#HowYouTagTweets: Learning User Hashtagging Preferences via Personalized Topic Attention

Yuji Zhang[1] Yubo Zhang[1] Chunpu Xu[1] Jing Li[1*] Ziyan Jiang[2†] Baolin Peng[3]

[1]Department of Computing, The Hong Kong Polytechnic University, HKSAR, China
[2]Amazon Alexa AI, Seattle, WA   [3]Microsoft Research, Redmond, WA
[1]{yu-ji.zhang, yubo.zhang, chun-pu.xu}@connect.polyu.hk
[1]jing-amelia.li@polyu.edu.hk   [2]ziyjiang@amazon.com
[3]bapeng@microsoft.com

Abstract

Millions of hashtags are created on social media every day to cross-refer messages concerning similar topics. To help people find the topics they want to discuss, this paper characterizes a user's hashtagging preferences via predicting how likely they will post with a hashtag. It is hypothesized that one's interests in a hashtag are related to what they said before (user history) and the existing posts present the hashtag (hashtag contexts). These factors are married in the deep semantic space built with a pre-trained BERT and a neural topic model via joint training. In this way, user interests learned from the past can be customized to match future hashtags, which is beyond the capability of existing methods assuming unchanged hashtag semantics. Furthermore, we propose a novel *personalized topic attention* to capture salient contents to personalize hashtag contexts. Experiments on a large-scale Twitter dataset show that our model significantly outperforms the state-of-the-art recommendation approach without exploiting latent topics.[1]

1 Introduction

Virtual communications are playing an increasingly crucial role in our daily activities. In this profound revolution in interpersonal communications, social media has become the key channel to connect an individual with human society. People are now turning to the online world to exchange viewpoints, voice opinions, and engage in topics they are interested in. Nonetheless, the deluge of the online contents streaming through social media every day has presented the concrete challenges for users to catch what they essentially need.

Sample tweets in H's hashtag contexts.
Love thriller and mystery? Check out: URL
She is writing the end for a long time.
85-5 star review!
Why not visit for Sunday share?

Sample tweets in U's user history.
Darpocalypse is now available as an ebook! Read book 2 in the *epic living dead series*. *Zombie Thriller Apocalypse* Reader: *Zeke* is a skilled lover, so easy to fall for. But he has a *dark* & *twisted* nature.

Figure 1: The hashtag context snippet of *#book* (H) is shown on the top and the target user U's hashtagging history snippet on the bottom. U later tweeted and tagged H. The topic words indicating U's hashtagging preferences are in italic and blue, so is the tweet in H's context implying its potential to be engaged with U.

To deal with this issue, many platforms encourage users to tag their messages with *hashtags* (e.g., "*#COVID19*" and "*#NLPeople*") — topic labels referring to similar posts and allowing easier microblog search. However, users may only want to explore the hashtags (and the linked messages) they are interested in, whereas platforms tend to display popular hashtags for all the users. In light of this concern, how to curate personalized hashtags to draw better user engagements?

This paper studies **hashtag personalization** which explores users' hashtagging preferences and predicts how likely a user will put a hashtag in their future posts (henceforth *user-hashtag engagements*). We hypothesize that one's hashtagging behavior is highly related to two factors: (1) the user's personal interests reflected by their past hashtagging history (*user history*) and (2) the contents appear with the hashtag in posts (*hashtag contexts*).

To illustrate this intuition, Figure 1 shows the sample tweets tagged with *#book* (H) and some history tweets of a user U. As can be seen, U was a big fan of mystery books, who later tweeted to share a book and tagged H for topic indication. Such future engagements can be signalled by the contexts of H from words like "thriller" and "mys-

tery" which suggest U's potential interests to H.

In previous research, most studies (Wu et al., 2017) ignore the possible semantic change of hashtags and hence insurmountable to tackle a ubiquitous issue in real life — new hashtags might appear after the model is trained and even the old ones may exhibit different meanings in the future (and are hence likely to engage different user groups).

Another similar-by-name concept is "hashtag recommendation", which attempts to fit a hashtag to a post and is thus post-oriented (Li et al., 2016; Gong and Zhang, 2016). Our task is oriented at users, whose past hashtagging behavior is explored to foresee their preferences to the future hashtags. If we consider a broader area of online content recommendation (e.g., tweets (Chen et al., 2012), conversations (Zeng et al., 2019)), many studies rely on user-content interaction history with pairwise labels for supervised learning. They are hence unable to adapt the user interests trained from history data to personalize future contents, due to the semantic gap between past and future data. This is however unignorable when studying social media contents which commonly exhibit evolving topics.

To leverage user hashtagging history (in the past) to hashtag contexts in the future, we first conduct unsupervised learning to encode hashtags, whose embeddings are later used for joint training to gain user interests from hashtagging history. For hashtag modeling, a pre-trained BERT (Devlin et al., 2019) and a neural topic model (NTM) (Miao et al., 2017) work together to couple the effects of local contexts on message level (handled by BERT) and global contexts observed from word statistics (handled by NTM).

Furthermore, we propose a novel mechanism of *personalized topic attention* over hashtag contexts to capture the keypoints therein, where the useful features might be sparse and the tremendous noise is likely to hinder the model's capability to learn anything helpful. For instance, in Figure 1, only the first tweet in H's contexts explicitly hint the connection of H with U's interests, while the others are relatively useless. To address such issues, we leverage latent topics and user embeddings to attend salient contextual messages exhibiting the positive match with the user hashtagging history.

To the best of our knowledge, *we are the first to personalize hashtags via joint training to adapt user interests gained in history to future hashtag semantics, where the joint effects of latent hash-tag topics and user hashtagging preferences are explored in a novel personalized topic attention.*

For the experiments, we gather a large-scale Twitter dataset and use absolute time to separate the training and validation data (before the time) from the test (after the time). In this way, hashtags available for user preference modeling (in training) will exhibit a gap compared with the hashtags (in test). Moreover, we focus on hashtags that do not appear in the target user's history, considering that users may prefer new and unseen hashtags.

Main experimental results show that our model significantly outperforms the state-of-the-art recommendation model. For example, we present 0.311 MAP compared with 0.277 obtained by adapting Zeng et al. (2020). We then examine the effects of varying user history and hashtag contexts, where we observe consistently better results on varying users, while hashtag modeling might be benefited from richer contexts. Next, an analysis of our model indicates that all its components collaborate to effectively capture user hashtagging preferences and meaningful topics can be discovered to help attend salient contents for personalizing hashtags. At last, we compare the attended history tweets and the future tweets presenting successful engagements and shed light on our potential to not only answer "yes-or-no" to whether a user will be engaged with a hashtag but also how it happens.

2 Related Work

This paper is in line with prior studies for personalizing hashtags. They adopt shallow word statistics (e.g., latent topics (Zhao et al., 2016) and content factors (Wu et al., 2017)) or handcrafted features (Alkouz and Al Aghbari, 2020) to characterize user behavior. Different from them, we employ neural networks to explore users' hashtagging preferences in the deep semantic space, which enables natural inclusion of richer information and better language understanding ability.

Our work is also related to hashtag recommendation, which "recommends" a hashtag to a microblog post via hashtag ranking (Li et al., 2016; Gong and Zhang, 2016; Huang et al., 2016) or generation (Ding et al., 2012; Wang et al., 2019) unaware of the personalized information. Others (Zhang et al., 2014; Gong et al., 2015) encode authors' writing styles for text consistency with the inserted hashtag. Bayesian graphical models are adopted for user modeling, which requires massive
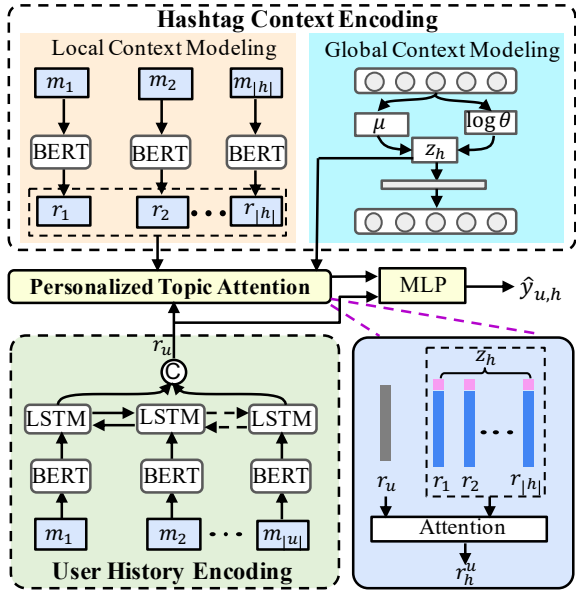
Figure 2: Our model that couples the effects of hashtag context encoding (top) and user history encoding (bottom left) with a personalized topic attention (bottom right) to predict user-hashtag engagements.

manual efforts to customize inference algorithms.

In a broader line of online content recommendation, collaborative filtering (CF) is popular to explore user interaction history for recommendation, e.g., tweets (Chen et al., 2012), topics (Lu et al., 2015), conversations (Zeng et al., 2019). For other text-based recommenders, contents (to be recommended) are explored barely with supervised learning (based on user history) (Yu et al., 2016; Zeng et al., 2018). These approaches are therefore unable to cater for contents with evolving semantics. Our work employs a pre-trained BERT and neural topic model to gain the preliminary context understanding of future hashtags, which has never been studied before. Zeng et al. (2020) also capture the dynamic user interests to recommend conversations. Our task exhibits a new challenge to encode fragmented and noisy hashtag contexts and we propose a novel personalized topic attention to capture salient contents to predict user's future behavior.

3 Hashtag Personalization Framework

This section presents how to predict a user u's future engagement with a hashtag h via leveraging u's hashtagging history and h's contextual messages. Figure 2 shows our model architecture.

Input and Output. Before we start the story about how our model works, we formulate the task

of hashtag personalization here. The inputs are a user u and a hashtag h, where h was not tagged by u in the history messages. To capture u's hashtagging preferences, we employ u's past chatting messages with other hashtags for user modeling. For the modeling of h, the messages sharing the hashtag h are adopted as the hashtag contexts. The output is how likely u will tag h in a future post – h forms a user-hashtag engagement pair with u.

3.1 User History Encoding

As shown above, the user history of u is encoded from a sequence of u's messages $\langle m_1, m_2, ..., m_{|u|} \rangle$ in chronologically order. $|u|$ is the number of u's history messages. User embeddings are hence learned in the two-level hierarchy on message and user.

Message-level Modeling For a message m in user history, we explore the semantics from its word sequence $\mathbf{w}_m$ with a pre-trained BERT encoder and map m into a latent vector space. The learned message embedding is denoted as $\mathbf{r}_m$ and will be further delivered to conduct user-level modeling with other history messages.

User-level Modeling Here, inspired by Zeng et al. (2020), temporal patterns of hashtagging history are explored to capture the possible user interest change over time. To that end, the embeddings of user history messages ($\mathbf{r}_m$) are sequentially encoded by a Bidirectional LSTM network. For the i-th message, u's current preferences $\mathbf{h}_i^u$ are updated based on the previous interests $\mathbf{h}_{i-1}^u$ and the current behavior $\mathbf{r}_{m_i}$. The last hidden states of the two directions are then concatenated to represent u's overall hashtagging preferences $\mathbf{r}_u$.

3.2 Hashtag Context Encoding

Following the aforementioned description, the contexts of a hashtag h are formed with h_{set}: a set of messages $\{m_1, m_2, ..., m_{|h|}\}$, each hashtagged with h. $|h|$ is the message number used to explore hashtag contexts, which are learned globally from inter-message word co-occurrence patterns and locally based on intra-message semantics.

Global Context Modeling The previous discussions concern the severe data sparsity in hashtag contexts attributed to their fragmented and noisy nature. To help our model capture essential features to characterize hashtags, we employ a neural topic

model (NTM) to discover the latent topics (distributional word clusters) via exploring the inter-message word co-occurrences on hashtag level.

Here we adopt the NTM design based on variational auto-encoder (VAE) following Miao et al. (2017). Here h's hashtag contexts are represented by a bag-of-word (BoW) vector $\mathbf{v}_h$ over the vocabulary V, which is first encoded into a latent topic vector $\mathbf{z}_h$ (in K dimension and K for the topic number), followed by the decoding process re-generates $\mathbf{v}_h$ conditioned on $\mathbf{z}_h$. For encoding, $\mathbf{v}_h$ is mapped into the latent topic space via Gaussian sampling:

$$\mu = f_\mu(f_e(\mathbf{v}_h)), log\sigma = f_\sigma(f_e(\mathbf{v}_h)), \mathbf{z}_h = \mathcal{N}(\mu, \sigma) \quad (1)$$

where $f_*(\cdot)$ denotes a ReLU-activated neural perceptron. In decoding, we first conduct a softmax transformation over the latent topic vector $\mathbf{z}_h$ to yield the hashtag-topic distribution θ_h for h. Then, we reconstruct $\mathbf{v}_h$ via predicting $\hat{\mathbf{v}}_h$ below:

$$\hat{\mathbf{v}}_h = \text{softmax}(f_\phi(\theta_h)) \quad (2)$$

$f_\phi(\cdot)$ is another ReLU-activated perceptron mapping topic space to V-dim BoW space. The weights of $f_\phi(\cdot)$ (after softmax normalization) are adopted to represent the topic-word distributions and the latent topic vector $\mathbf{z}_h$ is employed as the global context representation for h.

Local Context Modeling. Here, the focus is to explore the word semantics inside a message. Similar to the way we conduct message-level modeling for user history, pre-trained BERT is first employed to yield latent embeddings $\mathbf{r}_m$ for a message m in hashtag contexts. Then, to enable the model to focus on essential messages for hashtag modeling, a personalized topic attention (in aware of the user embedding $\mathbf{r}_u$ and latent hashtag topic $\mathbf{z}_h$) is put over the message embeddings to generate the personalized hashtag embedding $\mathbf{r}_h^u$ (Section 3.3).

3.3 User-Hashtag Engagement Prediction with Personalized Topic Attention

To capture user hashtagging preferences, personalized topic attention is put over hashtag contexts to explore user u's potential interests to hashtag h and the attended context vector is further delivered to learn the prediction of user-hashtag engagement.

Personalized Topic Attention. This mechanism is designed to indicate messages in hashtag contexts whose topics exhibit better consistency with

the target user's past hashtagging behavior. For all h's contextual messages, we first concatenate their embeddings with h's latent topic vector $\mathbf{z}_h$ to present a topic-aware message embedding $[\mathbf{r}_m; \mathbf{z}_h]$, which helps to provide the local message representation with the global view of hashtag-level topics.

Then, u's user embedding $\mathbf{r}_u$ is employed to query h's contextual messages to further inject personalized information to the learning of attention weights. Concretely, we capture the semantic relations between user embedding ($\mathbf{r}_u$) and the message m in h's context using the formula below:

$$f_{att}(\mathbf{r}_u, \mathbf{r}_m^h) = \mathbf{W}_{att}(\mathbf{r}_u, [\mathbf{r}_m; \mathbf{z}_h]) + \mathbf{b}_{att} \quad (3)$$

where $\mathbf{W}_{att}$ and $\mathbf{b}_{att}$ are learnable parameters. Next, attention weights (in aware of topic and user representations) are computed as following to indicate the messages in h's context set h_{set} which indicate better match with u's user interests:

$$\alpha_{u,m} = \frac{\exp(f_{att}(\mathbf{r}_u, [\mathbf{r}_m; \mathbf{z}_h]))}{\sum_{n \in h_{set}} \exp(f_{att}(\mathbf{r}_u, [\mathbf{r}_n; \mathbf{z}_h]))} \quad (4)$$

Afterwards, we conduct weighted sum to attend messages in h_{set} and produce the context vector:

$$\mathbf{r}_h^u = \sum_{m \in h_{set}} \alpha_{u,m} \cdot \mathbf{r}_m^h \quad (5)$$

$\mathbf{r}_h^u$, employed to represent h, carries both topic information and the personal hashtagging interests of u (henceforth personalized hashtag embedding).

User-Hashtag Engagement Prediction. At output layer, we explore how similar the hashtag contexts (represented by $\mathbf{r}_h^u$) are to the historical user hashtagging behavior (encoded in $\mathbf{r}_u$). Here MLP (He et al., 2017) is adopted to measure the potential engagements of user u and hashtag h:

$$\mathbf{r}_{u,h} = \gamma(\mathbf{W}_{MLP}[\mathbf{r}_h^u; \mathbf{r}_u] + \mathbf{b}_{MLP}) \quad (6)$$

where $\gamma(\cdot)$ is the ReLU activation function. $\mathbf{W}_{MLP}$ and $\mathbf{b}_{MLP}$ are both learnable parameters.

Finally, $\mathbf{r}_{u,h}$ is used to predict $\hat{y}_{u,h}$ which signals how likely u will hashtag h (positive engagement):

$$\hat{y}_{u,h} = \text{sigmoid}(\mathbf{W} \cdot \mathbf{r}_{u,h} + \mathbf{b}) \quad (7)$$

$\mathbf{W}$ and $\mathbf{b}$ are learnable parameters for training.

3.4 Training Losses and Joint Training

The entire framework in joint training fashion is trained via minimizing multiple losses together.

Training Losses. We design two training objectives — an engagement loss for predicting user-hashtag engagement and an NTM loss for exploring latent topics in hashtag contexts.

The former (engagement loss) is designed based on the weighted binary cross-entropy following (Zeng et al., 2020). Given a training set τ of user-hashtag pairs (u, h), we minimize the loss $\mathcal{L}_{eng}$ via learning from their pairwise ground-truth label $y_{u,c}$ (1 for positive engagement and 0 for negative):

$$\mathcal{L}_{eng} = -\sum_{(u,h)\in\tau} (\lambda\cdot y_{u,c}\log(\hat{y}_{u,c})+(1-y_{u,c})\log(1-\hat{y}_{u,c})) \quad (8)$$

The hyperparameter $\lambda > 1$ trades off the weights of positive and negative hashtag-user engagement pairs. Intuitively, more weights should be put on positive pairs, which is relatively more reliable, whereas negative pairs might be affected by many unpredictable things (e.g., users' busy schedule). For the same consideration, negative sampling is adopted to speed up the training (He et al., 2017).

For the NTM loss ($\mathcal{L}_{ntm}$) in hashtag modeling, we follow Miao et al. (2017) to use variational inference (Blei et al., 2017) to approximate a posterior distribution over hashtag h's latent topic $\mathbf{z}_h$ given the statistics of words observed in its contexts.

$$\mathcal{L}_{ntm} = D_{KL}(p(\mathbf{z}_h)\,||\,q(\mathbf{z}_h|h)) - \mathbb{E}_{p(\mathbf{z}_h)}[p(h|\mathbf{z}_h)] \quad (9)$$

$D_{KL}(\cdot)$ is the Kullback-Leibler divergence loss and $\mathbb{E}_*[\cdot]$ measures the VAE reconstruction.[2]

Joint Training. To gain the preliminary understanding for hashtags with limited user engagement history for training (hashtag coldstart), we optimize NTM loss first for pre-training. Then, we joint train NTM and user-engagement prediction in an unified framework. The joint-training loss of the overall framework to learn hashtagging preferences is defined as following:

$$\mathcal{L} = \mathcal{L}_{eng} + \mu \cdot \mathcal{L}_{ntm} \quad (10)$$

where μ is the weight balancing the two effects.

4 Experimental Setup

Datasets and Setup. A large Twitter dataset was first gathered with the official streaming API in

Feb 2013, which contains 900M tweets.[3] We then filtered out tweets without hashtags and capped the user history at 50 tweets. Next, hashtag texts are hidden from both history and contexts to avoid the trivial features learned by the models, and the tweets presenting hashtags in the middle were ignored to enable better semantic learning (following Wang et al. (2019)). At last, we removed users who posted original hashtags only (never tagged by others) because these users cannot be taken for prediction. Here we distinguish hashtag first-use from reuse because the latter is dominant (taking 82% of hashtags) yet relatively valueless (users tend to see new hashtags), while prior settings tend to mix them together (Zhao et al., 2016; Wu et al., 2017).

For training and evaluation, we rank the tweets by time and take the earliest 80% for training, the latest 10% for test, and the remaining 10% for validation. In this way, the model is trained with past data and tested for the future, which reflects a more realistic scenario compared to random split (Wu et al., 2017). Based on our setup, users newly appearing in the test will not be taken away because they have no history to learn user embedding.

Number of Tweets	33,881
Number of Users	2,571
Number of Hashtags	22,320
Average tweet number per user	13
Average hashtags number per user	12
Average tweet number per hashtag	3
New Hashtag Rate (%)	55

Table 1: Statistics. New hashtag: newly created hashtags appear in test but absent in training for all users.

Data Analysis. Table 1 shows the data statistics. It exhibits more than half of "future hashtags" which do not appear in any user's history. This demonstrates the prominence of coldstart hashtags resulting from dynamic social media topics. We also observe the severe sparsity of hashtag contexts, exhibiting the variability of hashtags caused by the casual writings on social media. User history, on the other hand, seems to dense, providing rich contents to learn dynamic interests yet presenting another challenge of how to match hashtags (with sparse context) to users (with dense history).

To further probe into user and hashtag statistics, we plot the tweet number distribution in hashtag contexts in Figure 3a and user history in 3b. Most

---

[2]Due to the space limitation, we leave out the details of the derivation and refer the readers to Miao et al. (2017).

[3]https://developer.twitter.com/en/docs/tutorials/stream-tweets-in-real-time

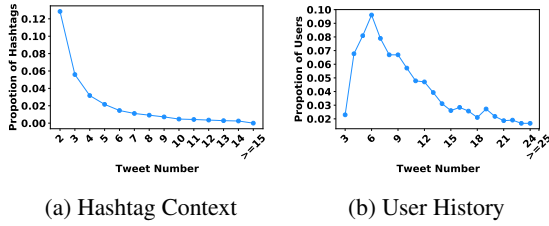|	(a) Hashtag Context	(b) User History

Figure 3: The distribution of tweet number (x-axis) in hashtag contexts (left) and user history (right). Y-axis indicates the proportion of instances.

hashtags appear in only a few tweets (Fig. 3a), exhibiting a power-law distribution. User history has relatively richer contents (Fig. 3b), which implies users who hashtag once are likely to do it again.

Preprocessing and Model Settings. The tweet pre-processing was first conducted via open-source toolkit NLTK (Bird, 2006) for tokenization, stemming, and lemmatization. Then, meaningless tokens (e.g., links, punctuation, mentions) were removed. At last, a vocabulary of the most frequent 10K tokens was maintained for both word sequence input (to BERT) and BoW (to NTM).

In training, hashtag contexts were capped at 30 tweets via sampling. And to avoid problem degeneration, we only consider hashtags appear in ≥ 5 tweets for context modeling, though others may engage in user history modeling. The sampling ratio was set to 5 for negative sampling for training while all test hashtags will be ranked for evaluation.

For model setup, we took BERTweet (Nguyen et al., 2020) as the pre-trained BERT for message encoding. We adopted two layers of Bi-LSTM whose hidden states were set to 768 for each direction. The models were trained with Adam optimizer, initialized with a learning rate of 1e-3, and training batch size 128. The hyperparameters were set via grid search on the validation set, where the topic number $K = 100$, the positive and negative sample tradeoff $\lambda = 100$ (Eq. 6), and $\mu = 1e - 4$ (Eq. 10) for balancing training losses.

For the joint training of our model, we first pre-trained NTM for 20 epochs, warmed up other parameters for 20 epochs, and then updated all parameters simultaneously for 100 epochs.

Evaluation and Comparison. Our evaluation metrics follow the recommendation practice (Zeng et al., 2020) to measure the hashtag ranks predicted for each target user. Here we employ the popular information retrieval metrics precision@N (P@N),

mean average precision (MAP) over the top N predictions, and normalized Discounted Cumulative Gain at N (nDCG@N). In the experimental results (Section 5), the scores reported are measured given $N = 5$ and similar trends hold for other N settings.

For baselines, we first consider a simple method which rank hashtags randomly (RANDOM), by their frequency in training set (POPULARITY), and the cosine similarity between user and hashtag embeddings (based on BERT) (henceforth BERT-SIM). Then, we examine features learned by LDA (Blei et al., 2001) and TF-IDF, where the popular learning-to-rank model GBDT (Friedman, 2001) is used for hashtag ranking.[4] Moreover, collaborative filtering (CF) is compared, which recommend hashtags based on user-hashtag interaction history (without content modeling).

For neural comparisons, we tailor-make Zeng et al. (2020), the state-of-the-art conversation recommendation model employing user history encoder based on LSTM with attention (henceforth LSTM-ATT), to personalize hashtags encoded by pre-trained BERT. Its two ablations — one without attention (LSTM) and the other simply adopting (MLP) to match BERT-generated user and hashtag embeddings are also considered.

5 Experimental Results

In this section, we first discuss the main results in Section 5.1 and quantify the effects of varying lengths of user history and hashtag contexts for measuring their engagement potential in Section 5.2. Then, Section 5.3 further interprets our superiority with an ablation study and a case study. At last, we analyze our ability to predict users' future hashtagging behavior in Section 5.4.

5.1 Main Comparison Results

Table 2 shows the main comparison results, where the following observations can be drawn.

First, personalizing hashtags is not trivial. All baselines obtain poor results, though BERT-SIM performs relatively better thanks to its language understanding ability priorly gained by pre-trained BERT. Then, CF yields worse results than BERT-SIM, which showcase content features' better ability to indicate the future user-hashtag engagements than user interaction history. This is because future

---

[4]The GBDT implementation is based on RankLib: https://sourceforge.net/p/lemur/wiki/RankLib/, while LDA and Tf-Idf feature are extracted by scikit-learn: https://scikit-learn.org/.

Dataset	Twitter		
Models	MAP	P@5	nDCG@5
Baselines			
RANDOM	0.0741	0.0200	0.0467
POPULAR	0.0814	0.0233	0.0564
BERT-SIM	0.2264	0.0800	0.2251
Non-neural			
LDA	0.1107	0.0360	0.0882
TF-IDF	0.1555	0.0560	0.1424
CF	0.1085	0.0424	0.1009
Neural			
MLP	0.2624	0.0994	0.2677
LSTM	0.2659	0.1004	0.2763
LSTM-ATT (*SOTA*)	0.2765	0.1043	0.2873
OUR MODEL	**0.3114**	**0.1148**	**0.3257**

Table 2: The results for hashtag ranking (average over users). For all the metrics, the higher the better. Our model obtains the best results (in boldface) and significantly outperform others (p<0.01, paired t-test).

hashtags will exhibit a different semantic space compared to the past, while CF is unable to leverage future hashtag contexts for prediction.

Next, neural features are more useful than the non-neural counterparts. This suggests the ability of neural encoders to explore deep semantics for effectively characterizing user and hashtag factors, as opposed to LDA and TF-IDF which rely on shallow word statistics to handle a challenging task. In neural comparison models, LSTM and LSTM-ATT slightly perform better than MLP because user hashtagging behavior may change over time which can be captured by LSTM encoders; LSTM-ATT outperforms LSTM thanks to the ability of attention mechanism to explore salient user history message for personalization. At last, all results from our model are significantly better than others by a large margin. This implies personalized topic attention can indeed effectively encode noisy hashtag contexts and allow a better learning of user hashtagging preferences.

5.2 Engagement Prediction with Varying User History and Hashtag Contexts

Here we further examine how varying lengths of user history and hashtag contexts affect the prediction of their future relations.

Results with Varying User History. Figure 4 shows the P@5 yielded by our model and the LSTM-ATT comparison for the prediction on users with varying history tweet number.

It is observed that both models exhibit a performance drop when handling very long user history, probably resulting from the small-scale training

samples for user modeling (shown in Figure 3b). Nevertheless, our model performs consistently better for users with varying sparsity degrees of hashtagging history, especially for those having limited data to capture their past hashtagging preferences. This is probably because unsupervised topic modeling will enable our model to make use of the unlabeled data (future hashtag contexts), which enrich the limited features allowed to be captured from the sparse user interaction history.
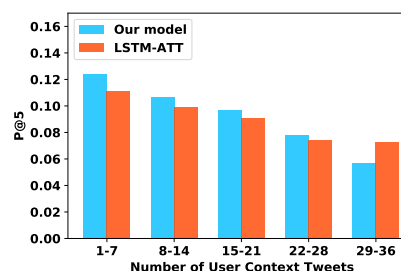

Figure 4: P@5 (y-axis) for predicting on users with number of varying history tweets (x-axis). For each group, the results of our model is shown on the left and those of LSTM-ATT is on the right.

Results with Varying Hashtag Contexts. Here we discuss the effects of hashtag contexts on the prediction and display the P@5 scores of our model and LSTM-ATT in Figure 5. It is observed that LSTM-ATT performs slightly better for hashtags with sparse contexts, while our model exhibits large-margin performance gain for hashtags with rich context.

The reasons are two-fold. First, the latent topics are discovered in word co-occurrences in hashtag contexts, while more tweets will result in richer statistical patterns for topic modeling and better representation learning for hashtags. Second, in addition to richer contents, more contextual tweets are likely to bring in more noisy information. The noise may hinder LSTM-ATT to effectively explore the context because it relies on local contexts (captured by BERT) for hashtag modeling. On the contrary, our model is able to couple local contexts with the global ones (latent topics), which will enable the personalized topic attention to concentrate on what is essentially helpful for prediction.

5.3 Further Discussions to Our Model

Here we probe into our model output to examine how it works to characterize user hashtagging preferences and its current limitation.
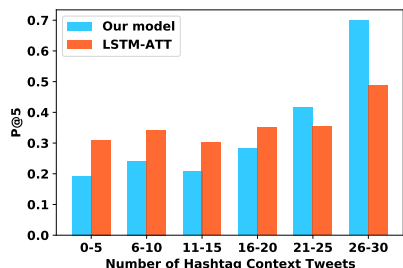
Figure 5: P@5 (y-axis) for predicting on hashtags with varying number of hashtag context tweets (x-axis).

Ablation Study. To examine the relative contributions of personalized topic attention and joint training, Table 3 shows the ablation study results.

We observe that simple concatenation of latent topics (⊕ TOPICS) can already exhibit the performance gain compared with LSTM-ATT (the SOTA without topics). This again demonstrates the usefulness of latent topics, which can signal topic words and enable the better modeling of hashtag factors.

Our variants	MAP	P@5	nDCG@5
LSTM-ATT	0.2765	0.1043	0.2873
⊕ TOPICS	0.2912	0.1063	0.2993
W/O JOINT	0.2947	0.1092	0.3067
FULL MODEL	**0.3114**	**0.1148**	**0.3257**

Table 3: Comparison of our FULL MODEL to the SOTA LSTM-ATT (Zeng et al., 2020) without latent topic modeling and its ablations — ⊕ TOPICS: joint training with simple topic concatenation and without personalized topic attention; W/O JOINT: personalized topic attention based on pre-trained topics.

Moreover, employing personalized topic attention can further boost the prediction results. This implies that latent topics can enable the user-aware attention to focus on keypoints in hashtag contexts that indicate the potential match with user interests.

In addition, joint training is crucial for the personalized topic attention to capture salient contents, which is seen from the much better results from FULL MODEL than W/O JOINT. It is probably because hashtag content change over time will result in different topics of the test hashtag contexts from those used for training user hashtagging preferences; joint modeling of latent topics and user-hashtag engagements shows its benefits to mitigate such topic gap.

Case Study. To further interpret how our personalized topic attention learns to characterize a user's hashtagging preferences, we take the user U and

hashtag H (#book) in Figure 1 as an example to interpret our learning results. Recall that U loves mystery books and engaged with H in a future tweets, though H did not appear in U's history.
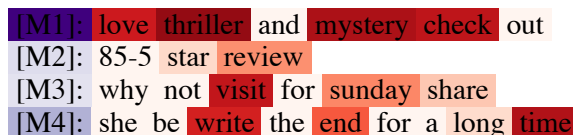

Figure 6: The visualization of personalized topic attention learned for H (#book) and U in Figure 1 over the example hashtag context tweets M_1 - M_4 (on the left in purple) and the topic likelihood assigned to each word (on the right in red). Darker color, higher weights.

Figure 6 shows the heatmap visualizing the attention weights over H's messages and the topic likelihood over their words. Personalized topic attention gives much higher weight to the first tweet M_1 compared others (M_2 to M_4) to indicate M_1's strong effects to predict H's potential to draw U's engagement. We also find that our NTM are able to highlight topic words with higher topic likelihood, e.g., "thriller", "mystery", and "write". This observation shows our capability to learn meaningful topic features to represent "#book".

5.4 Future Hashtagging Behavior Prediction

As discussed above, joint exploration of neural topics and user hashtagging preferences may enable better prediction of users' future hashtagging behavior. For an in-depth analysis, we first select the tweets in test, where the engagements of their hashtags and authors are correctly predicted by an attention-based model, and then measure their cosine similarity with the tweets in hashtag contexts assigned highest attention weights (queried by the user). Here we adopt our full model and the SOTA comparison LSTM-ATT, both employing user-aware attention, and show the test tweet frequency over similarity measures in Figure 7.

As can be seen, our model is able to predict more correct engagement attributed to its better ability to capture tweets in noisy hashtag contexts which exhibit more similar semantics to the future tweets presenting the successful user-hashtag engagements. This suggests our potential to not only personalize future hashtags but also predict their possible contexts by aligning the hashtagging behavior learned from the past to fit the future.
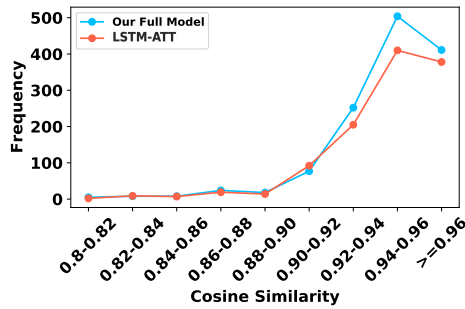
Figure 7: Y-axis shows frequency of test tweets whose user-hashtag engagements are correctly predicted by our full model (blue) and the LSTM-ATT (red). X-axis indicates the cosine similarity of BERT embeddings of the test tweet and the one in its hashtag contexts given the highest attention weights.

6 Conclusion

This paper has studied the learning of user hashtagging preferences from history data to predict their future trajectory. Neural topic model is jointly trained with the prediction of user-hashtag engagements in a novel personalized topic attention. Experimental results demonstrate our effectiveness benefited from the ability to align user hashtagging interests gained from the history to customize their future behavior in the future.

Acknowledgements

Ethical Considerations

We declare our dataset will cause no ethics problem. First, we followed the standard data acquisition process regularized by Twitter official API, where the data was downloaded for a purpose of academic research and is consistent with the Twitter terms of use. Second, we thoroughly navigated the data and ensured that no content will rise any ethics concerns, e.g. toxic languages, hate speeches, user privacy and data anonymization was conducted on the dataset (usernames were removed and IDs were newly and randomly assigned). Third, we only keep the posts with English text and all the tweets are posted in the year of 2013, 8 years ago.

References

Balsam Alkouz and Zaher Al Aghbari. 2020. Detection and visualization of bilingual trending topics. *Journal of Advances in Information Technology Vol*, 11(2).

Steven Bird. 2006. NLTK: the natural language toolkit. In *ACL 2006, 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, Sydney, Australia, 17-21 July 2006*. The Association for Computer Linguistics.

David M Blei, Alp Kucukelbir, and Jon D McAuliffe. 2017. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2001. Latent dirichlet allocation. In *Advances in Neural Information Processing Systems 14 [Neural Information Processing Systems: Natural and Synthetic, NIPS 2001, December 3-8, 2001, Vancouver, British Columbia, Canada]*, pages 601–608. MIT Press.

Kailong Chen, Tianqi Chen, Guoqing Zheng, Ou Jin, Enpeng Yao, and Yong Yu. 2012. Collaborative personalized tweet recommendation. In *The 35th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '12, Portland, OR, USA, August 12-16, 2012*, pages 661–670. ACM.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Zhuoye Ding, Qi Zhang, and Xuanjing Huang. 2012. Automatic hashtag recommendation for microblogs using topic-specific translation model. In *COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference: Posters, 8-15 December 2012, Mumbai, India*, pages 265–274. Indian Institute of Technology Bombay.

Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Yeyun Gong, Qi Zhang, and Xuanjing Huang. 2015. Hashtag recommendation using dirichlet process mixture models incorporating types of hashtags. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 401–410. The Association for Computational Linguistics.

Yuyun Gong and Qi Zhang. 2016. Hashtag recommendation using attention-based convolutional neural network. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pages 2782–2788. IJCAI/AAAI Press.

Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, pages 173–182. ACM.

Haoran Huang, Qi Zhang, Yeyun Gong, and Xuanjing Huang. 2016. Hashtag recommendation using end-to-end memory networks with hierarchical attention. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 943–952. ACL.

Yang Li, Ting Liu, Jing Jiang, and Liang Zhang. 2016. Hashtag recommendation with topical attention-based LSTM. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 3019–3029. ACL.

Zhongqi Lu, Zhicheng Dou, Jianxun Lian, Xing Xie, and Qiang Yang. 2015. Content-based collaborative filtering for news topic recommendation. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA*, pages 217–223. AAAI Press.

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, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 2410–2419. PMLR.

Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. 2020. Bertweet: A pre-trained language model for english tweets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, EMNLP 2020 - Demos, Online, November 16-20, 2020*, pages 9–14. Association for Computational Linguistics.

Yue Wang, Jing Li, Irwin King, Michael R. Lyu, and Shuming Shi. 2019. Microblog hashtag generation via encoding conversation contexts. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 1624–1633. Association for Computational Linguistics.

Liang Wu, Xia Hu, and Huan Liu. 2017. Early identification of personalized trending topics in microblogging. In *Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM 2017, Montréal, Québec, Canada, May 15-18, 2017*, pages 692–695. AAAI Press.

Yang Yu, Xiaojun Wan, and Xinjie Zhou. 2016. User embedding for scholarly microblog recommendation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 2: Short Papers*. The Association for Computer Linguistics.

Xingshan Zeng, Jing Li, Lu Wang, Nicholas Beauchamp, Sarah Shugars, and Kam-Fai Wong. 2018. Microblog conversation recommendation via joint modeling of topics and discourse. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pages 375–385. Association for Computational Linguistics.

Xingshan Zeng, Jing Li, Lu Wang, Zhiming Mao, and Kam-Fai Wong. 2020. Dynamic online conversation recommendation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 3331–3341. Association for Computational Linguistics.

Xingshan Zeng, Jing Li, Lu Wang, and Kam-Fai Wong. 2019. Neural conversation recommendation with online interaction modeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 4632–4642. Association for Computational Linguistics.

Qi Zhang, Yeyun Gong, Xuyang Sun, and Xuanjing Huang. 2014. Time-aware personalized hashtag recommendation on social media. In *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*, pages 203–212. ACL.

Feng Zhao, Yajun Zhu, Hai Jin, and Laurence T. Yang. 2016. A personalized hashtag recommendation approach using lda-based topic model in microblog environment. *Future Gener. Comput. Syst.*, 65:196–206.