Hierarchical Dirichlet Process Mixtureの導出

電気通信大学　電気通信学研究科
中村 友昭　(naka_t@apple.ee.uec.ac.jp)

平成23 年2 月10 日

1　Hierarchial Dirichlet Process

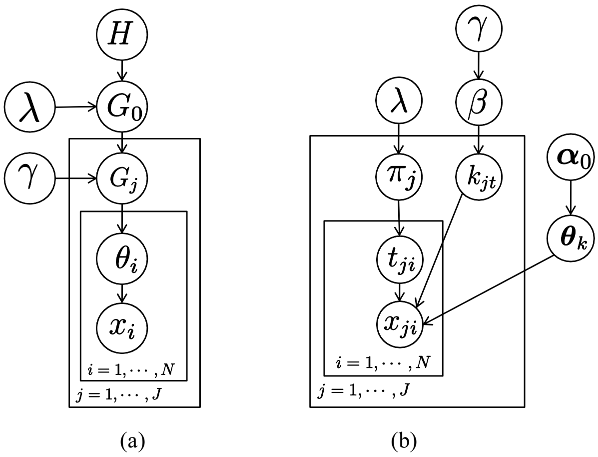

図 1: (a)HDP のグラフィカルモデル　(b) 文書分類における HDPM のグラフィカルモデル

図1(a) が Hierarchical Dirichlet Process (HDP)[1] のグラフィカルモデルである．HDP では，ディリクレ過程により生成された確率分布 G_0 が，基底分布になることで階層化されている．本稿では，実際に HDP を文書分類に適用し，文書分類における Hierarchical Dirichlet Process Mixture(HDPM) の導出を行う．

図1(b) が具体的な文書分類における HDPM のグラフィカルモデルである．HDPM では，各文書内の単語がクラスタリングされており，そのクラス内の単語の集合に対してトピックを割り当てることで，文書のトピックを推定する．x_{ji} は j 番目の文書内の i 番目の単語であり，そのクラスが t_{ji} である．単語 x_{ji} のクラスが $t_{ji} = t$ となる事前確率は，Chinese Restaurant Process (CRP) によって生成され，

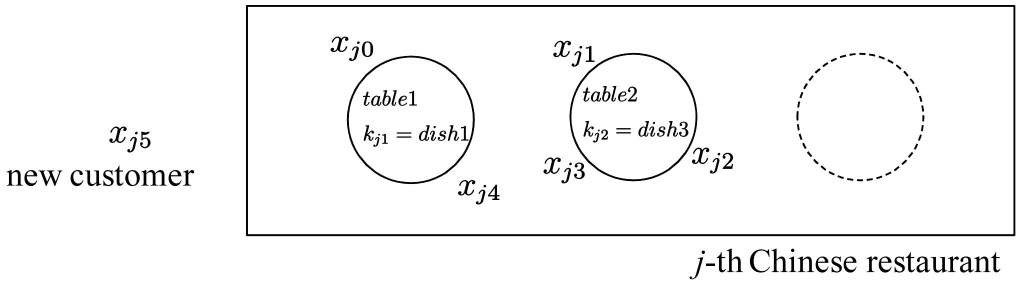

図 2: CRF におけるテーブルの選択

$$P(t_{ji} = t|\lambda) = \begin{cases} \frac{N_{jt}}{\gamma+N_j-1} & (t = 1, \cdots, T_j) \\ \frac{\lambda}{\gamma+N_j-1} & (t = T_j + 1) \end{cases} \tag{1}$$

で表される．ただし，N_j は j 番目の文書内の全単語数，N_{jt} は j 番目の文書内でクラス t が割り当てられている単語数である．$t = T_j + 1$ とは，新たなクラスを生成することを意味している．さらに，各クラス $t = \{1, \cdots, T_j\}$ にはトピック k_{jt} が割り当てられる．クラス t にトピック $k_{jt} = k$ が割り当てられる事前確率も CRP によって生成され，

$$P(k_{jt} = k|\lambda) = \begin{cases} \frac{M_k}{\gamma+M-1} & (k = 1, \cdots, K) \\ \frac{\gamma}{\gamma+M-1} & (k = K + 1) \end{cases} \tag{2}$$

と表される．ただし，M は全文書のクラスの総数 $M = \sum_j^J T_j$ であり，M_k は全文書でトピック k が割り当てられているクラスの総数である．

このようにして，j 番目の文書の i 番目のクラスタ $t_{ji} = t$ が決定され，そのトピックが $k_{jt} = k$ となるとき，単語 x_{ji} は，$\boldsymbol{\theta}_k$ をパラメタとする多項分布から生成される．この多項分布の事前分布として $\boldsymbol{\alpha}_0$ をパラメタとするディリクレ事前分布を仮定している．

HDPM ではこのように二段階で CRP により事前確率を生成する．このような過程は，Chinese Restaurant Franchise (CRF) によって表現される [1].

2 Chinese Restaurant Franchise

CRF では，文書を中華料理のチェーン店の 1 つと考え，各店舗 j にはテーブルが T_j 個用意されている．各テーブルには料理 k_{jt} が置かれており，店舗 j に来た i 番目の客 (単語 x_{ji}) は，テーブル t_{ji} に座り，置かれている料理を食べる．CRF では，客がテーブルを選択する過程と，テーブルに料理を置く過程の二つの過程に分けて考えることができる．

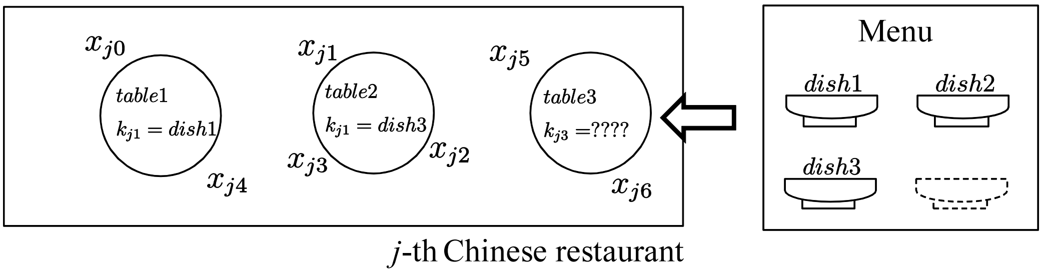

図 3: CRF における料理の選択

2.1 テーブル選択過程

今，図 2 の状態を考える．j 番目の店に新たな客が来店した際には，客 x_{ji} がテーブル t に座る事前確率は，式 (5) より次のようになる．

$$P(t_{ji} = t|\lambda) = \begin{cases} \frac{N_{jt}}{\gamma+N_j-1} & (t = 1, \cdots, T_j) \\ \frac{\lambda}{\gamma+N_j-1} & (t = T_j + 1) \end{cases} \tag{3}$$

この式において，N_{jt} は j 番目のチェーン店においてテーブル t に座っている人の人数であり，そのテーブルの人気に相当する．すなわち，客は人気のあるテーブルに座る確率が高くなる．(CRF の Chinese は，中国人が人気のあるテーブルに好んで座ることから，名付けられている．) また，その客がそのテーブルに置かれている料理を好む確率は $P(x_{ji}|\boldsymbol{X}_{k=k_{jt}})$ で表される．ただし，$\boldsymbol{X}_k$ は全店で料理 k を食べている客の集合である．したがって，客 x_{ji} がテーブル t に座る事後確率は，ベイズの定理より

$$P(t_{ji} = t|\boldsymbol{X}, \lambda) \propto P(x_{ji}|\boldsymbol{X}_{k=k_{jt}})P(t_{ji}|\lambda) \tag{4}$$

$$= \begin{cases} P(x_{ji}|\boldsymbol{X}_{k=k_{jt}})\frac{N_{jt}}{\gamma+N_j-1} & (t = 1, \cdots, T_j) \\ P(x_{ji}|\boldsymbol{X}_{k=k_{jt}})\frac{\lambda}{\gamma+N_j-1} & (t = T_j + 1) \end{cases} \tag{5}$$

となり，この確率にしたがって客はテーブルを選択する．ただし，$\boldsymbol{X}$ は全客の集合である．図 2 の状態において，$t = 3$ のテーブルを選択することは，新たなテーブルを生成することを意味する．新たなテーブルが選択された際には，次に説明する料理選択過程により料理が決定される．

2.2 料理選択過程

次に，図 3 の状態を考える．テーブル $t = 3$ に料理 k を置かれる事前確率は，式 (2) より次のようになる．

$$P(k_{jt}=k|\lambda)=\begin{cases}\frac{M_k}{\gamma+M-1} & (k=1,\cdots,K)\\ \frac{\gamma}{\gamma+M-1} & (k=K+1)\end{cases} \tag{6}$$

この式において，M_k は全店で料理 k が置かれているテーブルの数であり，料理の人気に相当する．すなわち，人気のあるメニューは選択される確率が高くなる．(CRF の Franchise とは，全店で共通のメニューから，全店での人気によって料理が選択されることから名付けられている．) また，既にテーブル t に座っている客の集合 $\boldsymbol{X}_{jt}$ が料理 k を好む確率は $P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)$ で表される．したがって，テーブル $t=3$ に料理 k を置く事後確率は，ベイズの定理より

$$P(k_{jt}=k|\boldsymbol{X},\gamma)=P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)P(k_{jt}=k|\gamma) \tag{7}$$

$$=\begin{cases}P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\frac{M_k}{\gamma+M-1} & (k=1,\cdots,K)\\ P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\frac{\gamma}{\gamma+M-1} & (k=K+1)\end{cases} \tag{8}$$

となり，この確率にしたがって料理が選択される．図 3 の状態において，$k=4$ の料理を選択することは，新たな料理を生成することを意味する．

3 HDPM の学習

HDPM でのクラスタリングは，Gibbs Sampling によって，モデルのパラメタ θ_k，テーブルの割り当て t_{ji}，料理の割り当て k_{jt} を推定するこによって実現される．Gibbs Sampling では，x_{ji} を除いた客の集合 $\boldsymbol{X}^{-ji}$ を条件とした条件付確率からテーブル t_{ji} をサンプリングする．

$$t_{ji}\sim P(t_{ji}|\boldsymbol{X}^{-ji},\lambda) \tag{9}$$

テーブル t に置かれる料理 k_{jt} は，テーブル t に座っている客 $\boldsymbol{X}_{jt}$ を除いた客の集合 $\boldsymbol{X}^{-jt}$ を条件とした条件付き確率からサンプリングする．

$$k_{jt}\sim P(k_{jt}|\boldsymbol{X}^{-jt},\gamma) \tag{10}$$

式 (9) を全店の全客に対して繰り返し，式 (10) を全店の全テーブルに対して繰り返すことで，パラメタが推定される．

4 HDPM の具体的導出

実際に HDPM を文書分類に適用する際に，式 (5) と式 (8) 内の $P(\boldsymbol{X}'|\boldsymbol{X}_k)$ を導出する必要がある．文書分類では，単語 x_{ji} はパラメタが $\boldsymbol{\theta}_k$ の多項分布から生成

される．

$$P(x_{ji} = w|\boldsymbol{\theta}_k) = \mathcal{M}(x_{ji} = w|\boldsymbol{\theta}_k) = \theta_{kw} \tag{11}$$

ただし，w は単語のインデックスを表している．次に，トピック (料理)k が割り当てられている全データを $\boldsymbol{X}_k$ と置くと，その尤度は，

$$P(\boldsymbol{X}_k|\boldsymbol{\theta}_k) = \prod_n^{N_k} P(x_{kn}|\boldsymbol{\theta}_k) = \prod_w^{W} \theta_{kw}^{N_{kw}} \tag{12}$$

となる．ただし，N_{kw} は，$\boldsymbol{X}_k$ 内で単語 w が発生した回数である．さらに，$\boldsymbol{\theta}_k$ はディリクレ事前分布

$$P(\boldsymbol{\theta}_k) = \mathcal{D}(\boldsymbol{\theta}_k|\boldsymbol{\alpha}_0) = \frac{1}{Z(\boldsymbol{\alpha}_0)} \prod_w^{W} \theta_{kw}^{\alpha_{0w}-1} \tag{13}$$

によって生成される．ただし，$Z(\boldsymbol{\alpha}_0)$ は正規化項を表し以下のようになる．

$$Z(\boldsymbol{\alpha}_0) = \frac{\prod_w^W \Gamma(\alpha_{0w})}{\Gamma(\sum_w^W \alpha_{0w})} \tag{14}$$

以上より，$\boldsymbol{X}_k$ と $\boldsymbol{\theta}_k$ の同時確率はは次式のようになる．

$$\begin{aligned} P(\boldsymbol{X}_k, \boldsymbol{\theta}_k) &= P(\boldsymbol{X}_k|\boldsymbol{\theta}_k)P(\boldsymbol{\theta}_k) = \prod_w^{W} \theta_{kw}^{N_{kw}} \frac{1}{Z(\boldsymbol{\alpha}_0)} \prod_w^{W} \theta_{kw}^{\alpha_{0w}-1} && (15) \\ &= \frac{1}{Z(\boldsymbol{\alpha}_0)} \prod_w^{W} \theta_{kw}^{N_{kw}+\alpha_{0w}-1} && (16) \end{aligned}$$

次に，$\boldsymbol{\theta}_k$ で周辺化し，周辺尤度を求める．

$$\begin{aligned} P(\boldsymbol{X}_k) &= \int P(\boldsymbol{X}_k, \boldsymbol{\theta}_k) d\boldsymbol{\theta}_k && (17) \\ &= \frac{1}{Z(\boldsymbol{\alpha}_0)} \int \prod_w^{W} \theta_{kw}^{N_{kw}+\alpha_{0w}-1} d\boldsymbol{\theta}_k && (18) \\ &= \frac{Z(\boldsymbol{\alpha}'_k)}{Z(\boldsymbol{\alpha}_0)} \int \frac{1}{Z(\boldsymbol{\alpha}'_k)} \prod_w^{W} \theta_{kw}^{\alpha'_{kw}-1} d\boldsymbol{\theta}_k && (19) \end{aligned}$$

ただし，$\alpha'_{kw} = N_{kw} + \alpha_{0w}$ である．この式の積分内がディリクレ分布と見なせ，

$$\int \frac{1}{Z(\boldsymbol{\alpha}'_k)} \prod_w^{W} \theta_{kw}^{\alpha'_{kw}-1} d\boldsymbol{\theta}_k = \int \mathcal{D}(\boldsymbol{\theta}_k|\alpha'_{kw}) d\boldsymbol{\theta}_k = 1 \tag{20}$$

が成り立ち，最終的に以下の式を得る．

$$P(\boldsymbol{X}_k) = \frac{Z(\boldsymbol{\alpha}'_k)}{Z(\boldsymbol{\alpha}_0)} \tag{21}$$

DPM で必要となるのは，新たな単語 x_{new} がクラス k に属する尤度であり，以下のようになる．

$$\begin{aligned}
P(x_{new}|\boldsymbol{X}_k) &= \frac{P(x_{new}, \boldsymbol{X}_k)}{P(\boldsymbol{X}_k)} = \frac{Z(\{\alpha'_{k0}, \cdots, \alpha'_{k,w=x_{new}} + 1, \cdots, \alpha'_{kW}\})}{Z(\boldsymbol{\alpha}'_k)} && (22)\\
&= \frac{\Gamma(\alpha'_{k0}) \cdots \Gamma(\alpha'_{k,w=x_{new}} + 1) \cdots \Gamma(\alpha'_{kW})}{\Gamma(\sum_w^W \alpha'_{0w} + 1)} \frac{\Gamma(\sum_w^W \alpha'_{0w})}{\prod_w^W \Gamma(\alpha'_{0w})} && (23)\\
&= \frac{\alpha'_{k,w=x_{new}} \prod_w^W \Gamma(\alpha'_{kw}))}{\sum_w^W \alpha'_{0w} \Gamma(\sum_w^W \alpha'_{0w})} \frac{\Gamma(\sum_w^W \alpha'_{0w})}{\prod_w^W \Gamma(\alpha'_{0w})} && (24)\\
&= \frac{\alpha'_{k,w=x_{new}}}{\sum_w^W \alpha'_{0w}} = \frac{N_{k,w=x_{new}} + \alpha'_{k,w=x_{new}}}{\sum_w^W N_{kw} + \alpha_{0w}} && (25)
\end{aligned}$$

ただし，式変形にはガンマ関数の性質 $\Gamma(a+1) = a\Gamma(a)$ を利用した．

以上より，Gibbs Sampling によるパラメタ推定アルゴリズムは次のようになる．

[HDPM algorithm]――――――――――――――――――――――――――

while 収束するまで

全店の全客のテーブルをサンプリング

for $j = 0, \cdots, J$

for $i = 0, \cdots, N$

1. $w = x_{ji}$ をテーブル $t = t_{ji}$ から削除し，$k = k_{jt}$ のパラメタを更新

$$\alpha_{kw} -- \tag{26}$$
$$N_{jt} -- \tag{27}$$
$$M_k -- \tag{28}$$

2. 新たな t を事後分布からサンプリング

$$t \sim \begin{cases} P(x_{ji}|\boldsymbol{X}_{k=k_{jt}})\frac{N_{jt}}{\gamma+N_j-1} & (t = t, \cdots, T_j) \\ P(x_{ji}|\boldsymbol{X}_{k=k_{jt}})\frac{\lambda}{\gamma+N_j-1} & (t = T_{j+1}) \end{cases} \tag{29}$$

3. if $t = T_j + 1$：新たなテーブルに置く料理をサンプリング

$$k_{jt} \sim \begin{cases} P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\frac{M_k}{\gamma+M-1} & (k = 1, \cdots, 3) \\ P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\frac{\gamma}{\gamma+M-1} & (k = 4) \end{cases} \tag{30}$$
$$T_j = T_j + 1 \tag{31}$$

4. $t_{ji} = t$ とし，トピック $k = k_{jt}$ のパラメタを更新

$$\alpha_{kw} ++ \tag{32}$$
$$N_{jt} ++ \tag{33}$$
$$M_k ++ \tag{34}$$

5. 空のテーブルを削除

end for

end for

全店の全テーブルに置かれる料理をサンプリング

for $j = 0, \cdots, J$

for $t = 0, \cdots, T_j$

1. $\boldsymbol{X}_t$ を料理 $k = k_{jt}$ を食べている客から除外し，のパラメタを更新

$$\alpha_{kw} -= N_{tw} \qquad \text{for all } w \tag{35}$$
$$M_k -- \tag{36}$$

ただし，N_{tw} は $\boldsymbol{X}_t$ のうち単語 w が発生した回数である．

2. 料理をサンプリング

$$k_{jt} \sim \begin{cases} P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\frac{M_k}{\gamma+M-1} & (k = 1, \cdots, K) \\ P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\frac{\gamma}{\gamma+M-1} & (k = K+1) \end{cases} \tag{37}$$

3. トピック $k = k_{jt}$ のパラメタを更新

$$\alpha_{kw} += N_{tw} \tag{38}$$
$$M_k ++ \tag{39}$$

新たな料理が生成された場合は $K = K + 1$ とする．

4. テーブルに置かれていない料理を削除

end for
end for
end while

参考文献

[1] Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei, "Hierarchical Dirichlet processes," Journal of the American Statistical Association, vol.101, no.476, pp.1566-1581, 2006.