Travel Route Recommendation Using Geotags in Photo Sharing Sites

Takeshi Kurashima*, Tomoharu Iwata+, Go Irie* and Ko Fujimura*
*NTT Cyber Solutions Labs., NTT Corporation
+NTT Communication Science Labs., NTT Corporation
*1-1 Hikari-no-oka, Yokosuka-Shi, Kanagawa, 239-0847 Japan
+2-4 Hikaridai, Seika-Cho, Soraku-gun, Kyoto, 619-0237 Japan
*{kurashima.takeshi,irie.go,fujimura.ko}@lab.ntt.co.jp +iwata@cslab.kecl.ntt.co.jp

ABSTRACT

The ability to create geotagged photos enables people to share their personal experiences as tourists at specific locations and times. Assuming that the collection of each photographer's geotagged photos is a sequence of visited locations, photo-sharing sites are important sources for gathering the location histories of tourists. By following their location sequences, we can find representative and diverse travel routes that link key landmarks. In this paper, we propose a travel route recommendation method that makes use of the photographers' histories as held by Flickr. Recommendations are performed by our photographer behavior model, which estimates the probability of a photographer visiting a landmark. We incorporate user preference and present location information into the probabilistic behavior model by combining topic models and Markov models. We demonstrate the effectiveness of the proposed method using a real-life dataset holding information from 71,718 photographers taken in the United States in terms of the prediction accuracy of travel behavior.

Categories and Subject Descriptors

H.4.m [**Information Systems**]: Miscellaneous; H.5 [**Information Interface and Presentation**]: General

General Terms

Algorithms, Experimentation

Keywords

Geolocation, Geo-referenced photographs, Travel route recommendation, Photographer behavior model

1. INTRODUCTION

Due to the proliferation of small digital cameras and mobile phone cameras, there has been great interest in online photo sharing services such as Flickr [1] and Google Picasa [2]. These services allow users to upload photographs and attach informative tags to them, and they have succeeded in collecting large-scale-sets of tagged photographs from huge numbers of users. *Geotag*s indicate where the photographs were taken, and are automatically captured by the aforementioned photo devices or location-aware devices, or alternatively are specified by the user. Geotags are powerful metadata for introducing spatial information into Web applications; recent research efforts have shown the potential of geotags by developing various geotag-associated applications such as georeferenced image search [7, 8], automatic image geolocation [9, 10, 11], and georeferenced content browsing [15, 16, 17, 18, 19].

This paper focuses on describing a travel route recommendation method that utilizes geotagged images in photo sharing services. By sorting the photographs by their timestamps for each user, the geotagged photographs yield personal travel route histories. The value of these travel route histories in geotagged photographs can be emphasized in terms of their *quality* and *quantity*. The photographed geolocations are good recommendations in terms of finding attractive locations because the action of uploading a photograph can be regarded as a positive vote that the location is worth visiting. Moreover, according to our research, at least 40,000,000 geotagged photographs uploaded by over 400,000 users are available on Flickr. This tremendous volume of travel data is a rich source of travel routes that can match various user preferences.

The major research problem addressed in this paper is generating recommended travel routes from a given geotag collection. This consists of two sub-issues; one is learning the model of travel routes from the geotag dataset, and the other is generating recommended routes from the model. The desired properties of recommended travel routes are not only that they be easily accessible, but also that they suit the users' interest. We then propose a method that trains a probabilistic *photographer behavior model*; it takes account of both the user's current location and user's personal interest. We then formalize the sub-issue of recommended route generation. Given past and/or future itineraries and the amount of spare time to spend on future travel, our method outputs a set of customized landmark sequences that match the user's preference, and the user's present location and spare time. We present an effective algorithm for finding recommended travel routes based on the photographer behavior models.

We also implement an online application that helps the user to plan a new trip. Figure 1 shows the current user interface of the application. Given a location history including the current location and desired time to spend on future travel, this application suggests a set of travel routes that might suit the user's interest. Each tour plan consists of travel route (a sequence of one or more landmarks) and the travel time between landmarks. The total time satisfies the time condition given by the user. The results are ranked by the

CIKM'10, October 26–30, 2010, Toronto, Ontario, Canada.

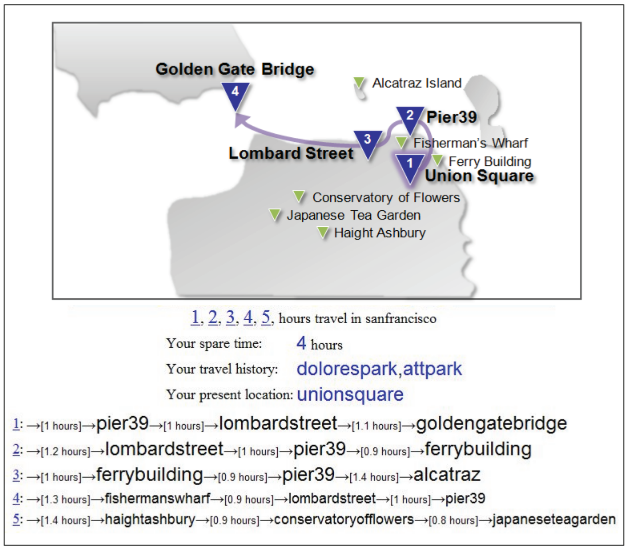


Figure 1: The interface of the route recommender system. Upon being given a location history including present location, and free time duration, returns a ranked list of travel plans. The best route is plotted on the map interface.

likelihood scores assigned by the photographer behavior model (in the lower part of Figure 1).

Our major contributions in this paper thus include:

- A model for travel route recommendation that takes account of both the user's current location and user's personal interest;
- A method and an application to provide a set of personalized travel routes that match the user's current location, user's interests, and user's spare time;
- Detailed evaluations of the performance achieved against an actual large-scale geotag dataset held by Flickr.

This paper is organized as follows: The next section describes related works. Section 3 defines the data model and our research problem and presents a method for travel route recommendation. Section 4 shows the effectiveness of the proposed method using Flickr data and Section 5 concludes this paper.

2. RELATED WORK

The important research areas related to our work are geolocation recommender systems and geotag-based applications. We study the latest reports that address these areas.

2.1 Recommender Systems

There are a number of local Web search systems that can recommend location-specific content to users. The main characteristic of these systems is that they can personalize their recommendations to the user. Given opening times and user ratings, Yahoo Travel enables the user to search for local information about sights to visit [3]. Horozov et al. proposed a collaborative-filtering-based method to recommend restaurants [12]. The difference between our work and theirs is that ours generates recommendations without explicit user ratings. The information used is a summary of knowledge acquired by photographers; such data includes landmarks, representative textual descriptions, photographs, routes, and travel times.

Data mining from GPS trajectories gathered by mobile devices is also related to the recommendation task because its main goal is to predict where a person may be going. Ashbrook et al. apply a Markov model to GPS data in an attempt to model traveler behavior; their traveler model always recommends the most chosen locations adjacent to the current location [26]. Krumm et al. also proposed a method for predicting the driver's destination based on multiple drivers' GPS trajectories [13]. Zheng et al. applied a graph mining method to a GPS dataset generated by 107 users in order to extract a region-of-interest and classical travel sequences between them [14]. Spatio-temporal sequential pattern mining from GPS trajectories was also performed by Giannotti et al. [24]. The major differences between our work and theirs lie in two aspects; one is that this work is the first to utilize geotagged photos in social photo-sharing sites for the purpose of route recommendation. Our method, which is based on large-scale personal databases automatically extracted from photo-sharing sites, dramatically increases the opportunity for the user to draw upon peoples' experiences. The other is that this work shows how to learn the photographer's personal interests from geotagged data; travel routes that are not only popular but also suitable for the user can be recommended.

2.2 Geotag-Associated Web Applications

Various types of Web applications such as image search [7, 8], content browsing [15, 16, 17, 18, 19], and geographic image annotation [11, 10, 9] leverage geotags in order to facilitate and to enrich photo browsing. Different from these existing applications, this work presents a new direction in the use of geotagged and time-stamped photographs, travel route recommendation. It goes beyond browsing to support human activities in the real world.

Image Search: In the current multimedia information retrieval research field, geolocation information is taking an important role in Web image search. Kennedy et al. proposed methods to extract representative and diverse location-specific photographs from Flickr based on tag co-occurrence, image similarity, and geolocation information [7, 8].

Content Browsing: Managing large-scale content collection based on geographic interface (e.g. Google Maps [4]) is a promising approach to facilitate georeferenced content browsing. Recent research has introduced methods that arrange landmark tags and geotagged images according to spatial coordinates [15, 16, 17, 18, 19]. One important work is *World Explorer* proposed in [15, 16]; it finds representative tags attached to location-specific photographs, and visualizes them on a map interface. Similar work performed by Crandall et al. allows users to browse all of the world's photographs on a single world map [17]. Snavely et al. proposed *Photo tourism* to enable full 3D navigation and exploration of a set of images [18, 19]. Their method reconstructs 3D structures of several landmarks from their related geotagged photos.

Geographic Image Annotation: A central goal is geolocating one or more Web images that have no original geotags. The key is to estimate geometric location by analyzing image content. [11] proposes a data-driven approach to exploit the Flickr geotagged photo collection. Their approach computes scene similarity between a target photo and each of the geotagged photos, and attaches the geotags of similar photos to the target photo. Different from the data-driven approach, the model-based approach assumes a probability (likelihood) model of image features conditioned by geometric location. Cao et al. proposed a method that uses kernel canonical correlation and logistic regression to enhance semantic annotation and geo-location information based on image content [10]. The method proposed in [9] introduces the photographers'

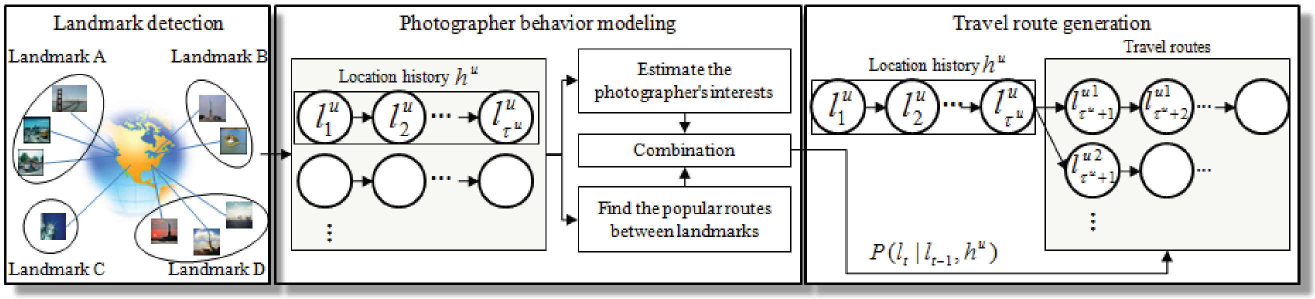


Figure 2: Travel route recommendation framework

Table 1: Notation

Symbol	Description
$\boldsymbol{U}$	set of users
u	user, $u \in \boldsymbol{U}$
l_i^u	geolocation of the ith photograph of user u represented by latitude and longitude coordinates in bi-axial space
t_i^u	universal time of the ith photograph of user u
v_i^u	set of annotation tags of the ith photograph of user u
τ^u	number of locations in the history of user u
h^u	location history of user u, where geolocations are sorted by their timestamps, $h^u = \langle l_1^u, \cdots, l_{\tau^u}^u \rangle$
K	number of routes to be recommended
T^{uk}	number of locations in the kth recommended route

travel sequence prior. Combinatorial use of the prior and image likelihood model offers better performance.

As far as we know, travel itinerary construction proposed in [20] is the only prior work that uses geotags and time-stamps for arranging a trip. The popularity of landmarks, stay times and transit times between landmarks are used to construct representative travel itineraries linking popular landmarks within a city. The popularity of each landmark is measured by the number of photos taken. The main difference between our work and theirs is that ours can recommend personalized travel routes/itineraries based on the interests of the individual (what kind of landmarks did he/she visit in the past). Personalization is not performed by [20].

3. TRAVEL ROUTE RECOMMENDATION METHOD

3.1 Formulation and Problem Definition

In this section, we describe the proposed recommendation method. Our approach is based on the fact that the data uploaded to photo sharing sites are significant information sources for finding landmarks and setting travel routes between them. By using these information sources, we can build photographer behavior models and apply them to make and recommend a plan to newly-arrived tourists. This recommendation is performed by understanding and modeling the processes that underlie a photographer's decision about what to do next throughout his/her journey.

Suppose that we have a set of geotagged and time-stamped photographs $\{(l_i^u, t_i^u, v_i^u)\}_{i=1}^{\tau^u}$ for each user in user set $\boldsymbol{U}$. Our notation is summarized in Table 1. Our research problem can be described as follows; given location history h^u and the amount of spare time d to spend on future travel, the task is to recommend a set of personalized travel routes which are sequences of geolocations $\{\langle l_{\tau^u+1}^{uk}, \cdots, l_{\tau^u+T^{uk}}^{uk} \rangle\}_{k=1}^{K}$ where the estimated time needed to fully follow a sequence is d.

We note that geolocation l_i^u refers to the location of the ith photo of user u, but it can also mark the geolocation of the photographed object. We also assume that the intervals between images are correct; the original image timestamps are not assumed to be correct because the camera's clock may not be set correctly.

The general approach of our travel route recommendation framework is illustrated in Figure 2. First, given a large collection of geotagged photos, we automatically find the landmarks that were photographed often. These landmarks are interesting in terms of route planning since they are assumed to represent a positive vote by previous visitors. Found landmarks are thus appropriate candidates for recommendation.

Second, we model the behavior of the photographer to estimate the probability that the photographer will visit a landmark. Our model is based on two assumptions;

1. the geolocation that the tourist will move to next largely depends on the current and recently visited locations.
2. the choice of what to do next is also determined by the traveler's interests.

For example, a traveler currently in Times Square is more likely to visit the Statue of Liberty than the Golden Gate Bridge in San Francisco because the Statue of Liberty is much closer. A traveler who is interested in art is likely to visit the Metropolitan Museum of Art while a traveler who is interested in sports is likely to visit Yankee Stadium.

Finally, we use the probabilistic photographer behavior model built in the previous step to recommend a set of landmark sequences. We recommend the route that is most likely to suit the user. The following subsection details each step.

3.2 Step 1: Photo Clustering by Mean-shift Procedure

Given a collection of photos that contain their geolocations, represented by latitude and longitude, we automatically extract often-photographed landmarks in the city because we want to recommend popular landmarks. We define a landmark as a uniquely represented specific location within the city; such as a sightseeing spot, a store, a building, a bridge, an outlet, and so on. Examples in the New York area include Union Square, Statue of Liberty, and Madison Square Garden.

The previous work by Crandall et al. [17] provides several techniques for the analysis of geotagged photo collections. They showed that the *mean-shift* procedure was very effective for landmark extraction from spatial data such as the information we consider here. Mean-shift is a non-parametric feature-space analysis technique that uses kernel density estimation and has been successfully applied to a wide range of vision and image processing applications. The issue tackled here can be regarded as the problem of clustering points in a two-dimensional feature space, latitude and longitude. The benefit of this approach is that it requires only the scale of observation, unlike most clustering methods that require choosing some number of clusters or making underlying distributional assumptions. Thus we use mean-shift clustering to mine often-photographed landmarks.

The method automatically estimates the modes of underlying unobservable probability distributions from a collection of observed geotagged photographs. The mode of each cluster indicates a landmark with high photo density. For given location l, the mean shift vector is defined as follows,

$$m_{w,g}(l) = \frac{\sum_u \sum_i l_i^u g\left(\| (l - l_i^u)/w\|^2\right)}{\sum_u \sum_i g\left(\| (l - l_i^u)/w\|^2\right)} - l, \tag{1}$$

where g is the weight of each data point corresponding to some chosen kernel function, and w is a bandwidth parameter. We adopted a Gaussian distribution function as the kernel function. The mean-shift procedure computes a sequence starting from some initial location $l^{(1)}$ where

$$l^{(c+1)} = l^{(c)} + m_{w,g}(l^{(c)}), \tag{2}$$

which converges to a location that corresponds to a local maximum of the underlying distribution as the mean-shift vector approaches zero. A single landmark thus can be regarded as a virtual location characterized by a group of several geolocations.

Once we have extracted a set of landmarks, each location history, which is a sequence of geolocations, is transformed based on the results as follows:

1. Convert each geolocation in a location history into one of the landmarks found by the mean-shift procedure. We use notation l_i^u to indicate the landmark captured by the ith photo of user u as well as the geolocation in the remainder of this paper.
2. Cluster successive geolocations in a location history; which represent that a photographer took two or more photos successively at the same landmark without visiting other landmarks.
3. If 2 is performed, time value of a set of successive geolocations is updated as the median of the photographer's arrival and leaving times at the cluster.

The result of Step 1 is thus a set of landmarks and a set of location histories.

3.3 Step 2: Photographer Behavior Model

We want to estimate the probability that user u at landmark l_{t-1} at time $t-1$ visits l_t at time t, $P(l_t|l_{t-1}, h^u)$, which represents the photographer behavior. In our model, we assume that the landmark to be visited next depends on the present location and the traveler's interests. This dependence on location and interest can be treated by using Markov and topic models, respectively. We model the photographer behavior by combining Markov and topic models in a probabilistic framework.

3.3.1 Markov model

Markov models are widely used as probabilistic models that can handle sequential information. For simplicity, we use a first-order Markov model but Markov models of any order can be employed in a similar way. In the first-order Markov model, the next landmark depends on the previous landmark as follows,

$$P(l_t|l_{t-1}, l_{t-2}, \cdots, l_1) = P(l_t|l_{t-1}). \tag{3}$$

Probability $P(l_t|l_{t-1})$ can be calculated by using maximum likelihood estimation as follows,

$$P(l_t|l_{t-1}) = \frac{N(l_{t-1}, l_t)}{N(l_{t-1})}, \tag{4}$$

where $N(l_{t-1}, l_t)$ is the number of times l_t is visited after to l_{t-1} in the data set, and $N(l_{t-1})$ is the number of times l_{t-1} is visited.

3.3.2 Topic model

A topic model is a hierarchical probabilistic model, in which a user is modeled as a mixture of topics, and a topic is modeled as a probability distribution over landmarks. Topic models have been successfully used in a wide variety of applications such as information retrieval and language modeling [22, 21] as well as modeling user interests [5, 6]. We assume that there are Z topics, and take $\boldsymbol{Z}$ to be a set of topics. In topic modeling, the probability that a user with location history h^u visits location l_t is calculated by the following equation under the assumption of the conditional independence of h^u and l_t given latent topic $z \in \boldsymbol{Z}$,

$$P(l_t|h^u) = \sum_{z \in \boldsymbol{Z}} P(z|h^u) P(l_t|z), \tag{5}$$

where $P(z|h^u)$ represents the interest of the user, which is the probability that user u is interested in topic z, and $P(l_t|z)$ represents trends in a topic, which is the probability that landmark l_t is chosen from topic z.

Several topic analysis techniques such as latent semantic analysis (LSA), probabilistic latent semantic analysis (PLSA) [21], and latent dirichlet allocation (LDA) [22] have been developed in the area of natural language processing. We adopt PLSA as the topic model in order to estimate the topics of each sequence.

We use the EM algorithm to infer topic proportions $P(z|h^u)$ for $z \in \boldsymbol{Z}$ and $u \in \boldsymbol{U}$, and landmark probabilities $P(l|z)$ for $l \in \boldsymbol{L}$ and $z \in \boldsymbol{Z}$. Here, $\boldsymbol{L}$ represents a set of locations. In the E-step, we estimate the posterior probabilities for the latent topics as follows,

$$P(z|l_t, h^u) = \frac{P(z|h^u) P(l_t|z)}{\sum_{z' \in \boldsymbol{Z}} P(z'|h^u) P(l_t|z')}. \tag{6}$$

In the M-step, we update the parameters so as to maximize the likelihood as follows,

$$P(z|h^u) \propto \sum_{l \in \boldsymbol{L}} N(l, h^u) P(z|l, h^u), \tag{7}$$

$$P(l|z) \propto \sum_{u \in \boldsymbol{U}} N(l, h^u) P(z|l, h^u), \tag{8}$$

where $N(l, h^u)$ is the number of times landmark l occurs in history h^u, and U is the set of users. By iterating E- and M-steps until convergence, we can obtain a topic model that maximizes the likelihood of the given data.

3.3.3 Combining Markov and topic models

We combine Markov and topic models in order to incorporate both present location and user interests into the photographer behavior model. Under the assumption that the history h^u and l_{t-1} are independently conditioned on l_t, the following approximation formula can combine the two models,

$$P(l_t|l_{t-1}, h^u) = \frac{P(l_t|l_{t-1})}{C(l_{t-1}, h^u)} \frac{P(l_t|h^u)}{P(l_t)}, \tag{9}$$

where $P(l_t|l_{t-1})$ and $P(l_t|h^u)$ are obtained by Markov and topic models, respectively, $P(l_t)$ is the probability that landmark l_t is visited, and $C(l_{t-1}, h^u)$ is the normalization factor. This technique is called unigram rescaling, and has been used for language modeling [23]. $P(l_t)$ is calculated as follows,

$$P(l_t) = \frac{N(l_t)}{N}, \tag{10}$$

where N is the number of photographs.

3.4 Step 3: Generating Travel Routes

In the previous step, we described a probabilistic photographer behavior model that predicts the next location from the user's present location and interests. We present below an effective travel route recommendation method that is based on the behavior model. A travel route is represented as a sequence of landmarks and the time intervals between them.

We generate and recommend K travel routes $\{\langle l^{uk}_{\tau^u+1}, \cdots, l^{uk}_{\tau^u+T^{uk}} \rangle\}_{k=1}^{K}$ that have high probabilities for given user u,

$$P(\langle l^{uk}_{\tau^u+1}, \cdots, l^{uk}_{\tau^u+T^{uk}} \rangle | l^u_{\tau^u}, h^u). \tag{11}$$

The routes selected coincide with the present location and the interest of user u. If we calculate the probabilities for all possible routes and sort them, we are sure of finding travel routes to recommend. However, this naive method requires excessive computational time, and it cannot be used in an interactive recommender system. Our solution is an effective route-finding method that enables users to search travel routes online. It is based on the *best-first search* algorithm.

Algorithm 1 describes how to generate travel routes. The input to Algorithm 1 consists of the location history h^u associated with user u, the amount of spare time to spend on future travel d, acceptable range ϵ, and the number of travel routes K. The function of Algorithm 1 is to generate an array that stores K travel routes, where s is the sequence of visited locations, d^s is the traveling time of s, p^s is the chosen probability of s, s_{last} is the last visited location in s, and s_{+l} is the updated sequence when location l is visited.

First we insert the present location $l^u_{\tau^u}$ into priority queue Q (line 4). A priority queue is a data structure. When popping elements from the queue, the highest-priority one is retrieved first. Our queue is a special variant of the priority queue where an element's priority is probability value p^s. We implement this by *max heap*. We get the highest-probability route s from Q and then decide whether it meets the time conditions specified by the user, lines 6 to 10 in Algorithm 1. By using the max heap data structure, we can efficiently find the highest probability route. If the travel time of the current sequence does not satisfy the time conditions, we search for a new route taken to other landmarks from the current location (last visited in the current sequence) and generate updated sequences, lines 11 to 17 in Algorithm 1. The chosen probability and traveling times of updated sequences are newly calculated by our proposed photographer model (line 13) and traveling time estimation (line 14), respectively. $TravelTime$ function returns the value of the traveling time between two locations (line 14). This function can be implemented using existing conventional local search services provided by portal sites and car navigation systems. They can tell us the expected traveling times between any two spots. In our approach, however, we use average traveling time between two

Algorithm 1 Generate travel routes

Require: $K > 0$, $d > 0$ and $\epsilon > 0$
1: Set an array $\boldsymbol{A} \leftarrow \boldsymbol{\Phi}$
2: Set $k \leftarrow 0$
3: Set a priority queue $\boldsymbol{Q} \leftarrow \boldsymbol{\Phi}$
4: Insert $l^u_{\tau^u}$ into $\boldsymbol{Q}$
5: **repeat**
6: $\quad s \leftarrow$ get the highest-probability one from $\boldsymbol{Q}$
7: $\quad$ **if** $d - \epsilon \le d^s \le d + \epsilon$ **then**
8: $\quad\quad$ Push s into A
9: $\quad\quad$ $k \leftarrow k + 1$
10: $\quad$ **end if**
11: $\quad$ **if** $d^s < d + \epsilon$ **then**
12: $\quad\quad$ **for** $l \in \boldsymbol{L}$ **do**
13: $\quad\quad\quad$ Set $p^{s+l} \leftarrow p^s \times P(l|s_{\text{last}}, h^u)$
14: $\quad\quad\quad$ Set $d^{s+l} \leftarrow d^s + TravelTime_{s_{\text{last}},l}$
15: $\quad\quad\quad$ Insert s_{+l} into $\boldsymbol{Q}$
16: $\quad\quad$ **end for**
17: $\quad$ **end if**
18: **until** $k = K$
19: Output $\boldsymbol{A}$

locations as estimated from the location histories. Average traveling time may be reasonable for travel planning because it includes the average time spent at each location, not just transfer time. This process is repeated until the number of routes approaches K (line 18). Since probability $P(l|s_{\text{last}}, h^u)$ is always less than or equal to one, $p^{s+l} \le p^s \times P(l|s_{\text{last}}, h^u)$. Therefore, the set of K sequences found by this best-first search algorithm holds the K highest probability sequences from among all sequences that satisfy the condition.

4. EXPERIMENTS

This section evaluates the performance of the proposed method by conducting three experiments on a Flickr-sourced geotag dataset. In the first experiment, we analyze the performance of our method with different parameter settings. The second experiment evaluates the one-step prediction accuracy of our probabilistic photographer behavior model, by comparing it to those of three other probabilistic models. In the third experiment, we analyze the multiple-step prediction accuracy. Moreover, we show some examples of recommended travel routes on our online application.

4.1 Flickr Dataset

The datasets used in these experiments were collected by downloading photo metadata from Flickr [1] using the site's public API. The crawled data consists of 696,394 photographs and their associated metadata, which were taken by 71,718 unique users. All were taken in the East Coast and the West Coast of the United States between January 1st, 2006 and June 31st, 2009.

Details of the data collection procedure are as follows. We first chose several major cities on each coast; Washington D.C., New York City, Philadelphia and Boston on the West Coast; Los Angeles, San Francisco and Las Vegas on the East Coast. For each city, geotagged photos taken within 20km from it's center were crawled. Given geo queries, which consisted of a latitude and a longitude, the Flickr API returns a collection of geotagged photos taken in the specified region. This is an efficient way to crawl the geotagged photos within each region because these major cities have been reported to be the top metropolises in the United States by a previous work [17]. We then extracted landmarks in each city

Table 2: Information about generated travel histories and the landmarks found by mean-shift procedure. The users (sequences) who visited fewer than five landmarks and landmarks that were captured by fewer than three users were omitted.

Region-scale	The num. of sequences	The num. of landmarks	Ave. length of sequences
East - 50m	9,267	414	14.02
West - 50m	6,450	316	14.05
East - 10m	11,354	1,419	15.82
West - 10m	7,913	1,119	15.65

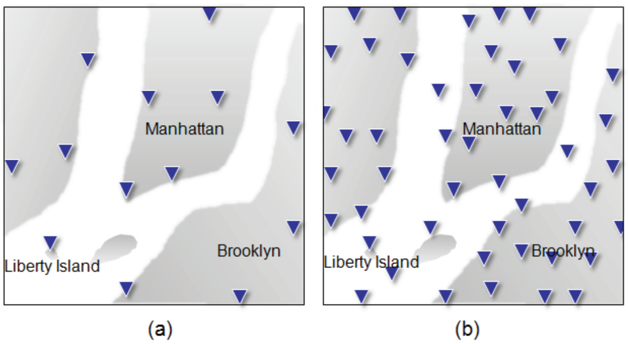


Figure 3: Landmark visualization at difference bandwidth parameters. Each icon on the map represents a landmark. (a) bandwidth is 50m and (b) bandwidth is 10m.

by applying the mean-shift algorithm (see Section 3.2). In this algorithm, as bandwidth parameter w is decreased, more landmarks are detected (Figure 3). In order to evaluate the performance of our method at several different scales, we set two different bandwidth parameters: $w = 0.0005$ (50m) and 0.0001 (10m). Finally, for each region and scale, the geotags (longitude and latitude) sequences of all users were translated into landmark sequences, i.e. travel routes. Detailed information for the number of users (sequences) and the number of landmarks are shown in Table 2. Note that we omitted the users (sequences) who visited fewer than five landmarks and landmarks that were captured by fewer than three users.

4.2 Experiment 1: Parameter Influence

One parameter which can influence the performance of our method is *the number of topics* introduced in the topic model. Therefore, we first studied its influence on the performance of our method. We used the data shown in Table 2 for training and testing of the model; training data consisted of landmark sequences excluding the last visited landmarks of all users, and the test data was the last visited landmarks of all users. Thus, the number of training data equals the number of sequences (users) in Table 2, and the number of test data also equals the number of sequences (users) in Table 2. Since the model predicts the next landmark likely to be visited by the user, we used the precision of predicted landmark as the performance measure; i.e., we calculated the percentage of correct predictions over all test examples. According to a previous survey, the precision measure is by far the most commonly used measure for evaluating the performance of recommender systems [25]. We applied the proposed Markov-topic model (photographer behavior model) to the test data, and changed the number of topics from 5

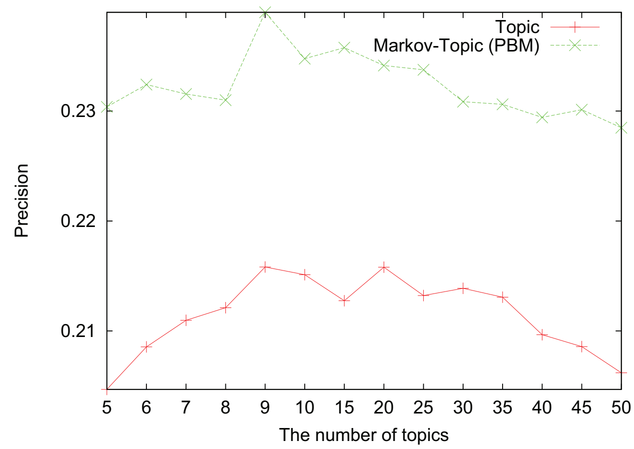


Figure 4: Precision at different numbers of topics.

to 9 and from 10 to 50 in steps of 5. In addition, the topic model (PLSA) was tested using the same condition.

The results are shown in Figure 4. In the figure, the X-axis plots the number of topics and the Y-axis plots the average precision score in four datasets (for each region and scale). The proposed model and topic model have similar shapes. When the number of topics is small, the average precision is low. The average precisions of both methods are maximized at 9, and after that, it gradually decreases with further increases in the number of topics. The result shows that the performance of our photographer behavior model depends on the topic model. It suggests that setting appropriate number of topics, i.e. user's interests, contributes to yielding better recommendation performance. The optimal parameters derived in this experiment were used in subsequent performance comparisons.

4.3 Experiment 2: One-Step Prediction Accuracy

In this experiment, we compared the following four probabilistic models including our photographer behavior model:

Multinomial model: predicts the next landmark based on its popularity. Popularity can be calculated by using the multinomial probability distribution over landmarks $P(l)$. The most visited landmark within the region, except for the current location, is always recommended. This model does not consider either the user's current location or interests.

Markov model: predicts the next landmark based on the user's current location, which is calculated by the Markov model $P(l_t|l_{t-1})$. The most chosen landmark, given the current landmark, is always recommended. This model considers the user's current location but does not consider the user's interest.

Topic model: predicts the next landmark based on user interest, which is calculated by the PLSA model $P(l|h^u)$. This model considers the user's interest, but not the user's current location.

Markov-Topic model (proposed photographer behavior model): predicts the next landmark based on both the user's current location and interest, which is calculated by combining the Markov model and the topic model using the unigram rescaling technique.

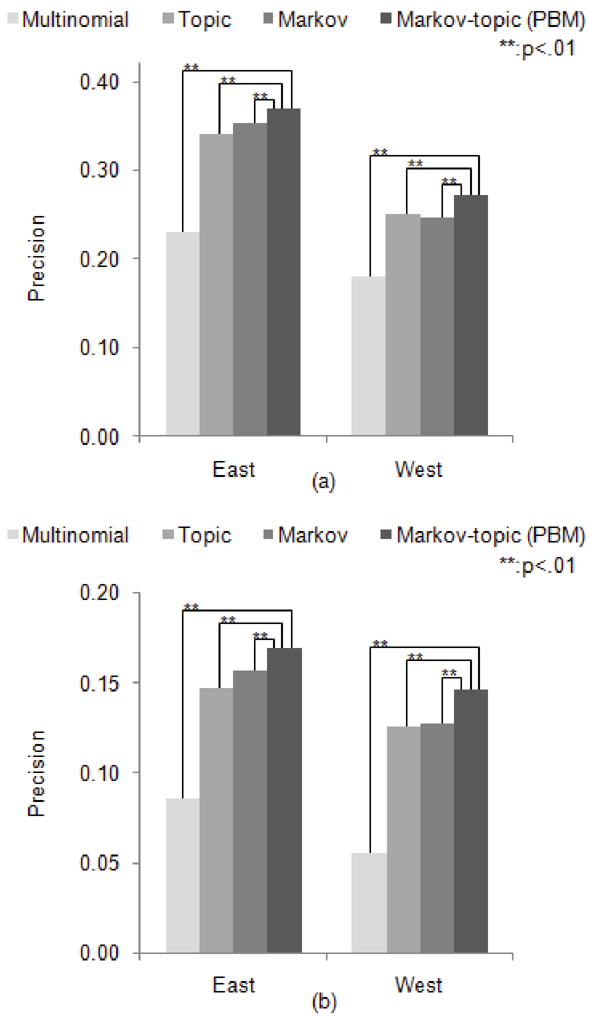


Figure 5: Comparison of precisions. The datasets are (a) the East coast and the West coast at bandwidth=50 m and (b) the East coast and the West coast at bandwidth=10 m.

The performance measure, training and the test data used were the same as in the first experiment. The results are shown in Figure 5. In this figure, the X-axis plots the dataset and the Y-axis plots the precision score. We tested the statistical significance between the proposed model and the other baselines using sign test[1]. To sum up, for both data sets, the proposed model (Markov-Topic model) yields better precision than Multinomial, Topic, and Markov model, and the differences are significant (two-sided sign test: $p < 0.01$). The result shows that the proposed method can appropriately predict the traveler's location since it uses both the user's current location and his/her interest.

4.4 Experiment 3: Multiple-Step Prediction Accuracy

To cover practical scenarios, our method generates and outputs travel routes depending on given spare time; each of them is a certain length of landmark sequence (see part 3.4). This experiment evaluates the appropriateness of the recommended travel routes under different spare time conditions. This experiment is based on the four datasets used in the experiments above. For each dataset, the test data was created by collecting the last part of each sequence within the given spare time s. The training dataset consisted of the set of sequences excluding test data part. To be exact, the number of training data (the number of test data) used in this experiment was slightly different from that of previous experiments since only the sequences whose traveling times exceeded the given spare time, were used as training data (test data). Detailed information about sequences and landmarks used in this experiment is shown in Table 3.

[1]In this paper, all of the statistical tests were conducted by using the R and R libraries at http://www.r-project.org/.

Table 3: Information about travel histories and the landmarks used in experiment 3. The users (sequences) who visited fewer than five landmarks and landmarks that were captured by fewer than three users were omitted.

Region-scale	Time period (hours)	The num. of sequences	The num. of landmarks
East - 50m	2	9,142	414
	3	9,124	414
	4	9,115	414
	5	9,102	414
West - 50m	2	6,349	316
	3	6,346	316
	4	6,333	316
	5	6,323	316
East - 10m	2	11,139	1,419
	3	11,102	1,418
	4	11,074	1,418
	5	11,049	1,418
West - 10m	2	7,763	1,117
	3	7,741	1,117
	4	7,723	1,117
	5	7,704	1,117

We measured the difference between the generated routes and each test sequence. The *edit distance* [2] is thus applied as the evaluation metric since it measures the distance between two sequences in terms of the minimum number of edit operations required to transform one sequence into the other [27]. The allowable edit operations are insert into a sequence, delete from a sequence, and replace one landmark with another. We compared our Markov-topic model with the other models, multinomial model, topic model, and Markov model. Figure 6 shows the performance of each model at four different spare times (2, 3, 4 and 5 hours). The results show that the proposed method (Markov-Topic model) offers the highest accuracy (i.e. the lowest edit distance). We also tested the statistical significance of the difference between the average edit distances of the proposed and the baselines using the Wilcoxon signed-rank test. For each region, bandwidth, and spare time condition, the result of the Wilcoxon signed-rank test is $p < 0.05$ (two-sided test). The results show that the proposed is significantly better than other baselines in terms of route prediction accuracy.

4.5 Example of Route Recommendation

We implemented our proposed method and constructed a route recommender system that would help the user in planning trips. We would like to show some examples of recommended travel routes output by the system, and discuss the effect of user information (i.e.

[2]Edit distance is also known as Levenshtein distance.

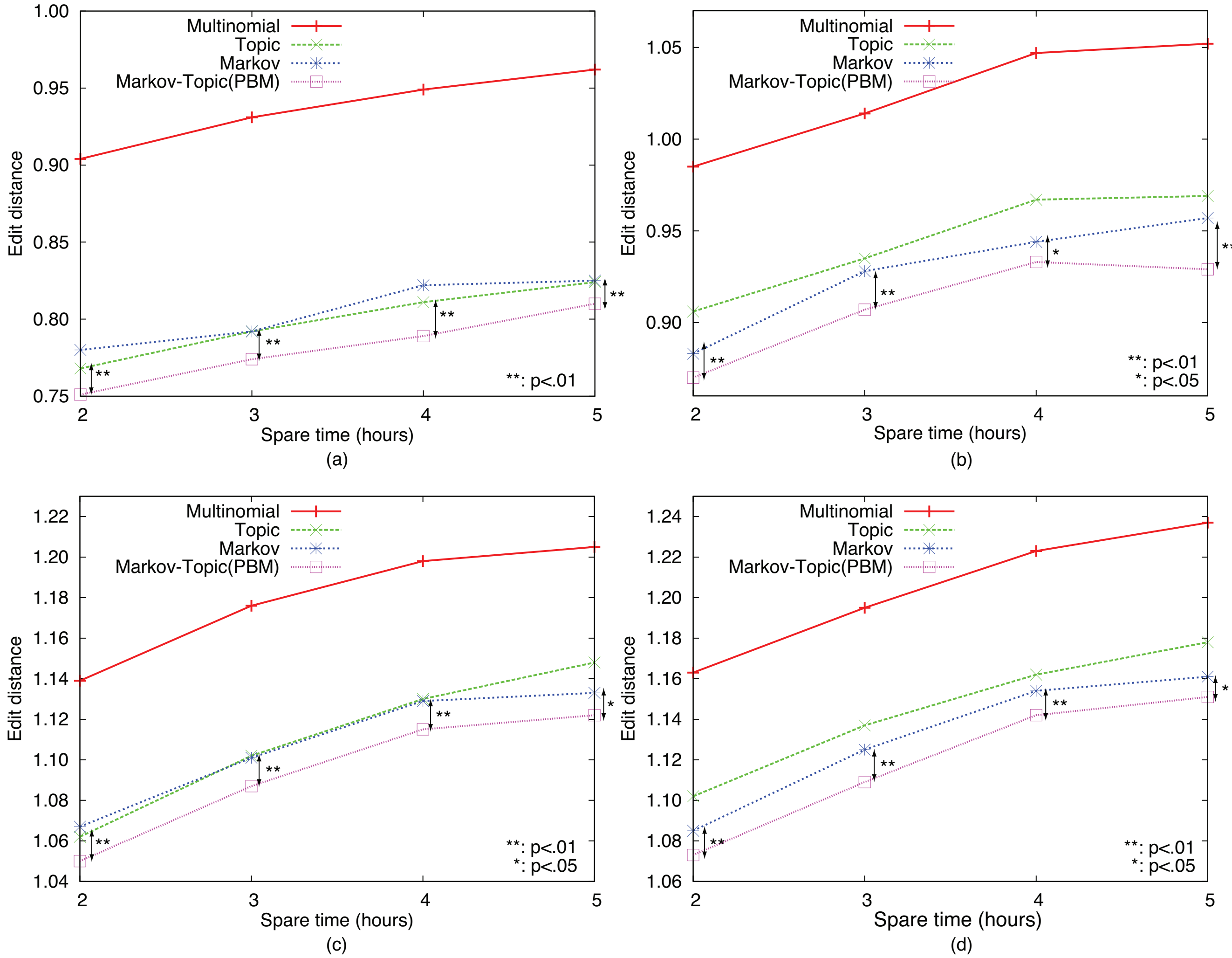


Figure 6: Average edit distances of recommended travel route at user-specified time periods. The datasets are (a) the East coast at bandwidth = 50 m, (b) the West coast at bandwidth = 50 m, (c) the East coast at bandwidth = 10 m and (d) the West coast at bandwidth = 10 m.

travel history including present location and time to spend on future travel) on personalization.

As shown in Figure 1, our recommender system is implemented as a map-based system. Given a geographical region on the map interface, the system extracts landmarks where many photographers took photos within the area specified, and visualizes them on the map interface. Each icon on the map represents a landmark, and a travel history and a current location is identified by choosing landmarks in turn.

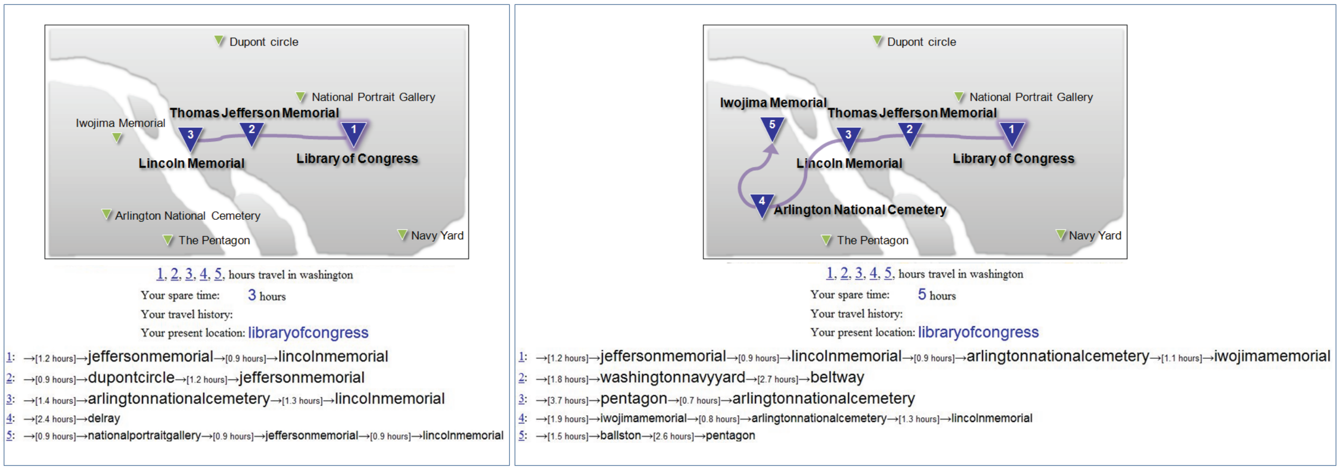
Note that representative textual descriptions for each landmark such as *pier35*, *goldengatebridge* and *brooklynbridge* are automatically extracted from the set of annotation tags. Crandall et al. [17] showed a simple method for extracting representative tags of each landmark cluster; we follow their strategy. For each landmark l, the score of tag v is calculated by;

$$TagScore_l(v) = P(l|v) = \frac{N(v,l)}{N(v)}, \tag{12}$$

where $N(v,l)$ is the number of photos at landmark l that have tag v, and $N(v)$ is the number of photos that have tag v in the dataset. Note that we discard any tags that do not occur in at least 5% of the photos in the landmark cluster because they are a significant source of noise.

4.5.1 Recommendation based on user's situation

Figure 7 shows recommendation examples when the user is at the Library of Congress in Washington D.C. As shown in Figure 7, our recommendation algorithm can provide different plans that well match the user's spare time. When the user has only a little time (user-specified time is 3 hours), the system recommends landmarks in the center of town such as Lincoln Memorial, the Thomas Jefferson Memorial and so on. When circumstances allow (user-specified time is 5 hours), the system arranges a tour of places in the region (i.e. longer travel routes are likely to be generated). The current approach to acquiring local information is to read tour guides, magazine articles, local portal sites, and other commercial sources. These media certainly introduce famous travel routes, but they may require much more time than the user can spare. Our recommender system can flexibly recommend travel routes that suit the user's available time.



3 hours 5 hours

Figure 7: Examples of travel routes at different user-specified time periods. The top ranked plan is plotted on the map interface. The departure point of the user is the Library of Congress in Washington D.C. User-specified time periods are 3 and 5 hours. Bandwidth parameter is 10m.

4.5.2 Recommendation based on user's interest and situation

We show an example of a travel route generated for a user who is interested in art, and discuss the impact of user interest on the likelihood attribute of the photographer behavior model. In this experiment, five landmarks connected with art, Chelsea, SoHo, DUMBO, Brooklyn Museum and the Lower East Side in the New York City, are set as the user's travel history in order to specify the user's interest. User's spare time is 3 hours. User's current location is Times Square because it is the largest landmark in our dataset. Figure 8 contrasts the recommended routes output by the Markov model to those of the proposed method (Markov-Topic model). As shown in Figure 8, the Markov model recommends typical routes including famous places such as ground zero, the West Village and Brooklyn Bridge because it considers only the current location and spare time (i.e. it does not consider user's interest). In our experiment, these recommended landmarks are actually ranked among the top five major landmarks in New York City sorted by the number of visitors. The proposed method (Markov-Topic model), on the other hand, recommends routes that include landmarks concerned with art such as the American Museum of Natural History, the Metropolitan Museum of Art, and Broadway. Furthermore, they are all relatively accessible to the user because our proposed is combined with the Markov model.

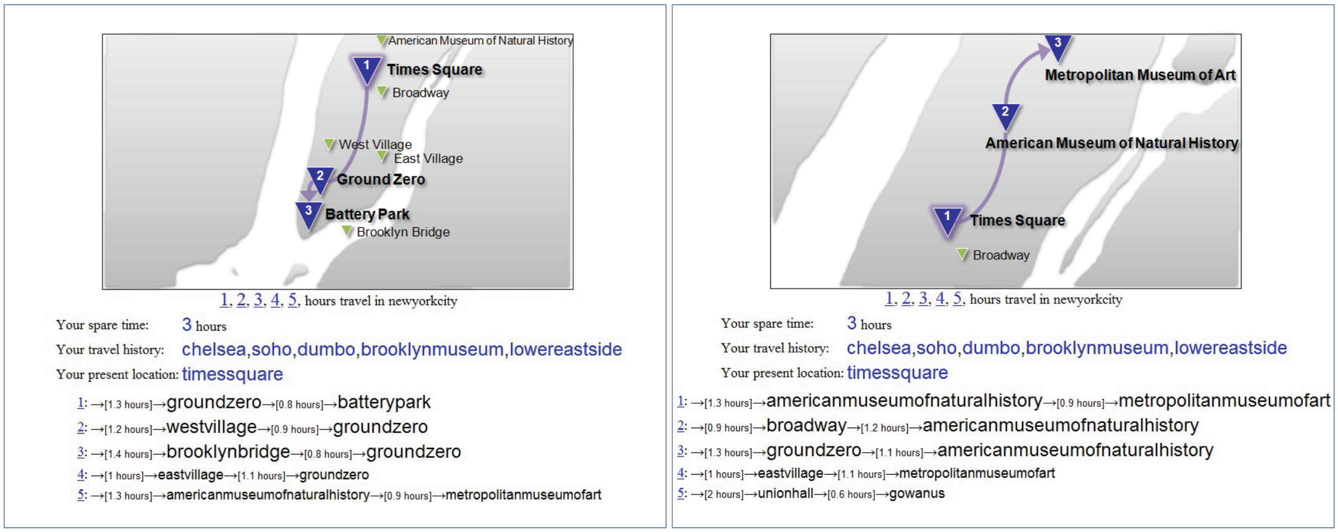
5. CONCLUSIONS

This paper introduced a framework for travel route recommendation based on the large-scale sets of geotagged and time-stamped photographs held by photo sharing sites. We assume that the geotagged photographs represent personal travel route histories and sort the locations indicated by the photographs according to their timestamps. To learn from these personal histories, we present a new technique to construct probabilistic photographer behavior models that can estimate the probability of a photographer visiting a landmark. Based on insights into photographer behavior, two models are combined in the photographer behavior model: one is the topic model, which estimates the user's own personal preference; the other is a Markov model, which can find typical routes of photographers. We conducted quantitative experiments on a large-scale dataset to compare our probabilistic photographer behavior model against three probabilistic models in terms of one-step and multiple-step prediction accuracy. The results demonstrate the effectiveness of the proposal in terms of its prediction accuracy. Finally, we demonstrated that our recommendation method outputs a set of landmark sequences, or travel routes, that coincide with the user's preference (interest), present location, and spare time. Different from previous works, which used this metadata in order to enhance the user's experience when photo browsing, this work represents a new direction in the use of the many geotagged and time-stamped photos available on the Web; we use them to enhance the experience of each traveler in the real world.

Our major future work is to utilize the attributes of the social network of photographers and each photographer's profile (where he/she lives). These attributes are characteristic of social media sites and there is the possibility that these sources will improve the accuracy with which photographer behavior can be predicted. We will also attempt to recommend more rich representations of location in combination with other image and video content analysis techniques.

6. REFERENCES

[1] Flickr: http://www.Flickr.com/
[2] Google Picasa: http://picasa.google.com/
[3] Yahoo! TRAVEL: http://www.travel.yahoo.com/
[4] Google Maps: http://maps.google.com/
[5] T. Hofmann. Collaborative Filtering via Gaussian Probabilistic Latent Semantic Analysis, Annual ACM SIGIR Conference (SIGIR), pp.259–266, 2003.
[6] T. Iwata, S. Watanabe, T. Yamada, and N. Ueda. Topic Tracking Model for Analyzing Consumer Purchase Behavior, in Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI), pp.1427–1432, 2009.
[7] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury. How Flickr Helps us Make Sense of the World: Context and Content in Community-Contributed Media Collections, in Proc. ACM Int. Conf. on Multimedia (ACMMM), pp.631–640, 2007.



(a) (b)

Figure 8: Examples of personalized routes in New York city. (a) shows recommended routes output by Markov model. (b) shows recommended routes output by Markov-Topic model (proposed model). User's current location is Times Square and spare time is 3 hours. Chelsea, SoHo, DUMBO, Brooklyn Museum and the Lower East Side are set as the user's travel history. Bandwidth parameter is 10m.

[8] L. Kennedy and M. Naaman. Generating Diverse and Representative Image Search Results for Landmarks, in Proc. Int. Conf. on World Wide Web (WWW), pp.297–306, 2008.

[9] E. Kalogerakis, O. Vesselova, J. Hays, A.A. Efros, and A. Hertzmann. Image Sequence Geolocation with Human Travel Priors, in Proc. Int. Conf. on Computer Vision (ICCV), pp.253–260, 2009.

[10] L. Cao, J. Yu, J. Luo, and T.S. Huang. Enhancing Semantic and Geographic Annotation of Web Images via Logistic Canonical Correlation Regression, in Proc. ACM Int. Conf. on Multimedia (ACMMM), pp.125–134, 2009.

[11] J. Hays and A.A. Efros. IM2GPS: Estimating Geographic Information from a Single Image, in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp.1–8, 2008.

[12] T. Horozov, N. Narasimhan and V. Vasudevan. Using location for personalized POI recommendations in mobile environments, in Proc. Int. Symp. on Applications and the Internet(SAINT), pp.124–129, 2006.

[13] J. Krumm and E. Horvitz. Predestination: Inferring Destinations from Partial Trajectories, in Proc. Int. Conf. on Ubiquitous Computing (UbiComp), pp.243–260, 2006.

[14] Y. Zheng, L. Zhang, X. Xie and W. Ma. Mining Interesting Locations and Travel Sequences from GPS Trajectories, in Proc. Int. Conf. on World Wide Web (WWW), pp.791–800, 2009.

[15] S. Ahern, M. Naaman, R. Nair, and J. Yang. World Explorer: Visualizing Aggregate Data from Unstructured Text in Geo-Referenced Collections, in Proc. ACM/IEEE-CS Joint Conf. on Digital Libraries (JCDL), pp.1–10, 2007.

[16] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. Generating Summaries and Visualization for Large Collections of Geo-Referenced Photographs, in Proc. ACM Int. Workshop on Multimedia Information Retrieval (MIR), pp.89–98, 2006.

[17] D. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the World's Photos, In Proc. Int. Conf. on World Wide Web (WWW), pp.761–770, 2009.

[18] N. Snavely, S.M. Seitz, and R. Szeliski. Photo Tourism: Exploring Photo Collections in 3D, ACM Trans. on Graphics, vol.25, no.3, pp.835–846, 2006.

[19] N. Snavely, S.M. Seitz, and R. Szeliski. Modeling the World from Internet Photo Collections, Int. J. of Computer Vision (IJCV), vol.8, no.2, pp.189–210, 2008.

[20] M.D. Choudhury, M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel and C. Yu. Constructing Travel Itineraries from Tagged Geo-Temporal Breadcrumbs, in Proc. Int. Conf. on World Wide Web (WWW), pp.1083–1084, 2010.

[21] T. Hofmann. Probabilistic Latent Semantic Analysis, in Proc. Conf. on Uncertainty in Artificial Intelligence (UAI), pp.289–296, 1999.

[22] D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent Dirichlet Allocation, Journal of Machine Learning Research (JMLR), vol. 3, pp.993–1022, 2003.

[23] D. Gildea and T. Hofmann. Topic-based Language Models Using EM, in Proc. EUROSPEECH, pp.2167–2170, 1999.

[24] F. Giannotti, M. Nanni, D. Perdreschi and F. Pinelli. Trajectory Pattern Mining, In Proc. Int. Conf. on Knowledge Discovery and Data Mining (KDD), pp.330–339, 2007.

[25] J.L. Herlocker, J.A. Konstan, L.G. Terveen and J.T. Riedl. Evaluating collaborative filtering recommender systems, ACM Trans. on Information systems (TOIS), vol.22, no.1, pp.5–53, 2004.

[26] D. Ashbrook and T. Starner. Using GPS to Learn Significant Locations and Predict Movement across Multiple Users, Personal and Ubiquitous Computing, vol.7, no.5, pp.275–286, 2003.

[27] G. Navarro. A guided tour to approximate string matching, ACM Comput. Surv. (CSUR), vol. 33, no.1, pp.31–88, 2001.