It Is Not Just What We Say, But How We Say Them: LDA-based Behavior-Topic Model
Supplementary Material

Minghui Qiu * Feida Zhu * Jing Jiang *

1 The Task

In this paper, we propose a LDA-based [1] behavior-topic model (B-LDA) which jointly models user topic interests and behavioral patterns for micro-blogs like Twitter. Our model is an extendsion of Twitter-LDA [2] which is catered for Twitter setting. The following presents the model and discuss the related inference details. Note that, this is a supplementary material for our paper on "It Is Not Just What We Say, But How We Say Them: LDA-based Behavior-Topic Model".

2 Model

Table 1 summarizes the set of notations and descriptions of our model parameters.

Notations	Descriptions
U	the total number of users
N_u	the total number of tweets in user u
$L_{u,n}$	the total number of words in u's n-th tweet
T	the total number of topics
b	a behavior in $\mathcal{B} = \{post, retweet, reply, mention\}$
y	a switch
z	a topic label
w	a word label
ϕ_t	topic-specific word distribution
ψ_t	topic-specific behavior distribution
ϕ'	background word distribution
θ_u	user-specific topic distribution
φ	Bernoulli distribution
α, η, β', β, γ	Dirichlet priors

Table 1: Notations and descriptions.

We now present our B-LDA model. First, we assume that there are T hidden topics, where each topic has a multinomial word distribution ϕ_t and a multinomial behavior distribution ψ_t. Each tweet has a single hidden topic which is sampled from the corresponding user's topic distribution θ_u ($1 \leq u \leq U$). We further assume that given a tweet with hidden topic t ($1 \leq t \leq T$), the words in this tweet are generated from two multinomial distributions, namely, a background model and a topic specific model. The background model ϕ' generates words commonly used in many tweets. The topic specific model ϕ_t generates words related to topic t. When we sample a word w ($1 \leq w \leq V$), we use a switch $y \in \{0, 1\}$ according to Bernoulli distribution φ, to decide which word distribution the word comes from. Specifically, if $y = 0$, the word w is sampled from ϕ'; otherwise, it is sampled from ϕ_t. We also assume the behavior pattern b ($b \in \mathcal{B}$) is sampled from the behavior distribution ψ_t. Lastly, we assume θ_u, ψ_t, ϕ', ϕ_t and φ have Dirichlet priors α, η, β', β and γ respectively. The plate notation and the generative process of the model

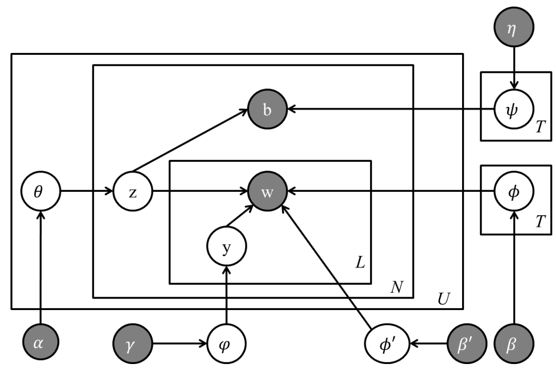

Figure 1: LDA-based behavior-topic model (B-LDA)

- For each topic $t = 1, \cdots, T$
  - Draw $\psi_t \sim \text{Dir}(\eta)$, $\phi_t \sim \text{Dir}(\beta)$
- Draw $\phi' \sim \text{Dir}(\beta')$, $\varphi \sim \text{Dir}(\gamma)$
- For each user $u = 1, \cdots, U$
  - Draw topic distribution $\theta_u \sim \text{Dir}(\alpha)$
  - For u's n-th tweet, $n = 1, \cdots, N_u$
    - ∘ Draw a topic $z_{u,n}$ from θ_u
    - ∘ For each word $l = 1, \cdots, L_{u,n}$
      - Draw $y_{u,n,l}$ from Bernoulli(φ)
      - Draw $w_{u,n,l} \sim \phi'$ if $y_{u,n,l} = 0$, otherwise draw $w_{u,n,l} \sim \phi_{z_{u,n}}$
    - ∘ Draw a posting behavior $b_{u,n} \sim \psi_{z_{u,n}}$

Figure 2: The generative process for all posts in B-LDA.

*School of Information System, Singapore Management University

are shown in Figure 1 and Figure 2.

2.1 Inference

We use Gibbs Sampling to estimate the parameters in the model. The Gibbs Sampling process is described in Algorithm 1.

Algorithm 1 Gibbs Sampling for B-LDA.

procedure GibbsSampling
    for each user $u = 1, \cdots, U$ **do**
        for u's n-th tweet, $n = 1, \cdots, N_u$ **do**
            Randomly assign a topic to $z_{u,n}$
            for each word $l = 1, \cdots, L_{u,n}$ **do**
                Randomly assign 0 or 1 to $y_{u,n,l}$
            end for
        end for
    end for
    for each Gibbs Sampling iteration **do**
        for each user $u = 1, \cdots, U$ **do**
            for u's n-th tweet, $n = 1, \cdots, N_u$ **do**
                Draw a topic $z_{u,n}$ according to Eqn. 2.1
                for each word $l = 1, \cdots, L_{u,n}$ **do**
                    Draw $y_{u,n,l}$ according to Eqn. 2.2
                end for
            end for
        end for
    end for
    Estimate model parameters θ, φ, ϕ', ϕ and ψ
end procedure

Hence, the problem is to compute the following two Gibbs updating rules.

To sample topic $z_{u,n}$, we use the following equation:

$$\begin{aligned} (2.1)\ & p(z_{u,n}|\mathbf{Z}_{\neg\{u,n\}}, \mathbf{W}, \mathbf{Y}, \mathbf{B}) \\ = \quad & \frac{p(\mathbf{Z}, \mathbf{W}, \mathbf{Y}, \mathbf{B}|\alpha, \beta, \beta', \gamma, \eta)}{p(\mathbf{Z}_{\neg\{u,n\}}, \mathbf{W}, \mathbf{Y}, \mathbf{B}|\alpha, \beta, \beta', \gamma, \eta}) \\ \propto \quad & \frac{p(\mathbf{Z}, \mathbf{W}, \mathbf{Y}, \mathbf{B}|\alpha, \beta, \beta', \gamma, \eta)}{p(\mathbf{Z}_{\neg\{u,n\}}, \mathbf{W}_{\neg\{u,n\}}, \mathbf{Y}_{\neg\{u,n\}}, \mathbf{B}_{\neg\{u,n\}}|\alpha, \beta, \beta', \gamma, \eta)}, \end{aligned}$$

where $\mathbf{Z}_{\neg\{u,n\}}$ denotes the set of all the topics in the data sets not including the topic of user u's n-th tweet.

Similarly, to sample topic $z_{u,n}$, we use the following equation.

$$\begin{aligned} (2.2) \qquad & p(y_{u,n,l}|\mathbf{Y}_{\neg_{u,n,l}}, \mathbf{Z}, \mathbf{W}, \mathbf{B}) \\ \propto \quad & \frac{p(\mathbf{Z}, \mathbf{W}, \mathbf{Y}, \mathbf{B}|\alpha, \beta, \beta', \gamma, \eta)}{p(\mathbf{Z}, \mathbf{W}_{\neg\{u,n,l\}}, \mathbf{Y}_{\neg\{u,n,l\}}, \mathbf{B}|\alpha, \beta, \beta', \gamma, \eta)}. \end{aligned}$$

After this process, we can obtain model parameters ϕ', ϕ, φ, ψ and θ.

2.2 Sampling topic $z_{u,n}$

We discuss how to derive Eqn. 2.1 for sampling topic $z_{u,n}$ in this section.

The problem is to compute $p(\mathbf{Z}, \mathbf{W}, \mathbf{Y}, \mathbf{B}|\alpha, \beta, \beta', \gamma, \eta)$. According to the model, we derive it as:

$$\begin{aligned} & p(\mathbf{Z}, \mathbf{W}, \mathbf{Y}, \mathbf{B}|\alpha, \beta, \beta', \gamma, \eta) \\ = \quad & p(\mathbf{W}|\mathbf{Z}, \mathbf{Y}, \beta, \beta')p(\mathbf{Y}|\gamma)p(\mathbf{B}|\mathbf{Z}, \alpha)p(\mathbf{Z}|\eta). \end{aligned}$$

We further derive $p(\mathbf{Z}|\eta)$ as:

$$p(\mathbf{Z}|\eta) = \int_\theta p(\mathbf{Z}|\theta)p(\theta|\eta)d\theta = \prod_{u=1}^{U} \frac{B(n_u^{\mathbf{t}} + \eta)}{B(\eta)},$$

where $n_u^{\mathbf{t}}$ is a vector in which each element denotes number of times that the corresponding topic occurs in user u's tweets.

Similarly, we can derive $p(\mathbf{B}|\mathbf{Z}, \alpha)$ and $p(\mathbf{Y}|\gamma)$ as:

$$\begin{aligned} p(\mathbf{B}|\mathbf{Z}, \alpha) \quad &= \quad \int_\psi p(\mathbf{B}|\psi)p(\psi|\mathbf{Z}, \alpha)d\psi = \prod_{t=1}^{T} \frac{B(n_t^{\mathbf{b}} + \alpha)}{B(\alpha)}, \\ p(\mathbf{Y}|\gamma) \quad &= \quad \frac{B(n_{(.)}^{\mathbf{y}} + \gamma)}{B(\gamma)}, \end{aligned}$$

where $n_t^{\mathbf{b}}$ is a vector in which each element denotes number of times that the corresponding posting behavior co-occurs with topic t, $n_{(.)}^{\mathbf{y}}$ has two elements denoting number of time $y = 0$ and $y = 1$ occurs.

We assume each word has a corresponding label y that indicates which model it is sample from. Specifically, if $y = 0$, the word is sample from background model, if $y = 1$, it is from topic specific model. To derive $p(\mathbf{W}|\mathbf{Z}, \mathbf{Y}, \beta, \beta')$, we then need to consider two types of word distribution ϕ and ϕ'. Specifically, we derive it as:

$$\begin{aligned} & p(\mathbf{W}|\mathbf{Z}, \mathbf{Y}, \beta, \beta') \\ = \quad & \int_\phi \int_{\phi'} p(\mathbf{W}|\mathbf{Z}, \mathbf{Y}, \phi, \phi')p(\phi|\beta)p(\phi'|\beta')d\phi d\phi' \\ = \quad & \frac{B(n_{y=0}^{\mathbf{w}} + \beta')}{B(\beta')} \prod_{t=1}^{T} \frac{B(n_{t,y=1}^{\mathbf{w}} + \beta)}{B(\beta)}, \end{aligned}$$

where $n_{y=0}^{\mathbf{w}}$ is a count vector in which each value denotes number of times the word is sampled and its label is $y = 0$, and each element in $n_{t,y=1}^{\mathbf{w}}$ means number of time the word is sampled when its topic is t and label is $y = 1$.

Given the above formulas, we can compute $p(\mathbf{Z}, \mathbf{W}, \mathbf{Y}, \mathbf{B}|\alpha, \beta, \beta', \gamma, \eta)$:

$$\begin{aligned} (2.3) \qquad & p(\mathbf{Z}, \mathbf{W}, \mathbf{Y}, \mathbf{B}|\alpha, \beta, \beta', \gamma, \eta) \\ = \quad & p(\mathbf{Y}|\gamma)p(\mathbf{W}|\mathbf{Z}, \mathbf{Y}, \beta, \beta')p(\mathbf{B}|\mathbf{Z}, \alpha)p(\mathbf{Z}|\eta) \\ = \quad & \frac{B(n_{(.)}^{\mathbf{y}} + \gamma)}{B(\gamma)} \frac{B(n_{y=0}^{\mathbf{w}} + \beta')}{B(\beta')} \prod_{t=1}^{T} \frac{B(n_{t,y=1}^{\mathbf{w}} + \beta)}{B(\beta)} \\ & \times \prod_{t=1}^{T} \frac{B(n_t^{\mathbf{b}} + \alpha)}{B(\alpha)} \prod_{u=1}^{U} \frac{B(n_u^{\mathbf{t}} + \eta)}{B(\eta)}. \end{aligned}$$

With Eqn 2.3, we are ready to derive 2.1. Let c denote $\{u, n\}$, we can derive it as follows.

$$\begin{aligned}(2.4)\quad & p(z_c|\mathbf{Z}_{\neg_c}, \mathbf{W}, \mathbf{Y}, \mathbf{B}) \\ = & \frac{p(\mathbf{Z}, \mathbf{W}, \mathbf{Y}, \mathbf{B}|\alpha, \beta, \beta', \gamma, \eta)}{p(\mathbf{Z}_{\neg_c}, \mathbf{W}_{\neg_c}, \mathbf{Y}_{\neg_c}, \mathbf{B}_{\neg_c}|\alpha, \beta', \beta, \gamma, \eta)} \\ = & \frac{B(n^{\mathbf{y}}_{(.)} + \gamma)}{B(n^{y}_{\neg_c} + \gamma)} \frac{B(n^{\mathbf{w}}_{y=0} + \beta')}{B(n^{\mathbf{w}}_{y=0,\neg_c} + \beta')} \prod_{t=1}^{T} \frac{B(n^{\mathbf{w}}_{t,y=1} + \beta)}{B(n^{\mathbf{w}}_{t,y=1,\neg_c} + \beta)} \\ & \times \prod_{t=1}^{T} \frac{B(n^{\mathbf{b}}_{t} + \alpha)}{B(n^{\mathbf{b}}_{t,\neg_c} + \alpha)} \prod_{u=1}^{U} \frac{B(n^{\mathbf{t}}_{u} + \eta)}{B(n^{\mathbf{t}}_{u,\neg_c} + \eta)}.\end{aligned}$$

When we sample a topic for z_c, we assume $\mathbf{Y}$, $\mathbf{W}$ and $\mathbf{Z}_{\neg_c}$ are fixed. Hence the above equation could be further simplified.

$$\begin{aligned}(2.5)\quad & p(z_c|\mathbf{Z}_{\neg_c}, \mathbf{W}, \mathbf{Y}, \mathbf{B}) \\ = & \prod_{t=1}^{T} \frac{B(n^{\mathbf{w}}_{t,y=1} + \beta)}{B(n^{\mathbf{w}}_{t,y=1,\neg_c} + \beta)} \\ & \times \prod_{t=1}^{T} \frac{B(n^{\mathbf{b}}_{t} + \alpha)}{B(n^{\mathbf{b}}_{t,\neg_c} + \alpha)} \prod_{u=1}^{U} \frac{B(n^{\mathbf{t}}_{u} + \eta)}{B(n^{\mathbf{t}}_{u,\neg_c} + \eta)}.\end{aligned}$$

To estimate the probability of assigning topic z to z_c, we need to compute $p(z_c = z|\mathbf{Z}_{\neg_c}, \mathbf{W}, \mathbf{Y}, \mathbf{B})$, which can be derived as follows.

$$\begin{aligned}(2.6)\quad & p(z_c = z|\mathbf{Z}_{\neg_c}, \mathbf{W}, \mathbf{Y}, \mathbf{B}) \\ = & \frac{B(n^{\mathbf{w}}_{z,y=1} + \beta)}{B(n^{\mathbf{w}}_{z,y=1,\neg_c} + \beta)} \\ & \times \frac{B(n^{\mathbf{b}}_{z} + \alpha)}{B(n^{\mathbf{b}}_{z,\neg_c} + \alpha)} \frac{B(n^{\mathbf{t}}_{u} + \eta)}{B(n^{\mathbf{t}}_{u,\neg_c} + \eta)},\end{aligned}$$

where the first component is computed as the following.

$$\begin{aligned}(2.7)\quad & \frac{B(n^{\mathbf{w}}_{z,y=1} + \beta)}{B(n^{\mathbf{w}}_{z,y=1,\neg} + \beta)} \\ = & \frac{\prod_{w=1}^{V} \Gamma(n^{w}_{z,y=1} + \beta)}{\prod_{w=1}^{V} \Gamma(n^{w}_{z,y=1,\neg_c} + \beta)} \frac{\Gamma(\sum_{w=1}^{V} n^{w}_{z,y=1,\neg_c} + V\beta)}{\Gamma(\sum_{w=1}^{V} n^{w}_{z,y=1} + V\beta)} \\ = & \frac{\prod_{w=1}^{V} \prod_{p=1}^{n^{w}_{c}} (n^{w}_{z,y=1} + \beta - p)}{\prod_{q=1}^{n^{\mathbf{w}}_{c}} (\sum_{j=1}^{V} n^{w}_{z,y=1} + V\beta - q)},\end{aligned}$$

where $n^{\mathbf{w}}_{c}$ and n^{w}_{c} denotes number of words and number of times word w occurs in user u's n-th tweets respectively.

For the rest two components, we can derive them similarly:

$$(2.8)\quad \frac{B(n^{\mathbf{b}}_{z} + \alpha)}{B(n^{\mathbf{b}}_{z,\neg_c} + \alpha)} = \frac{n^{b_c}_{z,\neg_c} + \alpha}{\sum_{b=1}^{B} n^{b}_{z,\neg_c} + B\alpha},$$

$$(2.9)\quad \frac{B(n^{\mathbf{t}}_{u} + \eta)}{B(n^{\mathbf{t}}_{u,\neg_c} + \eta)} = \frac{n^{z}_{u} + \eta - 1}{\sum_{t=1}^{T} n^{t}_{u} + T\eta - 1},$$

where n^{b}_{z} denotes number of times topic z co-occurs with behavior b, n^{z}_{u} denotes number of times topic z is sampled in user u's tweets.

Given Eqn. 2.6, 2.7, 2.8 and 2.9, we can then compute Eqn. 2.1.

$$\begin{aligned}(2.10)\quad & p(z_c = z|\mathbf{Z}_{\neg_c}, \mathbf{W}, \mathbf{Y}, \mathbf{B}) \\ = & \frac{\prod_{w=1}^{V} \prod_{p=1}^{n^{w}_{c}} (n^{w}_{z,y=1} + \beta - p)}{\prod_{q=1}^{n^{W}_{c}} (\sum_{j=1}^{V} n^{w}_{z,y=1} + V\beta - q)} \frac{n^{b_c}_{z,\neg_c} + \alpha}{\sum_{b=1}^{B} n^{b}_{z,\neg_c} + B\alpha} \\ & \times \frac{n^{z}_{u} + \eta - 1}{\sum_{t=1}^{T} n^{t}_{u} + T\eta - 1}.\end{aligned}$$

2.3 Sampling label $y_{u,n,l}$

We discuss how to derive Eqn. 2.2 to update $y_{u,n,l}$ for each word in the tweet in this section.

Let d be $\{u, n, l\}$, similar to Eqn. 2.4, we have the following equation.

$$\begin{aligned}& p(y_d|\mathbf{Y}_{\neg_d}, \mathbf{Z}, \mathbf{W}, \mathbf{B}) \\ \propto & \frac{B(n^{\mathbf{y}}_{(.)} + \gamma)}{B(n^{y}_{\neg_d} + \gamma)} \frac{B(n^{\mathbf{w}}_{y=0} + \beta')}{B(n^{\mathbf{w}}_{y=0,\neg_d} + \beta')} \prod_{t=1}^{T} \frac{B(n^{\mathbf{w}}_{t,y=1} + \beta)}{B(n^{\mathbf{w}}_{t,y=1,\neg_d} + \beta)} \\ & \times \prod_{t=1}^{T} \frac{B(n^{\mathbf{b}}_{t} + \alpha)}{B(n^{\mathbf{b}}_{t,\neg_c} + \alpha)} \prod_{u=1}^{U} \frac{B(n^{\mathbf{t}}_{u} + \eta)}{B(n^{\mathbf{t}}_{u,\neg_d} + \eta)}.\end{aligned}$$

When we sample y_d, we assume $\mathbf{Y}_{\neg_d}$, $\mathbf{Z}$, $\mathbf{W}$ and $\mathbf{B}$ are fixed, similar to Eqn. 2.6, we have:

$$\begin{aligned}& p(y_d|\mathbf{Y}_{\neg_d}, \mathbf{Z}, \mathbf{W}, \mathbf{B}) \\ \propto & \frac{B(n^{\mathbf{y}}_{(.)} + \gamma)}{B(n^{\mathbf{y}}_{\neg_d} + \gamma)} \frac{B(n^{\mathbf{w}}_{y=0} + \beta')}{B(n^{\mathbf{w}}_{y=0,\neg_d} + \beta')} \prod_{t=1}^{T} \frac{B(n^{\mathbf{w}}_{t,y=1} + \beta)}{B(n^{\mathbf{w}}_{t,y=1,\neg_d} + \beta)},\end{aligned}$$

where to derive each component is similar to Eqn. 2.7, 2.8 and 2.9.

We show the derived results as follows:

$$\begin{aligned}(2.11)\quad & p(y_d = 0|\mathbf{Y}_{\neg_d}, \mathbf{Z}, \mathbf{W}, \mathbf{B}) \\ = & \frac{n^{y_d=0}_{\neg_d} + \gamma}{\sum_{y=0}^{1} n^{y}_{\neg_d} + 2\gamma} \frac{n^{w_d}_{y=0,\neg_d} + \beta'}{\sum_{w=1}^{V} n^{w}_{y=0,\neg_d} + V\beta'}.\end{aligned}$$

$$\begin{aligned}(2.12)\quad & p(y_d = 1|\mathbf{Y}_{\neg_d}, \mathbf{Z}, \mathbf{W}, \mathbf{B}) \\ = & \frac{n^{y_d=1}_{\neg_d} + \gamma}{\sum_{y=0}^{1} n^{y}_{\neg_d} + 2\gamma} \frac{n^{w_d}_{z_c,y=1,\neg_d} + \beta}{\sum_{w=1}^{V} n^{w}_{z_c,y=1,\neg_d} + V\beta}.\end{aligned}$$

With Eqn. 2.10 and Eqn. 2.11, we can perform Gibbs Sampling for B-LDA as in Algorithm 1.

2.4 Parameter estimation

With Gibbs sampling, we can make the following estimation:

$$\phi'_w = \frac{n^w_{(.)} + \beta'}{\sum_{w=1}^{V} n^w_{(.)} + V\beta'}, \tag{2.13}$$

$$\phi_{t,w} = \frac{n^w_t + \beta}{\sum_{w=1}^{V} n^w_t + V\beta}, \tag{2.14}$$

$$\psi_{t,b} = \frac{n^b_t + \alpha}{\sum_{b=1}^{B} n^b_t + B\alpha}, \tag{2.15}$$

$$\varphi_y = \frac{n^y_{(.)} + \gamma}{\sum_{y=0}^{1} n^y_{(.)} + 2\gamma}, \tag{2.16}$$

$$\theta_{u,t} = \frac{n^t_u + \eta}{\sum_{t=1}^{T} n^t_u + T\eta}, \tag{2.17}$$

where $n^w_{(.)}$ is number of times w appears, n^w_t is, when given the topic t, number of times w is sampled, n^b_t is number of time posting behavior b co-occurs with topic t, $n^y_{(.)}$ is number of times y appears, where n^t_u is, when given the user u, number of times t is sampled.

3 Time Complexity

We compare the training time of our proposed B-LDA model against LDA in Table 2 on different number of users.

Model	Number of users			
	1k	2k	5k	All
LDA	1.73	2.73	6.00	10.00
B-LDA	2.20	3.07	6.27	10.61
B-LDA/LDA	1.26	1.12	1.05	1.06

Table 2: Comparison of training time of B-LDA model against LDA on 1k, 2k, 5k and all the users in terms of hour. Note that, 1k, 2k, 5k users are randomly selected from all the users.

Table 2 shows the running time ratio of B-LDA over LDA is from 1.05 $\sim$ 1.26, which means our proposed model B-LDA has a comparable time complexity with LDA.

References

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[2] W. X. Zhao, J. Jiang, J. Weng, J. He, E.-P. Lim, H. Yan, and X. Li, "Comparing twitter and traditional media using topic models," in *ECIR*, 2011, pp. 338–349.