Author Topic Model-Based Collaborative Filtering for Personalized POI Recommendations

Shuhui Jiang, Xueming Qian, *Member, IEEE*, Jialie Shen, *Member, IEEE*, Yun Fu, *Senior Member, IEEE*, and Tao Mei, *Senior Member, IEEE*

***Abstract*—From social media has emerged continuous needs for automatic travel recommendations. Collaborative filtering (CF) is the most well-known approach. However, existing approaches generally suffer from various weaknesses. For example, sparsity can significantly degrade the performance of traditional CF. If a user only visits very few locations, accurate similar user identification becomes very challenging due to lack of sufficient information for effective inference. Moreover, existing recommendation approaches often ignore rich user information like textual descriptions of photos which can reflect users' travel preferences. The topic model (TM) method is an effective way to solve the "sparsity problem," but is still far from satisfactory. In this paper, an author topic model-based collaborative filtering (ATCF) method is proposed to facilitate comprehensive points of interest (POIs) recommendations for social users. In our approach, user preference topics, such as cultural, cityscape, or landmark, are extracted from the geo-tag constrained textual description of photos via the author topic model instead of only from the geo-tags (GPS locations). Advantages and superior performance of our approach are demonstrated by extensive experiments on a large collection of data.**

***Index Terms*—Data mining, recommendation system, text mining, travel recommendation.**

I. INTRODUCTION

WHEN planning to visit a new city, many travel guide websites like IgoUgo.com can provide a lot of content such as travelogues and photos for users to arrange tours. However, gaining useful information from fussy raw materials via manual analysis can be very time consuming [28], resulting in automated travel planning receiving increased attention the by data mining and multimedia systems community. In particular, there is a growing concern about personalized travel recommendations, which can effectively integrate user preferences (e.g., cultural, cityscape, or landscape).

Collaborative filtering (CF) based recommendation is the most well-known approach, and is widely utilized in products, services [16], [29], and travel recommendations [8], [28], [30], [32], [38]–[40]. Location based collaborative filtering travel recommendation methods first mine POIs in a city which has been visited by social users using geo-tags or GPS trajectories [32]. Then similar users are detected by calculating the location co-occurrences from users' travel history. Then similar users are detected by calculating the location co-occurrences from users' travel history. Finally, the POIs of a new city are recommended according to similar users' visiting history. CF-based recommendation approaches are effective and efficient, but suffer from the well-known "sparsity problem" in recommendation systems, due to travel data being very sparse. In this circumstance, it makes accurate similar user identification very difficult if the user has only visited a small number of POIs.

Major efforts have been made to solve the CF sparsity problem. Recently, topic model method (TM) has been introduced into personalized travel recommendations [1], [13]. TM is similar to the content-based method in product recommendation systems [29]. TM analyzes tourist's travel preferences (such as culture, urban landscape, or landscape) and recommends POIs which match the themes of user preferences. Through interest category mapping, even if the user has visited very few points of interest, we can still analyze user preferences.

The category topics is usually determined by the naive category information from recommended systems in TM [1], [23]. For example, the original category information of social media websites, such as Foursquare [23], ODP [5], and Yelp [23], serve as topics. From the predetermined categories, it is convenient to calculate user preferences. Unfortunately, for rich photo sharing networks like Flickr and Panoramio, there is no such defined category information. Thus the naive topic-based recommendation approach cannot be utilized directly in travel recommendations.

To solve the above problems, we propose an author topic model-based collaborative filtering method (ATCF) to recommend personalized POI when users plan to visit a new city. In contrast to existing location based collaborative filtering methods, we learn users' travel preferences from the text descriptions associated with their shared photos on social

Manuscript received October 06, 2014; revised January 12, 2015; accepted February 17, 2015. Date of publication March 26, 2015; date of current version May 13, 2015. This work was supported by Program 973 under Grant 2012CB31640, by the NSFC under Grant 60903121, Grant 61173109, and Grant 61332018, by Microsoft Research Asia, and by the Ministry of Education Academic Research Fund Tier 2 of Singapore under Grant MOE2013-T2-2-156. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Vasileios Mezaris. *(Corresponding author: Xueming Qian.)*

S. Jiang was with the School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China. She is now with the Department of Electrical and Computer Engineering, College of Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: shjiang@coe.neu.edu).

X. Qian is with the School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: qianxm@mail.xjtu.edu.cn).

J. Shen is with the School of Information Systems, Singapore Management University, Singapore 188065 (e-mail: jlshen@smu.edu.sg).

Y. Fu is with the Department of Electrical and Computer Engineering, College of Engineering and the College of Computer and Information Science, Northeastern University, Boston, MA 02115 USA (e-mail: yunfu@ece.neu.edu).

T. Mei is with Microsoft Research Asia, Beijing 100190, China (e-mail: tmei@microsoft.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2015.2417506

Fig. 1. Example of Flickr image information.

media, instead of from GPS trajectories or check-in records. In addition, users' similarities are measured with author topic model (ATM) instead of location co-occurrence.

Users' photos on social media record their travel history. As shown in Fig. 1, a typical Flickr user's photo contains metadata like "User Id", "tags", "Taken data", and geo-tag including "Latitude" and "Longitude". Except for geo-tags, the textual description (such as tags and comments) that users attach to photos when sharing them on social media networks are important for inferring a user's travel preferences. For example, if a user visits a gym, the information about where he or she has been can be observed from the GPS trajectory data of the shared photos. His or her detailed preference information such as "football" or "vocal concert" can be determined via visual analysis of the images and related tags. Consequently, it is rational to apply the user's tags attached to photos to explore a user's travel preferences instead of GPS trajectories.

An important part of ATCF is learnIing user travel interest from ATM textual descriptions [27]. ATM is an extended version of latent Dirichlet allocation (LDA) [2], [12], [33] by considering author information for document collections with user information [21], [26], [27], [31]. In our proposed author topic collaborative filtering (ATCF) based approach, the ATM directly annotates the user's travel preference with automatically divided semantic topics corresponding to the distribution of the tags. The system framework of the proposed approach is shown in Fig. 2.

This paper is an extension of our previous conference paper on the same topic [9]. In this journal version, we propose a coarse-to-fine based approach to mine city-level POIs and map user travel history mining. In the experiment part, more discussions and experiments in each part of the method are shown. The four main contributions of this work can be summarized as follows.

- We propose an effective author topic model-based collaborative filtering method for travel recommendations by making full use of rich textual descriptions and user information. Tags contain richer information about users' latent travel preferences than GPS trajectories and are much easier to obtain.
- We focus on solving the sparse problem of classical location-based collaborative filtering (LCF). In our proposed ATCF-based personalized travel recommendation system, we utilize users' topic preferences as the law for collaborative filtering instead of location co-occurrences. Even for the user with very sparse POI records, our ATCF can still mine more related resource than LCF to carry out travel recommendation.
- We introduce an author topic model to adaptively elicit topic categories from tags associated with photos. Using the scheme, topics about user preferences can be accurately extracted and applied to personalized travel recommendations.
- We propose a coarse-to-fine based approach to mine city level POIs and map user travel history mining. In POI mining, we first coarsely cluster city-scale photos by geo-tags and then refine POIs from clusters using visual features. In travel history mining, we set a transition of mapping user's geo-tagged photos between coarse clusters and refined POIs.

The rest of paper is organized as follows: Section II reviews the related work on travel recommendations. In Section III we introduce the offline "Coarse-to-fine POIs and User History Mining" method. Section IV introduces the "Author Topic Model Learning" method, followed by "Author Topic Model based Collaborative Filtering" in Section V. Evaluation and the visualization of the system are shown in Section VI. Conclusions are drawn in Section VII.

II. Related Work

The flourishing of social media has promoted research on travel recommendations. There are four different major kinds of data used for travel recommendations: blogs [11], [14], GPS trajectories [30], [32], check-ins [28], and geo-tags [4], [7], [36]. Zheng *et al.* [32] use GPS trajectories to mine and recommend travel routes. However, a GPS trajectory is relatively difficult to obtain. Blogs and travelogues are also used to mine landmarks and travel routes. In [11], Kori *et al.* proposed a route extraction system based on entries. Multimedia contents describing those routes are also presented. There is rich information in user-generated travelogues. However, work with blogs and travelogues does not consider users' preferences or automatically recommended personalized travel plans. Check-in data are very helpful for mining service like restaurant [6], [20], [28], [37]. Yuan *et al.* focused on time-aware POI recommendation problem. They proposed a Geographical-Temporal influences Aware Graph which aims at recommending POIs to a user when he or she wants to visit at a given time [37]. It mainly focuses on local inhabitant instead of travelers and does not have rich landmark information. Besides travelogues, GPS, and check-in data, community-contributed geo-tagged photos would be good sources to facilitate the travel recommendation system [7], [36].

Recently, personalized travel recommendation has attracted much attention. Collaborative filtering (CF) and Content-based (CB) are the most widely used methods. Clements *et al.* recommended POIs according to user's travel history via collaborative filtering [4], [8], [38]. They used the number of users who have visited both places to measure the similarity between two landmarks. However, with an increasing number of locations and users, the complexity also increases. Moreover, if the location is not famous and, to be more extreme, if no geo-tag is marked here, this location will never be recommended. CF methods may face a "data sparsity" problem.

Applying a topic model method is effective to the "data sparse" problem [1], [13]. Bao *et al.*, presented a location-based

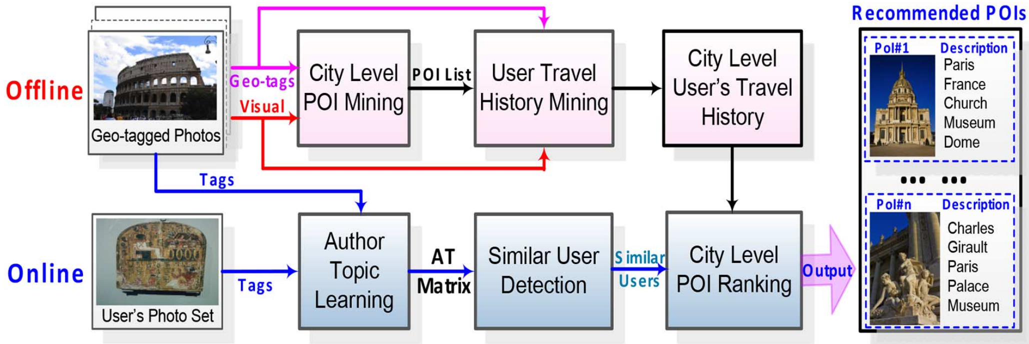


Fig. 2. System containing offline module (pink) and online module (blue). The offline module aims at mining city level POIs and users' travel histories from social community contributed photos and their coherent descriptions including tags and geo-tags by a coarse-to-fine mapping method. The online module is to recommend POIs to the query user who plans to visit a new city based on his or her travel preferences. The input of the online module is the query user's textual description (i.e., tags) of his or her shared photos. The online recommendation module consists of the following three steps: 1) author topic model learning. The category of latent travel topics of *city* 1 are adaptively mined, and travel topic distributions of a new user are mined simultaneously; 2) similar users detection according to their topic distributions; and 3) city-level POI ranking.

and preference-aware travel recommendation system [1]. They used a weighted category hierarchy to model each individual's personal preferences from learning an iterative learning model in their offline module. Probabilistic photographer behavior is mined by topic model via PLSA and Markov models [13].

In contrast to existing POIs recommendation methods based on CF or TM, our author topic model-based collaborative filtering method learns users' travel preferences from text descriptions associated with geo-tagged photos by author topic models. Users' similarities are then measured by the learned author topic model instead of location co-occurrence, meaning accurate similar users can still be mined even if users have visited very few POIs.

III. Coarse-to-Fine POIs and User History Mining

Most existing works related to the mining of POIs apply density-based clustering (e.g. mean-shift clustering) towards geo-tags attached to community-contributed photos [3]. For example, Cheng *et al.*, used mean-shift clustering with a bandwidth of 0.001 to find locations where many photos are taken. The reason to set the bandwidth as 0.001 is that the radiuses of a large amount of POIs are roughly at that value. However, it is known that the radiuses of different POIs are different. The basic mean-shift based POIs mining approaches may face the following two problems: 1) POIs with small radii may not to be mined when the bandwidth is set to be too large, and 2) several POIs are close to each other may be clustered into one POI when the bandwidth is set to too small. To solve these problems, we propose coarse-to-fine city-level POI mining and coarse-to-fine user history mining. In this section, we will introduce the gathering of travel data first. Then the POI mining and user history mining methods are described in detail.

A. Travel Data Gathering

Flickr.com is a famous photo sharing website with more than 10 billion photos uploaded by more than 80 million users. We have collected more than 7 million social images from Flickr by using its open API [15], [17], [24]. According to our statistics, 79.19% of photos have tags, 31.94% photos have geo-tags, and 28.34% photos have both tags and geo-tags [24]. Tags are one of the most important sources for social media mining and retrieval. Thus, we use tags instead of geo-taggs to mine user travel preferences and carry out personalized travel recommendations.

It is likely that user-shared photos have noises. For example, some photos have incorrect tags and geo-tags. The geo-tag can be corrected by image location estimation [15] and the tags can be enhanced by tag enrichment [22]. In this paper, a simple strategy is utilized. In our city level POI mining, we use both tags and geo-tags to process the crawled dataset.

Let l denote an image. Each l has three attributes $l = \{\text{tags, geo-tags, visual}\}$. First, we compare each tag of the photo with the name of the city like "London". We select the photos whose tags contain the city name. Then, we use mean-shift clustering [3], [15] towards all the geo-tags of the crawled photos at a very large bandwidth to get photos for each city. We set the bandwidth of mean-shift to be 0.5, which is almost the radius of a city [3], [15]. We only select the cluster whose center is close to the real location of the city to construct the dataset. Photos with incorrect city names or geo-tags are removed.

Let N_l denote the number of photos that meet our requirements. Let τ, g, and v denote tags, geo-tags, visual features of the image separately. We use $L = \{l_1, l_2, \dots, l_{N_l}\}$ to denote the collection of images of a city, $T = \{\tau_1, \tau_2, \dots, \tau_{N_l}\}$ to denote the collection of all the tags of L, $G = \{g_1, g_2, \dots, g_{N_l}\}$ to denote the collection of all the geo-tags of L, and $V = \{v_1, v_2, \dots, v_{N_l}\}$ to denote the visual feature of L.

B. Coarse-to-Fine City Level POIs Mining

In this section, we propose a coarse-to-fine method to mine POIs by carrying out the following two steps: (a) city level coarse POI clustering, and (b) visual matching based city level POI refinement. After that, for each city χ_m we get a set of POIs of the city which is denoted as $P_m = \{p_1, \dots, p_j, \dots, p_{N_p^{(m)}}\}$, where $N_p^{(m)}$ is the number of POIs in χ_m, and p_j denotes the j-th POI of the city.

1) City Level Coarse POI Clustering: For a set of geo-tagged images $L_m = \{l_1, \dots, l_2, \dots, l_{N_l^{(m)}}\}$, first

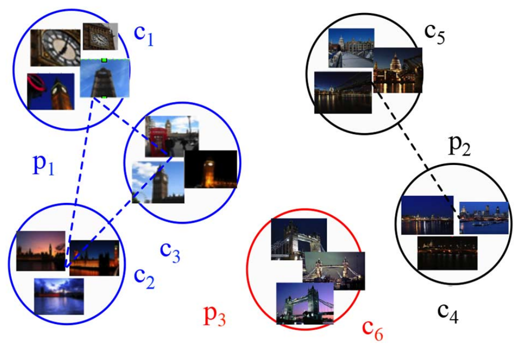


Fig. 3. Sketch map of coarse-to-fine method of POI mining.

mean-shift clustering is utilized towards their geo-tags $G_m = \{g_1, \ldots, g_2, \ldots, g_{N_l^{(m)}}\}$ at a very small bandwidth to discover landmarks with a small radius. In [3], Cheng *et al.*, set the bandwidth as 0.001, which is roughly the radius of a landmark. In our method, the bandwidth is set as 0.0005 instead of 0.001. That is to say, for most landmarks, the bandwidth of mean-shift is much smaller than the radius. In this circumstance, POIs with a large radius may be scattered to several clusters. Each cluster contains a specified view of the POI [10]. As shown in Fig. 3, $c_1 - c_3$ are the three clusters of the POI "Big Ben". Considering that community-contributed photos are noisy and crowded with erroneous geo-tags and tags, for a valid POI, it should contain enough users and have sufficient photos. In this paper, a simple post processing approach is utilized in POI verification. For example, each cluster must contain at least 20 photos and the number of different users who upload photos into this cluster must be larger than 10. After coarse clustering, we get a list of clusters of a city denoted as $C_m = \{c_1, \ldots, c_n, \ldots, c_{N_c^{(m)}}\}$, where $N_c^{(m)}$ denotes the number of clusters of χ_m, and c_n denotes the n-th cluster. $\xi_x^{(n)}$ and $\xi_y^{(n)}$ are latitudes and longitudes of the center of c_n.

2) Visual Matching Based City Level POIs Refinement: From the coarse clustering, we get a list of clusters $C_m = \{c_1, \ldots, c_n, \ldots, c_{N_c^{(m)}}\}$ of χ_m. It is possible that photos from one landmark may be divided into several clusters. We are required to merge them with the help of visual features $V_m = \{v_1, \ldots, v_l, \ldots, v_{N_l^{(m)}}\}$ of L_m. The visual matching based POIs refinement approach consists of the following two steps: "visual feature extraction and cluster representation", and "cluster merging".

Visual feature extraction and cluster representation: SIFT performs robustly in landmark image description especially for buildings like towers and cathedrals [15], [18], [29]. In this paper, we use SIFT to carry out a visual feature matching-based cluster merging. First, we extract the 128D SIFT features for each image $l_k (i = 1, \ldots, N_l^{(m)})$. Then we use bag-of-words (BoW) to present the SIFT descriptor via hierarchical quantization [35]. The size of the codebook is 61,724. Finally, each image l_k is represented by its BoW histogram h_k. The dimension of h_k is 61,724.

Cluster merging: Assume that the total number of POIs of χ_m is $N_p^{(m)}$. The corresponding POIs list is denoted as $P_m = \{p_1, \ldots, p_j, \ldots, p_{N_p^{(m)}}\}, N_p^{(m)} \le N_c^{(m)}$. For each p_j, its center of location is denoted by $\sigma_x^{(j)}$ and $\sigma_y^{(j)}$. For simplicity, we introduce a binary matrix $\Gamma_{c,p}(n, j)$ to present the mapping from c_n to p_j. $\Gamma_{c,p}(n, j) = 1$ if c_n is mapping to p_j, otherwise, $\Gamma_{c,p}(n, j) = 0$.

For each p_k, first we randomly pick up a cluster c_k to initialize $\hat{p}_k$. The basic idea is to find an optimal cluster c_o for $\hat{p}_k$ as

$$o = \arg\min_n d(\hat{p}_k, c_n), n = \{1, \ldots, N_c^{(m)}\} \tag{1}$$

where $d(\hat{p}_k, c_n)$ is the geo-distance of $\hat{p}_k$ and c_n, which is determined as

$$d(\hat{p}_k, c_n) = \sqrt{\left(\sigma_x^{(k)} - \xi_x^{(n)}\right)^2 + \left(\sigma_y^{(k)} - \xi_y^{(n)}\right)^2} \tag{2}$$

where $\sigma_x^{(k)}$ and $\sigma_y^{(k)}$ is center of $\hat{p}_k$ which is updated iteratively. $\xi_x^{(n)}$ and $\xi_y^{(n)}$ are latitudes and longitudes of the center of c_n.

Then we calculate the visual similarity between c_o for $\hat{p}_k$ as

$$S_v(\hat{p}_k, c_n) = \sqrt{(\hat{h}_k - \hat{h}_n) \times (\hat{h}_k - \hat{h}_n)'} \tag{3}$$

where $\hat{h}_k$ and $\hat{h}_n$ are the average of the BoW histograms of the images in c_n and p_k. If $S_v(\hat{p}_k, c_n) < \nu_v$, we update $\hat{p}_k = \{\hat{p}_k, c_o\}$, otherwise we find another c_o from the rest of the clusters.

It is very time consuming to visually match all the clusters for each $\hat{p}$. Actually, we find that if the clusters which are geographically near $\hat{p}_k$ do not belong to $\hat{p}_k$ according to visual matching, thus it is very likely that the clusters which are far away from $\hat{p}_k$ do not belonging to $\hat{p}_k$ either. Based on this observation, we introduce a fast cluster matching method. We introduce a variable t_{rej} to record the number of clusters which are geographically near $\hat{p}_k$, but failed in visual matching. If t_{rej} is larger than the threshold ν_t, we terminate the progress for finding other far away clusters for $\hat{p}_k$ and we start to find clusters for $\hat{p}_{k+1}$. In this paper, we set ν_t, by considering the eight orientations (each with 45 degrees) of the POI. The impact of parameter ν_t to POI mining is discussed in our experiments. The final updated $\hat{p}_k$ is defined as p_k and the POIs set in χ_m is denoted as $P_m = \{p_1, \ldots, p_j, \ldots, p_{N_p^{(m)}}\}$.

To illustrate the above steps intuitively, we give a sketch map as shown in Fig. 3. There are six clusters obtained by coarse clustering before merging, i.e. $C = \{c_1, \ldots, c_6\}$. After visual similarity based merging, we have three POIs, i.e. $P = \{p_1, p_2, p_3\}$. According to the merging, we find that $c_1 - c_3$ belonging to p_1 are all photos with "Big Ben" (blue), and $c_4 - c_5$ belonging to p_2 are "St Paul's Cathedral" (black). No other cluster is merged with c_6, which is "London Tower" (red). So only c_6 belongs to p_3.

C. Coarse-to-Fine User Travel History Mining

Assume there are $N_u^{(m)}$ users in city χ_m. Each u_i has a set of image $L_i = \{l_1, \ldots, l_k, \ldots, l_{N_u^{(k)}}\}$. By vast observation, we find that a photo on the border of a POI with a large radius is often assigned to its neighboring POI. To solve this problem, we make full use of the coarse layer clusters C and assignment vector $\Gamma_{c,p}$ to get an accurate travel history. We propose using a coarse-to-fine mapping method to mine user travel history, which contains two steps: (1) coarse user's travel history mapping from user's images to clusters, and (2) refined user's travel history mapping from clusters to POIs.

1) Coarse User's Travel History mapping: Assume that there are $N_p^{(m)}$ POIs $P_m = \{p_1, \ldots, p_j, \ldots, p_{N_p{(m)}}\}$ and $N_c^{(m)}$ clusters $C_m = \{c_1, \ldots, c_n, \ldots, c_{N_c{(m)}}\}$ in χ_m. We define a binary assignment matrix $\Gamma_{u,c}$ to record the mapping from a user's photos to clusters. $\Gamma_{u,c}(i,n) = 1$ means at least one image of L_i belongs to C_n as

$$\Gamma_{u,c}(i,n) = \begin{cases} 1, & \text{if } \sum_{k=1}^{N_l^{(i)}} \gamma(k,n) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $\gamma(k,n) = 1$ means that the k-th image l_k belongs to c_n, while $\gamma(k,n) = 0$ means l_k does not belong to c_n. $\gamma(k,n)$ is determined by the geo-distance between image l_k and the coarse clusters as

$$\gamma(k,n) = 1; n = \arg\min_t \hat{d}(k,t), t \in \{1, \ldots, N_c^{(m)}\} \tag{5}$$

where $\hat{d}(k,t)$ is calculated as

$$\hat{d}(k,t) = \sqrt{\left(g_x^{(k)} - \xi_x^{(t)}\right)^2 + \left(g_y^{(k)} - \xi_y^{(t)}\right)^2} \tag{6}$$

where $g_x^{(k)}$ and $g_y^{(k)}$ are the latitudes and longitudes of l_k. $\xi_x^{(t)}$ and $\xi_y^{(t)}$ are latitudes and longitudes of the center of c_t.

In this paper, a simple and efficient approach is utilized by introducing ε-ball based constrains on (7) to ensure only the photos belonging to the specified POI are utilized for travel recommendations. When all the $\hat{d}(k,t)$ between l_k and C_m are larger than ε, we determine that l_k is irrelevant to any POIs of the city. We then remove l_k for representing the travel history of the city by setting the distance as ∞, and we have

$$\hat{d}(k,t) = \begin{cases} \infty, & \text{if } \hat{d}(k,t) > \varepsilon \\ \hat{d}(k,t), & \text{if } \hat{d}(k,t) \le \varepsilon \end{cases}. \tag{7}$$

2) Refined User's Travel History Mapping: Assume that there are $N_p^{(m)}$ POIs $P_m = \{p_1, \ldots, p_j, \ldots, p_{N_p{(m)}}\}$ and $N_c^{(m)}$ clusters $C_m = \{c_1, \ldots, c_n, \ldots, c_{N_c{(m)}}\}$ in χ_m. The basic idea of refined user's travel history mapping is that of all the clusters belonging to a POI, if at least one image of the user belongs to these clusters, then the user is regarded as having visited this POI.

We define $Q_m^{(i)} = \{q_1^{(i)}, \ldots, q_j^{(i)}, \ldots q_{N_{p(m)}}^{(i)}\}$ to present the travel history of u_i. $q_j^{(i)} = 1$ means the u_i has visited p_j, while $q_j^{(i)} = 0$ means the user does not visited p_j. For each p_j, $q_j^{(i)}$ is calculated as

$$q_j^{(i)} = \sum_{n=1}^{N_c^{(m)}} \Gamma_{u,c}(i,n) \times \Gamma_{c,p}(n,j) \tag{8}$$

where $\Gamma_{u,c}(i,n)$ is the mapping from u_i to c_n and $\Gamma_{c,p}(n,j)$ is the mapping from c_n to p_j.

The performance of coarse-to-fine POIs and the user history mining method are discussed in Section VI-D in our experiments.

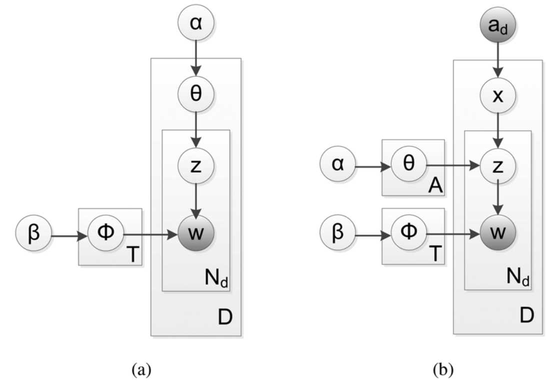


Fig. 4. Graphical model for the (a) LDA and (b) ATM [27].

IV. Author Topic Model Learning

For the users who have geo-tagged photos in their social communities, it is comparatively easier to carry out travel recommendations by mining GPS trajectories. However, according to our figures, only 1/3 of images have GPS records, while more than 90% users in social media sharing websites provide textual descriptions for their photos [17], [24]. Thus, it is practical for us to carry out travel recommendations by utilizing tags rather than GPS.

In this paper, we propose an ATM based approach to model social users to carry out personalized travel recommendations. ATM is an extended version of latent Dirichlet allocation (LDA) [2], [12], [25], [33] by considering author information, for document collections with user information [19], [21], [26], [27], [31]. The graphical models of LDA and ATM are shown in Fig. 4(a) and Fig. 4(b) respectively [27]. In this figure, shaded variables indicate observed variables and unshaded variables indicate latent variables. The arrow indicates a conditional dependency between the variables and plates represented by the box. Different from the LDA based model that only discovers which topics are expressed in a document, for the ATM model w_i is associated with two latent variables: the author (i.e. user) x_i and the topic z_i. Through ATM, both the category and user's travel preferences are mined by eliciting the latent model simultaneously.

A. Terminologies of ATM

To better describe the ATM model in the recommendation system, the original terms of ATM [31] are used to define the terminologies of LDA and ATM in this paper as follows.

- The *vocabulary* V denotes a set of N_d unique tags. Each tag is presented by the corresponding label $\{1, 2, \ldots, N_d\}$.
- The *word* $w_i \in \{1, 2, \ldots, N_d\}$ represents the label of one tag of the photo. Each tag of an image is mapped to V whose size is N_d through character matching.
- The *document* d corresponds to a tag set τ_i of the image. A user with N_f number of images in the photo set has N_f documents.
- The *author* a_d is a set of users who uploads the document d. In our paper, each a_d has only one element, because each

photo could only be uploaded by one user. Each user has a unique label $\{1, 2, \ldots, N_u\}$.

Among these four terminologies, a_d is only used in ATM. Three other terminologies are also used in LDA.

B. Data Preprocessing for ATM

For each city χ_m, first, in order to construct the vocabulary V, we filter all the tags with both "stop words" [34] and "Flickr-style words". A stop word can be identified as a word that has the same likelihood of occurring in those documents not relevant to a query as in those documents relevant to the query. For example, the "stop words" could be "his", "on" and etc. "Flickr-style words" are a list of words that frequently appear in Flickr tags like "Canon", but not in ordinary "stop words". After ranking the words according to frequency and deleting the "stop words", we manually define the tags with higher frequency but it is useless to distinguish categories as "Flickr-style words". After tag filtering, all the unique tags are constructed in the vocabulary V. Each tag in V has a label $w_i \in \{1, 2, \ldots, N_d\}$.

Second, for each user a_d who has uploaded document d (corresponds to an image), we map his or her tag set τ image to the V to get the label w for each tag. Thus all the tags of the city have been mapped to corresponding labels.

Finally, we use three sparse binary matrixes $\Gamma_{d,a}$, $\Gamma_{w,d}$ and $\Gamma_{w,a}$ to record whether a document belongs to a user, whether a word belongs to a document and whether a word belongs to a user. For example, if $\Gamma_{w,a}(i,j) = 1$ means w_i belongs to u_j.

C. User ATM Learning

For each city χ_m, the input of this model contains two parts: the query user u_k's photo set L_k with the corresponding tags with photos, and community users' photo sets with tags. The output is the topic preference distribution for each user.

The generative process of ATM mainly consists of two main steps: First is the probabilistic generative model and second is the Bayesian estimation of the model parameters [31].

First, to generate the probabilistic model, we assume the author x is associated with a multinomial distribution over the N_z topics. The author-topic distribution is denoted as θ. Meanwhile, each topic $z \in \{1, 2, \ldots, N_z\}$ is associated with a multinomial distribution over the words. The topic-word distribution is denoted as ϕ.

For a word w, we first choose an author x uniformly from a_d conditioned on a_d. Second, we choose a topic z from θ conditioned on x. Third, we choose a word w from ϕ conditioned on the topic distribution z_i. We repeat this three steps for N times to generate all the words in the document.

Second, to estimate the author-topic distribution θ and topic-word distribution ϕ, a Morkow chain Monte Carlo algorithm, the Gibbs sampler, is used to sample from the posterior distribution over θ and ϕ. A more detailed method of ATM can be inferred from [31].

Through ATM, we can determine the probabilities of each word to different topics. We also get author topic matrix A for all users. A is a sparse $N_x \times N_z$ matrix, where N_x is the number of authors and N_z is the number of topics. $A(i,j)$ contains the times that the words associated with author x_i have been assigned to topic z_j. The normalized $A(i)$ presents the topic distractions of each user u_i.

V. Author Topic Model Based Collaborative Filtering

Our recommendation system aims to recommend a series of POIs for a user when he or she plans to visit a new city. To simplify the description, let u_k denote the query, χ_1 denote the city which u_k has already visited and χ_2 denote the new city which u_k plans to visit. In traditional location-based collaborative filtering, first the user-POI matrix is generated to record the users' travel histories. Then to each two users u_k and u_i, the similarity between them is calculated by the cosine standard measurement towards their corresponding vectors in the user-POI matrix. In our author topic model based collaborative filtering, we use the users' author topic models to measure their similarity. Our ATCF based POI recommendation approach consists of the following two steps: 1) similar user detection, and 2) POI ranking.

1) Similar User Detection: Assuming $\hat{U} = \{u_1, u_2, \ldots, u_{N_u}\}$ are N_u users who have both visited χ_1 and χ_2. $A(i)$ is the topic distribution of $u_i, i \in \{1, 2, \ldots, N_u\}$ in χ_1. We calculate the similarity between u_k and u_i from their author topic vectors with the cosine similarity measurement as

$$\hat{S}_a(u,i) = \frac{A(u) \times A'(i)}{\|A(u)\| \cdot \|A'(i)\|}. \tag{9}$$

Other similarity measurement approaches are also discussed in Section 6.5 in our experiments. According to similar scores, we then rank the users according to $\hat{S}_a(u,i)$. The top ranked N_s users ($N_s \leq N_u$) are selected as the set of similar users to carry out travel recommendations.

2) POI Ranking: Assuming $P_2 = \{p_1, p_2, \ldots, p_{N_p^{(2)}}\}$ the set of POIs in χ_2. We rank P_2 according to travel history $\{Q_1^{(2)}, \ldots, Q_i^{(2)}, \ldots, Q_{N_u}^{(2)}\}$ of all the N_u users in χ_2, $i \in \{1, 2, \ldots, N_u\}$. For each u_i, $Q_i^{(2)} = \{q_1^{(i)}, \ldots, q_j^{(i)}, \ldots, q_{N_p^{(2)}}^{(i)}\}$. $q_j^{(i)} = 1$, if the u_i has visited p_j, otherwise $q_j^{(i)} = 0$. We define $S(j,k)$ as the relevant score of p_j for u_k by summarizing the visiting information of N_u POIs in the city as

$$S(j,k) = \sum_{i=1}^{N_u} q_j^{(i)}. \tag{10}$$

Then we rank the POIs in P_2 according to the relevant scores S_j in descending order, and we recommend top-ranked POIs for the user u_k.

VI. Experiments

In this section, first we introduce the dataset and criteria of the experiments. We compare our proposed method (ATCF) with Recommendation by Popularity (PO), Collaborative Filtering (CF) and Recommendation by LDA (LDA) to evaluate the robustness of ATCF. Then we show the discussions of ATCF such as the "sparsity problem". At last, the example of a real Flickr user is provided.

The detailed descriptions of PO, CF, and LDA-based travel recommendation approaches are described as follows.

Recommendation by Popularity: All the POIs of one city are ranked according to degree of popularity. The popularity of each POI is measured by how many users have uploaded photos of this POI. Then to each user, we recommend the top ranked POIs. To all users, the recommendation results are the same ones.

Collaborative Filtering: Location-based Collaborative Filtering (LCF) is the most popular method that can be easily realized [28], [32]. First a user-POI matrix is generated to record user travel history. Then for a given two users, their similarity is calculated by the cosine standard measurement towards their corresponding vectors in the user-POI matrix. Finally, locations are recommended based on similar users' visiting histories.

Recommendation by LDA: We replace the ATM with the LDA model [2] to mine user travel topic preferences. Unlike ATM, in LDA, we need to carry out an additional step to get a user's topic distribution. In the first step, all the tags of the photos in the city are allocated to different topics using the LDA model [2]. However, in this step the relationship between authors and words, and authors and documents are not yet considered. Therefore, in the second step, we calculate the proportion of user's tags allocated to each topic, which is mined in the first step. The other settings of the LDA-based approach are the same as ATM.

A. Dataset

We have collected more than 7 million Flickr photos through Flickr's open API. These photos are uploaded by 7,387 users and heterogeneous metadata are associated with the photos [17], [24].

We only retain photos with both tags and geo-tags from the original Flickr dataset. Though only tags are used to mine a user's travel preferences, geo-tags are also important to the recommendation system and for evaluating experiments. On one hand, in the offline module, geo-tags are involved in "city level POIs mining" and "users' travel history mining". On the other hand, in the experiments, the geo-tags, which users labeled originally, are regarded as the ground truth of what the user has actually visited.

We select nine popular cities to evaluate travel recommendation performances [4]. Actually, in [4], there are 10 cities in total. However, the data we crawl from Honolulu are far less than the other 9 cities, so we remove this city. These nine cities are *Barcelona*, *Berlin*, *Chicago*, *London*, *Los Angeles*, *New York*, *Pairs*, *Rome* and *San Francisco*. We use the method introduced in Section III-B. to mine POIs of these nine cities. Table I shows the corresponding number of users, POIs and photos in each city. There are 2,892 users, 307 POIs and 150,101 photos in total.

We further select users who have visited at least two cities among these nine cities. Each user should have more than 5 photos with geo-tags and tags. These two cities are defined as a city group $< \chi_1, \chi_2 >$. The label of the city conforms to Table I. For example, city group $< 1, 2 >$ means that the photos are uploaded by the users who have visited both Barcelona and Berlin. After filtering, there are 1405 users conform to the requirements. In order to find enough similar users, we further remove the city group whose numbers of users are less than 20.

Finally, there are 23 city groups and 1156 users retained. Actually, to each city group $< \chi_1, \chi_2 >$, we could both use the history of χ_1 to do recommendation for χ_2 and use the history of χ_2 to do recommendation for χ_1. So the number of city group is doubled and there are 46 city groups in total. Thus, the processing of the test dataset is completed. Fig. 5 and Fig. 6 show the final number of users and photos in each city group in our test dataset.

TABLE I
NUMBER OF USERS, POIS, AND PHOTOS IN EACH CITY

No	Name	User Num	POI Num	Photo Num
1	Barcelona	222	31	10,534
2	Berlin	226	32	11,083
3	Chicago	203	28	12,304
4	London	568	47	33,838
5	Los Angeles	118	25	4,907
6	New York	527	37	31,401
7	Pairs	524	37	20,196
8	Rome	204	34	7,594
9	San Francisco	300	35	18,244

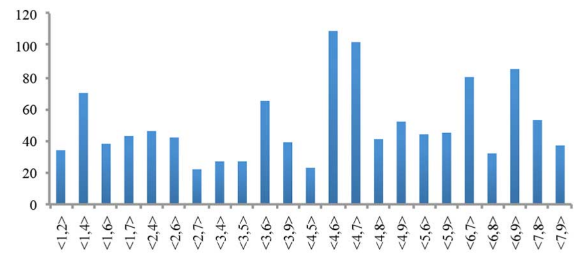


Fig. 5. Number of users in the city group.

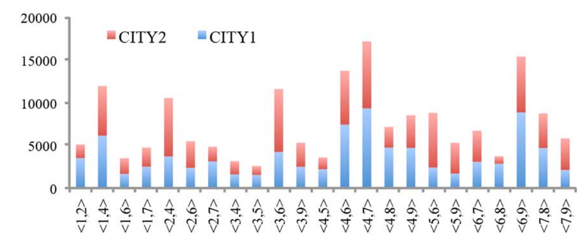


Fig. 6. Number of photos in each city group. The blue part of each bar is the number of photos of city1 uploaded by the users and the red part is of city2.

B. Ground Truth and Criteria

Ground Truth: We use the user's history in χ_1 to predict which POIs he or she would visit in χ_2. And compare the recommendation results of different methods with what the user actually visited in χ_2. As described in Section VI-B, all the user photos contain both tags and geo-tags. We only use tags as the online input to do recommendation. And geo-tags, which records which POIs the user actually visited, are regarded as the ground truth. We compare the POIs recommended by different methods and the ground truth to measure the performance of each method.

Criteria: We use MAP@n [4] to estimate the performance of our method and four comparative methods. n denotes the number of POIs which we recommend to the user. MAP@n is the mean average precision for a set of m users in the test data as

$$MAP@n = \left(\sum_{i=1}^{m} AP_i\right) / m \tag{11}$$

TABLE II
POI RECOMMENDATION PERFORMANCES OF PO, CF, LDA, AND ATCF

Performance	PO	CF	LDA	ATCF
MAP	0.3408	0.4137	0.4166	**0.4225**
MAP@1	0.4861	0.5595	0.5678	**0.5876**
MAP@5	0.3557	0.4312	0.4361	**0.4483**
MAP@10	0.3076	0.4059	0.4005	**0.4115**
MAP@20	0.2642	0.3519	**0.3545**	**0.3545**
MAP@30	0.2438	0.3151	0.3163	**0.3184**

where AP_i is Average Precision as

$$AP@n = \left(\sum_{i=1}^{n} \left(\sum_{j=1}^{i} rel_j\right) / i\right) / n \qquad (12)$$

where rel_i is a relevance value. $rel_i = 1$ if user have actually visited the recommended POI. Otherwise, $rel_i = 0$.

We also provide the performance under MAP (without@n). In MAP, the number of recommended POIs is the same as the number that the user actually visited.

C. Performance

Table II shows the MAP and MAP@n of PO, CF, LDA and ATCF respectively. Note that for all four methods, we use the coarse-to-fine method to mine POIs and user travel history, where ε is set as 0.01. In CF, LDA and ATCF, the number of similar users is set to $N_s = 40$, and the distance metric is the cosine distance. In LDA and ATCF, the number of topics N_z is set to 50.

The performance on MAP of ATCF is 0.4225, which outperforms PO, CF, and LDA by 8.17%, 0.88%, and 0.59% respectively. As on MAP, the number of recommended POIs is the same as those the user has actually visited. These results could better reflect the performance of different methods.

Table II also shows performance under MAP@n with $n = 1, 5, 10, 20$ and 30 for PO, CF, LDA, and ATCF. We can see the performances of ATCF and LDA are higher than PO and CF. We observe that when n increases, the performance of all methods decreases. As most users visit nearly five POIs in a city, if n is too large, the proportion of relevant POIs among all the recommended POIs is declined.

D. Discussion of the Impacts of Coarse-to-Fine POIs and User History Mining

In order to evaluate the performance of the coarse-to-fine method, we compare the performance of ATCF under three cases: M1, M2 and M3. The impact of ε- ball and the threshold of stop time t_ν of coarse-to-fine mapping are also discussed in this section. The methods of M1, M2, and M3 are described as follows.

M1: Both the POIs and user travel history are mined based on the coarse-to-fine approach.

M2: POIs are mined by a coarse-to-fine approach. User travel history is calculated by directly comparing the geo-tag of each image with the geo-center of all the POIs, instead of applying coarse-to-fine method.

M3: Neither POI and user travel history mining use the coarse-to-fine approach. We directly regard the clusters which contain more than 20 images uploaded by more than 10 users as POIs.

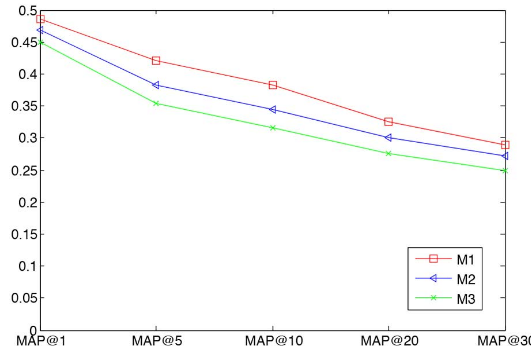


Fig. 7. Performance of ATCF under M1, M2, and M3.

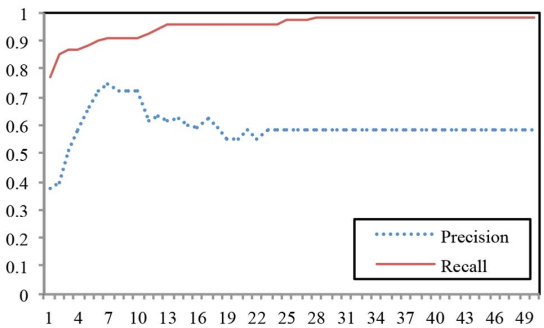


Fig. 8. Precision and recall of POI mining with the x-axis is ν_t in the range [1, 50].

1) Performances of ATCF Under M1, M2 and M3: In Fig. 7, we offer performance of ATCF under M1, M2 and M3 under MAP@n with $n = 1, 5, 10, 20$ and 30. The settings of other parameters are the same as Section VI-C. For all three methods, we do not use the ε- ball that constraints the geo-distance in user history mining. We set $\varepsilon = 0$.

From Fig. 7, we can see that under various n of. MAP@n., the performance of M1 is highest. M1 is almost.2%-4% higher than M2 and 4% -6% higher than M3. It proves that the coarse-to-fine user history mining method is effective. Also, M2 is higher than M3, which proves that the coarse-to-fine POI mining method is also effective.

2) Impacts of ν_t in the Coarse-to-fine Method: ν_t is the threshold of t_{rej} in POI mining. ν_t is set to be 8 in the baseline method. In this section, we discuss the impact of ν_t. We use precision and recall to measure the performance of POI mining under different ν_t.

From Fig. 8, we find that recall is higher than 0.9, when $\nu_t \leq 8$, indicating that only 10% clusters relevant to the POIs haven't been allocated to these POIs. Meanwhile, it greatly decreases the time consumed by almost 30%. The main reason is that if the nearest 8 clusters do not belong to the same POI, the calculation of this POI will stop.

Fig. 8 also shows that precision reaches the highest value at $\nu_t = 7$, which is 0.7465. When $\nu_t > 25$, precision reaches a stable level of 0.5812, meaning that the effect of ν_t decreases when ν_t is too large. When $\nu_t < 4$, precision is lower than 0.5812. There are many noise clusters around POIs, making it

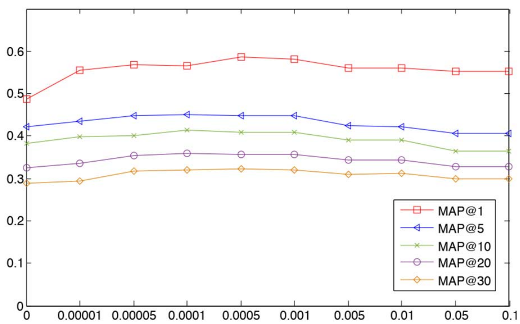


Fig. 9. Performance of ATCF on MAP@n under ε (x-axis) ranging from 0 to 0.1.

improper to stop the process if there are only one or two clusters nearby that don't belong to the POI ($\nu_t = 1$ or 2). However, when ν_t is around 8, better performances are achieved.

3) Impacts of ε in Coarse-to-Fine Method : ε is the threshold for coarse-to-fine based user travel history mining. When $\varepsilon = 0$, all images of users would be mapped to the cluster whose center is the closest to the geo-tag of the image. In Fig. 9, we show the result of ε ranging from 0-0.1. We can see that better performances are achieved when ε ranges from 0.0001-0.001. It meets the radiuses of almost all POIs. If the distances between a photo and all the centers of POIs are larger than the radii of POIs, the image is more likely to be a noisy image, instead of the photo of POIs.

E. Impacts of Different Similarity Measurements and Number of Similar Users

In the section, we discuss the impact of the number of similar users N_s and the standard of distance metric to the performances of ATCF. In our baseline algorithm, we use 'cosine standard' to measure the similarity and set $N_s = 40$. Here, we evaluate the performances of ATCF under the following distance metric standards: 'cosine' (denoted as CO), 'cityblock'(denoted as CI), 'hamming'(denoted as HA), and 'euclidean'(denoted as EU) with various numbers of similar users N_s. We utilize the ε-ball constraint and set it as 0.01. We take the dataset of $<$ London, New York $>$ to test performance. Fig. 10 shows performance of our ATCF method under CO, CI, HA, and EU. The topic number N_z is set to be 50 and NS ranges from 10 to 100 at an interval of 10. Fig. 10 shows the MAP curves of ATCF under N_s from 10 to 100 with different similarity measurement methods: EU, HA, CI, and CO.

Fig. 10 demonstrates that the difference in performance of ATCF under different similarity measurement approaches is not obvious. When $N_s > 30$, the differences are less than 10%. When $N_s < 30$, the performances increase violently when N_s increases. When $N_s > 40$, performance is gradually stable. From Fig. 10, we can also see that when N_s is very small like $N_s = 10$ or 20, the performance is not good. We see that as travel data is very sparse, the number of POIs that similar users have visited is not large, and thus it is difficult to recommend relevant POIs. When $N_s > 40$, the increase of N_s does not impact the performance significantly. This phenomenon shows that finding enough relevant users can improve recommendation performances.

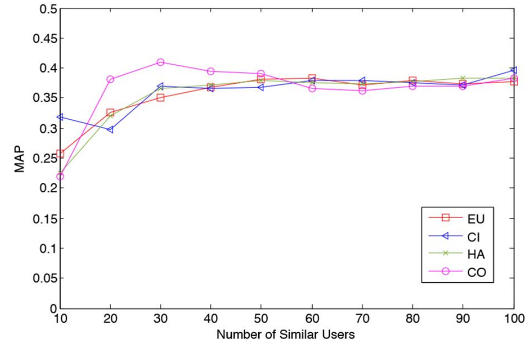


Fig. 10. MAP curves of ATCF under various number of similar users with different similarity measurement methods: EU, HA, CI, and CO.

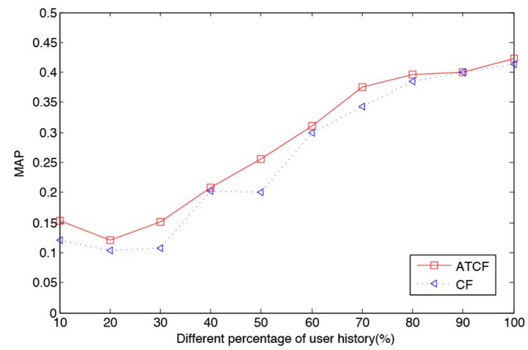


Fig. 11. MAP curves of CF and ATCF under sampled user travel history of different percentages.

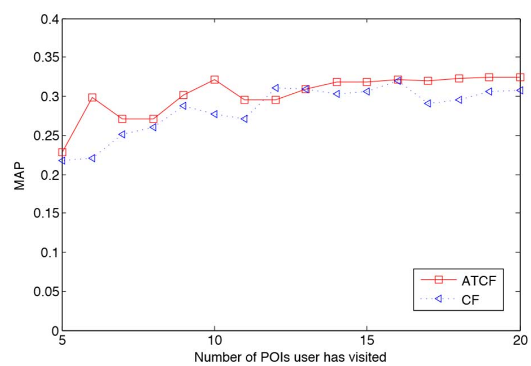


Fig. 12. MAP curves of CF and ATCF under users with sparser travel history.

F. Discussion of the Sparsity Problem

We conduct two experiments in order to evaluate the robustness of ATCF under "sparsity" conditions. In the first experiment, we randomly sample POIs from the user travel history we mine in the baseline experiment. User travel history in this experiment will be much sparser than the baseline experiment. In Fig. 11, the x-coordinate means the proportion that we sampled towards the user travel history.

In the second experiment, we select the users whose travel histories are much sparser to construct the experimental dataset. In Fig. 12, the x-coordinate means users whose number of visited POIs is less than the certain value would be selected as experimental dataset. In both experiments, $\varepsilon = 0.01$, $N_s = 40$, $N_z = 50$, and the distance metric is the cosine distance.

TABLE III
FAMOUS AND NON-FAMOUS POIS IN PARIS

Latitude	Longitude	POI Names
Five Famous POIs		
48.8577	2.2953	Eiffel tower
48.8422	2.3472	Notredame
48.8604	2.3387	Louvre
48.8601	2.3260	Museedorsay
48.8865	2.3406	Montmartre
Five Non-Famous POIs		
48.862	2.3473	Eglise saint-eustache
48.863	2.3295	Tuileries gardens
48.872	2.3003	Champs Elyses
48.849	2.3566	Mosque grande
8.8552	2.3125	Les invalides

TABLE IV
POIS IN NEW YORK

ID	POI	ID	POI
1	Brooklyn Bridge	20	Stature of Liberty
2	Washington Square Park	21	Ellis Island
3	Times Square	22	Coney Island
4	Dakota Building	23	Metropolitan Museum of Art
5	Madison Square Park	24	Apple Store
6	Guggenheim Museum	25	Fort Greene Park
7	Natural History Museum	28	Tompkins Square Park
8	Central Park	27	Williams Burg
9	Metropolitan Museum	28	Tompkins Square Park
10	Central Park Zoo	29	High Line Park
11	Museum of Modern Art	30	Flatiron Building
12	City Hall Theater	31	Korea Town
13	Union Square Theater	32	Empire State Building
14	Little Italy	33	Midtown
15	China Town	34	Rockefeller Center
16	World Trade Center	35	New York Central Station
17	St Pauls Chapel	36	NYC Apartment View
18	Trinity Church	37	New York Public Library
19	New York Stock Exchange		

From Fig. 11 and Fig. 12, we find that the performances of both CF and ATCF decrease when the data is much sparser. Under the "sparsity" condition, the performances of ATCF is higher than CF in the two experiments.

G. Example POIs in Two Cities

Table III shows ten example POIs with their latitude and longitude we mine in Paris. The top ranked five POIs are famous POIs, such as Eiffel Tower, Notredame, Louvre, Musee Dorsay, and Montmartre respectively. They could also be mined by basic mean-shift clustering under the bandwidth of 0.001 set as the paper [3]. However, by [3], non-famous POIs can not be mined, as shown in the last five rows, such as Eglise Saint-eustache. In our proposed "coarse-to-fine POI mining" approach, not only the famous but also the non-famous POIs can be mined. This shows that the coarse-to-fine POI mining method is better than the basic mean-shift clustering method [3].

In Table IV, we show the mined 37 POIs in New York with POI ID and POI names, including "Brooklyn bridge", "Washington Square park", "Times Square", and etc. This table is also used in Section VI-I.

H. Example Topics of ATM

Table V illustrates three example topics, which are selected from 100 topics (i.e. the topic number $N_z = 100$) learned by ATM in London. Each topic is illustrated with its top 10 words with the highest probabilities such as "Thames" and "River", and the probabilities of these words (in short PROB).

TABLE V
ILLUSTRATION OF THREE MINED TOPICS IN LONDON

TOPIC 76		TOPIC 85		TOPIC 41	
WORD	**PROB**	**WORD**	**PROB**	**WORD**	**PROB**
Thames	0.28421	Holiday	0.03955	Flashmob	0.03531
River	0.08950	Fish	0.03921	Underground	0.03399
Londoneye	0.04950	Beach	0.03887	City	0.03377
Tower	0.04890	Water	0.03870	Drinks	0.03268
Stpauls	0.04298	Hotel	0.03785	Youth	0.03246
Bigben	0.03468	Sand	0.03684	Tube	0.03202
Parliament	0.02075	Food	0.03650	Party	0.03136
Cathedral	0.02016	Flying	0.03650	Drinking	0.03114
Light	0.01927	Sun	0.03633	Alcohol	0.03114
Trafalgar	0.01867	Fun	0.036616	Cocktail	0.03004

Fig. 13. Photos with tags of the Flickr users.

TABLE VI
POIS RECOMMENDED UNDER GT, PO, CF, LDA, AND TPM FOR USER "15960635@N05"

	POI ID
GT	20, 32, 34, 35, 4, 17, 21, 11, 26, 3
PO	**3**, **34**, 27, 1, **20**, 28, 2, 14, 15, 19
CF	**3** ,1, **35**, **34**, **20**, 24, **11**, 10, 2, **32**
LDA	**3**, **20**, **35**, **32**, 10, 1, **34**, 24, **11**, 37
ATCF	**3**, **20**, **32**, **34**, **35**, 1, 24, 11, 4, 21

From Table V we can see that some topics correspond to ordinary categories like "museum", "beach" or "mountain". For example, in Topic 41, words like "alcohol" and "drinks" are related to "bar" category. Also, we see that some topics extracted with ATM do not like ordinary categories. For example, in Topic 76, there are some landmarks near the "Thames River" like "Stpauls" and "Bigben". As these landmarks are quite near each other, users may visit them together when they are traveling to London. It is also rational that these words often appear at the same time in the textual descriptions of their photos.

I. Example of a Flickr User

In this section, we show an example of a real Flickr user. The user's ID is "15960635@N05". This user has visited both London and New York. We show four example photos with corresponding tags in his or her album in London in Fig. 13. The top four POIs in London detected by the system of this user are "Canary Wharf", "London Aquarium", "Tower Bridge" and "St Paul's Cathedral". They are compliant with the POIs which we evaluated ourself from his album.

In the experiment, we recommend POIs in New York for a user based on his or her travel history in London. What the user actually visits is shown in the row "Ground Truth" in Table VI. In Table VI, we show the IDs of POIs in New York recommended for this user under different methods: PO, CF, LDA, and ATCF. The corresponding 37 names of these IDs are shown in Table III in Part VI-G. As the user has visited

10 POIs, we show the top 10 POIs we recommended of each method. Bold font is used to mark POIs that also appear in ground truth.

According to Table VI, among PO, CF, LDA, and ATCF, ATCF has the best performance. It recommends 8 POIs that appear in the ground truth. Both CF and LDA recommend 6 POIs in the ground truth but the rank orders are different. To PO, only 3 POIs are in the ground truth, which shows that personalized recommendations outperform generalized recommendations.

VII. Conclusions

In this paper, we proposed an author topic model-based collaborative filtering (ATCF) method for personalized travel recommendations. User's topic preference can be mined from the textual descriptions attached with his/her photos via author topic model (ATM). Through ATM, travel topics and a user's topic preference can be elicited simultaneously. In ATCF, POIs are ranked according to similar users, who share similar travel topic preferences, instead of raw GPS (geo-tag) data as is the case of most previous works. Unlike location-based collaborative filtering, even without GPS records, similar users can still be mined accurately according to the similarity of users' topic preferences. What's more, the coarse-to-fine city level POIs and user history mining approaches are both contributive.

In future work, we will first deeply combine the tags and GPS coordinates to mine user travel preferences. In our current work, we use the textual information of geo-tag photos to carry out travel recommendations. The geo-tags are only served as constraints in our recommendation. Combining tags and geo-tags would be an interesting and challenging task. What's more, we continue enlarging our dataset, especially by adding some less famous places. To less famous places, the data would be even sparser and noisier, which poses an even greater research challenge.

References

[1] J. Bao, Y. Zheng, and M. F. Mokbel, "Location-based and preference-aware recommendation using sparse geo-social networking data," in *Proc. 20th Int. Conf. Adv. Geographic Inf. Syst.*, 2012, pp. 199–208.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[3] A. Cheng, Y. Chen, Y. Huang, W. Hsu, and H. Liao, "Personalized travel recommendation by mining people attributes from community-contributed photos," in *Proc. 19th ACM Int. Conf. Multimedia*, 2011, pp. 83–92.

[4] M. Clements, P. Serdyukov, A. de Vries, and M. Reinders, "Using flickr geo-tags to predict user travel behaviour," in *Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2010, pp. 851–852.

[5] H. Feng and X. Qian, "Mining user-contributed photos for personalized product recommendation," *Neurocomput.*, vol. 129, pp. 409–420, 2014.

[6] H. Gao, J. Tang, X. Hu, and H. Liu, "Content-aware point of interest recommendation on location-based social networks," in *Proc. 29th Int. Conf. AAAI*, Austin, TX, USA, Jan. 2015, pp. 1–7.

[7] Y. Gao, J. Tang, R. Hong, Q. Dai, T. Chua, and R. Jain, "W2go: A travel guidance system by automatic landmark ranking," in *Proc. Int. Conf. Multimedia*, 2010, pp. 123–132.

[8] H. Huang and G. Gartner, "Using trajectories for collaborative filtering–based poi recommendation," *Int. J. Data Mining, Model., Manage.*, vol. 6, no. 4, pp. 333–346, 2014.

[9] S. Jiang, X. Qian, J. Shen, and T. Mei, "Travel recommendation via author topic model based collaborative filtering," in *Proc. 21st Int. Conf., MMM 2015*, Sydney, N.S.W., Australia, 2015, pp. 392–402.

[10] S. Jiang, X. Qian, Y. Xue, F. Li, and X. Hou, "Generating representative images for landmark by discovering high frequency shooting locations from community-contributed photos," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, Jul. 2013, pp. 1–6.

[11] H. Kori, S. Hattori, T. Tezuka, and K. Tanaka, "Automatic generation of multimedia tour guide from local blogs," *Adv. Multimedia Model.*, vol. 4351, no. 1, pp. 690–699, 2006.

[12] R. Krestel, P. Fankhauser, and W. Nejdl, "Latent Dirichlet allocation for tag recommendation," in *Proc. 3rd ACM Conf. Recommender Syst.*, 2009, pp. 61–68.

[13] T. Kurashima, T. Iwata, G. Irie, and K. Fujimura, "Travel route recommendation using geo-tags in photo sharing sites," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage.*, 2010, pp. 579–588.

[14] T. Kurashima, T. Tezuka, and K. Tanaka, "Mining and visualizing local experiences from blog entries," in *Database and Expert System Applications*. New York, NY, USA: Springer, 2006, pp. 213–222.

[15] J. Li, X. Qian, Y. Y. Tang, L. Yang, and T. Mei, "GPS estimation for places of interest from social users' uploaded photos," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 2058–2071, Dec. 2013.

[16] Q. Liu, E. Chen, H. Xiong, C. H. Ding, and J. Chen, "Enhancing collaborative filtering by user interest expansion via personalized ranking," *IEEE Trans. Cybern.*, vol. 42, no. 1, pp. 218–233, Feb. 2012.

[17] X. Liu, X. Qian, D. Lu, X. Hou, and L. Wang, "Personalized tag recommendation for Flickr users," in *IEEE Int. Conf. Multimedia Expo*, Jul. 2014, pp. 1–6.

[18] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th Int. Conf. Comput. Vis.*, 1999, vol. 2, pp. 1150–1157.

[19] W. Luo, H. Li, and G. Liu, "Automatic annotation of multispectral satellite images using author–topic model," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 634–638, Jul. 2012.

[20] Y. Lyu, C.-Y. Chow, R. Wang, and V. C. Lee, "Using multi-criteria decision making for personalized point-of-interest recommendations," in *Proc. SIGSPATIAL*, Dallas, TX, USA, Nov. 2014, pp. 1–4.

[21] M. Pennacchiotti and S. Gurumurthy, "Investigating topic models for social media user recommendation," in *Proc. 20th Int. Conf. Companion WWW*, 2011, pp. 101–102.

[22] X. Qian, X.-S. Hua, Y. Tang, and T. Mei, "Social image tagging with diverse semantics," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2493–2508, Dec. 2014.

[23] X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized recommendation combining user interest and social circle," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 7, pp. 1763–1777, Jul 2014.

[24] X. Qian, X. Liu, C. Zheng, Y. Du, and X. Hou, "Tagging photos using users' vocabularies," *Neurocomput.*, vol. 111, pp. 144–153, 2013.

[25] D. Ramage, S. T. Dumais, and D. J. Liebling, "Characterizing microblogs with topic models," *ICWSM*, vol. 5, no. 4, pp. 130–137, 2010.

[26] M. Rosen-Zvi, C. Chemudugunta, T. Griffiths, P. Smyth, and M. Steyvers, "Learning author-topic models from text corpora," *ACM Trans. Inf. Syst.*, vol. 28, no. 1, pp. 1–38, 2010.

[27] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth, "The author-topic model for authors and documents," in *Proc. 20th Conf. Uncertainty AI*, 2004, pp. 487–494.

[28] J. Sang, T. Mei, J.-T. Sun, C. Xu, and S. Li, "Probabilistic sequential POIs recommendation via check-in data," in *Proc. 20th Int. Conf. Adv. Geographic Inf. Syst.*, 2012, pp. 402–405.

[29] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proc. 10th Int. Conf. WWW*, 2001, pp. 285–295.

[30] Y. Shi, P. Serdyukov, A. Hanjalic, and M. Larson, "Personalized landmark recommendation based on geo-tags from photo sharing sites," *ICWSM*, vol. 11, pp. 622–625, 2011.

[31] M. Steyvers, P. Smyth, M. Rosen-Zvi, and T. Griffiths, "Probabilistic author-topic models for information discovery," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 306–315.

[32] L.-Y. Wei, Y. Zheng, and W.-C. Peng, "Constructing popular routes from uncertain trajectories," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 195–203.

[33] X. Wei and W. B. Croft, "LDA-based document models for ad-hoc retrieval," in *Proc. 29th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2006, pp. 178–185.

[34] W. J. Wilbur and K. Sirotkin, "The automatic identification of stop words," *J. Inf. Sci.*, vol. 18, no. 1, pp. 45–55, 1992.

[35] Y. Xue, X. Qian, and B. Zhang, "Mobile image retrieval using multi-photos as query," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, Jul. 2013, pp. 1–4.

[36] H. Yin, C. Wang, N. Yu, and L. Zhang, "Trip mining and recommendation from geo-tagged photos," in *Proc. IEEE Int. Conf. Multimedia and Expo Workshops*, Jul. 2012, pp. 540–545.

[37] Q. Yuan, G. Cong, and A. Sun, "Graph-based point-of-interest recommendation with geographical and temporal influences," in *Proc. 23rd ACM Int. Conf. Inf. Knowl. Manage.*, 2014, pp. 659–668.

[38] C. Zhang and K. Wang, "POI recommendation through cross-region collaborative filtering," *Knowl. Inf. Syst.*, pp. 1–19, 2015.

[39] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang, "Collaborative location and activity recommendations with gps history data," in *Proc. 19th Int. Conf. WWW*, 2010, pp. 1029–1038.

[40] Y. Zheng, L. Zhang, X. Xie, and W. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *Proc. 18th Int. Conf. WWW*, 2009, pp. 791–800.

Shuhui Jiang received the B.S. and M.S. degrees from Xi'an Jiaotong University, Xi'an, China, in 2007 and 2011, respectively, and is currently working toward the Ph.D. degree in electrical and computer engineering from Northeastern University, Boston, MA, USA.

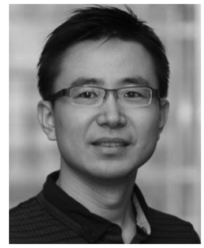
Xueming Qian (M'09) received the B.S. and M.S. degrees from the Xi'an University of Technology, Xi'an, China, in 1999 and 2004, respectively, and the Ph.D. degree from the School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, China, in 2008.

From 1999 to 2001, he was an Assistant Engineer with Shannxi Daily, Xi'an, China. He was a Visiting Scholar with Microsoft Research Asia, Beijing, China, from August 2010 to March 2011. From 2008 to 2011, he was an Assistant Professor with the School of Electronics and Information Engineering, Xi'an Jiaotong University, where he has been a Full Professor since 2014. He is also the Director of the SMILES LAB, Xi'an, China. He has authored or coauthored over 70 journal and conference papers. His research interests include social mobile multimedia mining learning and search.

Dr. Qian was a TPC Member of ICME, Multimedia Modeling, ICIMCS, and is the Session Chair/Organizer of VIE '08, ICME '14, and MMM '14. He is a Member of the ACM, and a Senior Member of CCF. He was awarded a Microsoft Fellowship in 2006.

Jialie Shen (M'07) received the Ph.D. degree in computer science (in the area of large-scale media retrieval and database access methods) from the University of New South Wales, Sydney, N.S.W., Australia.

He is an Assistant Professor with the School of Information Systems, Singapore Management University, Singapore. His recent work has been published or is forthcoming in various leading journals and international conferences. His main research interests include information retrieval in the textual and multimedia domain, economic-aware media analysis, and multimedia systems.



Yun Fu (S'06–M'08–SM'11) received the B.Eng. degree in information engineering and the M.Eng. degree in pattern recognition and intelligence systems from Xi'an Jiaotong University, Xi'an, China, and the M.S. degree in statistics and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA.

From 2008 to 2010, he was a Scientist with BBN Technologies, Cambridge, MA, USA. He was a Part-Time Lecturer with the Department of Computer Science, Tufts University, Medford, MA, USA, in 2009. He was a tenure-track Assistant Professor with the Department of Computer Science and Engineering, State University of New York, Buffalo, NY, USA, from 2010 to 2012. Since 2012, he has been an interdisciplinary Faculty Member with the College of Engineering and the College of Computer and Information Science, Northeastern University, Boston, MA, USA. He has authored or coauthored papers appearing in leading journals, books/book chapters, and international conferences/workshops. His research interests include interdisciplinary research in machine learning and computational intelligence, social media analytics, human–computer interaction, and cyber-physical systems.

Dr. Fu is a Lifetime Member of the ACM, AAAI, SPIE, and Institute of Mathematical Statistics, a Member of INNS, and was a Beckman Graduate Fellow from 2007 to 2008. He serves as Associate Editor, Chair, PC Member, and Reviewer of many top journals and international conferences/workshops. He is currently an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEANING SYSTEMS and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He was the recipient of five Best Paper Awards (SIAM SDM 2014, IEEE FG 2013, IEEE ICDM-LSVA 2011, IAPR ICFHR 2010, and IEEE ICIP 2007), three Young Investigator Awards (2014 ONR Young Investigator Award, 2014 ARO Young Investigator Award, 2014 INNS Young Investigator Award), two Service Awards (2012 IEEE TCSVT Best Associate Editor, 2011 IEEE ICME Best Reviewer), the 2011 IC Postdoctoral Research Fellowship Award, the 2010 Google Faculty Research Award, the 2008 M. E. Van Valkenburg Graduate Research Award, the 2007–2008 Beckman Graduate Fellowship, the 2007 Chinese Government Award for Outstanding Self- Financed Students Abroad, the 2003 Hewlett-Packard Silver Medal and Science Scholarship, multiple Edison Cups of the 2002 GE Fund Edison Cup Technology Innovation Competition, and the 2002 Rockwell Automation Master of Science Award.

Tao Mei (M'06–SM'11) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2001 and 2006, respectively.

He is a Researcher with Microsoft Research Asia, Beijing, China. He has authored or coauthored over 140 papers in journals and conferences and eight book chapters, and has edited two books. He holds six U.S. patents with more than 30 pending. His current research interests include multimedia information retrieval and computer vision.

Dr. Mei is a Senior Member of the ACM. He is an Associate Editor of *Neurocomputing* and the *Journal of Multimedia*, a Guest Editor of the IEEE TRANSACTIONS ON MULTIMEDIA, the *IEEE MultiMedia*, *ACM/Springer Multimedia Systems*, and the *Journal of Visual Communication and Image Representation*. He is the Program Co-Chair of MMM 2013 and the General Co-Chair of ACM ICIMCS 2013. He was the recipient of several awards from prestigious multimedia conferences, including the Best Paper Award and the Best Demonstration Award at ACM Multimedia 2007, the Best Poster Paper Award at the IEEE MMSP 2008, the Best Paper Award at ACM Multimedia 2009, the Top 10% Paper Award at the IEEE MMSP 2012, the Best Paper Award at ACM ICIMCS 2012, the Best Student Paper Award at the IEEE VCIP 2012, and the Best Paper Finalist at ACM Multimedia 2012. He was the Principle Designer of the automatic video search system that achieved the best performance in the worldwide TRECVID evaluation in 2007. He was the recipient of the Microsoft Gold Star Award in 2010 and the Microsoft Technology Transfer Award in 2010 and 2012.