Dual Topics to Bicluster Model

Inference Equation

DANIEL RUGELES
danrugeles@gmail.com

[Revised on 7 November 2016]

In this document we show how to integrate out the multinomial parameters for Dual Topics to Bicluster Model (DT2B). Without loss of generality, we derive the equations applied to a matrix representing the visits of users to places. Collapsing the parameters supports fully stochastic Gibbs sampling where the model iteratively updates the user-topic/place-topic pair assigned to each user/place visit. Typically, stochastic sampling leads to quicker convergence to the stationary state of the Markov chain made up of the Gibbs samples. We thank Bob Carpenter for writing with high-detail how to integrating out multinomial Parameters in the latent Dirichlet allocation and naive Bayes models. (`https://lingpipe.files.wordpress.com`). To begin, we present the description of the Dual Topics to Bicluster model in Figure 1

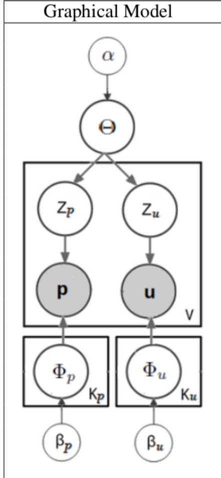
Graphical Model	Variable	Dimension	Description	Generative Process
	v	1	Index to a visit record	- Draw a distribution over topics Θ $\Theta \sim Dirichlet(\alpha)$ For i in $1..Ku$ - Draw a distribution over the users Φ_i $\Phi_i \sim Dirichlet(\beta_u)$ For j in $1..Kp$ - Draw a distribution over the places Φ_j $\Phi_j \sim Dirichlet(\beta_p)$ For v in 1..V - Draw simultaneously a user-topic - and a place-topic $z_{p(v)}, z_{u(v)}$ $z_{p(v)}, z_{u(v)} \sim Categorical^2(\Theta)$ - Draw a user u $u \sim Categorical(\Phi_{z_{u(v)}})$ - Draw a user p $p \sim Categorical(\Phi_{z_{p(v)}})$
	V	1	Number of places	
	K_p	1	Number of place-topics	
	K_u	1	Number of user-topics	
	U	1	Number of Users	
	P	1	Number of Places	
	Φ_p	$P \times K_p$	Place per place-topic distribution $Dirichlet(\beta_p)$	
	β_p	$P \times 1$	Hyper-parameter of Φ_p	
	p_v	1	Observed place in the v-th visit record $Categorical(\Phi_p, K_p)$	
	Φ_u	$U \times K_u$	User per user-topic distribution $Dirichlet(\beta_u)$	
	β_u	$U \times 1$	Hyper-parameter of Φ_u	
	u_v	1	Observed user in the v-th visit record $Categorical(\Phi_u, K_u)$	
	θ	$K_u \times K_p$	Joint distribution of user-topic and place-topics $Dirichlet(\alpha)$	
	α	$K_u \times K_p$	Hyper-parameter of θ	
	$z_{u(v)}, z_{p(v)}$	1,1	User-topic and Place-topic assigned to the v-th visit record $Categorical^2(\Theta, K_p, K_u)$	

FIG. 1. Graphical representation, description and generative process of the DT2B

Our objective is to obtain an analytic form of the posterior distribution of $P(z_u, z_p|u, p)$. Based on this distribution we can sample a user-topic/place-topic pair given a user/place pair.

1. Derivation of the Inference Equation using Collapsed Gibbs Sampling

We start by deriving the posterior distribution used to assign the topics. The derivation includes two main sections. First, we apply collapsed gibbs sampling to derive analytically the posterior distribution given the conditions of the model. Second, we use the fact that every record is independent and identically distributed so that we can find the posterior derivation for each particular record in the dataset.

1.1 *Posterior Derivation given the conditions of the model*

First we write down the target distribution that we want to derive:

$$P(z_{u,v}, z_{p,v} | z_{u,-v}, z_{p,-v}, u, p, \alpha, \beta_u, \beta_p) \tag{1.1}$$

The idea is to find the updating topics z_p and z_u for a given visit record v, given the training data. The definition of the variables used in this derivation can be found in figure 1.

Now, we start by formulating the joint distribution over all the variables in the model.

$$P(u, p, z_u, z_p, \Phi_u, \Phi_p, \Theta | \alpha, \beta_u, \beta_p) = \tag{1.2}$$

To obtain the target distribution using collapsed gibbs sampling, we need to integrate out the unknown parameters Φ_u, Φ_p and Θ.

$$\int_{\Theta}\int_{\Phi_u}\int_{\Phi_p} P(u, p, z_u, z_p, \Phi_u, \Phi_p, \Theta | \alpha, \beta_u, \beta_p)\, \mathrm{d}\Phi_p \mathrm{d}\Phi_u \mathrm{d}\Theta = \tag{1.3}$$

Next, we apply DT2B's independence assumptions.

$$\int_{\Theta}\int_{\Phi_u}\int_{\Phi_p} P(\Theta|\alpha) P(z_p|\Theta) P(z_u|\Theta) P(p|\Phi_p) P(\Phi_p|\beta_p) P(u|\Phi_u) P(\Phi_u|\beta_u)\, \mathrm{d}\Phi_p \mathrm{d}\Phi_u \mathrm{d}\Theta \tag{1.4}$$

We split the integrals into three independent cases:
Case 1 (Θ):

$$\int_{\Theta} P(\Theta|\alpha) P(z_p|\Theta) P(z_u|\Theta)\, \mathrm{d}\Theta = \tag{1.5}$$

We use the template to describe the V number of visit records:

$$\begin{aligned}\int_{\Theta} P(\Theta|\alpha) \prod_{v=1}^{V} \left(P(z_{p,v}|\Theta) P(z_{u,v}|\Theta)\right) \mathrm{d}\Theta &= \\ \int_{\Theta} P(\Theta|\alpha) \prod_{v=1}^{V} \left(P(z_{p,v}, z_{u,v}|\Theta)\right) \mathrm{d}\Theta &= \end{aligned} \tag{1.6}$$

Θ is a $K_u x K_p$ dimensional variable. As demonstrated in section **??**, Θ_k can be transformed into $\Theta_{i,j}$ using a bijection function g where $i \in \mathbb{N}_{1:K_u}, j \in \mathbb{N}_{1:K_p}$.

$$\int\limits_{\Theta_k} P(\Theta_k|\alpha) \prod_{v=1}^{V} \left(P\left(z_{p,v}, z_{u,v}|\Theta_{i,j}\right)\right) \mathrm{d}\Theta_k = \tag{1.7}$$

We replace the probability distribution with the definition of the Dirichlet distribution with equal number of parameters α as the dimensionality of Θ.

$$\iint\limits_{\Theta_{i,j}} \frac{\Gamma\left(\sum_{k=1}^{K_u x K_p}(\alpha_k)\right)}{\prod_{k=1}^{K_u x K_p}(\Gamma(\alpha_k))} \prod_{k=1}^{K_u x K_p} \left(\Theta_k^{\alpha_k - 1}\right) \prod_{v=1}^{V} (\Theta_{i,j}) \, \mathrm{d}\Theta_{i,j} = \tag{1.8}$$

For simplicity, We define the counting function $c(a,b,c,d)$ using an SQL query:

```
Select count(*)
From VisitRecords
Where usertopic=a
      and placetopic=b
      and user=c
      and place=d
```

If *a,b,c* or *d* are equal to *, then the correspondent filter is removed from the *Where* clause.

$$\int\limits_{\Theta_{i,j}} \frac{\Gamma\left(\sum_{k_u,k_p=1}^{K_u x K_p}\left(\alpha_{ku,kp}\right)\right)}{\prod_{k_u,k_p=1}^{K_u x K_p}\left(\Gamma\left(\alpha_{ku,kp}\right)\right)} \prod_{k_u,k_p=1}^{K_u x K_p} \left(\Theta_{k_u,k_p}^{\alpha_{ku,kp}-1}\right) \prod_{k_u,k_p=1}^{K_u x K_p} \left(\Theta_{k_u,k_p}^{c\left(k_p,k_u,*,*\right)}\right) \mathrm{d}\Theta_{k_p} \propto \tag{1.9}$$

Now, we arrange the variables so that we can describe a Dirichlet distribution which will integrate to one. After the integration we find the following:

$$\frac{\prod_{k_u,k_p=1}^{K_u x K_p}\left(\Gamma\left(c\left(k_p,k_u,*,*\right)+\alpha_{ku,kp}\right)\right)}{\Gamma\left(\sum_{k_u,k_p=1}^{K_u x K_p}\left(c\left(k_p,k_u,*,*\right)+\alpha_{ku,kp}\right)\right)} \tag{1.10}$$

Case 2 (Φ_u):
Following a similar formulation that the one used for case 1.

$$\int\limits_{\Phi_u} P(u|\Phi_u) P(\Phi_u|\beta_u) \, \mathrm{d}\Phi_u \tag{1.11}$$

$$\prod_{k_u=1}^{K_u} \left(\int\limits_{\Phi_u} P(u|\Phi_u) P(\Phi_u|\beta_u) \, \mathrm{d}\Phi_u \right) \tag{1.12}$$

We obtain:

$$\prod_{k_u=1}^{K_u}\left(\frac{\prod_{v=1}^{V}\left(\Gamma\left(c\left(*,k_u,*,u_v\right)+\beta_{u,v}\right)\right)}{\Gamma\left(\sum_{v=1}^{V}\left(c\left(*,k_u,*,u_v\right)+\beta_{u,v}\right)\right)}\right) \tag{1.13}$$

Case 3 (Φ_p):
Following the same formulation that the one used for case 2.

$$\int\limits_{\Phi_p} P\left(p|\Phi_p\right)P\left(\Phi_p|\beta_p\right)\mathrm{d}\Phi_p = \tag{1.14}$$

$$\prod_{k_p=1}^{K_p}\left(\int\limits_{\Phi_p} P\left(p|\Phi_p\right)P\left(\Phi_p|\beta_p\right)\mathrm{d}\Phi_p\right) = \tag{1.15}$$

We obtain:

$$\prod_{k_p=1}^{K_p}\left(\frac{\prod_{v=1}^{V}\left(\Gamma\left(c\left(k_p,*,p_v,*\right)+\beta_{p,v}\right)\right)}{\Gamma\left(\sum_{v=1}^{V}\left(c\left(k_p,*,p_v,*\right)+\beta_{p,v}\right)\right)}\right) \tag{1.16}$$

In Summary:

$$P\left(z_u,z_p,u,p|\alpha,\beta_u,\beta_p\right) \propto$$

$$\begin{aligned}\frac{\prod_{k_u,k_p=1}^{K_u x K_p}\left(\Gamma\left(c\left(k_p,k_u,*,*\right)+\alpha_{ku,kp}\right)\right)}{\Gamma\left(\sum_{k_u,k_p=1}^{K_u x K_p}\left(c\left(k_p,k_u,*,*\right)+\alpha_{ku,kp}\right)\right)}\prod_{k_u=1}^{K_u}\left(\frac{\prod_{v=1}^{V}\left(\Gamma\left(c\left(*,k_u,*,u_v\right)+\beta_{u,v}\right)\right)}{\Gamma\left(\sum_{v=1}^{V}\left(c\left(*,k_u,*,u_v\right)+\beta_{u,v}\right)\right)}\right)\\ \prod_{k_p=1}^{K_p}\left(\frac{\prod_{v=1}^{V}\left(\Gamma\left(c\left(k_p,*,p_v,*\right)+\beta_{p,v}\right)\right)}{\Gamma\left(\sum_{v=1}^{V}\left(c\left(k_p,*,p_v,*\right)+\beta_{p,v}\right)\right)}\right)\end{aligned} \tag{1.17}$$

1.2 *Posterior Derivation for each record*

Since every visit record is considered to be independent, we can update every single record's user-topic and place-topic instead, we refine our objective probability function to the following:

$$P\left(Z_u^{(v)}, Z_p^{(v)} | Z_u^{(-v)}, Z_p^{(-v)}, u_v, p_v, \alpha, \beta_u, \beta_p\right) \tag{1.18}$$

Where $Z_u^{(v)}, Z_p^{(v)}$ represents the user and place topic for the *v-th* record and $Z_u^{(-v)}, Z_p^{(-v)}$ represent the user and place topic for all other records. Once again, we split equation 1.17 into three cases.

Case 1 (Θ): Given $u_v, p_v, Z_u^{(-v)}, Z_p^{(-v)}$:

$$\frac{\prod_{k_u,k_p=1}^{K_u x K_p} \left(\Gamma\left(c\left(k_p, k_u, *, *\right)^{-v} + \alpha_{ku,kp}\right)\right)}{\Gamma\left(\sum_{k_u,k_p=1}^{K_u x K_p} \left(c\left(k_p, k_u, *, *\right)^{-v} + \alpha_{ku,kp}\right)\right)} \tag{1.19}$$

We use the property of the gamma distribution $\Gamma(x+1) = \Gamma(x)\dot{x}$ in order to extract the components that correspond to our *v-th* record of interest.

$$\frac{\prod_{k_u,k_p \neq z_{p,v}, z_{u,v}}^{K_u x K_p} \left(\Gamma\left(c\left(k_p, k_u, *, *\right)^{-v} + \alpha_{ku,kp}\right) \Gamma\left(c\left(z_{p,v}, z_{u,v}, *, *\right)^{-v} + \alpha_{zu_v,zp_v}\right)\right)}{\Gamma\left(1 + \sum_{k_u,k_p=1}^{K_u x K_p} \left(c\left(k_p, k_u, *, *\right)^{-v} + \alpha_{ku,kp}\right)\right)}$$
$$\frac{c\left(z_{p,v}, z_{u,v}, *, *\right)^{-v} + \alpha_{zu_v,zp_v}}{\Gamma\left(1 + \sum_{k_u,k_p=1}^{K_u x K_p} \left(c\left(k_p, k_u, *, *\right)^{-v} + \alpha_{ku,kp}\right)\right)} \propto$$

Finally, after simplifying all terms that are not proportional to the topics $Z_u^{(v)}, Z_p^{(v)}$

$$c\left(z_{p,v}, z_{u,v}, *, *\right)^{-v} + \alpha_{zu_v,zp_v} \tag{1.20}$$

We use the same formulation for the other cases and we obtain the updating equation for $Z_u^{(v)}, Z_p^{(v)}$

$$P\left(Z_u^{(v)}, Z_p^{(v)} | Z_u^{(-v)}, Z_p^{(-v)}, u_v, p_v, \alpha, \beta_u, \beta_p\right) \propto$$

$$\frac{\left(c\left(z_{p(v)}, z_{u(v)}, *, *\right)^{(-v)} + \alpha_{z_{u(v)},z_{p(v)}}\right)\left(c\left(z_{p(v)}, *, p_{(v)}, *\right)^{(-v)} + \beta_{p(v)}\right)\left(c\left(*, z_{u(v)}, *, u_{(v)}\right)^{(-v)} + \beta_{u(v)}\right)}{\sum_{j=1}^{|P|}\left(c\left(z_{p(v)}, *, p_{(j)}, *\right)^{(-v)} + \beta_{p(j)}\right)\sum_{i=1}^{|U|}\left(c\left(*, z_{u(v)}, *, u_{(i)}\right)^{(-v)} + \beta_{u(i)}\right)} \tag{1.21}$$

2. Inference algorithm using the inference equation

Equation 2.1 correspond to the posterior distribution that we use to update the user-topic $z_{u(v)}$ and place-topic $z_{p(v)}$ assigned to the *v-th* visit record. For each visit record, we must sample from this distribution to update its topic assignments, the posterior distribution must be recomputed after each assignment. Updating all the records corresponds to one iteration. After several iterations, the posterior distribution will converge, and we have found the topic assignment for each record.

$$P\left(z_{u(v)}, z_{p(v)} | z_{u(-v)}, z_{p(-v)}, u, p, \alpha, \beta_u, \beta_p\right) \propto \frac{\mathscr{Z}\,\mathscr{P}(v)\,\mathscr{U}(v)}{\sum_{j=1}^{|P|} \mathscr{P}(j) \sum_{i=1}^{|U|} \mathscr{U}(i)} \tag{2.1}$$

$$\mathscr{Z} = \sum_{m=1}^{|U|} \sum_{n=1}^{|P|} c\left(z_{p(v)}, z_{u(v)}, m, n\right)^{(-v)} + \alpha_{z_{u(v)}, z_{p(v)}}$$

$$\mathscr{P}(i) = \sum_{y=1}^{Kp} \sum_{n=1}^{|P|} c\left(z_{p(v)}, y, p_{(i)}, n\right)^{(-v)} + \beta_{p(i)}$$

$$\mathscr{U}(i) = \sum_{x=1}^{Ku} \sum_{m=1}^{|U|} c\left(x, z_{u(v)}, m, u_{(i)}\right)^{(-v)} + \beta_{u(i)}$$

$$c(x, y, m, n) = \sum_{j=1}^{|V|} I(z_{u(j)} = x, z_{p(j)} = y, u_{(j)} = m, p_{(j)} = n)$$

The inference equation depends on three types of counting. First, a normalized counting of the user assignment to the user-topics $\mathscr{U}$ capturing user co-occurrence. Second, a normalized counting of the place assignment to the place-topics $\mathscr{P}$ capturing place co-occurrence. And finally a counting over the current relationship between user-topics and place-topics $\mathscr{Z}$.