between lines in W direction
1.2) Search-k: moving (lines that are in W direction) in K direction
1.3) Merge-Table: merging lines that are in W direction

2) Decompose Dish: Fix K=k, Reconfig J-W plane
2.0) Decompose Restaurants $j \in \{j : \exists t\ s.t. k_{jt} = k\}$ without dish k: try to explain the J-W planes by others
If the J-W plane cannot be well explained by others:
2.1) Local-Dish: exchanging datas between lines in J direction with the same w coordinate
2.2) Merge-Dish: merge two J-W plane

a) z-m View:

$L = logp(x,z|\lambda)$
$=\sum_{j=1}^J [log\frac{\Gamma(\alpha)}{\Gamma(n_{j..}+\alpha)}] + log(\Gamma(\gamma))$
(constant)
$+\sum_{k=1}^K [log(\frac{\Pi_{w=1}^W \Gamma(\lambda_0+n^w_{..k})\ \Pi_{j=1}^J \alpha\Gamma(n_{jkt_k})}{\Gamma(n_{..k}+W\lambda_0)}) + log(\gamma\frac{\Gamma(W\lambda_0)}{\Gamma(\lambda_0)^W})]$
(z-term:J-W plane,counts in J,W direction)
$+log(\frac{\Pi_{k=1}^K \Gamma(m_{.k})}{\Gamma(m_{..}+\gamma)})$
(m-term:)

b) t-k View:

$L = logp(x,z|\lambda)$
$=\sum_{j=1}^J [log\frac{\Gamma(\alpha)}{\Gamma(n_{j..}+\alpha)}] + log(\Gamma(\gamma))$
(constant:)
$+\sum_{j=1}^J \{\sum_{t=1}^{m_{j.}} [log(\Gamma(n_{jt.}) + log\alpha]\}$
(t-term:K-W plane,counts in W direction)
$+log\frac{1}{\Gamma(m_{..}+\gamma)} + \sum_{k=1}^K [log(\frac{\Pi_{w=1}^W \Gamma(\lambda_0+n^w_{..k})}{\Gamma(n_{..k}+W\lambda_0)}) + log(\frac{\Gamma(W\lambda_0)}{\Gamma(\lambda_0)^W}) + log(\Gamma(m_{.k}) + log\gamma]$
(k-term:J-W plane,counts in J direction)