4-2014

On Modeling Community Behaviors and Sentiments in Microblogging

Tuan-Anh Hoang
Singapore Management University

William W. Cohen
Carnegie Mellon University, wcohen@cs.cmu.edu

Ee-Peng Lim
Singapore Management University

Published In
Proceedings of the 2014 SIAM International Conference on Data Mining, 479-487.

On Modeling Community Behaviors and Sentiments in Microblogging

Tuan-Anh Hoang* William W. Cohen† Ee-Peng Lim*

Abstract

In this paper, we propose the CBS topic model, a probabilistic graphical model, to derive the user communities in microblogging networks based on the sentiments they express on their generated content and behaviors they adopt. As a topic model, CBS can uncover hidden topics and derive user topic distribution. In addition, our model associates topic-specific sentiments and behaviors with each user community. Notably, CBS has a general framework that accommodates multiple types of behaviors simultaneously. Our experiments on two Twitter datasets show that the CBS model can effectively mine the representative behaviors and emotional topics for each community. We also demonstrate that CBS model perform as well as other state-of-the-art models in modeling topics, but outperforms the rest in mining user communities.

1 Introduction

Microblogging sites such as Twitter and Weibo have become highly popular media services for social communication. These sites allow users to publish short messages, which are called *tweets*, to exchange information of different topics, and to express their emotional reaction on these topics. Other than tweeting, users on these sites may adopt a wide range of behaviors. For example, a user may follow other users to quickly receive information of her interest, mention some terms in her biography, mention hashtags in her tweets to indicate topic of the tweets, or forward (or *retweet*) tweets of other users. As microblogging have been heavily used for information sharing [13], product broadcasting [11], and political campaigning [7, 8], analyzing the content, network structure, and user behavior in microblogging has therefore attracted a lot of research works in different fields including social science, computer science and marketing science.

Recent empirical works have shown that, in microblogging, there is exists some strong dependencies between a user's community affiliation and the topic and sentiment expressed in her tweets, as well as her microblogging behaviors [10, 6, 19, 9]. Previous research works have attempted to analyze user community affiliations based on one or some subset of the above factors [16, 3, 21]. However, to the best of our knowledge, there is no work that considers all above factors in modeling user communities.

In this work, we postulate that, other than the textual content generated by users, sentiments expressed on topics and other microblogging behaviors of a user can be shaped by her community affiliations. For example, users belonging to a political community may be more interested in retweeting each other, or express positive sentiment on issues they support but negative sentiment on those they oppose. We therefore aim to develop a new model that simultaneously derives the community of each user, and the common behaviors and common topic-specific sentiment of each community. This research task is however challenging due to the following reasons:

- There is a wide range of behaviors users may adopt. For example, a user may follow and retweet other users, at the same time using many hashtags in her tweets. These different behaviors have to be treated differently, but modeled in a consistent way.
- Topic and sentiment of tweets are not known before hand. One either has to first determine the topics and sentiments before using them in modeling user behaviors and communities, or to learn them as part of the model.

This paper addresses the first challenge by developing a general framework that allows different types of behaviors to be modeled as different bag-of-behaviors. We address the second challenge by coupling with an existing sentiment analysis tool for microblogging. Lastly, we develop a probabilistic graphical model that simultaneously infers latent topics, users' topic interests, latent communities and their associated behaviors and topic-specific sentiments. Our main contributions in this work consist of the following.

- We propose a probabilistic graphical model, called CBS, for mining topics and user communities, as well as mining behaviors and topic-specific sentiments associated with the communities.
- We develop a sampling method to infer the model's parameters.
- We apply CBS model on two real politics related Twitter datasets and show that it outperforms other baseline topic models.

*Living Analytics Research Centre, Singapore Management University

†Machine Learning Department, Carnegie Mellon University

- An empirical analysis of behaviors and topic-specific sentiments for the two datasets has been conducted to demonstrate the efficacy of the CBS model.

While CBS model does not explicitly capture the links and interactions among users, it can be easily extended to model linking and interacting behaviors, such as following other users, or mentioning other users in tweets.

The rest of the paper is organized as follows. We discuss the related works on modeling topics, user behaviors, and user communities in Section 2. Our proposed model is presented in Section 3. We describe two experimental datasets in Section 4. The experimental evaluation of the model on the two datasets is reported in Section 5 and Section 6 respectively. Finally, we give our conclusions and discuss future work in Section 7.

2 Related Works

Analyzing topics together with network structures in microblogging has been widely studied. However, most of the existing works are based on the assumptions that: (1) users/documents having the similar topic distributions are more likely to connect with one another, e.g., [15, 4]; or (2) users/documents within a community have similar topic distributions, e.g., [1, 21]. Our model, on the other hand, does not assume similarity of topic between users within a community, but assume similarity of their behaviors. Moreover, different from the existing works that only consider network among the users and their associated text, our model also takes into account sentiments expressed in the text.

Mrinmaya *et. al.* proposed to use communicating behaviors in modeling topics and communities in communication networks [18]. Similarly, Qiu *et. al.* proposed to jointly modeling topics of tweets and their associated behavioral patterns [17]. However, these works only considers only one type of behavior that is associated with text, while our model allows different type of user behaviors to be modeled simultaneously.

Lastly, there are works that use some supervised approach to determine user communities based on a vast variety of features including text and behaviors, e.g., [16, 3]. Despite of high performance reported in these works, Cohen *et. al.* recently showed that learning to classify users in microblogging is not transferable due to the diversity in users' tweeting topics and tweeting behaviors [5]. In contrast, our model can be used as both unsupervised or semi-supervised learner.

3 The Proposed Model

In this section, we present our proposed model in detail. We first introduce notations used in our paper. Next, we describe the model and the sampling method for learning the model's parameters.

3.1 Notations. Consider a set of Twitter users together with their posted tweets and behavior traces. We use U and L to denote the number of users and the number of behavior types in the dataset respectively. For each user u_i, we denote the set of M_i tweets she posts by $\mathcal{T}_i = \{t_1^i, \cdots, t_{M_i}^i\}$; and denote the set of all the tweets in the dataset by $\mathcal{T}$, i.e., $\mathcal{T} = \bigcup_i \mathcal{T}_i$. Each tweet t_j^i is a bag-of-words with length N_{ij}, i.e., $t_j^i = \{w_1^{ij}, \cdots, w_{N_{ij}}^{ij}\}$, where each word w_n^{ij} is drawn from a common vocabulary of W words $\mathcal{V} = \{w_1, \cdots, w_W\}$. Also, for each tweet t_j^i, we denote its topic and sentiment by z_j^i and s_j^i respectively. The bag-of-topics and the bag-of-sentiments of all the tweets is denoted by $\mathcal{Z}$ and $\mathcal{S}$ respectively. Similarly, for each user u_i, and each behavior type l, we use $\mathcal{B}_i^l$ to denote the length-B_{il} bag-of-behaviors of type l that u_i adopts, i.e., $\mathcal{B}_i^l = \{b_1^{il}, \cdots, b_{B_{il}}^{il}\}$, and use $\mathcal{B}$ to denote the bag-of-all-behaviors (of all types) of all the users.

3.2 CBS Model. The basic assumption of our model is that while users within a community may have different topical interest in tweeting, they should adopt similar behaviors. We therefore assume that, for *each type of behaviors*, each community has a certain interest in some behaviors of the type, and all the users within the community adopt the behaviors following this interest. For example, a Christian often mentions religion in her biography, or a football fan often follows and retweets from her supporting team's pages. Moreover, different communities may express different sentiments on the same topic, e.g., Democrats are more positive about healthcare issues while Republicans are more negative. Hence, behaviors a user adopted and sentiment she expressed in her tweets are useful in identifying the community that she belongs to.

The CBS model has K latent topics, where each topic k has a multinomial distribution ϕ_k over the vocabulary $\mathcal{V}$. As tweets are short with no more than 140 characters, we assume that each tweet has only one topic. Each user u belongs to one of C communities, following the (global) community distribution π. Each user u has a topic distribution θ_u, while each community c has a topic-specific sentiment distribution σ_{ck} for each topic k. Moreover, for each behavior type l, each community c has a multinomial distribution λ_{cl} over the set of all type-l behaviors. Lastly, we assume that π, θ_u, σ_{cz}, and λ_{cl} have Dirichlet priors τ, α, η, and γ_l respectively.

In summary, the CBS model has the plate notation as shown in Figure 1 and the generative process as follows.

- Sample the community distribution vector $\pi \sim Dirichlet(\tau)$
- For each $k = 1, \cdots, K$, sample the k-th topic $\phi_k \sim$

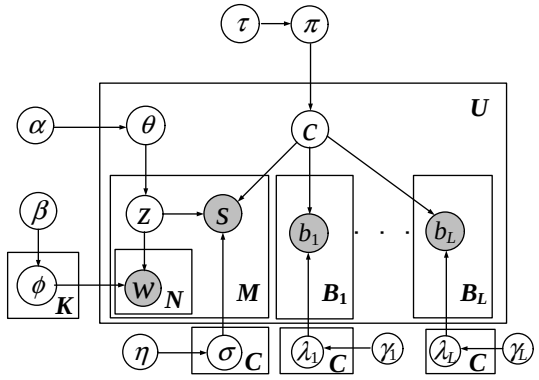


Figure 1: Plate notation for CBS model

$Dirichlet(\beta_k)$

- For each community c and each topic k, sample the topic-specific sentiment distribution $\sigma_{ck} \sim Dirichlet(\eta_{ck})$
- For each community c, and each type of behavior l, sample type-l behavior distribution $\lambda_{cl} \sim Dirichlet(\gamma_{cl})$
- For each user u, sample community indicator $c_u \sim Multinomial(\pi)$
- For user u, generate tweets for the user:
  1. Sample topic distribution $\theta_u \sim Dirichlet(\alpha)$
  2. For each tweet t:
     (a) Sample topic for the tweet $z_t \sim Multinomial(\theta_u)$
     (b) Sample tweet's words: for each word slot n, sample the word $w_{t,n} \sim Multinomial(\phi_{z_t})$
     (c) Sample tweet's sentiment: sample the sentiment $s_t \sim Multinomial(\sigma_{cz_t})$
- Generate behaviors for each user u suppose u is assigned community label c
  1. For each behavior of type l, sample the behavior $b \sim Multinomial(\lambda_{cl})$

Note that in CBS model, we currently determine the sentiments of tweets using Stanford's sentiment scoring API[1,2]. The widely used Stanford's sentiment scoring API implements a machine learning method to detect sentiment expressed in a tweet purely based on content of the tweet. For each tweet, the API returns a score of 4, 0, or 2 to indicate the tweet is positive, negative, or neutral respectively.

[1]http://help.sentiment140.com/api

[2]This also reduces the complexity of CBS model as sentiment mining itself is already well studied research problem.

3.3 Learning. Due to the intractability of LDA-based model [2], we make use of sampling method in learning and estimating the parameters in the model. More exactly, we use a collapsed Gibbs sampler to iteratively sample the latent community of every user, and latent topic of every tweet.

Assume that the current user we have to sample the community for is u_i. We use $\mathcal{C}_{-i}$ to denote the bag-of-communities of all other users in the dataset except u_i. Similarly, for each tweet t_j^i (of user u_i), we use $\mathcal{Z}_{-z_j^i}$ and $\mathcal{S}_{-s_j^i}$ to denote the bag-of-topics and bag-of-sentiments, respectively, of all other tweets in the dataset except t_i^j. Finally, for each behavior b_n^{il}, we use $\mathcal{B}_{-b_n^{il}}$ to denote the bag-of-behaviors excluding b_n^{il}. Then, the community of u_i is sampled according to Equation 3.1.

Now, we have to sample the topic for the current tweet denoted by t_j^i. Let $\mathcal{T}_{-t_j^i}$ denotes the set of all tweets in the dataset excluding t_j^i. Then topic of t_j^i is sampled according to Equation 3.2.

In Equations 3.1 and 3.2, $\mathbf{n_s}(s, z, c, \mathcal{S}, \mathcal{Z})$ records the number of times the sentiment s observed in the topic z in the set of tweets posted by users of community c for bag-of-sentiments $\mathcal{S}$ and bag-of-topics $\mathcal{Z}$. Similarly, $\mathbf{n_b}(b, c, \mathcal{B}, \mathcal{C})$ records the number of times the behavior b is adopted by users of community c for the bag-of-behaviors $\mathcal{B}$ and the bag-of-communities $\mathcal{C}$; $\mathbf{n_w}(w, z, \mathcal{T}, \mathcal{Z})$ records the number of times the word w is observed in the topic z for the set of tweets $\mathcal{T}$ and the bag-of-topics $\mathcal{Z}$; and $\mathbf{n_z}(z, i, \mathcal{Z})$ records the number of times the topic z is observed in the set of tweets posted by user u_i for the bag-of-topics $\mathcal{Z}$.

In our experiments, we used symmetric Dirichlet hyperparameters with $\alpha = 50/K$, $\beta = 0.01$, $\tau = 5$, $\eta = 5$, and $\gamma_l = 0.01$ for all $l = 1, \cdots, L$. Each time, we run the model for 300 iterations of Gibbs sampling. We take 20 samples with a gap of 5 iterations in the last 100 iterations to assign values to all the hidden variables.

4 Datasets

In order to get clear notions of communities and topics, the following two politically oriented datasets were used for evaluating the CBS model.

MoC Dataset. The first dataset consists of tweets posted by members of the 112th U.S congress. We manually identified the official Twitter accounts of 93 senators (47 Democrats and 46 Republicans) and collected their tweets in the duration of May 2012 - Feb 2013. In other words, we have the ground truth political affiliations of all users in this dataset.

One-Week Dataset. The second dataset is large set of tweets generated just before the 2012 US presidential election. We first manually selected 56 *seed users* who are popular political related figures with many followers on Twitter. These include major American politicians,

$$p(c_i = c|\mathcal{T}, \mathcal{S}, \mathcal{B}, \mathcal{C}_{-i}, \mathcal{Z}, \alpha, \beta, \tau, \eta, \lambda) \propto$$

$$\propto \prod_{j=1}^{M_i} \frac{\mathbf{n_s}(s_j^i, z_j^i, c, \mathcal{S}_{-s_j^i}, \mathcal{Z}_{-z_j^i}) + \eta_{cz_j^i s_j^i}}{\sum_{q=1}^{C}(\mathbf{n_s}(s_j^i, z_j^i, q, \mathcal{S}_{-s_j^i}, \mathcal{Z}_{-z_j^i}) + \eta_{qz_j^i s_j^i})} \cdot \prod_{l=1}^{L}\prod_{n=1}^{B_{il}} \frac{\mathbf{n_b}(b_n^{il}, c, \mathcal{B}_{-b_n^{il}}, \mathcal{C}_{-c_i}) + \lambda^l_{cb_n^{il}}}{\sum_{v=1}^{W_l^n} \mathbf{n_b}(v, c, \mathcal{B}_{-b_n^{il}}, \mathcal{C}_{-c_i}) + \lambda^l_{cv}} \cdot \frac{\mathbf{n_c}(c, \mathcal{C}_{-c_i}) + \tau_c}{\sum_{q=1}^{C}(\mathbf{n_c}(q, \mathcal{C}_{-c_i}) + \tau_q)} \tag{3.1}$$

$$p(z_j^i = z|\mathcal{T}, \mathcal{S}, \mathcal{B}, \mathcal{C}, \mathcal{Z}_{-z_j^i}, \alpha, \beta, \tau, \eta\lambda, \eta) \propto$$

$$\propto \prod_{n=1}^{N_{ij}} \frac{\mathbf{n_w}(w_n^{ij}, z, \mathcal{T}_{-t_j^i}, \mathcal{Z}_{-z_j^i}) + \beta_{zw_n^{ij}}}{\sum_{v=1}^{W}(\mathbf{n_w}(v, z, \mathcal{T}_{-t_j^i}, \mathcal{Z}_{-z_j^i}) + \beta_{zv})} \cdot \frac{\mathbf{n_s}(s_j^i, z, c_i, \mathcal{S}_{-s_j^i}, \mathcal{Z}_{-z_j^i}, \mathcal{C}) + \eta_{c_i z s_j^i}}{\sum_{p=1}^{P}(\mathbf{n_s}(p, z, c_i, \mathcal{S}_{-t_j^i}, \mathcal{Z}_{-z_j^i}, \mathcal{C}) + \eta_{c_i z p})} \cdot \frac{\mathbf{n_z}(z, i, \mathcal{Z}_{-z_j^i}) + \alpha_z}{\sum_{k=1}^{K}(\mathbf{n_z}(k, i, \mathcal{Z}_{-z_j^i}) + \alpha_k)} \tag{3.2}$$

such as 2012 US presidential candidates, e.g., Barack Obama, Mitt Romney, and Newt Gingrich; well known political bloggers in U.S., e.g., America Blog, Red State, and Daily Kos; and political sections of US news media, e.g., CNN Politics, and Huffington Post Politics. The set of users were then expanded by adding all users following at least three seed users. This resulted in 23,992 users whose biographies are collected. Based on their biographies, we were able to manually label the political affiliations of 2,319 of them, including 202 Democrats, 228 Neutrals and 1709 Republicans. The following links of these users were then collected. Since users in this dataset have different degree of political involvement, their tweets cover not only politics but also a variety of other topics. To focus on political topics, we extracted only the political tweets from all tweets posted in the first week of October 2012 using a keyword-based filter. The keywords are political hashtags and political topics' representative words/phrases indentified by the semi-automatic method presented in [9].

Data Preprocessing. We employed the following preprocessing steps to clean both the datasets. We first removed all stopwords from the tweets. Then, for **MoC** dataset, we removed all tweets containing stopwords only and users with less than 5 (remaining) tweets. For **One-Week** dataset, we removed all tweets with less than 3 non-stopwords and and users with less than 10 tweets. In **MoC** dataset, we consider the following behavior types for each user: (1) *user mention*, and (2) *hashtag*; while in **One-Week** dataset, behavior types a user may perform are: (1) *user mention*, and (2) *hashtag*, (3) *retweet*, (4) *followee*, and (5) *profile word* (i.e., non-stopwords in the user's biography). The *hashtag*, *retweet*, and *user mention* behaviors are further divided into positive, neutral, or negative depending on whether the behavior is contained in a positive, neutral, or negative tweet. For each of those behavior, we assign a *_(+)*, *_(0)*, or *_(-)*) suffix to indicate that if the behavior is positive, neutral, or negative respectively. For example, if user u mentions *BarrackObama* in a positive tweet (respectively neutral and negative), then we have *BarakObama_(+)* (respectively *BarakObama_(0)* and *BarakObama_(-)*) in the bag-of-user-mentions of the user. Lastly, for each behavior, for **MoC** we filtered out all the behaviors with less than 5 users performing the behavior, while for **One-Week** dataset we filtered out all the behaviors with less than 50 users performing the behavior.

Table 1: Statistics of the experimental datasets

Dataset			**MoC**	**One-Week**
#user	Total		93	23,992
	With political label	All labels	93	2,193
		Democrat	47	202
		Neutral	0	228
		Republican	46	1,709
#tweets			87,182	839,687
#behaviors	mention		14,609	68,804
	hashtag		26,152	561,098
	retweet		-	181,661
	followee		-	24,044,367
	profile word		-	64,107

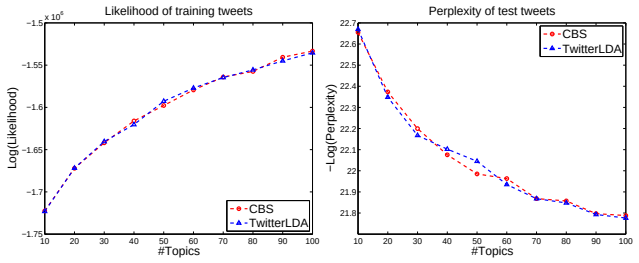
The reasons that, in the proprocessing steps, we used higher thresholds for **One-Week** dataset than for **MoC** dataset are: (1) we expected that the former contains much more noise than the latter, and (2) the former has a much larger number of users than the latter, and we wanted to focus on global behaviors rather than local behaviors. Table 1 shows the statistics of the two datasets after the preprocessing steps.

5 Experiments on MoC dataset

In this experiment, we evaluate the performance of CBS model and other baseline methods in topic modeling and user clustering tasks using **MoC** dataset.

Topic modeling task. Proposed by Zhao *et. al.* [22], TwitterLDA is a variant of LDA [2], a commonly used method for topic modeling. TwitterLDA constrains each tweet to have only one topic. This constraint is appropriate for short documents as well as tweets. We will therefore compare CBS and TwitterLDA based on their abilities to model topics as the number of topics is varied from 10 to 100.

User clustering task. To evaluate the performance of CBS in user clustering, we compare it with K-means clustering. To implement K-means clustering,



(a) (b)

Figure 2: **MoC** dataset: Likelihood and Perplexity of CBS and TwitterLDA

we represent each user as a vector of features, where the features include (1) topic distribution of tweets posted by the user, and (2) bags-of-behaviors of the users. The topic distribution of tweets posted by a user is discovered using TwitterLDA model with the number of topics is set to 70 as will be explained below.

5.1 Metrics. We adopt *likelihood* and *perplexity* for evaluating the topic modeling task. For each user, we randomly selected 90% of tweets of the user to form a training tweet set, and use the remaining 10% of the tweets as the test tweet set. Then for each method, we compute the likelihood of the training tweet set and perplexity of the test tweet set. The method with a higher likelihood, or lower perplexity is considered better for the task.

For user clustering task, we adopt *weighted entropy* as the performance metric. As we have two different political affiliations in the dataset, we run the methods with the number of communities set to 2. We finally computed the weighted entropy of the resultant communities as follows.

$$E = -\sum_{c=0}^{1} \frac{n_c}{U} * \left[\frac{n_c^D}{n_c} * log\frac{n_c^D}{n_c} + \frac{n_c^R}{n_c} * log\frac{n_c^R}{n_c}\right] \quad (5.3)$$

where n_c is the number of users assigned to community c, and n_c^D and n_c^R are the numbers of Democrats and Republicans assigned to community c respectively. Recall that $U = 93$ is the number of users in the dataset. The method with a lower entropy is the winner in the task.

5.2 Performance results. Figure 2 shows the performance of TwitterLDA and CBS model in topic modeling while Figure 3 shows the performance of K-mean and CBS model in user clustering. As expected, larger number of topics K gives larger likelihood and smaller perplexity, and the amount of improvement diminishes as K increases. Considering both time and space complexities, we set the number of topics to be 70 for the user clustering task. Figures 2 shows that CBS and TwitterLDA yield very similar performance in topic modeling. Figure 3, on the other hand, shows that CBS outperforms K-means in user clustering. CBS therefore is a better solution for user clustering than the combination of TwitterLDA and K-means.

Table 3: **MoC** dataset: top positive and negative topics per community

		Topic ID	Topic Label
Democrat	Positive	Topic 16	Greetings
		Topic 29	U.S. teams in Olympic 2012
		Topic 44	Live talks
	Negative	Topic 34	Shooting & terroism
		Topic 32	Legislative issues
		Topic 26	Economics issues
Republican	Positive	Topic 23	Live shows
		Topic 16	Greetings
		Topic 29	US teams in Olympic 2012
	Negative	Topic 34	Shooting & terroism
		Topic 25	Financial issues
		Topic 43	Recovering from Sandy hurrican

5.3 Topic Sentiment Analysis. We now analyze the topic sentiment results of CBS model on **MoC** dataset. For the two learnt communities, we assign each community to be Democrat or Republican if most users in the community are democrat or republican respectively. Table 3 shows the top positive topics and top negative topics of each community as obtained by CBS. Note that the topic labels are manually assigned based on examining the topics' top words and top tweets. For each topic, the topic's top words are the words having the highest likelihoods given the topic, and the topic's top tweets are the tweets having the lowest perplexities given the topic. Table 3 shows that those extreme topics are reasonable. On one hand, the two communities share the common sentiment on topic about broadcasting the talks/shows by senators of the same party (Topic 16, Topic 23), or nationalwide common topics like greetings for vacation and holidays (Topic 16), victories of U.S. team in Olympic 2012 (Topic 29), shooting and terrorism (Topic 34). On the other hand, the two communities are negative on different topics: the Democrat community is negative on topics on legislative issues and economics issues, which mostly under control of Republicans, while the Republican community is negative on the process of recovering from Sandy hurrican (Topic 43) and financial issues, which are mostly raised by Democrats.

5.4 Behavior Analysis. Next, we look into the community representative behaviors uncovered by CBS from the **MoC** dataset. Table 2 shows the top hashtags and top user mentions by users in each community. The table clearly shows that those extreme behaviors are also reasonable. For *hashtag*, all the top hashtags are neutral, and the top ones of each community are most popular hashtags among Twitter users of the community. For *user mention*, the top mentioned users in the Democrat community are democrat users (e.g., Barack-

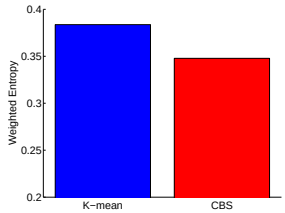


Figure 3: **MoC** dataset: weighted entropy of CBS and K-means

Table 2: **MoC** dataset: top behaviors per community

Hashtag		**User mention**	
Democrat	Republican	Democrat	Republican
#jobs_(0)	#tco_(0)	@speakerboehner_(0)	@wsj_(0)
#nj_(0)	#tcot_(0)	@barackobama_(0)	@foxnews_(+)
#vawa_(0)	#obamacare_(0)	@whitehouse_(0)	@foxnews_(0)
#senate_(0)	#sayfie_(0)	@fema_(0)	@johncornyn_(+)
#sandy_(0)	#fiscalcliff_(0)	@msnbc_(+)	@grahamblog_(0)
#veterans_(0)	#jobs_(0)	@markudall_(0)	@johncornyn_(0)
#budget_(0)	#libya_(0)	@senatorcollins_(0)	@mittromney_(+)
#gop_(0)	#gop_(0)	@nytimes_(0)	@senate_(0)
#job_(0)	#syria_(0)	@senatormenendez_(0)	@senatorayotte_(0)
#socialsecurity_(0)	#debt_(0)	@barackobama_(+)	@joelieberman_(0)

Obama), goverment officers (e.g., speakerboehner), or pro-democrat media (e.g., msnbc), while the top mentioned user in the Republican community are republican senators (e.g., johncornyn), and pro-republican media (e.g., foxnews).

6 Experiments on One-Week dataset

In this section, we report our experiments on **One-Week** dataset. Given the large number of users and tweets, and a partial ground truth of users' political affiliations in the dataset, we evaluate CBS and other comparative methods in topic modeling and user classification tasks.

Topic modeling task. Similar to the experiments presented in Section 5, we compare CBS with TwitterLDA based on their abilities to model topics as the number of topics is varied from 10 to 100.

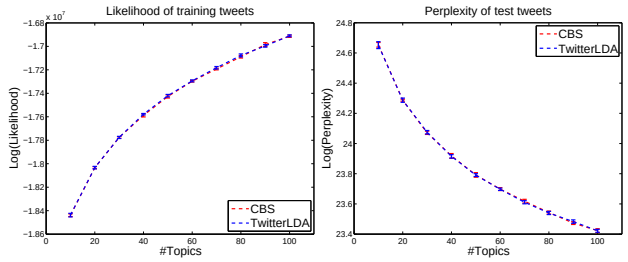
User classification task. We formulate the user classification task as a semi-supervised learning problem since: (1) we have ground truth of political affiliations for only 10% of the users in the dataset, and (2), as shown in [5], the supervised learning approach for users' political affiliation classification in microblogging is not practical given the users having different degree of political involvement like in **One-Week** dataset. To evaluate the performance of CBS in this task, we therefore compare it with semi-supervided learning methods provided in Junto toolbox[3], which are shown to be among state-of-the-art semi-supervised learning methods[20]. The Junto toolbox implements label propagation methods which iteratively update label for each (unknown label) user u based on labels of the other users who are most similar to u. Here, we choose to use the cosine similarity between pairs of users. To do this, we represent each user as a vector of features, where the features are: (a) tweet-based features, and (b) bags-of-behaviors of the users. We employ two ways to compute tweet-based features for each user: (1) Tf-Idf based: the features of each user are TF-IDF scores [14] of the terms contained in the user's tweets; and (2) TwitterLDA based: the features of each user are the components in topic distribution of the user's tweets discovered by TwitterLDA model. For computing the TwitterLDA based features, we set the number of topics in TwittterLDA model to 80 as will be explained below.

[3]https://github.com/parthatalukdar/junto

6.1 Metrics. Again, we adopt *likelihood* and *perplexity* for evaluating the topic modeling task. Similarly to the experiment in Section5, for each user, we randomly selected 90% of tweets of the user to form training tweets set, and use the remaining 10% of the tweets as the test tweets set. Then for each method, we computed the likelihood of the training tweets set and perplexity of the test tweets set. Method with a higher likelihood, or lower perplexity is considered better for the task.

For user classification task, we adopt *average* $F1$ *score* as the performance metric. To do this, we first evenly distributed the set of known political affiliation users in 10 folds such that the folds have the same fraction of Democrat/Neutral/Republican users. Then, for each method, we run 10-fold cross validation with number of communities set to 3 (corresponding to three different political affiliations in the dataset). More precisely, for each method and each time, we use 9 folds of known political affiliation users and all unknown political affiliation users as (semi-)training set, and use the remaining fold of known political affiliation users as test set. For CBS model, in the training phase, we set Democrat, Neutral, and Republican to be community 0, 1, and 2 respectively. We also fix the community indicators of the users in the 9 folds of the (semi-)training set according to their ground truth political affiliation (i.e., we do not sample community for those users). We then compute the average $F1$ score obtained by each method in all three classes (i.e., Democrat, Neutral, and Republican). The method with a higher score is the winner in the task.

6.2 Performance results. Figure 4 shows the performance of TwitterLDA and CBS model in topic modeling. The likelihood and perplexity values in the figure are averaged over 10 runs. Again, as we expected, more topics K gives larger likelihood and smaller perplexity,



(a) (b)

Figure 4: **One-Week** dataset: Likelihood and Perplexity of CBS and TwitterLDA

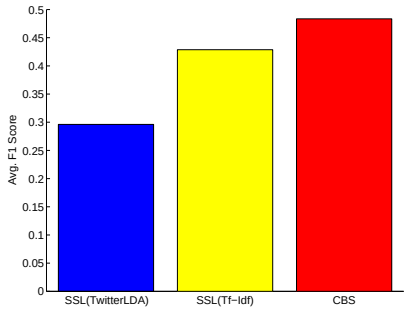


Figure 5: **One-Week** dataset: performance of comparative models in user classification

and the amount of improvement diminishes as K increases. Similar to what reported in Section 5, the figure shows that the topic modeling performance of CBS and TwitterLDA are very similar. This suggests that CBS model is robust against the changes in the bag-of-behaviors used.

Based on Figure 4 results, and in consideration of both time and space complexities, we set the number of topics to 80 for the user classification task. The performance of CBS and the SSL methods in user classification task is shown in Figure 5. The SSL(TwitterLDA) (respectively SSL(Tf-Idf)) is the best performance obtained by methods provided in the Junto toolbox where the users' tweet-based features are TwitterLDA-based features (repectively Tf-Idf based features). The fact that SSL(Tf-Idf) outperforms SSL(TwitterLDA) can be explained that TwitterLDA suffers from noise as, within only one week, many users do not have many tweets for their topic distribution to be inferred correctly by TwitterLDA model. Finally, the figure clearly shows that our CBS model is the best among all the methods.

6.3 Topic Sentiment Analysis. We now analyze the results obtained from applying CBS model on **One-Week** dataset. Table 4 shows the top positive topics and top negative topics of each community as obtained by CBS. Again, we have manually assigned labels for those topics by examining the topics' top words and top tweets. The table shows that those extreme topics are reasonable. In one end, while the two wings,

Table 4: **One-Week** dataset: top positive and most negative topics per community

		Topic	Topic Label
Democrat	Positive	Topic 57	Mr&Mrs Obama's anniversary
		Topic 58	Voting for national building
		Topic 3	Politics as a sport game
	Negative	Topic 26	Conservative issues
		Topic 66	Military issues
		Topic 20	Financial issues
Neutral	Positive	Topic 57	Mr&Mrs Obama's anniversary
		Topic 60	Protecting the country
		Topic 62	Economics changes
	Negative	Topic 20	Financial issues
		Topic 66	Military Policy
		Topic 21	Tax policy
Republican	Positive	Topic 3	Politics as a sport game
		Topic 47	Campaining
		Topic 58	Voting for national building
	Negative	Topic 20	Financial issues
		Topic 21	Tax policy
		Topic 66	Military Policy

i.e., the Democrat and the Republican communities, are positive on the election related topics, e.g., calling for vote for the one building the nation (Topic 58), or tweeting about politics using sport terms (Topic 3), the Neutral is more positive in tweeting about protecting the country (Topic 60), and changes in economics (Topic 62). Also, it is expected that both the Demorat and the Neutral community are positive on Mr&Mrs Obama's anniversary (Topic 57). On the other end, while all three communities are negative on financial issues (Topic 20) and military issues (Topic 66), the Democrat community is more negative on issues raised by the conservatives (Topic 26), but the Neutral and the Republican communities are more negative on the tax policy (Topic 21).

6.4 Behavior Analysis. Table 5 shows the top behaviors performed by users in each community of all five behavior types. The table clearly shows that those extreme behaviors are also reasonable. The top profile words of each community are representative ones for the community: *liberal, progressive, democrats*, etc. for the Democrat community; *conservative, christian, #tcot*, etc. for the Republican community; and *media, sport, music, editor*, etc. for the Neutral community, which including most of accounts of professional persons/associations. For *followee*, it is expected that the top followed users of the Democrat and the Republican communities are most popular ones in each community respectively, while the top ones of the Neutral community are mostly goverment office (e.g. WhiteHouse) and media (e.g., nytimes, BreakingNews, and AP). Similarly, for *retweet*, the top retweeted users of Democrat and Republican communities are most popular ones in each community respectively, while the top ones of Neutral community are mostly media. The top hashtags suggest that the two wings (i.e., the Democrat

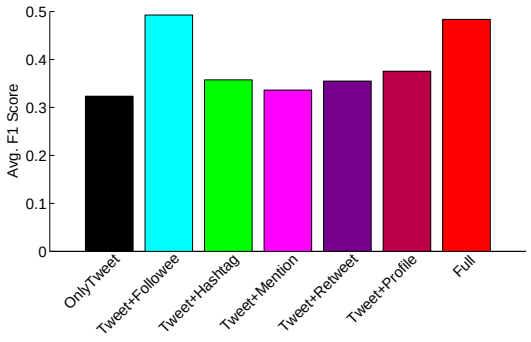


Figure 6: **One-Week** dataset: Performance of variants of CBS in user classification

and the Republican communities) tweet most about topics within their own community (e.g.,#p2 for the Democrat community, and #tcot for the Republican community) and then about the opposite one, while the Neutral community tweets more about topics related to international issues (e.g., #syria, #iran). For *user mention*, it is interesting that while the Neutral community mentions the two candidate equally, users of the two wings mention the opposite candidate more. This due to the fact that, during the campaign period, the wing users often mention the opposite candidate in their tweets for questioning about facts or issues that they do not support.

6.5 Usefulness of behavior types. Lastly, we examine the usefulness of the different behavior types in user classification task. To do this, we perform the same experiments on **One-Week** dataset using the following variants of CBS model

- **OnlyTweet**: the variant in which we do not take any behavior (of any type) into account, i.e., only tweets and sentiments are modeled.
- **Tweet+Followee**: the variant in which we only consider tweets, sentiments, and behaviors of *Followee* type. Similarly we have **Tweet+Hashtag**, **Tweet+Mention**, **Tweet+Retweet**, and **Tweet+Profile** variants.
- **Full**: the CBS model presented as above where all (5) types of behaviors are taken into account.

Figure 6 shows the performance of the different variants of CBS in user classification task. The figure suggests that adding behaviors improves the performance, and *Followee* is more useful than other behaviors. We further conducted McNemar's test [12] and showed that: (1) the behaviors are helpful in user classification as all the variants with behaviors added have performance that is statistically significantly higher than performance of **OnlyTweet** variant; and (2) among the behaviors, following behavior is the most useful as **Tweet+Followee** and **Full** have statistically significant higher performance than the other variants' performance. The test also showed that the difference between the **Tweet+Followee** and **Full** variants is not statistically significant.

7 Conclusion

In this paper, we propose a novel framework to model user communities in microblogging based on sentiments the users expressed on different topics, and behaviors they performed. Our framework allows different type behaviors can be modeled simultaneously. Our experiments on two real Twitter datasets show that the proposed model outperforms baseline methods.

Generally, for each community, our model does not capture behaviors expressed towards members within the community. The kind of behaviors we model here are of general nature and can be expressed towards users, items, or some groups of users/items. In the future, we would like to study more fine-grained factors having effects on user behaviors. These factors include interest of the user herself, the interest of communities the user belongs to, and interactions with other users in the network.

8 Acknowledgments

This research is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office, Media Development Authority (MDA).

References

[1] R. Balasubramanyan and W. W. Cohen. Block-lda: Jointly modeling entity-annotated text and entity-entity links. In *SDM*, 2011.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, Mar. 2003.

[3] A. Boutet, H. Kim, and E. Yoneki. What's in your tweets? i know who you supported in the uk 2010 general election. In *ICWSM*, 2012.

[4] J. Chang, J. Boyd-Graber, and D. M. Blei. Connections between the lines: augmenting social networks with text. In *KDD*, 2009.

[5] R. Cohen and D. Ruths. Classifying political orientation on twitter: It's not easy! In *ICWSM*, 2013.

[6] A. Feller, M. Kuhnert, T. O. Sprenger, and I. M. Welpe. Divided they tweet: The network structure of political microbloggers and discussion topics. In *5th ICWSM*, 2011.

[7] J. Golbeck, J. M. Grimes, and A. Rogers. Twitter use by the u.s. congress. *J. Am. Soc. Inf. Sci. Technol.*, 2010.

[8] J. A. Hendricks and R. E. D. Jr. *Communicator-In-Chief: How Barack Obama Used New Media Technology to Win the White House*. Lexington Books, 2010.

[9] T.-A. Hoang, W. W. Cohen, E.-P. Lim, D. Pierce,

Table 5: **One-Week** dataset: top behaviors per community

Profile			**Followee**		
Democrat	Neutral	Repulican	Democrat	Neutral	Repulican
liberal	politics	conservative	BarackObama	BarackObama	michellemalkin
love	media	love	maddow	WhiteHouse	PAC43
politics	love	christian	thinkprogress	politico	KatyinIndy
progressive	sports	god	WhiteHouse	nytimes	BraveLad
lover	music	american	MotherJones	mittromney	Heritage
obama	world	country	TheDailyEdge	BreakingNews	Miller51550
democrat	editor	#tcot	DavidCornDC	WSJ	SarahPalinUSA
mom	student	wife	billmaher	cnnbrk	AndyWendt
music	tweets	family	dccc	AP	seanhannity
rights	life	party	TheNewDeal	washingtonpost	marcorubio

Retweet		
Democrat	Neutral	Repulican
thedailyedge_(0)	thinkprogress_(0)	patdollard_(0)
barackobama_(0)	reuters_(0)	jjauthor_(0)
thinkprogress_(0)	barackobama_(0)	newsninja2012_(0)
lolgop_(0)	ap_(0)	mittromney_(0)
thenewdeal_(0)	drudge_report_(0)	slone_(0)
truthteam2012_(0)	thedailyedge_(0)	katyinindy_(0)
jeffersonobama_(0)	truthteam2012_(0)	connewsnow_(0)
chrisrockoz_(0)	huffpostpol_(0)	iowahawkblog_(0)
bluedupage_(0)	patdollard_(0)	keder_(0)
utaustinliberal_(0)	buzzfeedandrew_(0)	twitchyteam_(0)

Hashtag			**User mention**		
Democrat	Neutral	Repulican	Democrat	Neutral	Repulican
#p2_(0)	#tcot_(0)	#tcot_(0)	@mittromney_(0)	@barackobama_(0)	@barackobama_(0)
#romney_(0)	#obama_(0)	#obama_(0)	@cspanwj_(0)	@mittromney_(0)	@mittromney_(0)
#gop_(0)	#syria_(0)	#teaparty_(0)	@mittromney_(+)	@mittromney_(+)	@mittromney_(+)
#tcot_(0)	#p2_(0)	#p2_(0)	@barackobama_(0)	@barackobama_(+)	@barackobama_(+)
#obama_(0)	#iran_(0)	#tlot_(0)	@barackobama_(+)	@cnn_(0)	@youtub_(0)
#tco_(0)	#romney_(0)	#gop_(0)	@cspanwj_(+)	@barackobama_(-)	@breitbartnew_(0)
#obama2012_(0)	#gop_(0)	#tco_(0)	@thinkprogres_(0)	@mittromney_(-)	@sharethi_(0)
#p_(0)	#news_(0)	#romney_(0)	@edshow_(0)	@abc_(0)	@cnn_(0)
#p2b_(0)	#tco_(0)	#lnyhbt_(0)	@maddow_(0)	@paulryanvp_(0)	@seanhannity_(0)
#teaparty_(0)	#ndaa_(0)	#romneyryan2012_(0)	@ffhelper_(+)	@thedemocrats_(0)	@truthteam2012_(0)

and D. P. Redlawsk. Politics, sharing and emotion in microblogs. In *ASONAM*, 2013.

[10] R. Jacob, C. Michael, M. Mark, G. Bruno, P. Snehal, F. Alessandro, and M. Filippo. Truthy: mapping the spread of astroturf in microblog streams. In *WWW '11*, 2011.

[11] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11), 2009.

[12] N. Japkowicz and M. Shah. *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge University Press, 2011.

[13] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *WWW'10*, 2010.

[14] C. D. Manning, P. Raghavan, and H. Schutze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[15] R. M. Nallapati, A. Ahmed, E. P. Xing, and W. W. Cohen. Joint latent topic models for text and citations. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD, 2008.

[16] M. Pennacchiotti and A.-M. Popescu. Democrats, republicans and starbucks afficionados: user classification in twitter. In *KDD*, 2011.

[17] M. Qiu, J. Jiang, and F. Zhu. It is not just what we say, but how we say them: Lda-based behavior-topic model. In *SDM*, 2013.

[18] M. Sachan, D. Contractor, T. A. Faruquie, and L. V. Subramaniam. Using content and interactions for discovering communities in social networks. In *WWW*, 2012.

[19] S. Stieglitz and L. Dang-Xuan. Political communication and influence through microblogging–an empirical analysis of sentiment in twitter messages and retweet behavior. In *45th HICSS*, 2012.

[20] P. P. Talukdar and F. Pereira. Experiments in graph-based semi-supervised learning methods for class-instance acquisition. ACL, 2010.

[21] Z. Yin, L. Cao, Q. Gu, and J. Han. Latent community topic analysis: Integration of community discovery with topic modeling. *ACM Trans. Intell. Syst. Technol.*, 2012.

[22] 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.