Interaction-Aware Topic Model for Microblog Conversations through Network Embedding and User Attention

Ruifang He[1,2,*], **Xuefei Zhang**[1,2,*], **Di Jin**[1,2,†], **Longbiao Wang**[1,2], **Jianwu Dang**[1,2,3], **Xiangang Li**[4]
[1]School of Computer Science and Technology, Tianjin University, Tianjin, China
[2]Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin, China
[3]Japan Advanced Institute of Science and Technology, Ishikawa, Japan
[4]AI Labs, Didi Chuxing, Beijing, China
{rfhe,jindi,longbiao_wang}@tju.edu.cn,zhangxuefei29@163.com
jdang@jaist.ac.jp,lixiangang@didichuxing.com

Abstract

Traditional topic models are insufficient for topic extraction in social media. The existing methods only consider text information or simultaneously model the posts and the static characteristics of social media. They ignore that one discusses diverse topics when dynamically interacting with different people. Moreover, people who talk about the same topic have different effects on the topic. In this paper, we propose an Interaction-Aware Topic Model (IATM) for microblog conversations by integrating network embedding and user attention. A conversation network linking users based on reposting and replying relationship is constructed to mine the dynamic user behaviours. We model dynamic interactions and user attention so as to learn interaction-aware edge embeddings with social context. Then they are incorporated into neural variational inference for generating the more consistent topics. The experiments on three real-world datasets show that our proposed model is effective.

1 Introduction

The prosperity of microblog platforms, such as Twitter[1] and Sina Weibo[2] brings the large scale, noisy and user-generated short posts. Automatically extracting topics in social media aims to reveal thematic information of the underlying collection, which can be used in summarization (Zhuang et al., 2016), hashtag recommendation (Li et al., 2016b), response generation (Xing et al., 2017) and so on.

The conventional topic models, like Latent Dirichlet Allocation (LDA) (Blei et al., 2003), infer the hidden topics based on word co-occurrences in documents. They could not be directly transferred in social media due to the data sparsity and the noise of short texts.

The existing relevant researches can be roughly categorized into: (1) Methods exploit the content of short texts by aggregation strategy (Mehrotra et al., 2013) or modeling word-pair co-occurrence (Yan et al., 2013). (2) Models incorporating with word embedding try to deeply understand the posts by representation learning (Sridhar, 2015; Hu and Tsujii, 2016), yet they ignore the social context of microblog messages. Essentially, social media content and network structures influence each other (Bi et al., 2014), the only content analysis is insufficient. (3) (Li et al., 2016a; Chen and Ren, 2017) take into account the content and static characteristics of network structures to deduce topics. However, they ignore dynamic user behaviours.

Actually, a user may talk about various topics when interacting with diverse neighbours in a social network. For instance, Fig. 1(a) shows two conversation trees, where [U0] communicates with [U1] and [U2] about the topic of United Kingdom European Union membership referendum, while [U0] talks a game with [U3]. Hence, we need to infer topics according to the interactions between users. Moreover, although [U0] and [U1] argue about the same topic, [U0] contains salient content in topic description, e.g., *UK, EU, referendum* while the reply of [U1] is nothing but simply response to [U0], e.g., *Yes, agree*.

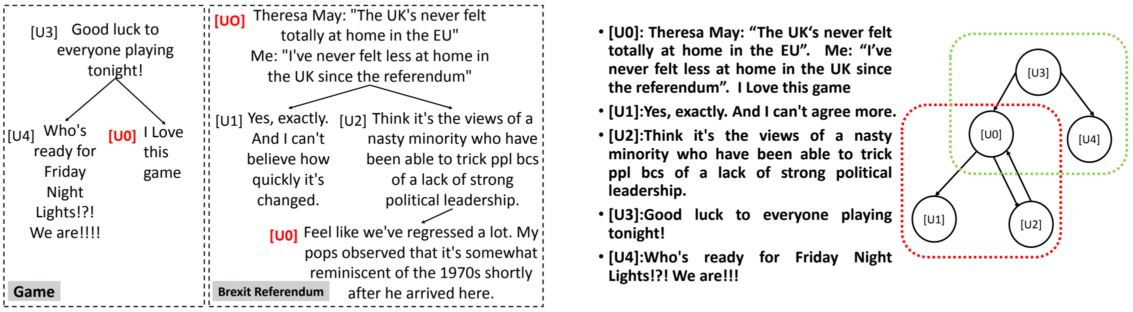
* Indicates equal contribution.

†Corresponding author.

[1] https://twitter.com

[2] https://weibo.com

Proceedings of the 27th International Conference on Computational Linguistics, pages 1398–1409
Santa Fe, New Mexico, USA, August 20-26, 2018.


[U3] Good luck to everyone playing tonight!
[U4] Who's ready for Friday Night Lights!?! We are!!!!
[U0] I Love this game
Game
[UO] Theresa May: "The UK's never felt totally at home in the EU" Me: "I've never felt less at home in the UK since the referendum"
[U1] Yes, exactly. And I can't believe how quickly it's changed.
[U2] Think it's the views of a nasty minority who have been able to trick ppl bcs of a lack of strong political leadership.
[U0] Feel like we've regressed a lot. My pops observed that it's somewhat reminiscent of the 1970s shortly after he arrived here.
Brexit Referendum
• [U0]: Theresa May: "The UK's never felt totally at home in the EU". Me: "I've never felt less at home in the UK since the referendum". I Love this game
• [U1]:Yes, exactly. And I can't agree more.
• [U2]:Think it's the views of a nasty minority who have been able to trick ppl bcs of a lack of strong political leadership.
• [U3]:Good luck to everyone playing tonight!
• [U4]:Who's ready for Friday Night Lights!?! We are!!!
[U3]
[U0]
[U4]
[U1]
[U2]

(a) Two conversation trees. [Ui]: the i-th user; Arrow lines: reposting/replying relations.

(b) An example of a conversation network. The green box shows the topic about a game; The red one exhibits the topic of Brexit referendum.

Figure 1: A conversation network is transformed from two conversation trees.

Thus, it is intuitive that during the topic extraction of United Kingdom European Union membership referendum, [U0] is more important than [U1]. The **dynamic interactions** and **diverse user attention** on the topic provide us heuristic insights for topic extraction.

Note that conversation trees can be transformed into conversation networks through user relations, as illustrated in the Fig. 1. Inspired by network embedding and social user behaviour researches, we propose an **I**nteraction **A**ware **T**opic **M**odel (IATM) integrating dynamic interactions and various effects of users on the topic. The contributions are as follows.

- We propose a novel Interaction-Aware Topic Model (IATM) for microblog conversations, which simultaneously considers social media content and network topology.

- Our method models dynamic interactions and various effects of users, and encodes them as interaction-aware edge embeddings through network embedding and user attention.

- The deeper edge representations are employed in neural variational inference for generating the more consistent topics.

- The experiments verify that the proposed IATM is effective. Dynamic user behaviours are useful for topic extraction of short texts in social media.

2 Related Work

Previous researches for topic extraction in social media can be mainly classified into two aspects.

2.1 Focusing on Content Information

They depend on pure content to generate document-topic distribution and topic-word distribution. (1) **Aggregation Strategy Based.** To solve the data sparsity short texts, some use different strategies to heuristically aggregate microblog messages based on authorship (Hong and Davison, 2010; Zhao et al., 2011) or hashtags (Mehrotra et al., 2013; Tang et al., 2013) before applying traditional topic model. Self-Aggregation based Topic Model (SATM) (Quan et al., 2015) combines short texts aggregation and topic induction into a unified model. Others use Biterm Topic Model (BTM) (Yan et al., 2013) and RNN-IDF based Biterm Short-text Topic Model (RIBSTM) (Lu et al., 2017) modeling biterm co-occurrence in the whole corpus to enhance topic discovery. (2) **Word Embedding Based.** Since word embeddings have been shown their ability to form clusters of conceptually similar words in the embedding space (Mikolov et al., 2013c), Gaussian Mixture Topic Model (GMTM) (Sridhar, 2015) and Latent Concept Topic Model (LCTM) (Hu and Tsujii, 2016) utilize word embeddings to improve topic generation. However, since social media content and network structures influence with each other, only focusing on content is insufficient.

2.2 Integrating Content and Network Structure information

This kind of researches considers social content and network structures together, including followship and repostship. (1) **Followship Based**. Lim et al.(2013) jointly model the text and user-follower network during topic inference. Yet the appearance of *Zombie Fans* purchased by those who would like to increase their influence, disturbs the followship network. (2) **Repostship Based**. LeadLDA (Li et al., 2016a) generates words according to topic dependencies derived from conversation trees by differentiating messages from leader and follower. Leader is the post that contains salient and new information in topic description while follower is the post that simply responds to its reposted or replied messages. Chen and Ren (2017) propose ForumLDA to infer topics by distinguishing response messages from relevant and irrelevant to the original post.

However, due to the zombie fans in followship network, or exploring the static characteristics in repostship network, these methods ignore the dynamic user behaviours latent in reciprocal interactions between users. Therefore we emphasize on exploring whether user interactions and user attention can help topic generation in social media.

2.3 Network Embedding and Neural Variational Inference

The further researches (Perozzi et al., 2014; Tu et al., 2017) about network embedding make it possible to suitably model the dynamic user behaviours. Meanwhile, comparing with topic models based on word embedding, Miao et al. (2017) and Srivastava and Sutton (2017) apply topic model with neural variational inference (Kingma and Welling, 2014) on traditional long documents, which consistently produces the better topics. A small change in the document will produce a small change in topics. Yet they have not taken into account sparse and noisy texts and network structures in social media.

Therefore, we model dynamic user behaviours and content information as interaction-aware edge embeddings, which combine network embedding and user attention. Further, they are incorporated into variational auto-encoder for topic extraction in social media, which can produce the more consistent and precise topics.

3 Interaction-Aware Topic Model

The large scale short and noisy texts in social media bring the more serious data sparseness and inconsistency in topics. Yet the dynamic user behaviours also provide us opportunities. Here, we propose two hypotheses and explore whether they are useful for topic extraction.

H1: Dynamic Interactions. One discusses various topics with different people.

H2: User Attention. Users who debate the same topic have diverse effects on the topic.

We need to construct a conversation network, shown in Fig. 1(b), transformed from conversation trees.

3.1 Conversation Network

Conversation networks are transformed from conversation trees through user relations, which are exhibited in Fig. 1(b). We first give the basic notations and definitions of a conversation network. Given $G = (V, E, T)$, where V is the set of vertices, $E \subseteq V \times V$ is the edges between vertices and T denotes text information of vertices. Each vertex $v \in V$ represents a user and every edge $e_{(u,v)} \in E$ stands for the reposting or replying relationship between two vertices (u, v). To solve the sparsity of short texts, we aggregate all the messages posted by the same user, including the original and the reposting messages[3]. Text information of the unique vertex $v \in V$ is the messages $M_v = (w_1, w_2, ..., w_n)$, where n is the count of words in M_v.

In order to illustrate the proposed IATM model, we firstly give two definitions:

D1: Edge Embedding. The edge embedding $\mathbf{e}$ for every edge $e_{(u,v)}$ is acquired by concatenating the vertex embeddings $\mathbf{u}$ and $\mathbf{v}$. It encodes user representations with respect to various neighbours.

D2: Interaction-aware Edge Embedding. It models the various effects of u and v on a topic via user attention mechanism, and makes edge embedding as an interaction-aware style.

[3] Both of replying and reposting messages are called reposting messages.

3.2 Model

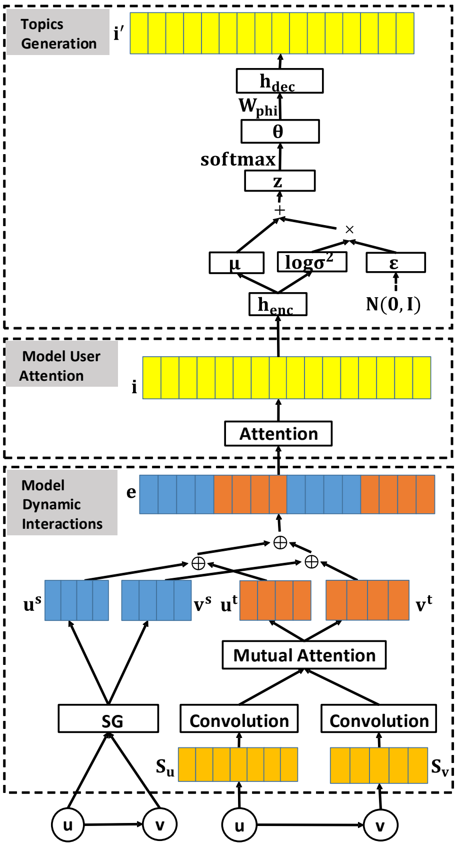


Figure 2: The IATM framework.

The proposed IATM framework is shown in Fig. 2. The generative process of IATM mainly includes three parts: (1) model dynamic interactions, (2) model user attention and (3) topics generation. A basis of three parts is that each word is represented as a low dimensional, continuous and real-valued vector, also known as word embedding (Mikolov et al., 2013c). Given text information of a vertex v, we take the embedding $\mathbf{w}_i \in \mathbb{R}^{d'}$ for each word to obtain embedding sequence as $\mathbf{S}_v = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_n)$. Here, d' indicates the dimension of word embeddings.

3.2.1 Model Dynamic Interactions

Corresponding to the assumption H1, a specific vertex should have its own diverse points with distinct neighbours, which leads to different edge embeddings.

To make full use of network structures and associated text information, we encode each vertex v as a concatenation of a structure-based embedding $\mathbf{v}^{(s)}$ and a text-based embedding $\mathbf{v}^{(t)}$, and get the vertex embedding $\mathbf{v} = \mathbf{v}^{(s)} \oplus \mathbf{v}^{(t)} (\mathbf{v} \in \mathbb{R}^d)$.

Structure-based Embedding. We adopt a neural language model (SkipGram, or SG for short) (Mikolov et al., 2013a) for $\mathbf{v}^{(s)}$. To maximize the probability of a node's neighbourhood co-occurrence, we define the objective function of structure-based embeddings as follows

$$L_s = -\log \sum_{-k \leq j \leq k} p(\mathbf{v}_{i+j}^{(s)} | \mathbf{v}_i^{(s)}) \tag{1}$$

where $\mathbf{v}_i^{(s)}$ is corresponding to the i-th vertex, the window size is k, and $p(\mathbf{v}_{i+j}^{(s)} | \mathbf{v}_i^{(s)})$ is defined using the *softmax* fuction

$$p(\mathbf{v}_{i+j}^{(s)}|\mathbf{v}_i^{(s)}) = \frac{\exp((\mathbf{v}_{i+j}^{(s)})^T \mathbf{v}_i^{(s)})}{\sum_{t=1}^{|V|} \exp((\mathbf{v}_t^{(s)})^T \mathbf{v}_i^{(s)})}. \tag{2}$$

In this way, nodes with similar neighbours share the similar structure-based embeddings.

Text-based Embedding. To discover the thematic information of the vertex pair in an edge, we utilize mutual attention (Santos et al., 2016; Tu et al., 2017) to obtain text-based embeddings, which allows the pooling operation to be aware of the topic of an edge $e_{(u,v)}$. To some extent, content information from a vertex can directly affect the text-based embedding of the other vertex, and vice versa.

Convolutional neural network (CNN) has gained great performance on the textual information encoding (Chen et al., 2015; Wang et al., 2017). Given the embedding sequence $\mathbf{S}_v$, we conduct convolution operation over $\mathbf{S}_v$ within the i-th window as follows

$$\mathbf{x}_i = \mathbf{C} \cdot (\mathbf{S}_v)_{i:i+l-1} + \mathbf{b} \tag{3}$$

where $\mathbf{C} \in \mathbb{R}^{d\times(l\times d')}$ is a convolution matrix, $\mathbf{b}$ is the bias vector and window size is l. The same operation is also on $\mathbf{S}_u$. The outputs of convolution, $\mathbf{P} \in \mathbb{R}^{d\times m}$ and $\mathbf{Q} \in \mathbb{R}^{d\times n}$ where m and n mean the length of $\mathbf{S}_u$ and $\mathbf{S}_v$ respectively, are as the input of mutual attention layer to compute the correlation matrix $\mathbf{F}$.

$$\mathbf{F} = relu(\mathbf{P}^{\mathbf{T}}\mathbf{A}\mathbf{Q})(\mathbf{F} \in \mathbb{R}^{m\times n}) \tag{4}$$

$\mathbf{A} \in \mathbb{R}^{d\times d}$ is a matrix to be learned by the neural network and we emlpoy $relu$ as activation function. Note that, the element $\mathbf{F}_{i,j} \in \mathbf{F}$ is the pair-wise correlation score of the hidden vectors $\mathbf{P}_i$ and $\mathbf{Q}_j$.

Then, we employ the mean-pooling along rows and columns of $\mathbf{F}$ to generate pooling vectors by

$$g_i^{(p)} = \mathbf{mean}(\mathbf{F}_{i,1}, ..., \mathbf{F}_{i,n}) \quad g_j^{(q)} = \mathbf{mean}(\mathbf{F}_{1,j}, ..., \mathbf{F}_{m,j}). \tag{5}$$

The pooling vectors of $\mathbf{P}$ and $\mathbf{Q}$ are obtained as

$$\mathbf{g}^{(p)} = (g_1^{(p)}, ..., g_m^{(p)})^T \quad \mathbf{g}^{(q)} = (g_1^{(q)}, ..., g_n^{(q)})^T. \tag{6}$$

Next, the $softmax$ function is operated on $\mathbf{g}^{(p)}$ and $\mathbf{g}^{(q)}$ to get the mutual attentive vectors $\mathbf{a}^{(p)}$ and $\mathbf{a}^{(q)}$. For instance, the i-th element of mutual attentive vector $\mathbf{a}^{(p)}$ is computed as

$$a_i^{(p)} = \frac{\exp(g_i^{(p)})}{\sum_{t=1}^{m} \exp(g_t^{(p)})}. \tag{7}$$

Finally, we get the text-based embeddings $\mathbf{u}^{(t)}$ and $\mathbf{v}^{(t)}$ by $\mathbf{u}^{(t)} = \mathbf{P}\mathbf{a}^{(p)}$ and $\mathbf{v}^{(t)} = \mathbf{Q}\mathbf{a}^{(q)}$. The objective function of text-based embeddings is as

$$L_t(e) = \alpha \log p(\mathbf{v}^{(t)}|\mathbf{u}^{(t)}) + \beta \log p(\mathbf{v}^{(t)}|\mathbf{u}^{(s)}) + \gamma \log p(\mathbf{v}^{(s)}|\mathbf{u}^{(t)}) \tag{8}$$

where α, β and γ control the weights of corresponding parts. Similarly, we employ $softmax$ function for calculating the conditional probabilities in Eq. (8) as in Eq. (2).

Edge Embedding. Here, we obtain the vertex embeddings $\mathbf{u}$ and $\mathbf{v}$ by $\mathbf{u} = \mathbf{u}^{(s)} \oplus \mathbf{u}^{(t)}$ and $\mathbf{v} = \mathbf{v}^{(s)} \oplus \mathbf{v}^{(t)}$. In this way, we operate $\mathbf{e} = \mathbf{u} \oplus \mathbf{v}$ to get the edge embedding $\mathbf{e} \in \mathbb{R}^{2d}$ for $e_{(u,v)}$, which is dynamic context-aware.

3.2.2 Model User Attention

As referred before, we assume that users who discuss the identical topic may take different effects on the topic. Attention mechanism is designed for mining the different significance of users on the topic. Since edge embedding is incorporated by the vertex-pair embeddings in an edge and each vertex denotes a user, the attention of users on the topic are transformed into the user attention vector $\mathbf{a}^{(e)} \in \mathbb{R}^{2d}$. $\mathbf{a}^{(e)}$

is computed by conducting $softmax$ function on the edge embedding $\mathbf{e}$. To formulate, the i-th element of $\mathbf{a}^{(e)}$ is computed as

$$a_i^{(e)} = \frac{\exp(\mathbf{e}_i)}{\sum_{t=1}^{2d}\exp(\mathbf{e}_t)}. \tag{9}$$

After that, through combining topic information of the vertex-pair in an edge with user attention on the topic, we obtain interaction-aware edge embedding $\mathbf{i} \in \mathbb{R}^{2d}$ by using element-wise product as $\mathbf{i} = (\mathbf{e}_1\mathbf{a}_1^{(e)}, ..., \mathbf{e}_{2d}\mathbf{a}_{2d}^{(e)})$.

3.2.3 Topics Generation

Neural variational inference (Miao et al., 2017) approximates the posterior of a generative model with a variational distribution parameterized by a neural network, which consistently generates the better topics. With respect to social media, we input the interaction-aware edge embedding, which is encoded with dynamic user behaviours and content information, into variational auto-encoder. To formulate, suppose $\hat{d}$ is a document, w is a word in $\hat{d}$ and the number of topics is K. Here, we adopt neural variational inference to infer the multinomial document-topic distribution $\theta_{\hat{d}} = (p(t_1|\hat{d}), ..., p(t_K|\hat{d}))$ and topic-word distribution $\phi_w = (p(w|t_1), ..., p(w|t_K))$, where t_i is the i-th topic.

Document-topic distribution. Precisely, given the interaction-aware edge embedding $\mathbf{i}$, we first encode it to a hidden space as

$$\mathbf{h}_{enc} = relu(\mathbf{W}^{(ih)}\mathbf{i} + \mathbf{b}^{(ih)}) \tag{10}$$

where $\mathbf{W}^{(ih)}$ and $\mathbf{b}^{(ih)}$ are the parameters of the neural network. Then the Gaussian parameters $\mu_{\hat{d}}$ and $\sigma_{\hat{d}}^2$ can be obtained as

$$\mu_{\hat{d}} = \mathbf{W}^{(h\mu)}\mathbf{h}_{enc} + \mathbf{b}^{(h\mu)} \quad \log(\sigma_{\hat{d}}^2) = \mathbf{W}^{(h\sigma)}\mathbf{h}_{enc} + \mathbf{b}^{(h\sigma)}. \tag{11}$$

The latent semantic vector $\mathbf{z} \in \mathbb{R}^K$ can be calculated using the reparameterization trick as $\mathbf{z} = \mu_{\hat{d}} + \epsilon \times \sigma_{\hat{d}}$ where $\epsilon \sim \mathcal{N}(\mu_0, \sigma_0^2)$ is the prior Gaussian distribution. The hyper-parameters μ_0 and σ_0^2 is set to a zero mean and unit variance Gaussian. Here we pass the Gaussian random vector $\mathbf{z}$ through the $softmax$ function to parameterize the multinomial document-topic distribution $\theta_{\hat{d}}$.

Topic-word distribution. The conventional topic models compute the conditional probability $p(w|\hat{d})$ as

$$p(w|\hat{d}) = \phi_w \times (\theta_{\hat{d}})^T. \tag{12}$$

Therefore, we compute the topic-word distribution ϕ_w as the parameter of neural network by

$$\mathbf{h}_{dec} = softmax(\phi_w \times (\theta_{\hat{d}})^T). \tag{13}$$

Thereafter, a new interaction-aware edge embedding $\mathbf{i}'$ is generated as

$$\mathbf{i}' = relu(\mathbf{W}^{(hi)}\mathbf{h}_{dec} + \mathbf{b}^{(hi)}) \tag{14}$$

where $\mathbf{W}^{(hi)}$ and $\mathbf{b}^{(hi)}$ are the parameters of neural network.

The objective for this part is as

$$L_{\theta,\phi}(e) = \mathbb{E}_{q(\theta,\mathbf{z}|\mathbf{i})}[\log p(\mathbf{i}|\mathbf{z},\theta,\phi)] - D_{KL}[q(\theta,\mathbf{z}|\mathbf{i})||p(\theta|\mu_{\hat{d}},\sigma_{\hat{d}}^2)] \tag{15}$$

where the variational distribution $q(\theta, \mathbf{z}|\mathbf{i})$ approximates the true posterior $p(\theta|\mu_{\hat{d}}, \sigma_{\hat{d}}^2)$ through Kullback-Leibler divergence.

3.3 Model Training

We need to minimize the overall objective function as

$$L = L_s + \sum_{e \in E}(L_t(e) + L_{\theta,\phi}(e)). \tag{16}$$

The conditional probability exploiting $softmax$ function is computationally expensive according to E-q. (1) and Eq. (8). Therefore, we employ negative sampling (Mikolov et al., 2013b) and Adam (Kingma and Ba, 2015) to optimize the overall objective function. In order to prevent overfitting, we also employ dropout (Srivastava et al., 2014) during the generation of document-topic and topic-word distribution.

4 Experiments

4.1 Datasets

We obtain the datasets shown in Tab. 1 based on the original microblog corpus used by Li et al., (2016a). They collected the posts of 50 frequent hashtags during May 1 -July 31, 2014 through Sina Weibo hashtag-search API[4]. Then they split the whole corpus into 3 datasets and each month represented a datasets. We further deal with the original datasets as follows: 1) Remove the posts whose length is less than 3 words or that have no poster username; 2) Filter users who have no reposting or replying relationship; 3) Aggregate all the original and the reposting posts from the same user to form text information of a user vertex.

Month	#Users	#Reposting
May	8907	10435
June	19293	25962
July	16990	20971

Table 1: Statistics of Datasets.

4.2 Evaluation Metrics

In previous work, a popular metric for topic model, perplexity, is evaluated based on the likelihood of held-out documents. Nonetheless, Chang et al. (2009) have proved that higher likelihood of held-out documents doesn't necessarily correspond to human perception of semantically coherent topics. Insteadly, we follow (Mimno et al., 2011) to calculate the coherence score of a topic given the top N words as

$$\mathcal{C} = \frac{1}{K} \cdot \sum_{k=1}^{K} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \log \frac{D(w_i^k, w_j^k) + 1}{D(w_j^k)} \tag{17}$$

where w_i^k refers to the i-th word in topic k ranked by topic distribution over words, $D(w_i^k, w_j^k)$ stands for the count of documents where w_i^k and w_j^k co-occur,and $D(w_j^k)$ indicates the number of documents that contain word w_j^k. In this paper, a document is the aggregated text of a user.

4.3 Baselines

To validate whether two assumptions mentioned in Section 3 are useful for topic extraction, we compare the proposed IATM with the following state-of-the-art baselines and its two variants:

Text Analysis Only:

(1) **SATM** (Quan et al., 2015) combines the aggregation of short texts and topic inference into an unified model.

(2) **BTM** (Yan et al., 2013) directly models the word co-occurrence patterns in the whole corpus.

(3) **LCTM** (Hu and Tsujii, 2016) uses word embeddings pretrained by a log-linear word2vec model[5] to deduce topics for tackling the data sparsity.

Text and Stucture Analysis:

(4) **LeadLDA** (Li et al., 2016a) generates words according to topic dependencies derived from conversation trees.

(5) **ForumLDA** (Chen and Ren, 2017) models the generation of topics by discriminating response messages relevant or irrelevant to the original post.

Two Variants:

(6) **IATM (-interaction)** is the IATM without dynamic interactions between users, which just takes text information into account and also uses neural variational inference to deduce topics.

[4] http://open.weibo.com/wiki/2/search/topics

[5] https://code.google.com/archive/p/word2vec/

(7) **IATM (-user attention)** considers the interactions between users, but not the various effects of users on the topic.

4.4 Experiment Settings

The hyperparameters of SATM, BTM, LCTM, LeadLDA and ForumLDA are set according to the best hyperparameters reported in their original papers. We run Gibbs samplings (in SATM, BTM, LCTM, LeadLDA and ForumLDA) with 1000 iterations to ensure convergence. For IATM (-interaction) and IATM (-user attention), the parameter settings are kept the same as IATM. Here, we set the vertex embedding dimension d as 200. We take the hyperparameters which achieve the best performance on datasets via an small grid search over combinations of the initial learning rate $[0.001, 0.0001]$, $\alpha \in [0.1, 1]$, $\beta \in [0.1, 1]$ and $\gamma \in [0.1, 1]$. Finally, learning rate is set as 0.001, α as 1.0, β as 0.3 and γ as 0.3. The other parameters are initialized by randomly sampling from normal distribution $\mathcal{N}(1.0, 0.3^2)$.

4.5 Performance Evaluation

Model	$K = 50$			$K = 100$		
	$N = 10$	$N = 15$	$N = 20$	$N = 10$	$N = 15$	$N = 20$
SATM	-76.10	-190.38	-341.22	-76.07	-190.48	-341.46
BTM	-79.27	-181.95	-326.77	-79.58	-183.00	-319.28
LCTM	-70.91	-165.37	-296.36	-58.65	-140.10	-261.40
LeadLDA	-53.91	-138.53	-258.38	-58.15	-141.34	-261.65
ForumLDA	-55.76	-129.57	-231.90	-55.84	-132.23	-236.89
IATM (-interaction)	-60.33	-145.98	-274.08	-64.26	-152.74	-280.40
IATM (-user attention)	-57.77	-126.06	-240.66	-59.57	-131.79	-233.54
IATM	**-43.34**	**-112.64**	**-228.27**	**-47.32**	**-121.46**	**-219.96**

Table 2: Coherence scores on May. Higher is better.

Model	$K = 50$			$K = 100$		
	$N = 10$	$N = 15$	$N = 20$	$N = 10$	$N = 15$	$N = 20$
SATM	**-31.04**	**-24.30**	**-58.39**	**-29.74**	**-22.02**	**-55.00**
BTM	-78.79	-179.99	-321.74	-75.77	-176.13	-315.43
LCTM	-91.72	-208.75	-367.76	-81.88	-181.57	-323.16
LeadLDA	-63.54	-150.18	-278.19	-72.07	-169.80	-309.40
ForumLDA	-78.22	-140.46	-229.62	-82.33	-160.46	-258.72
IATM (-interaction)	-74.61	-180.84	-340.83	-67.29	-161.30	-301.66
IATM (-user attention)	-69.75	-147.53	-234.73	-61.34	-148.01	-280.06
IATM	-46.69	-113.09	-213.61	-59.11	-133.96	-225.48

Table 3: Coherence scores on June. Higher is better.

Model	$K = 50$			$K = 100$		
	$N = 10$	$N = 15$	$N = 20$	$N = 10$	$N = 15$	$N = 20$
SATM	-128.28	-254.45	-431.88	-128.68	-254.74	-432.01
BTM	-73.26	-172.43	-313.12	-76.10	-176.67	-320.70
LCTM	-72.78	-160.08	-275.58	-63.56	-137.36	-238.31
LeadLDA	-70.40	-157.83	-268.23	-59.75	-130.83	-226.62
ForumLDA	-89.16	-215.47	-396.20	-89.96	-213.59	-386.65
IATM (-interaction)	-93.59	-224.17	-409.84	-91.96	-233.86	-396.22
IATM (-user attention)	-61.60	-144.75	-251.46	-57.00	-127.57	-227.81
IATM	**-50.75**	**-119.48**	**-212.26**	**-46.80**	**-110.27**	**-204.35**

Table 4: Coherence scores on July. Higher is better.

Following (Li et al., 2016a; Yan et al., 2013), we set the number of topics to 50 ($K = 50$) and 100 ($K = 100$). $K = 50$ is to match the count of hashtags and $K = 100$ is much larger than the real number of topics. As shown in Tab. 2, 3 and 4, we evaluate the top N = 10, 15, 20 words of $K = 50$ and $K = 100$. From these tables, we have the following overall observations:

(1) What call for special attention is that, SATM exhibits the unstable performance on various datasets. Specifically, SATM performs poorly on May, the worst on July and the best on June. It may be ascribed to its heavy reliance on the number of pseudo-documents. On the contrary, IATM has a stable performance and is only outperformed by SATM on June.

(2) Topic models that analyze text and structures perform better than the ones with only text except SATM on June. It indicates that considering texts and structures together is necessary due to their reciprocal influence in a social network.

(3) Our proposed IATM achieves crucial improvement in comparison with all the baselines on May and July, and it gives the greater coherence scores than the other baselines expect SATM on June. The reasons are two-fold: 1) It effectively identifies topics via simultaneously considering social media content and network topology. 2) It further produces interaction-aware edge embedding as a deeper edge representation combined dynamic user interactions and user attention.

Assumptions	$K = 50, N = 20$	$K = 100, N = 20$
Dynamic Interactions	27.32%	22.13%
User Attention	9.91%	11.87%

Table 5: The effects of two hypotheses by comparing IATM with its two variants.

To further evaluate the effectiveness of two assumptions, we compare IATM with its two variants, observations are as follows:

(4) IATM (-interaction) gives the worst coherence scores. The reason is that it has no consideration of social context. After introducing dynamic interactions, IATM (-user attention) obtains the considerable improvement. Moreover, IATM outperforms IATM (-user attention) on three datasets whatever different number of topics. It demonstrates the effectiveness of dynamic interactions (H1) and user attention (H2).

(5) For each component, we compute the average growth percentage under top $N = 20$ words of $K = 50$ and $K = 100$[6] in comparison of IATM and two variants. Seen from Tab. 5, two hypotheses are both useful for topic extraction in social media, and dynamic interactions modeling has greater influence than user attention.

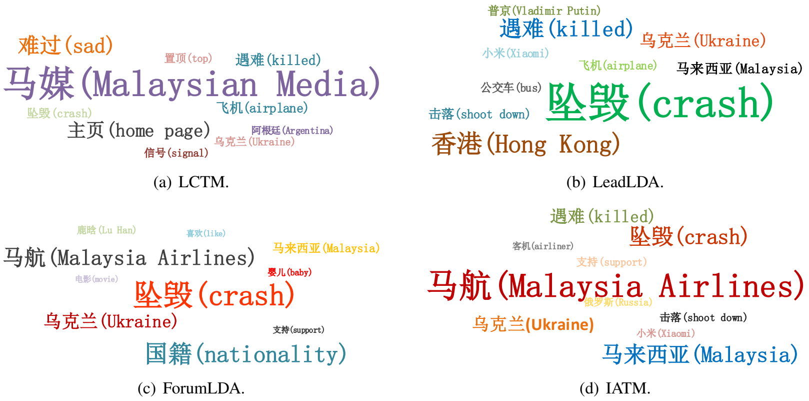
4.6 Case Study

To get an intuitive understanding of extracted topics, we design an experiment to visualize the top 10 words about "MH17 Crash" induced by the different models when $K = 50$, shown in Fig. 3. Due to "MH17 Crash" is included in the July dataset, LCTM gives the highest coherence score among the models of text analysis only. Then we choose LCTM, LeadLDA and ForumLDA as the competitors of our proposed method due to the limited space. Note that, individual words accompany with diverse font sizes. The larger the font size is, the more relevant the word is to the topic. We have the following observations from Fig. 3:

(1) As for LCTM based on word embeddings, "Malaysian Media" is the top one key word and "killed", "crash", "sad" and "Ukraine" is correlated to the topic. However, it mistakenly groups "Argentina lose the World Cup", which co-occurs with "sad" and "Argentina". Compared with LeadLDA, ForumLDA and IATM, LCTM performs the worst. It further testifies that the analysis of texts and structures in a social network is necessary due to their relevance in social media.

(2) With respect to LeadLDA, which distinguishes every message from leader and follower, "crash" is the top one key word and "killed", "Ukraine", "shoot down" and "Malaysia" is related to the topic. However, "bus" is falsely aggregated, which is relevant to the bus explosion in Guangzhou. Maybe it is

[6] The ones under other settings are not shown due to the limited space.



(a) LCTM.

(b) LeadLDA.

(c) ForumLDA.

(d) IATM.

Figure 3: Word clouds describing "MH17 Crash" of different models. Each word cloud represents the similar topic generated by the corresponding model with top 10 words. English translations for the original Chinese words are inside the brackets.

because it wrongly recognizes the reposting message about the bus explosion in Guangzhou as a leader message.

(3) ForumLDA differentiates reposting messages from relevant and irrelevant response posts. It generates the top one key word "crash", and "Malaysia Airlines", "Ukraine", "Malaysia" and "nationality" are correlated to the topic. However, "Lu Han" and "movie" about a movie starring Lu Han, are mistakenly mixed. Perhaps it is because no prior knowledge is given to ensure the validity of relevance distinction.

(4) The top words generated by IATM, "Malaysia Airlines", "Malaysia", "crash" and "Ukrain", are related to the topic. Moreover, some detailed information "airliner" is also produced. Nevertheless, the generated "Xiaomi" is not relevant to the topic. We check the corpus and find that there are lots of posts tagged one hashtag, but expressing the other topic. For instance, someone makes product promotion of "Xiaomi" by utilizing the hot event "MH17 Crash". This belongs to a spam message.

5 Conclusion and Future Work

We propose an Interaction-Aware Topic Model (IATM) for microblog conversations through integrating dynamic interactions and various user attention. Our method not only makes full use of content and structure information in social media, but also keeps the better consistency of generated topics by variational auto-encoder. We model user behaviours through transforming conversation trees into network. Further interaction-aware embeddings produced by adding diverse effects of different users on a topic, encode the content and structure information of two neighbour users. This helps to incorporate the conversation context, and is easy to absorb user attention. Then it is plugged in neural variational inference to generate topics. Experiments on three real-world microblog datasets demonstrate that the proposed IATM is effective.

Yet lots of people would like to utilize hot events to have a product promotion and post some irrelevant messages. These spams damagingly affect the performance of IATM. In the future, we will unify spam posts separation during topics deduction.

Acknowledgments

The work was supported by National Natural Science Foundation of China (61472277, 61772361, 61771333). We would like to thank anonymous reviewers for the valuable and detailed comments and suggestions.

References

Bin Bi, Yuanyuan Tian, Yannis Sismanis, Andrey Balmin, and Junghoo Cho. 2014. Scalable topic-specific influence analysis on microblogs. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 513–522.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.

Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L Boyd-Graber, and David M Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in Neural Information Processing Systems*, pages 288–296.

Chaotao Chen and Jiangtao Ren. 2017. Forum latent dirichlet allocation for user interest discovery. *Knowledge-Based Systems*, 126:1–7.

Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, Jun Zhao, et al. 2015. Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 167–176.

Liangjie Hong and Brian D Davison. 2010. Empirical study of topic modeling in twitter. In *Proceedings of the 1st Workshop on Social Media Analytics*, pages 80–88.

Weihua Hu and Junichi Tsujii. 2016. A latent concept topic model for robust topic inference using word embeddings. In *The 54th Annual Meeting of the Association for Computational Linguistics*, pages 380–386.

Diederik Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. *Proceedings of the International Conference on Learning Representations*.

Diederik P Kingma and Max Welling. 2014. Auto-encoding variational bayes. In *Proceedings of the International Conference on Learning Representations*.

Jing Li, Ming Liao, Wei Gao, Yulan He, and Kam-Fai Wong. 2016a. Topic extraction from microblog posts using conversation structures. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*.

Yang Li, Ting Liu, Jing Jiang, and Liang Zhang. 2016b. Hashtag recommendation with topical attention-based lstm. In *Proceedings of the 26th International Conference on Computational Linguistics*.

Kar Wai Lim, Changyou Chen, and Wray L. Buntine. 2013. Twitter-Network Topic Model: A full Bayesian treatment for social network and text modeling. In *Advances in Neural Information Processing Systems: Topic Models Workshop*, pages 1–5.

Hengyang Lu, Lu-Yao Xie, Ning Kang, Chong-Jun Wang, and Jun-Yuan Xie. 2017. Don't forget the quantifiable relationship between words: Using recurrent neural network for short text topic discovery. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pages 1192–1198.

Rishabh Mehrotra, Scott Sanner, Wray Buntine, and Lexing Xie. 2013. Improving lda topic models for microblogs via tweet pooling and automatic labeling. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 889–892.

Yishu Miao, Edward Grefenstette, and Phil Blunsom. 2017. Discovering discrete latent topics with neural variational inference. In *Proceedings of the 34th International Conference on Machine Learning*.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *CoRR, abs/1301.3781*, abs/1312.6114.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119.

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013c. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 13, pages 746–751.

David Mimno, Hanna M Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 262–272.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 701–710.

Xiaojun Quan, Chunyu Kit, Yong Ge, and Sinno Jialin Pan. 2015. Short and sparse text topic modeling via self-aggregation. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, pages 2270–2276.

Cicero Dos Santos, Ming Tan, Bing Xiang, and Bowen Zhou. 2016. Attentive pooling networks. *CoRR, abs/1602.03609*.

Vivek Kumar Rangarajan Sridhar. 2015. Unsupervised topic modeling for short texts using distributed representations of words. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pages 192–200.

Akash Srivastava and Charles Sutton. 2017. Autoencoding variational inference for topic models. *Proceedings of the International Conference on Learning Representations*.

Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958.

Jian Tang, Ming Zhang, and Qiaozhu Mei. 2013. One theme in all views: modeling consensus topics in multiple contexts. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 5–13.

Cunchao Tu, Han Liu, Zhiyuan Liu, and Maosong Sun. 2017. Cane: Context-aware network embedding for relation modeling. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 1722–1731.

Xuepeng Wang, Kang Liu, and Jun Zhao. 2017. Handling cold-start problem in review spam detection by jointly embedding texts and behaviors. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 366–376.

Chen Xing, Wei Wu, Yu Wu, Jie Liu, Yalou Huang, Ming Zhou, and Wei-Ying Ma. 2017. Topic aware neural response generation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pages 3351–3357.

Xiaohui Yan, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. 2013. A biterm topic model for short texts. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1445–1456.

Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. 2011. Comparing twitter and traditional media using topic models. In *Proceedings of the 33rd European Conference on Information Retrieval*, pages 338–349.

Hao Zhuang, Rameez Rahman, Xia Hu, Tian Guo, Pan Hui, and Karl Aberer. 2016. Data summarization with social contexts. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 397–406.