

# Dynamic User Attribute Discovery on Social Media

Xiu Huang, Yang Yang<sup>(✉)</sup>, Yue Hu, Fumin Shen, and Jie Shao

School of Computer Science and Engineering,  
University of Electronic Science and Technology of China, Chengdu, China  
{huangxiu,huyue}@std.uestc.edu.cn, dlyyang@gmail.com,  
fumin.shen@gmail.com, shaojie@uestc.edu.cn

**Abstract.** Social media service defines a new paradigm of people communicating, self-expressing and sharing on the Web. Users in today's social media platforms often post contents, inferring their interests/attributes, which are significant for many Web services such as social recommendation, personalized searching and online advertising. User attributes are temporally dynamic along with internal interest changing and external influence. Based on topic modeling, we present a probabilistic method for dynamic user attribute discovery. Our method automatically detects user attributes and models the dynamics using time windows and decay function, thereby facilitating more accurate recommendation. Evaluation on a Sina Weibo dataset shows the superiority in terms of precision, recall and F-measure as compared to baselines, such as static user attribute modeling.

**Keywords:** Dynamic user attribute · Topic model · Time window

## 1 Introduction

In recent years, we have witnessed dramatic growth of social media services such as Twitter<sup>1</sup> and Pinterest<sup>2</sup>, where people can publish, share and consume instant information. In China, as one of the leading microblogging service providers, Sina Weibo<sup>3</sup> receives significant attention from research area. Launched by Sina Corporation in August 2009, Sina Weibo has approximately 500 million registered users by December 2012, on which more than 4.6 million users are active on a daily basis, generating 100 million microblogs per day. On Sina Weibo, it allows people to create concise microblogs with a limitation of 140 characters, made up with a mix of Chinese and English characters as well as self-defined hashtags (e.g. # and @) and external URLs. Thus, users' activities in real time from Sina Weibo stream enable us to automatically discover user attributes by dynamically monitoring users' status, which would help us to timely detect and analyze users' opinions, sentiments and preferences.

<sup>1</sup> <http://www.twitter.com/>.

<sup>2</sup> <http://www.pinterest.com/>.

<sup>3</sup> <http://www.weibo.com/>.

Nevertheless, on account of the sparsity and noise of content in short text, diverse and fast changing topics, and large data volume, it is challenging to dynamically discover user attributes. As a result, addressing the specialty and uncertainty of microblogs is crucial for us to analyze changing tendency in user attributes and behaviors. In previous works [5, 9, 13, 20], users attributes or interests were constructed by using language models in a static manner. Actually, as newly-emerging elements keep occurring on the Web, user attributes may be temporally dynamic, i.e., some interests will be out-of-date while others may become popular attributes that are likely to better reflect current user requirements. Therefore, it is necessary to explore user attributes from the very recent Web contents.

In this work, a novel dynamic user attribute model (DUAM) is proposed to overcome the shortcomings of static attribute model. In particular, we leverage a topic model by the name of Biterm Topic Model (BTM) [19], which is capable of addressing the sparsity of content in short text. BTM is an extending model of Latent Dirichlet Allocation (LDA) [6], which generates topics over microblogs through modeling the biterns directly. As defined in [19], a biterm that is an unordered word-pair co-occurring in a short context can model the word co-occurrence patterns, and it also can avoid the problem of sparse patterns through aggregating word co-occurrence patterns so as to discover topics, thus distinguishing from adding external knowledge to content. Besides, since microblogs are normally input as a rapidly growing stream of prohibitively large volume, they require the user attribute model to dynamically update with the continuously arrived new data. Inspired by this analysis, we introduce a decay function over time windows to model the dynamics in user interests. We assume that microblog documents arrive in a batch mode and in our experiments we divide the whole dataset by a fixed time window (e.g., three months). In this way, our proposed model only needs to store a small part of microblog data online, which can be much more efficient than static attribute discovery. From the experiments on real dataset crawled from Sina Weibo, the dynamic property of user attributes can be detected according to DUAM, which outperforms static user attribute models. The major contributions of this work are summarized as follows:

1. We propose a Dynamic User Attribute Model (DUAM) based on a topic model named Biterm Topic Model (BTM), which can effectively address the sparsity of short text content and significantly overcome the shortcoming of static attribute.
2. Our model dynamically establishes topic-attribute mapping by introducing a decay function over time windows, and detects the shift over user attributes based on the microblog stream.
3. We construct a Sina Weibo microblog dataset by manually labeling user attributes. The promising results demonstrate that our proposed approach significantly outperforms static user profiles.

The rest of the paper is organized as follows. Section 2 illustrates the related work about user attribute modeling. In Sect. 3, we discuss how to generate user

attribute dynamically. Section 4 demonstrates the experimental results and evaluation. Finally, conclusions and future work are given in Sect. 5.

## 2 Related Work

In this section, we summarize some related work about user profiling and indicate the differences in relation to our own work.

Researchers have long been interested in studying mining user interests, which are established by extracting users' characteristics and preferences from posted content on social media [10, 12, 17, 18]. Most previous studies attempted to exploit external knowledge (e.g. Wikipedia, DBpedia) for semantic linking to enrich the presentation of microblogging. For instance, Abel et al. [1] analyzed methods for contextualizing Twitter activities via connecting Twitter posts with the related news articles. The proposed method semantically represents individual Twitter activities through extracting from tweets and the related news articles. Lim et al. [11] proposed a method that can automatically classify the relative interests of Twitter users with a weighting in relation to their other interests using information from Wikipedia. Besides, Ding et al. [8] studied user biographies from Twitter to indicate user interests and analyzed the extracted interest tags from biographies to enrich the information of tweets. However, their work relies heavily on the availability of users' biographies. In those works, exploiting external knowledge, which is widely leveraged for enriching semantics of microblogging, is only effective when auxiliary data are in a close correlation to the original data. On the contrary, our method exploits statistics of word co-occurrence in the microblog corpus with no need to infer external knowledge.

Besides, some previous works also attempted to exploit cross-OSN content to extract user interests. Abel et al. [2] studied form-based user profiles in social web services, e.g. Twitter, Facebook, and also investigated tag-based user profiles based on user tagging activities in some other social systems, e.g. Flickr, StumbleUpon and Delicious, in order to explore the benefits of building user profiles between different systems. Ottoni et al. [14] studied behavior and interests of users, whose accounts are associated Twitter with Pinterest. However, the majority of the existing works focus on tackling the sparse and noisy user-generated data, which represent static user attribute, leading to inconsistency with users' actual status.

Our work employs BTM algorithm to address the sparsity of short texts in microblogs, so that generate topics from microblog content. Then, we propose a novel DUAM model to dynamically discover user attributes over time. Yet, there are some prior works exploiting user attributes by extracting topics, which base on LDA-like model utilizing various inference algorithms. For example, Rosen-Zvi et al. [15] presented the author-topic model, which extends Latent Dirichlet Allocation (LDA) and is a generative model for authorship information and documents. Then, Xu et al. [18] proposed a twitter-user model, which is a modified author-topic model by using a latent variable to indicate author's interests, instead of constructing a "bag-of-words" user profile. Bhattacharya et al. [4] proposed a novel mechanism named Labeled LDA, which aims to generate topics of

interest for individual users on Twitter. Overall, Our work distinguishes from the above researches in that we focus on exploiting user dynamic attribute by automatically modeling topic-attribute mapping in time windows, thus overcoming the shortcoming of static user attribute.

### 3 The Proposed Approach

In this section, we first present the overall framework of dynamic user attribute discovery. Then, we elaborate how to utilize the Biterm Topic Model (BTM) to extract topics of user attributes. Finally, we formally present the algorithm of inferring Dynamic User Attribute Model (DUAM) in detail.

As depicted in Fig. 1, in the data collection process, we crawl microblogs of randomly selected users according to several different topics from Sina Weibo. After filtering out noisy microblogs, we conduct the following preprocessing steps: (1) We remove links from the microblogs; (2) We eliminate non-Chinese characters and self-defined characters (e.g. “@”); (3) We segment the crawled microblog documents into words; and (4) We remove stop words and non-sense words of high frequency in the microblogs. Subsequently, we leverage the BTM model to extract topics of attributes from the users’ microblogs and employ DUAM to dynamically learn attributes for general users, through which we can capture the changing tendency of attributes from an individual user.

#### 3.1 Biterm Topic Model

Biterm Topic Model (BTM) is a probabilistic topic model extending LDA. The underlying idea in BTM is that if two word co-occurrences appear more frequently in the same microblog document, there is a better chance for them to belong to the same topic. As defined in [19], a biterm denotes an unordered

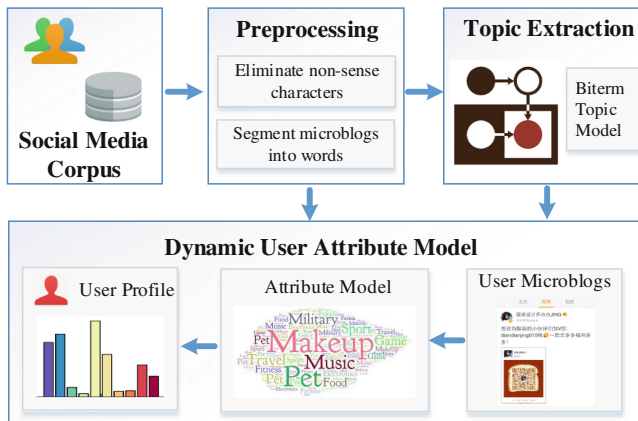


Fig. 1. The overall framework of dynamic user attribute discovery model.

word pair, which is composed of any two different words in a microblog document. For instance, if there are three distinct words in a microblog document, we can generate three biterns:

$$(w_1, w_2, w_3) \Rightarrow \{(w_1, w_2), (w_2, w_3), (w_1, w_3)\}.$$

where  $(\cdot, \cdot)$  is unordered combination. BTM considers the whole corpus of microblogs as a mixture of attributes (topics), where any pair of words are drawn from a specific topic independently and a topic submits to the topic mixture distribution over the whole microblog corpus. Particularly, the topics extracted from microblogs of a specific user indicate the user attributes, characteristics and preferences.

Suppose single-valued hyperparameters  $\alpha$  and  $\beta$  are Dirichlet priors for  $\theta$  and  $\phi_k$ , respectively. The specific generating process of BTM can be described as below:

1. Draw topic proportions  $\theta \sim \text{Dirichlet}(\alpha)$ ;
2. For each topic  $k$ , where  $k = 1, 2, \dots, K$   
draw word probability  $\phi_k \sim \text{Dirichlet}(\beta)$ ;
3. For each bitern  $b_i \in B$   
draw topic  $z_i \sim \text{Multinomial}(\theta)$ , and  
draw bitern  $(w_{i,1}, w_{i,2}) \sim \text{Multinomial}(\phi_{z_i})$ ;

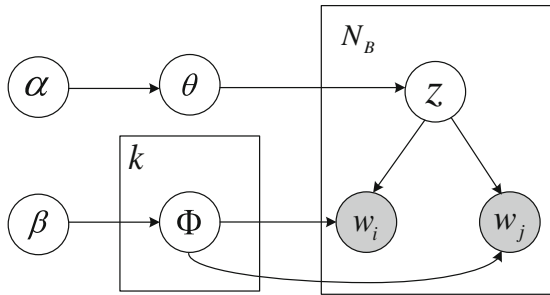


Fig. 2. The graphical presentation of Bitern Topic Model.

Figure 2 shows BTM graphical representation. Given the single-valued hyperparameters  $\alpha$  and  $\beta$ , the joint probability distribution of a bitern  $b_i = (w_{i,1}, w_{i,2})$  can be written as:

$$P(b_i|\alpha, \beta) = \sum_{k=1}^K P(w_{i,1}, w_{i,2}, z_i = k|\alpha, \beta) = \int \int \sum_{k=1}^K \theta_k \phi_{k,w_{i,1}} \phi_{k,w_{i,2}} d\theta d\Phi \quad (1)$$

Thus we can get the likelihood in the whole microblog corpus, where  $B = \{b_i\}_{i=1}^{N_B}$ ,  $N_B$  referring the number of biterns in documents.

$$P(B|\alpha, \beta) = \prod_{i=1}^{N_B} \int \int \sum_{k=1}^K \theta_k \phi_{k,w_{i,1}} \phi_{k,w_{i,2}} d\Theta d\Phi \tag{2}$$

We can see that BTM directly uses the word co-occurrence patterns as an unit revealing the latent semantics of attributes (topics), rather than a single word. In addition, BTM assigns a topic for every bitern in order to learn a global topic distribution. We can obtain the topic proportions of a document by follows:

$$P(z_i = k|d) = \sum_{b=b_i}^{N_B} P(z_i = k|b = b_i) P(b = b_i|d) \tag{3}$$

### 3.2 Inference Parameters

To perform approximate inference for  $\Theta$  and  $\Phi$ , we adopt Gibbs sampling, which is introduced detailedly in [3]. For BTM algorithm, in order to infer the topics, we are required to sample the topic assignment  $z$  for each bitern according to its conditional distribution, thus obtaining the following conditional probability:

$$P(z_i = k|z_{-b}, B, \alpha, \beta) \propto (n_{-b,k} + \alpha) \frac{(n_{-b,w_i|k} + \beta)(n_{-b,w_j|k} + \beta)}{[\sum_{w=1}^U (n_{-b,w|k} + \beta) + 1][\sum_{w=1}^U (n_{-b,w|k} + \beta)]} \tag{4}$$

where  $z_{-b}$  means the topic assignments for biterns except  $b$ , and  $n_{-b}$  denotes the number of biterns assignment over topic  $k$  except  $b$ , as well as  $n_{-b,w|k}$  is the number of times for a word  $w$  assignment over topic  $k$  excluding  $b$ .

Finally, the counts of the topic assignments of biterns and word occurrences are used to infer the distributions  $\Phi$  of topic-word and global topic distribution  $\Theta$  as follows:

$$\Phi_{k,w} = \frac{n_{w|k} + \beta}{\sum_{w=1}^U n_{w|k} + U\beta} \tag{5}$$

$$\Theta_k = \frac{n_k + \alpha}{N_B + K\alpha} \tag{6}$$

### 3.3 Dynamic User Attribute Model

DUAM model keeps user attribute dynamic where new microblog data arrive continuously. We assume microblog documents are divided by time windows, therefore, we can dynamically generate user attribute using topic-attribute mapping in time windows. By leveraging BTM algorithm, we can obtain  $m$  topics/attributes presenting  $m$  clusters of microblogs. At the same time, we use top- $k$  method to choose  $n$  words with highest scores as the representative keywords of a topic. The top  $n$  words' scores are presented by a  $n$ -dimension vector

$s = [s_1, \dots, s_n]$ . Relevantly, a user's attribute is represented by an  $m$ -dimension vector  $a(u, t)$ . Here, in time window  $t$ , our model generates attribute vector  $a(u, t)$  for user  $u$  by using obtained attributes in the previous time windows. Therefore, we utilize both previous data and new data to produce user attributes, which can better reflect up-to-date preferences of users.

For general users, after data preprocessing, we calculate the frequency of keywords  $f_m$  corresponding to the  $m$ th topic. Then, the product between  $f_{mi}$ , the frequency of  $i$ th keyword in topic  $m$ , and  $s_{mi}$ , the weight of  $i$ th keyword in topic  $k$ , denotes the weight of the topic  $m$  for a single user.

$$a_m = \sum_{i=1}^k f_{mi} s_{mi} \quad (7)$$

After normalizing  $a_m$ , we can obtain the relative value  $a'_m = a_m / \sum_{i=1}^m a_i$ , which is used to decide whether a user have the topic/attribute.

During the topic-attribute mapping, we set a threshold  $\theta$  for mapping rules, i.e., if the matching number is larger than the given threshold  $\theta$ , we consider that the topic successfully maps to user attribute.

After topic-attribute mapping, users' static attributes during various periods of time is obtained. However, the obtained attributes are restricted to the specific time, and the long-term attributes that do not appear in the specific time are ignored. User attributes have a continuous changing tendency, which attributes in every period of time have mutual connections. Inspired by [7, 16], in order to invoke user attribute in last time window, we define a decay coefficient  $0 < \lambda < 1$  to infer the influence of prior attribute as follows:

$$\lambda(t) = 1 - \mu t^v, \quad (8)$$

where  $0 < \mu < 1$  and  $v > 0$  are two decay parameters.  $a(u, t_i)$ , a  $m$ -dimension vector, indicates  $m$  attributes in time  $t_i$ . Then, we estimate the user attribute vector  $a(u, t_i)$  as below:

$$a(u, t_i) = \sum_{j=1}^i \lambda(t_j) a(u, t_j). \quad (9)$$

Here,  $a(u, t_0) = [a_{1_0}, a_{2_0}, \dots, a_{10_0}]$

Thus, we can obtain the latest attributes revealing users' current status, and by analyzing changing tendency of user attribute over time windows, we are able to make a prediction on users attributes in the near future.

## 4 Experiments

In this section, the collecting process of our experimental dataset from Sina Weibo is introduced firstly. Then, we demonstrate how to implement our experiment in details and give the evaluation metric. Finally, compared with static method with no time windows, we illustrate the effectiveness of our proposed DUAM model.

## 4.1 Experimental Dataset

We build up our dataset via crawling information streams over a one-year period from January to December, 2015, published by randomly selected 2100 users about 640 000 microblogs from Sina Weibo. In order to get sufficient text data, we filter out those who posted less than 200 microblogs. As our work aims to model the dynamic attribute on Sina Weibo for a single user, we employ a simple and effective means of user-selection through randomly selecting 100 active users who generate more than 5000 microblogs as the training set to obtain topics of attributes. Then, we also randomly select 100 active users to evaluate the experimental performance compared with static method.

After data collection and noise removal, we take the following preprocessing steps: (1) We remove links from the micrologs; (2) We eliminate non-Chinese characters and self-defined characters (e.g. “@”); (3) We segment the crawled microblog documents into words; and 4) We remove stop words and non-sense words of high frequency in the micrologs.

## 4.2 Implementation Details

In BTM algorithm, we set  $\alpha = 50/K$ , and  $\beta = 0.01$  empirically and use training dataset to generate  $m = 10$  topics of attributes, including  $n = 20$  top words as keywords in each topic. In such case, we can obtain 10 topics over the whole microblog document and the keywords are with scores  $s_{mi}$  over the corresponding topic. Showed in Table 1, we present top 5 keywords with scores for every topic. We can see that the keywords are closely related to the corresponding topics. In DUAM, we divide the microblog documents by 3 months as a time window and empirically set  $\mu = 0.56$ ,  $v = 0.06$  separately.

To show the effectiveness of DUAM, we compare it with the static method without time windows. For the static method, we conduct the experiment with fixed time in the first three months to generate user static attributes. While, in our proposed model, we slide the time window and leverage decay function to generate fresh attributes of users according to new data arriving continuously.

**Table 1.** 10 attributes and corresponding top 5 keywords extracted by BTM.

健身	动作(0.0231)	健身(0.0083)	马甲(0.0075)	教程(0.0072)	瑜伽(0.0068)
美食	美食(0.0171)	做法(0.0097)	吃货(0.0079)	美味(0.0065)	好吃(0.0064)
数码	手机(0.0174)	苹果(0.0083)	小米(0.0076)	摄像头(0.0068)	处理器(0.0063)
体育	篮球(0.0101)	曼联(0.0100)	比赛(0.0074)	范加尔(0.0068)	科比(0.0067)
美妆	化妆(0.0173)	教程(0.0173)	妆容(0.0104)	美妆(0.0074)	技巧(0.0061)
旅游	旅行(0.0224)	攻略(0.01881)	旅途(0.0102)	见闻(0.0060)	旅游(0.0057)
军事	中国(0.0096)	航母(0.0060)	美国(0.0057)	海军(0.0048)	南海(0.0045)
音乐	一首歌(0.0140)	音乐厅(0.0127)	翻唱(0.0104)	私人(0.0077)	演唱会(0.0077)
萌宠	主人(0.0110)	狗狗(0.0105)	汪星(0.0070)	主子(0.0062)	猫咪(0.0043)
游戏	游戏(0.0295)	玩家(0.0106)	发售(0.0067)	本作(0.0065)	地址(0.0052)



### 4.3 Evaluation Metric

In the dataset collection process, we also crawl the tags labeled by the authors themselves. Due to the crawled tags, which are lack of complete information, we manually annotate 100 users' tags. Similar to clustering, we define a topic of attribute as a cluster C. To evaluate the experimental result, we compare the obtained attributes from each microblog document with that provided both by authors and the manual work. The clustering performance is measured by frequently used evaluation methods, precision, recall and F-measure analysis.

Here, F-measure is the average value of recall and precision and used in our experimental evaluation as a measure of accuracy. Higher F-measure value reflects the algorithm is better. Higher the precision implies better quality of the algorithm in prediction as recall indicates quantitative analysis.

### 4.4 Experimental Results

We employ the proposed DUAM to generate dynamic user attribute. Based on BTM algorithm, we can obtain 10 attributes. Figure 3 shows quite different set of attributes for each of the randomly selected 3 users. Apparently, user 1 is most interested in makeup and food, and user 2 has an affection on fitness, while user 3 shows special preference to military. Due to the diversity of different users, we analyze a randomly selected user to see the changing tendency of user's attributes in a time span. As displayed in Fig. 4, we can see that in different time period the user has different preferences, which represent long-term attributes and short-term attributes. As we can see, the user scarcely has interest in sport, electronics, military and music. However, there is a rising trend for him or her on food and travel.

Compared with the static method, the precision, recall and F-measure values of the proposed DUAM based on Sina Weibo are showed in Table 2. As we can see there is a significant increase in precision, recall about 8.9% and 18.1%

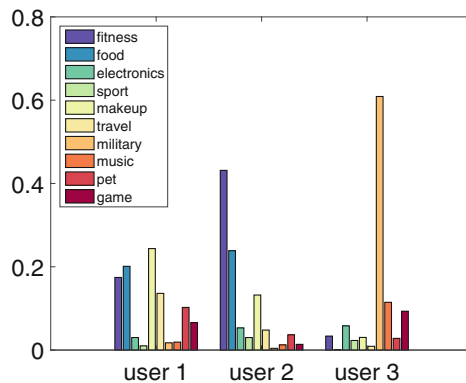


Fig. 3. The distribution of attributes from 3 randomly selected users.

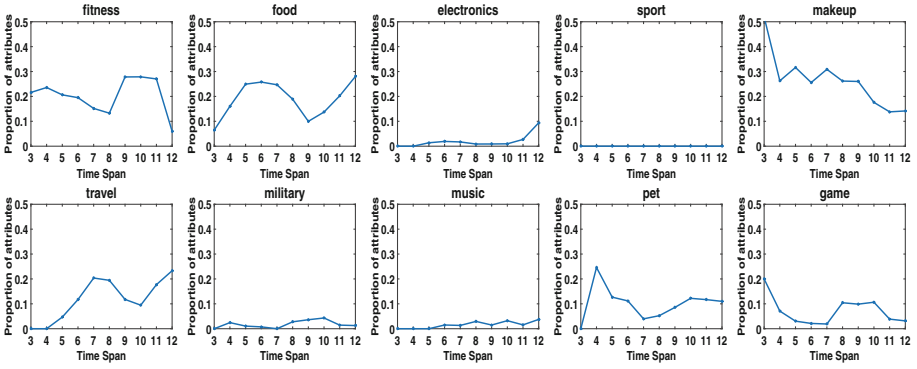


Fig. 4. The changes of a randomly selected user over 10 attributes.

Table 2. Precision, recall and F-measure.

	Precision	Recall	F-measure
the static method	0.4412	0.6317	0.5195
DUAM	0.5300	0.8125	0.6415

separately on the DUAM over time windows. The average value of precision and recall is also higher on DUAM. Accordingly, the result of DUAM is better for us to predict user attribute in the near future and consequently deliver personalized recommendation in line with users’ current preferences.

Obviously, different thresholds exert great influence on our model to generate attributes in accordance with users. To visually evaluate performance in the proposed DUAM and the static method, we utilize ROC curve which is typically used to evaluate binary classifier output quality and can also be applied to assess our model.

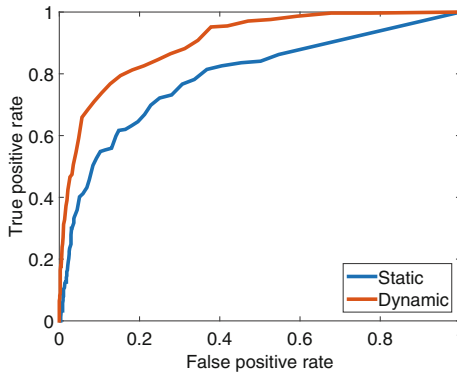


Fig. 5. ROC curve of the static method and DUAM.

The accuracy is denoted as the area under the ROC curve (AUC), which is ranging from 0 to 1. As presented in Fig. 5, the larger AUC implies the more accurate prediction. Hence, our proposed model significantly outperforms the static method.

## 5 Conclusions and Future Work

In this paper, we targeted at dynamically discovering user attribute on social media service. Based on Biterm Topic Model (BTM), we proposed a novel Dynamic User Attribute Model (DUAM) to analyze changing tendency of user attribute on Sina Weibo. As compared with the static method, which presents user attribute in a static process, our proposed model leveraged time windows and a decay function to describe fresh attributes that better meet user current demands. Extensive experiments on our crawled dataset from Sina Weibo showed the effectiveness of our model.

In future work, we will further research the problem with multi-data sources adding images or short videos. For our initial exploration, we only focus on the content-based data. However, in order to set up a real application running on a variety of social media platforms, we should further investigate how to automatically discover user attribute through user-generated contents, images and short videos. On all accounts, our proposed method has a great potential to stimulate future research in social network.

## References

1. Abel, F., Gao, Q., Houben, G.-J., Tao, K.: Semantic enrichment of Twitter posts for user profile construction on the social web. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., Leenheer, P., Pan, J. (eds.) ESWC 2011, Part II. LNCS, vol. 6644, pp. 375–389. Springer, Heidelberg (2011)
2. Abel, F., Herder, E., Houben, G.J., Henze, N., Krause, D.: Cross-system user modeling and personalization on the social web. *User Model. User-Adap. Inter.* **23**(2–3), 169–209 (2013)
3. Asuncion, A., Welling, M., Smyth, P., Teh, Y.W.: On smoothing and inference for topic models. In: *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. pp. 27–34. AUAI Press (2009)
4. Bhattacharya, P., Zafar, M.B., Ganguly, N., Ghosh, S., Gummadi, K.P.: Inferring user interests in the Twitter social network. In: *Proceedings of the 8th ACM Conference on Recommender Systems*, pp. 357–360. ACM (2014)
5. Bian, J., Yang, Y., Chua, T.S.: Multimedia summarization for trending topics in microblogs. In: *Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management*, pp. 1807–1812. ACM (2013)
6. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003)
7. Chen, J., Wang, C., Wang, J.: A personalized interest-forgetting Markov model for recommendations. In: *Twenty-Ninth AAAI Conference on Artificial Intelligence* (2015)

8. Ding, Y., Jiang, J.: Extracting interest tags from Twitter user biographies. In: Jaafar, A., Mohamad Ali, N., Mohd Noah, S.A., Smeaton, A.F., Bruza, P., Bakar, Z.A., Jamil, N., Sembok, T.M.T. (eds.) AIRS 2014. LNCS, vol. 8870, pp. 268–279. Springer, Heidelberg (2014)
9. Gao, Q., Abel, F., Houben, G.J., Tao, K.: Interweaving trend and user modeling for personalized news recommendation. In: Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, vol. 1, pp. 100–103. IEEE Computer Society (2011)
10. Geng, X., Zhang, H., Song, Z., Yang, Y., Luan, H., Chua, T.S.: One of a kind: user profiling by social curation. In: Proceedings of the ACM International Conference on Multimedia, pp. 567–576. ACM (2014)
11. Lim, K.H., Datta, A.: Interest classification of Twitter users using Wikipedia. In: Proceedings of the 9th International Symposium on Open Collaboration, p. 22. ACM (2013)
12. Michelson, M., Macskassy, S.A.: Discovering users' topics of interest on Twitter: a first look. In: Proceedings of the Fourth Workshop on Analytics for Noisy Unstructured Text Data, pp. 73–80. ACM (2010)
13. He, W., Liu, H., He, J., Tang, S., Du, X.: Extracting interest tags for non-famous users in social network. In: Proceedings of the 24th ACM International Conference on Information and Knowledge Management, pp. 861–870. ACM (2015)
14. Ottoni, R., Las Casas, D.B., Pesce, J.P., Meira Jr., W., Wilson, C., Mislove, A., Almeida, V.: Of pins and tweets: investigating how users behave across image-and text-based social networks (2014)
15. Rosen-Zvi, M., Griffiths, T., Steyvers, M., Smyth, P.: The author-topic model for authors and documents. In: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, pp. 487–494. AUAI Press (2004)
16. Sen, W., Xiaonan, Z., Yunnan, D.: A collaborative filtering recommender system integrated with interest drift based on forgetting function. *Int. J. u- and e- Serv. Sci. Technol.* **8**(4), 247–264 (2015)
17. Wang, T., Liu, H., He, J., Du, X.: Mining user interests from information sharing behaviors in social media. In: Pei, J., Tseng, V.S., Cao, L., Motoda, H., Xu, G. (eds.) PAKDD 2013, Part II. LNCS, vol. 7819, pp. 85–98. Springer, Heidelberg (2013)
18. Xu, Z., Lu, R., Xiang, L., Yang, Q.: Discovering user interest on Twitter with a modified author-topic model. In: 2011 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), vol. 1, pp. 422–429. IEEE (2011)
19. Yan, X., Guo, J., Lan, Y., Cheng, X.: A biterm topic model for short texts. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 1445–1456. International World Wide Web Conferences Steering Committee (2013)
20. Yin, H., Cui, B., Chen, L., Hu, Z., Huang, Z.: A temporal context-aware model for user behavior modeling in social media systems. In: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, pp. 1543–1554. ACM (2014)