Temporal Dynamics of User Interests in Tagging Systems

Dawei Yin Liangjie Hong Zhenzhen Xue Brian D. Davison
Department of Computer Science & Engineering, Lehigh University
Bethlehem, PA, USA
{day207, lih307, zhx208, davison}@cse.lehigh.edu

Abstract

Collaborative tagging systems are now deployed extensively to help users share and organize resources. Tag prediction and recommendation systems generally model user behavior as research has shown that accuracy can be significantly improved by modeling users' preferences. However, these preferences are usually treated as constant over time, neglecting the temporal factor within users' interests. On the other hand, little is known about how this factor may influence prediction in social bookmarking systems. In this paper, we investigate the temporal dynamics of user interests in tagging systems and propose a user-tag-specific temporal interests model for tracking users' interests over time. Additionally, we analyze the phenomenon of topic switches in social bookmarking systems, showing that a temporal interests model can benefit from the integration of topic switch detection and that temporal characteristics of social tagging systems are different from traditional concept drift problems. We conduct experiments on three public datasets, demonstrating the importance of personalization and user-tag specialization in tagging systems. Experimental results show that our method can outperform state-of-the-art tag prediction algorithms. We also incorporate our model within existing content-based methods yielding significant improvements in performance.

Introduction

Collaborative tagging systems have become widely used for sharing and organizing resources in recent years. In social bookmarking, as one type of collaborative tagging system, users can add metadata in the form of descriptive terms, called tags, to describe web resources. Social bookmarking systems have been utilized successfully in many areas, such as web search (Bao et al. 2007), personalized search (Xu et al. 2008), web resource classification (Yin et al. 2009) and clustering (Ramage et al. 2009). However, the major bottleneck for large scale applications of social tagging systems is that only a small number of web resources have manually assigned tags, compared to the size of the Web. Therefore, systems that can automatically tag web resources are needed. On the other hand, from the user's perspective, a system that can provide suggestions and recommendations when users are about to assign tags to new resources can improve human-computer interactions and organization of the knowledge base as well.

Motivated by the needs described above, researchers have considered how to build systems to recommend or predict tag usage. Early work in this area, such as Hotho et al. (2006) and Lipczak et al. (2009), has demonstrated two basic approaches—content-based and graph-based methods—to tackle the problem. Recent work, including Rendle et al. (2009) and Yin et al. (2010), show how personalized tag recommenders that take the user's previous tagging behaviors into account usually have better performance. Furthermore, some attempts (e.g., Halpin et al. (2007) and Zhang et al. (2009)) which study the temporal characteristics of tagging systems more broadly, suggest that more frequently and recently used tags should be favored for tag suggestion, due to the fact that users may re-use tags in a short time-frame and the scope of users' interests might change over time. This implies that it is not appropriate to make suggestions simply based on all past data, as most current methods do.

In this paper, we systematically investigate the temporal dynamics of user interests in social tagging systems and propose a novel approach for the tag prediction problem by modeling temporal preferences in a principled manner. Our method stands on techniques introduced to address "concept drift" (Lebanon and Zhao 2008), which imposes a continuous smoothing scheme over the timeline. However, we show that this smoothing scheme may lead to sub-optimal predictions due to the phenomenon that users may suddenly change interests and topics while using social bookmarking systems, as we suggest in Yin et al. (2011). We tackle the problem by explicitly modeling session-like behaviors and incorporate such models into our prediction process.

Our contributions in this paper are as follows: 1) we verify the existence of short-term user interests; 2) we present a novel personalized method to model temporal dynamics of users' interests; 3) our experiments show that our methods (based on personal historical tagging sequences) can outperform state-of-the-art tag predictors in the presence of concept drift; and, 4) by combining with our methods, state-of-the-art algorithms can realize significant improvements in tag prediction quality.

Related Work

Personalized tag recommendation is a recent topic in recommender systems. The two main directions for these systems are content-based approaches and graph-based approaches. Content-based methods, which usually encode users' preferences from textual information (e.g., web pages), can predict tags for new users and new items. One state-of-the-art content-based tag recommendation system (Lipczak et al. 2009) utilized several tag sources including item content and user history to build both profiles for users and tags. New tags are checked against user profiles, which are rich, but imprecise sources of information about user interests. The result is a set of tags related to both the resource and the user. Graph-based approaches such as Rendle et al. (2009), which often have stronger assumptions than content-based ones (e.g., requiring every user, every item and every tag to occur in at least p posts), can provide better performance. Recently, we presented a probabilistic model for personalized tag prediction (Yin et al. 2010) involving three factors—egocentric effects, environmental effects and item content—and demonstrated the use of an online evaluation mode (also adopted in this paper) which is more realistic than traditional evaluation modes of randomly selecting test data and even time-point splitting of test data.

An important factor not considered by any of the above methods is how users' interests change over time. Recent research (Zhang, Mao, and Li 2009) also shows that users are much more likely to use their recently used tags. Zhang et al. investigated the recurrence dynamics of social tagging. In recommender systems and collaborative filtering, temporal information has already shown its success (e.g., Ding and Li (2005), Koren (2009) and Xiang et al. (2010)). However, these methods are for user-item recommendation. It is difficult to directly use these methods on tag prediction because tagging systems are more complex and contain three elements: user, tag and item. Modeling temporal interests is also related to the problem of concept drift which needs to find the balance between varying temporal effects and long-term trends (Schlimmer and Granger 1986; Widmer and Kubat 1996). Lebanon et al. (2008) introduce a local likelihood model for concept drift which weights the local likelihood using a kernel function. Another similar method is the positional language model proposed by Lv and Zhai (2009). In this kind of method, the smoothing actually models the lifetime of users' short-term interests. However, in real life, different users will have different behaviors with respect to short-term interests.

Preliminaries

Data Sets

We use three public datasets. The first is the Bibsonomy dataset of the ECML PKDD 09 Challenge Workshop[1] which includes item content. The remaining two datasets are Delicious and Flickr datasets crawled by Gorlitz et al. (2008)[2]. There is no item content in the Delicious and Flickr datasets while all three contain timestamps. In order to observe the versatility of user interests on three datasets, for each user, we calculate and plot the total number of tags, and the total number of posts. In Figure 1(a), we can see that the three datasets have different properties and users form three clusters. In Bibsonomy, users typically apply a larger variety of tags across fewer posts, suggesting that their interests are more varied. In contrast, the users in Flickr use fewer tags and their interests are more focused, by reusing their tags many times. This implies that it may be easier to track the user interests in Flickr.

Table 1: Fractions of new users, items, or tags in samples from each data set.

	Bibsonomy	Delicious	Flickr
New/Total Users	41/668	16/1000	23/1000
New/Total Items	602/668	712/1000	1000/1000
New/Total Tags	321/2207	181/2920	175/4123

Time-Sensitive Sampling

As in our earlier work in tag prediction (Yin et al. 2010), we employ online evaluation[3] in which only training posts which have earlier timestamps than those of the test posts are used. Note that this implies that the available training data is different for each test post and for items tagged earlier in the timeline, fewer training data are available. While the online evaluation approach naturally fits the real-world case in which every post is used for testing a model trained on all prior posts, its feasibility depends highly on the efficiency of the training method as a new model may be necessary for each post. Instead, we can estimate the performance of the complete system by performing evaluation on only a sample of test posts, and largely avoid model-building efficiency concerns. We use the common F_1-measure as our principal metric.

We utilize the online evaluation model and conduct time-sensitive sampling experiments on three data sets. For the Bibsonomy dataset, we use the same sampling dataset as in Yin et al. (2010) which includes 668 test posts. For Delicious and Flickr, we randomly choose 1000 posts. In all cases we effectively simulate a system running—the tagging system operates in an incremental mode. The data set statistics (shown in Table 1) demonstrate that in Bibsonomy data, we face a new user (a user which is not in any prior data) in 6.1% of the cases, and in 90.1% of the time users are trying to bookmark a "new item" not previously seen by the system. In addition, there is 13.9% chance that users would use new tags (which do not appear in the system before).

This shows that most of the time (i.e., 86.1% of posts) it is feasible to predict tags based only on past tags. The other two datasets also show similar distributions. Thus, in

[1]http://www.kde.cs.uni-kassel.de/ws/dc09/

[2]https://www.uni-koblenz.de/FB4/Institutes/IFI/AGStaab/Research/DataSets/PINTSExperimentsDataSets/

[3]In this paper, online mode means an incremental mode of a real tagging system rather than real-time tag prediction.

the real world, the principal difficulty is to handle cases in which existing users try to tag new items and therefore strictly graph-based recommenders (e.g., (Rendle et al. 2009; Rendle and Schmidt-Thieme 2010)) will not be able to make recommendations most of the time.

The Baselines

Let U be the set of users u, I be the set of items i being tagged, T be the set of tags t and M be the set of timestamps τ. Additionally, S is the set of all records s, representing the relations among the four types of objects, $S \subseteq U \times I \times T \times M$. Each record $(u, i, t, \tau) \in S$ means that user u has tagged an item i with the tag t at time τ. Here, we also define P_s as all the distinct user-item-time combinations: $P_s = \{(u, i, \tau) | \exists t \in T : (u, i, t, \tau) \in S\}$.

Long-Term Interests Model. If we assume that users' interests are not drifting over time, then users' interests can be modeled as long-term interests. We assume that the users' interests—$P(t|u)$ the probability of tags occurring—follows a multinomial distribution, from which the MLE gives us a simple representation of $P_{\tau_p}(t|u) = \frac{\sum_{p' \in P'_u} c(t,p'|u)}{\sum_{t'} \sum_{p' \in P'_u} c(t',p'|u)}$, where where $c(t', p'|u)$ is the number of times that tag t' occurs on post p', and typically users use a tag only once per post. P'_u is the set of u's posts whose timestamps are earlier than the current time. Long-term interest models simply recommend the most frequent tags used in the past.

Short-Term Interests Model. Users' interests may change over time; thus users' recent behaviors can better represent users' current preferences. We model short-term interests using a sliding window which is common in temporal methods. $P_{\tau_p}(t|u)$ will be calculated only based on recent data (e.g., within three days). The size, σ, of the time window corresponds to the lifetime of short-term interests. Based on this Short-Term Interests model, we tune the parameter—the size of the time window.[4] The results are shown in Figure 1(b). We find that in Bibsonomy, the best performance is achieved when $\sigma = 30$ days. Overall, the more recent the data, the more accurate the estimate of users' interests.

Temporal Interests Model

The experiments using the Short-Term Interests Model show that the users' interests are continuous and similar within a time slot. However, the above time window methods may not fit the real case in which their interests are drifting over time, that is, $P_\tau(t|u)$ varies with changing τ. If we assume that the tagging behaviors of different users are independent, then for a specific user, we can only focus on the user's past behaviors. The occurrences of tags $P(t|u)$ can be generated by a multinomial distribution or n-gram extension. We further make the assumption that the lifetimes for different tags are independent. Then, in post p, the tags are generated by a multinomial distribution and from a definite set T. Let $\theta_{t,u}$ refer to $P(t|u)$.

$$P_\tau(p|u) \propto \prod_{t \in T_p} \theta_{t,u}^{c(t,p|u)}$$

[4]Dataset and evaluation are the same as in Yin et al. (2010)

To model the dynamics of users' interests, we use the standard kernel smoothing technique and the likelihood at time τ is smoothed or weighted on users' data D_u by a non-negative smoothing kernel $K : \mathbb{R} \to \mathbb{R}$. By further assuming that the number of tags on posts is independent of t, the local likelihood can be written as

$$\begin{aligned} l_\tau(\eta|D_u) &\stackrel{\text{def}}{=} \sum_{\tau' \in M} K(\tau - \tau') \sum_{p' \in P_{\tau,u}} \log P(p'; \eta) \\ &= \sum_{\tau' \in M} K(\tau - \tau') \sum_{p' \in P_{\tau,u}} \sum_{t \in T_{p'}} c(t, p'|u) \log \eta_t \end{aligned}$$

At each time τ, for user u, the estimation of each θ is derived by maximizing the local likelihood.

$$\hat{\theta}_{\tau,u} = \arg\max_{\eta \in \Theta_u} l_\tau(\eta|D_u)$$

There is a closed form expression for the local likelihood maximizer $\hat{\theta}_{\tau,u}$ which can be obtained by setting the gradient of the Lagrangian to 0.

$$0 = \frac{1}{[\hat{\theta}_{\tau,u}]_t} \sum_{\tau' \in M} K(\tau - \tau') \sum_{p' \in P_{\tau,u}} c(t, p'|u) + \lambda_t$$

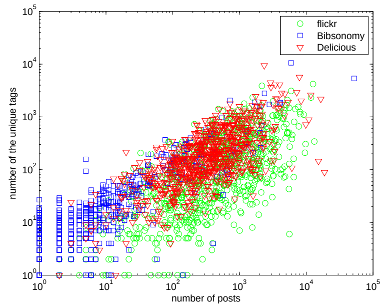
By solving the above equation, we obtain

$$[\hat{\theta}_{\tau,u}]_t = \frac{\sum_{\tau' \in M} K(\tau - \tau') \sum_{p' \in P_{\tau,u}} c(t, p'|u)}{\sum_{\tau' \in M} K(\tau - \tau') \sum_{p' \in P_{\tau,u}} \sum_{t \in T'_p} c(t, p'|u)} \tag{1}$$

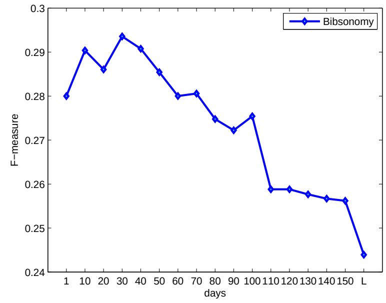
we can see that the present distribution $[\hat{\theta}_{\tau,u}]_t$ is actually the fraction of occurrences weighted by the kernel function. There are several choices for the kernel function (Lv and Zhai 2009; Lebanon and Zhao 2008). Usually, the kernel function is symmetric, like the uniform kernel ($K(\tau) = \mathbf{1}_{\{|\tau|<\sigma\}}$) and the Gaussian kernel. Because our task is to estimate the user's present distribution $[\hat{\theta}_{\tau,u}]_t$ based only on the past data, the kernel is only the right half of the symmetric kernel function and it can be also considered as decaying of interests. The speed of decay measures the probability of the user staying on the same topic over time. Unlike traditional approaches to concept drift which try to track global trends across the whole dataset and use a fixed kernel function, a very essential problem in social tagging systems is personalization. In particular, different users may have different decay speeds for short-term interests. Even for the same user, the behaviors on different tags are different. Thus, we propose a personalized method and moreover a personalized tag-specific model. It is more reasonable to model the problem as tag lifetime rather than as a simple kernel smoothing problem. Intuitively, once an interest appears, it will stay for a while and then become weaker and weaker. Assuming that the lifetime of the short-term interests follow the exponential distribution, then at time τ_i, the topics emerged and the probability of interests still staying at time τ_j is $P_\tau(t \ \texttt{stay}|u) = \int_\tau^\infty \frac{1}{\sigma_{u,t}} e^{-\tau'/\sigma_{u,t}} = e^{-\tau/\sigma_{u,t}}$. Using this equation as the kernel function results in:

$$K_t(\tau|u) = e^{-\tau/\sigma_{u,t}} \tag{2}$$

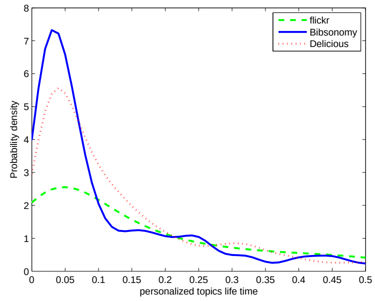
where $\sigma_{u,t}$ is the user-tag specific parameter. For each user-tag pair, there will be a specific $\sigma_{u,t}$ to control the decay of this tag for the user. Equation 1 can be interpreted



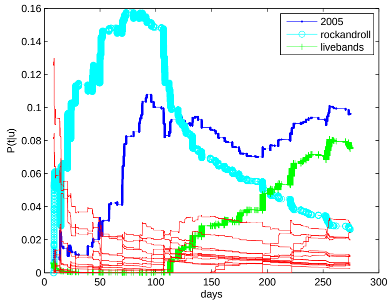
(a) The number of unique tags as a function of the number of posts for each user



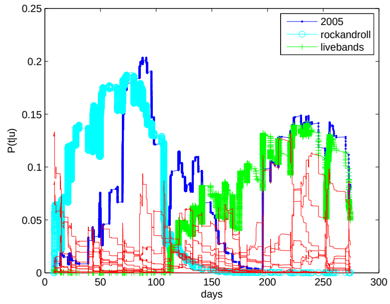
(b) F-measure as a function of the time window size



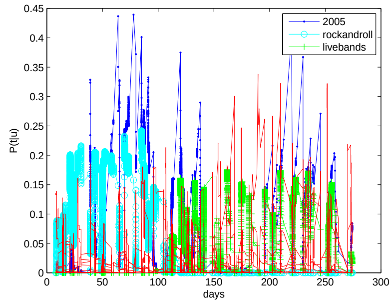
(c) Estimated probability density of personalized topic lifetime



(d) Tracking users interests $\epsilon = 100$



(e) Tracking users interests $\epsilon = 10$



(f) Tracking users interests $\epsilon = 1$

Figure 1: Experimental Analysis

as the fraction of the weighted remaining interests. If we make the assumption that the same user has the same decay and lifetime distribution on all different tags, that is, $\sigma_u = \sigma_{u,t_1} = \sigma_{u,t_2} \cdots = \sigma_{u,t_i}$, the model becomes a personalized model.

For simplicity, we rewrite Equation 1 as follows.

$$P_{\tau_p}(t|u) = \frac{c'(t,p|u)}{\sum_t' c'(t',p|u)}$$

where $c'(t,p|u) = \sum_{p' \in P_u, p' \neq p, \tau_p \geq \tau_{p'}} c(t,p'|u) K_t(\tau_p - \tau_{p'}|u)$ and τ_p is the timestamp on post p. The problems is how to estimate the parameters $\sigma_{u,t}$.

Estimation of parameters

From the assumption that the lifetime of the short-term interests follow the exponential distribution, we know that $\sigma_{u,t}$ is the mean lifetime of tag t for user u. We will consider a continuous sequence of posts where user u keeps using t as an event of topic t occurring.

Formally, for user u, let $p_1, p_2, ..., p_i, ..., p_n$ represent the sequence of u's posts in chronological order so far. Their timestamps $\tau_1 \leq \tau_2 \leq ...\tau_i \leq ... \leq \tau_n$. Let $s = (p_i, p_{i+1}...p_{j-1}, p_j)$ be a subsequence with maximum length where all posts contain tag t. At time τ_i, user u starts to use tag t, and at τ_j, and the user stops using tag t. In this event, the lifetime of tag t is $\tau_s = \tau_j - \tau_i$. Let $S_{u,t}$ represent the set of all such subsequences of tag t for user u. The parameter σ can be estimated as $\hat{\sigma}_{u,t} = \frac{1}{|S_{u,t}|} \sum_{s \in S_{u,t}} \tau_s$. It is consistent with the intuition that in the past, once user u starts to be interested in tag t and the interest always stays for a long while, then recent use of tag t will hold a strong signal that t will be used again. However, the above estimation may cause too much emphasis on personalization and so smoothing and controlling the weight of personalization are required:

$$\hat{\sigma}_{u,t} \quad = \quad \lambda \frac{1}{|S_{u,t}|} \sum_{s \in S_{u,t}} \tau_s + (1-\lambda)\tau_a + \epsilon$$

In the above equation, λ is a factor which controls the tradeoff between personalization and non-personalization and τ_a is the average tag lifetime over all users and all tags. ϵ is a smoothing factor and is usually set to a small value. In fact, it is not only a smoothing factor, but also controls the tradeoff between short-term and long-term interests. If it is infinity, the model will be equivalent to the long-term interests model. The larger the ϵ, the smaller the differences of decays among different tags. In Figure 1(d), 1(e) and 1(f), we track 20 tags for a random user in Flickr. The x-axis is the time (day), and the y-axis is the $P_\tau(t|u)$. Three tags "2005", "rockandroll" and "livebands" are highlighted. From Figure 1(d), we can see the change of $P_\tau(t|u)$ of these tags. Because the data is from 2004 to 2005, we can see from the middle, the tag "2005" emerged and because of continuous usage of "2005", $P_\tau(2005|u)$ grows higher and higher. The tag "livebands" in first half is zero, because the user never uses that tag before 2005, and later user u became very interested in "livebands". Comparing the three figures, we notice that from $\epsilon = 100$, $\epsilon = 10$ to $\epsilon = 1$, the tracking become more and more detailed. Because as ϵ becomes lower, the local interests start to outperform the global interests and $P_\tau(t|u)$ becomes more sensitive to the short-term behaviors. For larger ϵ, it can capture the long-term trends of tags, and for smaller ϵ, it may better predict tags for current posts. It is difficult to determine which one is better and it depends on the task: when you try to capture trends of user interests, larger ϵ is suitable, and when you want to find the accurate tags on the test posts, smaller ϵ may be more suitable.

Similarly, for the non-tag-specific model, σ is the overall mean lifetime on all tags, resulting in:

$$\hat{\sigma}_u = \lambda \frac{1}{\sum_t |S_{u,t}|} \sum_t \sum_{s \in S_{u,t}} \tau_s + (1-\lambda)\tau_a + \epsilon$$

When considering the whole data set, the variance of tag lifetime is large, making it difficult to determine a single lifetime for all users. Thus, we calculated a personalized tag lifetime for each user. Figure 1(c) shows the probability density of personalized tag lifetime. We can see that more users in Flickr hold longer tag lifetimes.

Capturing topic switches

Elsewhere we report (Yin, Hong, and Davison 2011) on the session-like behaviors in tagging systems. From our observations, in personal tagging data, there often exist some topic switches—session-like behaviors as users switch between several subtopics. For the task of capturing the trends of users interests, the effects of topic switches are not so important as in task of tag prediction which require more accurate models of short-term interests.

Users may become interested in some new topics suddenly or switch back to some older topics because of some unknown external effects. We first assume that the current post (from the test set) is not a topic switch post, meaning the user continues the most recent session of tags on a particular topic. As in Yin et al. (2011), we use a threshold on the tag similarity as measured by Jaccard's coefficient to define topic switches. For a given user, let p_{i-1}, p_i be two consecutive posts, whose timestamps are $\tau_{i-1} \leq \tau_i$ and tag sets are T_{i-1} and T_i. Use J_{p_{i-1},p_i} as the measurement of the possibility of a topic switch at post p_i: $J_{p_{i-1},p_i} = \frac{|T_{i-1} \cap T_i|}{|T_{i-1} \cup T_i|}$. The personalized session lengths for each user are controlled by a global threshold κ. If $J_{p_{i-1},p_i} < \kappa$, the post p_i is considered to be a topic switch. For each test post p, our method will find the post p_i from which the latest session begins, and then the kernel smoothing will be only effective from p_i. Although κ is a shared parameter among all users, it generates personalized session lengths for users.

In the above session model, we made an assumption that the current test post is not a topic switch post; however, in fact, the current post may be the start of a new session. We believe that the time interval from the current test post to the most recent post can help predict such a case. Intuitively, the longer the interval is, the higher the probability of a new session starting. To measure whether the current post p_c is the start of a new session, we propose a function $J_{p_c} = f(\tau_c), \mathbb{R} \to \mathbb{R}$ where J_{p_c} is the predicted tag similarity between the current test post p_c and the most recent post based on the elapsed time. For the current test post p_c of user u, we have all past posts of user u—P_u. For every two consecutive posts p_{i-1}, p_i, we have a time interval $\tau_i = \tau_{p_{i,u}} - \tau_{p_{i-1}}$ and their similarity value $J_i = J_{p_{i-1},p_i}$. Then we have a set of samples $(\tau_1, J_1), (\tau_2, J_2), ..., (\tau_n, J_n)$, from which we need to learn the function $J_{p_c} = f(\tau_c)$. While there are many regression methods, we use a non-parametric technique—the nearest neighbor method. Compared to kernel methods, the nearest neighbor method defines points local to τ_c not through the fixed kernel bandwidth, but instead on a set of points closest to τ_c, measured by the distance $d_{i,c} = |\tau_i - \tau_c|$. Then the regression at τ_c is calculated as $J_{p_c} = \frac{\sum_i w_i \cdot J_i}{\sum_i w_i}$ where w_i is a tri-cube weight function

$$w_i = \begin{cases} (1-(\frac{d_{i,c}}{d_{k,c}})^3)^3 & d_{i,c} \leq d_{k,c} \\ 0 & d_{i,c} > d_{k,c} \end{cases}$$

where only k of n points closest to τ_c are considered as the neighborhood and $d_{k,c}$ is the distance of the furthest τ_c. Following the previous definition: if $J_{p_c} \geq \kappa$, the current test post will still stay in the current session and the session-based prediction method will be employed while if $J_{p_c} < \kappa$, we will treat this test post as the start of a new session and so at this moment, other methods which do not depend on temporal information can be employed, such as content-only methods (Lipczak et al. 2009). In the following experiments, we will also discuss combinations of methods.

Table 2: Validation Results

Method	Bibsonomy	Delicious	Flickr
Long-term model	.245	.161	.369
TIM	.325	.258	.726
User-tag TIM	**.334**	**.283**	**.733**
User-tag TIM (w/o κ)	.302	.276	.726
LZ (uniform)	.291	.191	.448
LZ (triagular)	.301	.237	.616

Experiments

On all three data sets, we split the whole data into two parts: earlier data and test data (the last 30 days data). Validation data in which 1000 posts are sampled from earlier data at random is used to tune and analyze the parameters. Then based on the last 30 days data, we perform completely online evaluation to simulate the tagging system running (evaluate each post over time and after that the post will be treated as an additional training post). In our interests lifetime model, there are two models: the personalized temporal interests model which assumes the users' behaviors on different tags are the same, and the personalized user-tag-specific temporal interests model in which users have different behaviors on different tags. We call them TIM and User-tag TIM.

We compare our method with three kinds of leading algorithms, which are from Lebanon and Zhao's (2008) method of temporal document modeling (LZ), Yin et al.'s (2010) method of personalized tag prediction (YXHD), and Lipczak et al.'s (2009) method of content-only tag prediction (LHKM). Lipczak's method took the first place in the "content-based" recommendation task in ECML PKDD Discovery Challenge (Eisterlehner, Hotho, and Jäschke 2009).

Table 3: Results on 30 day test data

Method	Bibsonomy	Delicious	Flickr
Long-term model	.118	.163	.312
User-tag TIM	**.501**	**.267**	**.835**
LZ (uniform)	.431	.203	.419
LZ (triangular)	.497	.232	.701

Table 4: Results on Bibsonomy

Method	F_1	Method	F_1	p-value
LZ	.306	User-tag TIM	**.341**	.0498
LHKM	.136	LHKM w. TIM	**.369**	7.56e-004
YXHD	.309	YXHD w. TIM	**.357**	.0033

We use the common F-measure function of precision and recall to evaluate prediction performance as we used previously (Yin et al. 2010). F-Measure is measured in break even point.

Parameter Analysis

Here we describe the parameter tuning process using the validation data (prior to the final month). In the predictive model, there are three parameters: ϵ is a smoothing factor, λ controls the personalization weight and κ is the factor of session detection. If $\epsilon = \infty$ and $\kappa = \infty$, the model is exactly the long-term interests model. On all three data sets, the effects of the three parameters are similar: for the tag prediction task, smaller ϵ is more suitable and can capture local interests better. λ tends to be better near to one. In Bibsonomy, the maxima appears when $\lambda = 1.0, \epsilon = 0.001, \kappa = 0.1$. In Delicious and Flickr datasets, the maximas appear at $\lambda = 0.8, \epsilon = 0.0001, \kappa = 0.3$ and $\lambda = 0.9, \epsilon = 0.0001, \kappa = 0.6$ respectively.

We also compare several variations of our methods to analyze the effects of each part. At first, we compare user-tag TIM with TIM where all tags of the same user share the same σ. In Table 2, the results show that user-tag TIM can outperform the default personalized model. Because the computational cost for the two algorithms is the same, we will use user-tag TIM in the following experiments. We also find that session-like behaviors are an important factor. In the tag prediction task, performance can be improved significantly over the version without topic switch detection (w/o κ).

The comparison method LZ is also carefully tuned, resulting in $h = 5$ in Bibsonomy and $h = 1$ in Delicious and Flickr. The triangular kernel and uniform kernel are used in local likelihood: the uniform kernel—$K_h(\tau) = 2^{-1} \cdot \mathbf{1}_{\{r<h\}}$ and the triangular kernel $K_h(\tau) = \frac{(1-\frac{r}{h})}{h} \cdot \mathbf{1}_{\{r<h\}}$.

Simulating the Real System

We simulate the real tagging system running on the last 30 days of data—performing completely online evaluation on the test data. There are 4,742 posts and 17,785 records in Bibsonomy, 21,916 posts and 76,213 records in Delicious and 110,551 posts and 517,949 records in Flickr. The results are shown in Table 3.

The results on test data are better than the results on validation data because the system has more historical information. It show that our user-tag TIM is better than the baselines and LZ on all three data sets. In Flickr, the performance achieves over 80% which is consistent with the fact that Flickr users' interests are more focused and easier to be tracked. Interestingly, it suggests that in real tagging systems, we can make effective recommendation through users' temporal interests analysis only.

Incorporating content

In this section, we compare our temporal interests model with two successful content-related methods—YXHD (Yin et al. 2010) and LHKM (Lipczak et al. 2009). We use the Bibsonomy data set—the same data set as in (Yin et al. 2010), the same evaluation methods[5] and the same parameter tuning. Table 4 presents the results. Our temporal interests model can outperform the two content-related methods. The p-value is also calculated by two-sample t-test, compared to the state-of-art YXHD. We can see that TIM gets significant improvement.

YXHD's method treats the tag prediction problem as the reverse problem of web searching and start from the the basic Bayes rule, integrating three factors—an ego-centric effect, environmental effects and web page content. Because users' preferences on each tags are drifting over time, intuition suggests that temporally adjusting the prior can get better results. To incorporate the content, we combine the two methods by replacing the $P(t|u)$ with the temporal prior $P_\tau(t|u)$ which has already been shown to better capture users' current preferences. The combined methods achieve an F-measure of 0.357, which is significantly better than either YXHD or our temporal interests model.

LHKM only uses content to recommend tags. The advantage of the LHKM algorithm is when processing new items and during topic switches. Because it is a content-only method, it does not distinguish whether the item has appeared in training data or not. Even if the current user suddenly changes interests, the algorithm can also obtain stable performance. In the detecting topics switches section, we describe a non-parametric method for simply combining TIM with other method. The results show that the combined LHKM can achieve the best performance. We also notice that because YXHD has already involved a high weight on the ego-centric effect, the improvement is not as high as LHKM.

Conclusion

In this paper, we investigated the temporal dynamics of user interests in tagging systems, and proposed a user-tag-specific temporal interests model for tracking users' interests. Using three public datasets we showed the impact of personalization and user-tag specification.

[5] Under online evaluation mode, we also calculated Top-5 F-measure, and the results are similar.

Based on our experiments, we are able to conclude that our temporal user interests model, generated only from the temporal tag sequence, can achieve an F-measure of 0.341 and outperform the state-of-the-art which is 0.309 for Bibsonomy data. Combining with existing methods YXHD and LHKM, performance further improved to 0.357 and 0.369, respectively. All three methods incorporating TIM can outperform the state-of-the-art as well as a leading algorithm addressing concept drift.

Acknowledgments

This work was supported in part by a grant from the National Science Foundation under award IIS-0545875. We also appreciate helpful discussions on the model with Wei-min Huang and Lifeng Shang.

References

Bao, S.; Xue, G.; Wu, X.; Yu, Y.; Fei, B.; and Su, Z. 2007. Optimizing web search using social annotations. In *Proceedings of the 16th International Conference on World Wide Web*, 501–510. ACM Press.

Ding, Y., and Li, X. 2005. Time weight collaborative filtering. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, 485–492. ACM Press.

Eisterlehner, F.; Hotho, A.; and Jäschke, R., eds. 2009. *ECML/PKDD Discovery Challenge Workshop (DC09)*, volume 497 of *CEUR Workshop Proceedings*.

Görlitz, O.; Sizov, S.; and Staab, S. 2008. Pints: Peer-to-peer infrastructure for tagging systems. In *Proceedings of the Seventh International Workshop on Peer-to-Peer Systems (IPTPS)*.

Halpin, H.; Robu, V.; and Shepherd, H. 2007. The complex dynamics of collaborative tagging. In *Proceedings of the 16th International Conference on World Wide Web*, 211–220. ACM Press.

Hotho, A.; Jäschke, R.; Schmitz, C.; and Stumme, G. 2006. Information retrieval in folksonomies: Search and ranking. In *The Semantic Web: Research and Applications*, volume 4011 of *Lecture Notes in Computer Science*. Springer. chapter 31, 411–426.

Koren, Y. 2009. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 447–456. ACM Press.

Lebanon, G., and Zhao, Y. 2008. Local likelihood modeling of temporal text streams. In *Proceedings of the 25th International Conference on Machine Learning (ICML)*, 552–559. ACM Press.

Lipczak, M.; Hu, Y.; Kollet, Y.; and Milios, E. 2009. Tag sources for recommendation in collaborative tagging systems. In *ECML/PKDD Discovery Challenge Workshop (DC09)*.

Lv, Y., and Zhai, C. 2009. Positional language models for information retrieval. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 299–306. ACM Press.

Ramage, D.; Heymann, P.; Manning, C. D.; and Garcia-Molina, H. 2009. Clustering the tagged web. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining (WSDM)*, 54–63. ACM Press.

Rendle, S., and Schmidt-Thieme, L. 2010. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the 3rd International Conference on Web Search and Web Data Mining (WSDM)*, 81–90. ACM Press.

Rendle, S.; Balby Marinho, L.; Nanopoulos, A.; and Schmidt-Thieme, L. 2009. Learning optimal ranking with tensor factorization for tag recommendation. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 727–736. ACM Press.

Schlimmer, J. C., and Granger, R. H. 1986. Beyond incremental processing: Tracking concept drift. In *Proceedings of the 5th National Conference on Artificial Intelligence (AAAI)*, 502–507. AAAI Press.

Widmer, G., and Kubat, M. 1996. Learning in the presence of concept drift and hidden contexts. In *Machine Learning*, 69–101.

Xiang, L.; Yuan, Q.; Zhao, S.; Chen, L.; Zhang, X.; Yang, Q.; and Sun, J. 2010. Temporal recommendation on graphs via long- and short-term preference fusion. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 723–732. ACM Press.

Xu, S.; Bao, S.; Fei, B.; Su, Z.; and Yu, Y. 2008. Exploring folksonomy for personalized search. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 155–162. ACM Press.

Yin, Z.; Li, R.; Mei, Q.; and Han, J. 2009. Exploring social tagging graph for web object classification. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 957–966. ACM Press.

Yin, D.; Xue, Z.; Hong, L.; and Davison, B. D. 2010. A probabilistic model for personalized tag prediction. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 959–968. ACM Press.

Yin, D.; Hong, L.; and Davison, B. D. 2011. Exploiting session-like behaviors in tag prediction. In *Companion Proceedings of the 20th International Conference on World Wide Web*, 167–168. ACM Press.

Zhang, D.; Mao, R.; and Li, W. 2009. The recurrence dynamics of social tagging. In *Proceedings of the 18th International Conference on World Wide Web*, 1205–1206. ACM Press.