

PTHMM: Beyond Single Specific Behavior Prediction

Suncong Zheng, Hongyun Bao, Guanhua Tian, Yufang Wu, Bo Xu, Hongwei Hao
Institute of Automation Chinese Academy of Sciences
Beijing, 100190, China

{suncong.zheng,hongyun.bao,guanhua.tian,yufang.wu,bo.xu,hongwei.hao}@ia.ac.cn

ABSTRACT

Existing works on user behavior analysis mainly focus on modeling a single behavior and predicting whether a user will take an action or not. However, users' behaviors do not always happen in isolation, sometimes, different behaviors may happen simultaneously. Therefore, in this paper, we try to analyze the combination of basic behaviors, called behavioral state here, which can describes users' complex behaviors comprehensively. We propose a model, called Personal Timed Hidden Markov Model (PTHMM), to settle the problem by considering time-interval information of users' behaviors and users' personalization. The experimental result on sina-weibo demonstrates the effectiveness of the model. It also shows that users' behavioral state is affected by their historical behaviors, and the influence of historical behaviors declines with the increasing of historical time.

Categories and Subject Descriptors

H.0 [Information Systems]: General

Keywords

User behavior; User modeling; Microblogging; Sequential model; Social Media

1. INTRODUCTION

With the rise of social media, people are given more opportunities to exhibit different behaviors on internet such as sharing, posting, commenting, and befriending. Among the various successful social media, the microblogging service is one of the most popular social media and has great wealth of both textual and behavioral data. It provides us with a good platform to analyze users' diverse behaviors.

Motivating Example. People on microblogging platform-s can either share information or interact with each other. They can even experience different behaviors simultaneously. For example, Figure 1 illustrates the phenomenon in Sina-Weibo. When user "RainBowSun1" reposts a piece of

tweet from "BBCWorld", he also interacts with his friend "zsc" simultaneously and takes part in the topic of "The Missing Malaysia plane". How should we describe RainBowSun1's behaviors and what is RainBowSun1's real intention? In order to depict RainBowSun1's such complex behaviors, we define the behavioral state as the combination of basic behaviors, which can describes users' complex behaviors comprehensively. However, existing works on user behavior

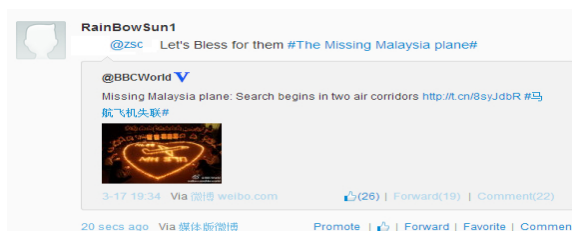


Figure 1: Example of behavioral state (the combination of basic behavioral information)

analysis mainly focus on modeling the specific behavior and predicting whether a user will take an action or not[1, 4, 5]. Different from these works, the task, in this paper, is to analyze users' various combinations of basic behaviors and to predict users' behavioral states in future.

Model Description. Many potential factors affect users' behavioral states, which may be users' habit, current mood, environmental events or others. These factors are cyclical and invisible, but they can trigger users' specific behaviors and are reflected from users' behavioral data. Thus, modeling users' historical behaviors based on a latent variable is beneficial to understand users' behavioral characteristics and predict future behaviors. We call the latent variable as behavioral pattern and it can represent these periodic factors. Besides, users' actions have strong time-relationship with their historical behaviors[4] and everyone has its own personality when using the microblogging service. Motivated by these intuitions, we propose an unsupervised model, Personal Timed Hidden Markov Model (PTHMM), to predict users' behavioral states. We first train a global model called THMM based on all users' behavioral data. When predicting users' behavioral state, we take a personal strategy to fine-tune the global model to make it more adaptable for all kinds of users. Different from classical HMM model, PTHMM not only captures chronological information of behavioral sequences but also considers the time-interval information between two adjacent behaviors and users' personalization.

2. BEHAVIORS ANALYSIS

The task in this paper is to analyze users' complex behaviors and to predict users' behavioral states. Therefore, the first work we need to determine is the basic behaviors selected to represent users' behavioral states in microblogging. The basic behaviors should have following characteristics: 1) **Typical**: All basic behaviors emphasize different aspects of users' behaviors and reflect one of typical characteristics of behavior; 2) **Elementary**: Each basic behavior must be as simple as possible; 3) **Comprehensive**: The combinations of basic behaviors should cover most of behaviors in microblogging.

Java et al.[2] and Zhang et al.[6] have analyzed the characteristics of users' behaviors and classified them into three main categories in microblogging. These categories can be summarized as following: sharing, interaction and mention oneself. The basic behaviors meet the constraint of "typical", when choosing from these three categories.

Sharing Because of the property of "self media", sharing information is the most common behaviors in microblogging platform. When user share information, he can either publish informative tweets or just re-tweet informative tweets from others. Hence, we divide sharing information into two parts: one is posting original informative tweets which reflects users' initiative to share information and the other is reposting other's tweets which reflect user's passive to share information. Besides, we can easily judge whether a user is sharing information or not in microblogging. Because a piece of informative tweet always contains URL, video or picture and it also can attract many users' attention to repost it or comment it.

Interaction Microblogging also has the property of social network. Users in microblogging usually interact with their friends or participate in group topics. Microblogging provides various mechanisms for users to interact with each other. Therefore, based on different kinds of interaction behaviors, we subdivide interaction behaviors into three parts: the first one is to use symbol '@' to attract friends' attention; the second is to repost or comment the tweet of friends; the third is to participate in topic group. Among them, the former two are both interacting with friends: one is initiative to seek friends conversation, the other is just to response to friends. The last one is interacting with a group of people.

Mention oneself Many users like to post tweets to talk about their daily routine or what they are doing and feeling currently in microblogging. A tweet about mentioning oneself is always the plain text generated by the user, and the tweet may also contains the information of location check-in.

However, due to the diversity of users' behaviors in microblogging, the above three categories of behaviors can not cover all kinds of behaviors. Apart from these, we append "the others" category to denote a kind of special behavioral state. We sum up the dataset which contains 4,018,012 tweets and obtain the distribution of the all basic behaviors shown in Table 1. It shows that the three categories cover most behaviors in microblogging and the remainder only occupy 8%, which means the choose behaviors can meet the requirement of comprehensive. Each behavioral state is represented by a 6-dimension vector, and each dimension of the vector represents one of the six behaviors. Accordingly, the behavior state of "the others" is a 6-dimension vector with all dimension equal to zero. Therefore, users' behavioral state in microblogging is fully depicted.

Table 1: Distribution of all basic behaviors.

Basic behavior	Category	Ratio
post information	sharing	6.64%
repost information	sharing	36.26%
seek conversation	interaction	10.48%
response to friends.	interaction	4.63%
participate in group	interaction	12.27
mention oneself	mention oneself	21.72
—	the others	8.00%

3. PERSONAL TIMED HIDDEN MARKOV MODEL

In this section, we formulate the problem of predicting users' behavioral states based on their historical data.

3.1 Notation

For the convenience of description, we firstly give the notations of model elements:

Behavioral state B: A behavioral state is obtained by parsing its corresponding tweet content, and it is denoted as a 6-dimension vector: $B = \{b_1, b_2, \dots, b_6\}$. When $b_i = 1$ means current tweet contains the basic behavior b_i else not.

Users U: The behavioral data of each user is represented by two sequences: one is behavioral state sequence and the other is time sequence accordingly. We denote user m as following form: $U_m = \{U_m(B), U_m(t)\}$, where $U_m(B) = \{B_1, B_2, \dots, B_n\}$ and $U_m(t) = \{t_1, t_2, \dots, t_n\}$, n is the total tweets' number of user m .

Behaviors pattern Z: To describe users' periodic factors that can affect users' behavioral states, we introduce a latent variable named behavioral pattern. The set of behavioral patterns shared by all users is denoted as $Z = \{z_1, z_2, \dots, z_k\}$, where k is the number of behavior patterns.

Model parameters Θ : PTHMM's parameters are the same as HMM. It concludes: transfer probability of behavioral patterns $p(z_j|z_i)$, the relationship between behavioral patterns and behavioral states $P(B_i|z_k)$ and the probability distribution of behavioral pattern $P(Z)$.

3.2 The Framework of PTHMM

The framework of PTHMM is shown in Figure 2, and it contains two main procedures: global training and personal inference. In the stage of global training, we take advantage of all users' behavioral data to train the global model: THMM, which is reflect the general behavioral characteristics that all users shared. However, each person has his own personality. When predicting user m 's state, we take a personal strategy to fine-tune the global model to make it more adaptable for all kinds of users and the personal inference process will be detailed in section 3.2.2.

3.2.1 Training of THMM

In this section, we present the global training procedure of our approach. As all we know, HMM is the classical model for inferring hidden state sequences from observed variables[3]. However users' behaviors have strong relationship with the time interval. Classical HMM only takes sequences' chronological information into consideration, which can't not fully capture the law of users' behavior. Therefore, the proposed THMM can make up this problem by taking time interval information into consideration.

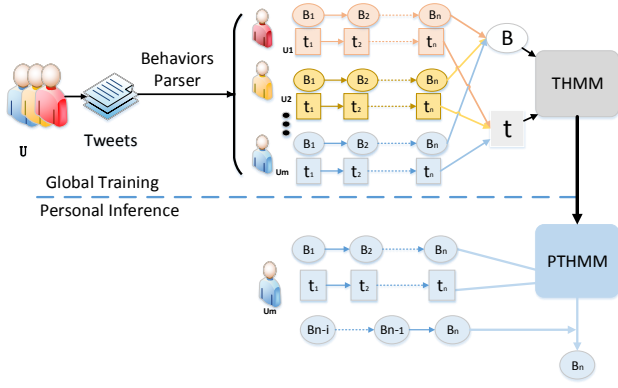


Figure 2: The Framework of PTHMM

We use $\gamma_l(z_k)$ to denote the distribution of behavioral pattern z_k at time l and $\xi_l(z_i, z_j)$ to denote the joint distribution of two successive behavioral patterns, which means behavioral pattern is z_i at time l and z_j at time $l+1$. Similar to the HMM, THMM also can be naturally cast in the framework of Expectation-Maximization(EM), and the difference is about updating $\gamma_l(z_k)$ and $\xi_l(z_i, z_j)$. We take the time interval factor into account. The smaller the two adjacent time interval is, the stronger influence the former behavior has on the later.

E-step: fix Θ , then update $\gamma_l(z_k)$ and $\xi_l(z_i, z_j)$. In order to update the parameters, we take the Baum-Welch algorithm[3] and $\xi_l(z_i, z_j)$ is computed as following:

$$\begin{aligned} \xi_l(z_i, z_j) &= P(z_i^l, z_j^{l+1} | \Theta, U(B), U(t)) \\ &= \frac{P(z_i^l, z_j^{l+1}, U(B) | \Theta, U(t))}{P(U(B) | \Theta, U(t))} \\ &= \frac{P(z_i^l, z_j^{l+1}, U(B) | \Theta, U(t))}{\sum_{k=1}^{k=K} \sum_{j=1}^{j=K} P(z_i^l, z_j^{l+1}, U(B) | \Theta, U(t))} \end{aligned} \quad (1)$$

where

$$\begin{aligned} P(z_i^l, z_j^{l+1}, U(B) | \Theta, U(t)) &= P(B_1, \dots, B_l, z_i^l) \cdot P(z_j | z_i) \\ &\cdot P(B_{l+1} | z_j) \cdot P(B_{l+1} \dots B_n | z_j^{l+1}) \cdot P(t_l, t_{l+1}) \end{aligned} \quad (2)$$

Especially, $P(t_l, t_{l+1})$ capture the influence of time interval by modifying two adjacent behaviors' relationship and we define the time interval factor as :

$$P(t_l, t_{l+1}) = \exp(-(t_{l+1} - t_l) / T_{max}) \quad (3)$$

where T_{max} is the largest time interval in the sequence. After computing $\xi_l(z_i, z_j)$, we just sum up $\xi_l(z_i, z_j)$ to get the variable of $\gamma_l(z_k)$ as following:

$$\gamma_l(z_k) = \sum_{z_i \in Z} \xi_l(z_i, z_k) \quad (4)$$

M-step: fix $\gamma_l(z_k)$ and $\xi_l(z_i, z_j)$, then update Θ . When giving $\gamma_l(z_k)$ and $\xi_l(z_i, z_j)$, the parameter Θ is computed as following :

$$P(z_k) = \gamma_l(z_k) \quad (5)$$

$$P(z_j | z_i) = \frac{\sum_{l=1}^{L-1} \xi_l(z_i, z_j)}{\sum_{l=1}^{L-1} \gamma_l(z_k)} \quad (6)$$

$$P(B_i | z_k) = \frac{\sum_{l=1}^L \gamma_l(k) \times \delta(B_l, B_i)}{\sum_{l=1}^L \gamma_l(k)} \quad (7)$$

3.2.2 Personal Inference

THMM captures the common behavioral rules that all people shared. However, each user has his own personalities. In order to combine the commonness and personalities, we take the personalized strategy to fine-tune the global THMM. When predicting user m 's behavioral state, we use m 's historical behavior data $P(H(U_m(B)))$ to re-estimate the behavioral pattern distribution $P_m(Z)$ of m and behavior patterns transformation law $P_m(z_j^{l+1} | z_i^l)$ based on formula 1 to 6. At last, a user's behavioral state at time $L+1$ can be computed as:

$$B_{L+1} = \underset{B}{\operatorname{argmax}} \sum_{j=1}^K \sum_{i=1}^K P_m(z_i^L) \cdot P_m(z_j^{L+1} | z_i^L) \cdot P(B | z_j^{L+1}) \quad (8)$$

Where $P(z_i^L)$ means the probability of behavioral pattern at time L . It can be calculated by decoding behavior state sequences with the help of viterbi algorithm[3].

4. EXPERIMENTS

4.1 Data Set and Experimental Setup

Dataset We constructed the dataset based on the Sina-Weibo platform, one of the most popular microblogging in China. The data used in this experiment is randomly choose from the ordinary accounts and filtered out the users who post less than 50 tweets. For each user, we crawl all of his tweets which can describe their behavioral track in Sina-Weibo completely. Finally, the dataset consists of 3500 users with 4,018,012 tweets. For evaluation, we randomly choose 500 users as the test set and the left 3000 users as the train set.

Metric The final task is to predict the combinations of basic behaviors. Based on the characteristics of the problem, we focus on the similarity between predicted state and actual state. Therefore, the evaluation metric we used is $acc@w$, where w refers to the wrong number we make and acc means the accuracy of predicted results. For example, $acc@0$ means the precision when we predict all behaviors correctly. Accordingly $acc@1$ means the precision when we miss to predict one of the 6 behaviors. The smaller of w , the more similar between prediction state and actual behavior state. In this paper, we just consider the results of $acc@0$ and $acc@1$, which are more useful and insightful.

Baselines Because of the unsupervised properties of the task, supervised sequential models such as CRF is invalid. To evaluate the performance of PTHMM, we consider three common unsupervised methods as baselines:

TF is the term frequency, which is wildly used in text retrieval. Here, it just account the behavioral state frequency information. We choose the most common behavioral state of user to represent his future behavioral state.

MM is the markov model. Compared with TF, MM considers the chronological information of behavioral sequence.

HMM have been proven to be effective in a variety of sequences applications. Different from the above models, it takes user's hidden behavioral pattern into consideration.

4.2 Prediction performance

The predicted results are reported in Table 2. The sequential models achieve better results than TF that does not consider chronological information. Besides, these models that apply the latent factor of behavioral pattern such as HMM, THMM and PTHMM can bring substantial performance improvement to MM, which neglects the mechanism behind behaviors. THMM outperforms the classical HMM that does not consider the information of time interval. It suggests that time-interval information is an important factor when analyzing users' behaviors. Finally, the PTHMM shows a better result than THMM by considering user's own personality. It means that PTHMM is more adaptable for different kinds of people.

Table 2: The predict results of different methods

	TF	MM	HMM	THMM	PTHMM
acc@0(%)	41.8	48.2	48.8	49.4	50.0
acc@1(%)	57.0	61.2	60.7	61.5	62.0

4.3 Analysis on time correlation

PTHMM is based on the intuition that user's action has a strong time-relationship with his historical behaviors. When predicting behavioral state, we apply the viterbi algorithm to decode the latest historical behavioral sequences. The length of decoded sequence means the number of historical behaviors considered in the model when predicting future behavioral state. Figure 3 shows that both $acc@0$ and $acc@1$ rise with the increasing of predict length and achieve stability at last. The results reflect that users' behavioral state is affected by their historical behaviors, and the influence of historical behaviors declines with the increasing of historical time, then it tend to be stable finally.

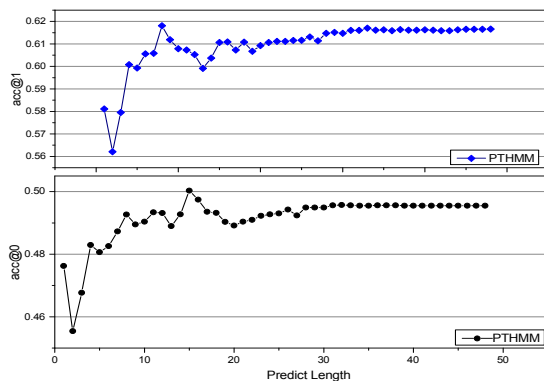


Figure 3: Effect of Predict Length

Besides, we also compute the similarity between current behavioral state and its former N behavioral states. The statistical result, shown in Figure 4, shows that users' current behavioral states are most similar to the nearest behavioral state, and the similarity decreases with the time distance increasing, then achieves stability at last. The result of Figure 4 demonstrates the same behavioral property as Figure 3 does, which also strongly support the assumption of our model.

5. CONCLUSIONS

In this paper, we formally formulate the problem of users' behavioral states, which are the different combinations of

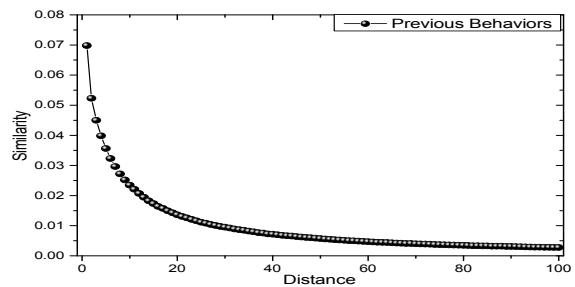


Figure 4: The similarity between historical behavioral states and current behavioral state

basic behavior information. Then we choose six basic behaviors to represent all kinds of behavioral states based on the analysis of users' behaviors in Sina-Weibo. We propose an unsupervised model called PTHMM to settle the problem by considering time-interval information of behaviors and users' personality. The experimental result on sina-weibo demonstrates the effectiveness of PTHMM, and it also shows that users' current behavioral state is affected by their historical behaviors, where the influence of the historical behaviors declines with the increasing of historical time. In future, we plan to examine the scalability of our method and experiment on more large-scale datasets. Besides, we also need further analysis of tweets' textual information to improve the representation of users' behavioral states.

6. ACKNOWLEDGMENTS

This research work was funded by the National Natural Science Foundation of China under Grant No.Y3FE191D61, No.2015AA015402, No.71402178 and No.61303172.

7. REFERENCES

- [1] C. Budak and R. Agrawal. On participation in group chats on twitter. In *Proceedings of the 22nd international conference on World Wide Web*, pages 165–176. International World Wide Web Conferences Steering Committee, 2013.
- [2] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 56–65. ACM, 2007.
- [3] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [4] C. Tan, J. Tang, J. Sun, Q. Lin, and F. Wang. Social action tracking via noise tolerant time-varying factor graphs. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1049–1058. ACM, 2010.
- [5] Z. Xu, Y. Zhang, Y. Wu, and Q. Yang. Modeling user posting behavior on social media. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 545–554. ACM, 2012.
- [6] J. Zhang, Y. Qu, J. Cody, and Y. Wu. A case study of micro-blogging in the enterprise: use, value, and related issues. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 123–132. ACM, 2010.