

Personalized Travel Recommendation by Mining People Attributes from Community-Contributed Photos

An-Jung Cheng[†], Yan-Ying Chen[†], Yen-Ta Huang[†], Winston H. Hsu[†]
and Hong-Yuan Mark Liao[‡]

[†] Dept. of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan

[‡] Institute of Information Science, Academia Sinica, Taipei, Taiwan

{anon,yanying,hyd0916}@cmlab.csie.ntu.edu.tw

winston@csie.ntu.edu.tw, liao@iis.sinica.edu.tw

ABSTRACT

Leveraging community-contributed data (e.g., blogs, GPS logs, and geo-tagged photos) for travel recommendation is one of the active researches since there are rich contexts and trip activities in such explosively growing data. In this work, we focus on personalized travel recommendation by leveraging the freely available community-contributed photos. We propose to conduct personalized travel recommendation by further considering specific user profiles or attributes (e.g., gender, age, race). In stead of mining photo logs only, we argue to leverage the automatically detected people attributes in the photo contents. By information-theoretic measures, we will demonstrate that such people attributes are informative and effective for travel recommendation – especially providing a promising aspect for personalization. We effectively mine the demographics for different locations (or landmarks) and travel paths. A probabilistic Bayesian learning framework which further entails mobile recommendation on the spot is introduced. We experiment on four million photos collected for eight major worldwide cities. The experiments confirm that people attributes are promising and orthogonal to prior works using travel logs only and can further improve prior travel recommendation methods especially in difficult predictions by further leveraging user contexts in mobile devices.

Categories and Subject Descriptors: H.4 [Information Systems Applications]: Miscellaneous; H.2.8 [Database Applications]: Data mining, Spatial databases and GIS

General Terms: Algorithms, Experimentation, Human Factors

Keywords: Personalized tourist recommendation, geo-tagged photos, route planning

1. INTRODUCTION

With the prosperity of the social media and the success of many photo-sharing websites, like Flickr and Picasa, the volume of community-contributed photos have increased dras-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.

Copyright 2011 ACM 978-1-4503-0616-4/11/11 ...\$10.00.

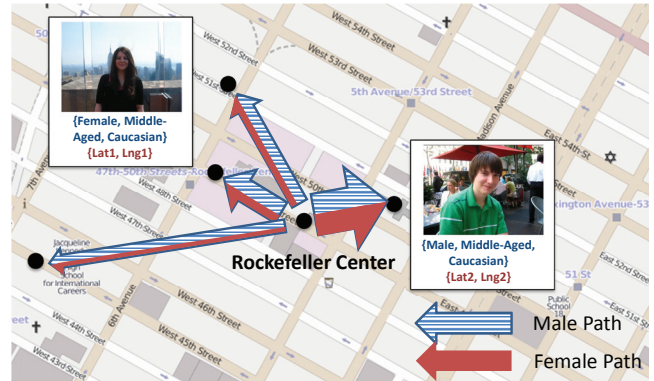


Figure 1: Community-contributed photos are freely available and with rich context information (e.g., geo-tags, time). We can mine the travel patterns for the users by associating their trips with the major landmarks. Meanwhile, we can indirectly mine demographic information (e.g., gender, race, age) from the photos along with the trips by automatically detecting people attributes in the photos. In this example, the width of an arrow denotes the travel frequency between the two locations. The color regions are proportional to the percentages of genders – male (blue) and female (red). Such minded demographic information is promising for personalized travel recommendation and planning; for example, suggesting the best route for Asian travelers in Manhattan or the next location for a family group after visiting Rockefeller Center.

tically. Such large-scale user-contributed photos contain rich metadata information such as tags, time, and geo-locations (or geo-tags). These overwhelming amounts of context data, though noisy, is tremendously useful for many multimedia applications including annotation, searching, advertising and recommendation [15].

Among all the applications, travel recommendation has been attracted by many researchers because of the importance and the intrinsic relationship between people's everyday lives. In general, a typical travel recommendation system consists of two aspects: generic recommendation and personalized recommendation [19]. For the generic recommendation, it contains the suggested travel information for the destination given by user when he/she is planning a trip,

which answers the question like *"I want to go to New York, what are the must-see attractions?"*. The personalized recommendation, on the other hand, takes user's profile into account such that it can provide a more appropriate recommendation result matching user preferences. Both aspects are to support route planning before the journey.

Millions of human-sensors [16] capture different aspects of the spatio-temporal information. In order to mine the travel knowledge automatically, a focus of recent interest is the use of user-contributed resources, including the textual travel-ogues (i.e., blogs or logs) ([10, 9, 8]) and photos taken during the trips ([2, 14]). Discovering and summarizing knowledge from these huge amount of user-contributed multimedia resources provide us useful tourist information.

However, the previous works solely consider the travel logs and ignore the rich people attributes in the photo contents. Such rich people attributes can be automatically detected and provide another important aspect regarding travel demographics (e.g., gender, race, age). Figure 1 illustrates the real attribute-oriented travel movements in our dataset in Manhattan¹. Rather than the plain travel frequencies between certain locations, we can further investigate their demographic distributions for the people attributes, which are promising for personalized travel recommendation and planning. Meanwhile, we had observed that such (detected) people attributes are correlated with traveling between locations (cf. Section 3.2).

Additionally, based on the mined people attributes and preferred travel patterns between locations, it is able to adopt the mobile devices where the user profile and context information (e.g., geo-locations) can be easily detected from mobile sensors and further entails the "location-aware" – recommending next travel location from his/her current location or even delivering context-related advertisements or services ([4, 13]).

In this work, we focus on the personalized recommendation framework to provide not only a context-aware recommendation system (i.e., *mobile travel recommendation*) but also a *route planning* application before the journey. The personalization is achieved by adopting specific user profiles with the automatically detected people attributes (e.g., gender, age and race) along with the trips. That is, the question we want to answer is, for example, *"For a male, what is the suggested travel sequence in Rome?"* (route planning), *"I am a female, I am now at Central Park in Manhattan, what is the next suggested destination?"* (mobile travel recommendation). Although users are reluctant to reveal their true identities or profiles in the community-contributed photos, we can actually indirectly mine the demographic information about these trips from their associated photos.

In this sense, we crawl 4 millions geo-tagged photos belonging to eight major worldwide cities, adopt the people attribute detectors on the generated photo trips and then provide a probabilistic recommendation model based on the user's profiles and the travel logs information. Experiments show that the people attributes are very informative and helpful for both mobile travel recommendation and route planning. To summarize, the contributions of this paper are:

- To our best knowledge, this is the first research work that uses the additional context in the photo, i.e., people

attributes, to support the personalized recommendation framework. We leverage these automatically detected people attributes in the large-scale photos for social media mining and uncover the differences in travel behaviors across demographic groups.

- We propose a probability-based personalized travel recommendation model based on user's attributes and the knowledge mined from travel logs. Such scheme is promising to apply in mobile environment. (Section 7)
- We conduct the experiments on eight major cities in the world and show that using people's attributes has the potential to improve the personalized travel recommendation, especially in the location where people have diverse choices of the next stops. (Section 9)

The remainder of this paper is structured as follows. In the next section, we discuss the related work. Section 3 are some attribute-related characteristics on photo trips and our system overview is described in Section 4. Section 5 introduces people attribute detection scheme. The proposed recommendation algorithms and the application are mainly in Section 6, Section 7 and Section 8. The experimental results and discussions are reported in Section 9, followed by the conclusion in Section 10.

2. RELATED WORK

Trip mining and recommendation has been shown important in recent years. Generally, the data sources for learning to recommend can be roughly classified into three categories: GPS trajectory data, travelogues (i.e., blogs), and geo-tagged photos.

GPS trajectory data which obtained by GPS receivers are mainly used at the early stage. Zheng et al. [18, 19, 20] utilize GPS trajectory data to extract the interesting locations, classical travel sequences and provide a personalized friend and location recommender using the similarity of users in terms of their location histories. The main obstacle for trajectories-based method is that the data resources is not easy to obtain from a large number of people.

[10, 9, 8] provide location-based travel recommendation which analyzes the blogs to obtain trip-related knowledge. [10] emphasizes on the mining of city landmarks by a graph-based method. [9] proposes a probabilistic topic model which discovers topics from travelogues and then represents locations with appropriate topics for further destination recommendation and summarization. The goal of [8] is to automatically recognize and rank the landmarks for travelers. They use geo-tag information, metadata of photos and user knowledge in Yahoo! Travel Guide to identify and rank landmarks for any location specified by the travelers. The travelogue-based method has the difficulty in determining the certain location of travelogues which are usually unstructured and contain much noisy metadata. They only play a role of destination recommendation which merely shows the information about a location.

Recently, there is an increasing tendency to adopt the information from geo-tagged photos. Crandall et al. [7] systematically adopt large-scale photos to discover important landmarks. They evaluate on many cities and indicate that the time-stamped and geo-tagged photos will construct the typical pathways of people movements. Arase et

¹Note that the photos in the paper are courtesy of Flickr users under the Creative Commons License.

Table 1: Relationship of people attribute and travel patterns measured by entropy and mutual information in 5 locations sampled from 5 major worldwide cities. Taking the nearest five locations (indexed by j) into considerations, as starting from location i , the entropy for the uniform choice is $H(L_{i \rightarrow j}) = 2.3219$. We show how the entropy (uncertainties in choosing next visiting location) will be reduced as introducing mined (from community-contributed photos) knowledges. Here F denotes the calculated visiting frequencies between location i and j from the photo trips (cf. Figure 4(d)); A denotes the (detected) people attributes along with the photo trips. We can see that the people attributes (A) are promising to predict travel locations $L_{i \rightarrow j}$. See more explanations in Section 3.2.

City	Location i	$H(L_{i \rightarrow j})$	$H(L_{i \rightarrow j} F)$	$H(L_{i \rightarrow j} A, F)$	$I(L_{i \rightarrow j}; A F)$
Manhattan	Madison Square	2.3219	1.9911	1.4582	0.5329
Rome	Castel Sant’Angelo	2.3219	1.8909	1.0068	0.8841
Berlin	Marx-Engels Forum	2.3219	0.8281	0.1883	0.6397
Barcelona	Casa de la Caritat	2.3219	0.9751	0.3907	0.5844
Taipei	Xinyi Square	2.3219	1.9796	0.8566	1.1229

al. [2] define the problem of photo trip pattern mining and show an application with which users can search frequent trip patterns given some preferences (e.g., destination, visit duration, trip theme). Although they also provide a similar route planning framework, the recommendation level is inter-cities, while our work is based on the intra-cities level for a one-day trip recommendation in a specific city. Most of all, we can further recommend travel routes satisfying the designated user profiles (e.g., gender, age, race) by leveraging people attributes in the user-contributed photos; for example, the best one-day trip in Manhattan for the Asian girls.

Another similar work is [14], which uses geo-tagged photos and textual blogs information on landmark generation, path discovering, and route planning. They merge incomplete paths inside a destination to suggest tourists with typical travel paths and stay times in a destination. For route planning, they model as a graph analysis problem and use dynamic programming as the solver. We also have the similar processes and extract travel patterns from pictures but we want to emphasize that the major difference between our work and other related researches described above is that we bring in the concept of people attributes in the travel photos and consider these demographic information with the movements of photographer into a personalized travel recommendation framework. Besides, we have fundamentally different definitions on the travel patterns. That is, the travel patterns in this study are measured from the face detection results and further augmented by the face attributes (e.g., gender, age, race) extracted from the photos. By this novel aspect, we can mine more human attributes for effective travel personalization. For example, we will have the travel preferences about Caucasian males or Asian kids. Additionally, in the previous researches (e.g., [14, 2]), the travel patterns are defined over travel objects such as landmarks, events only. We will demonstrate in the next section that such automatically detected people attributes are informative for mobile travel recommendation and route planning.

3. OBSERVATIONS

3.1 Rich People Attributes in User-Contributed Photos

Using geo-tagged and time-stamped photos from social media as a resource for travel information mining have attracted a lot of researchers in recent years (e.g., [7, 2, 14,

5]). These large amounts of photo trajectories not only reveal the users’ travel movements but are also promising for mining the demographic information about the locations by detecting people attributes on the photos. Authors in [1] have shown that users are willing to share photos for organization and (social) communication purposes – especially for travel photos. Meanwhile, such photos can be treated as the *social pixels* by travelers’ cameras among travel locations (cf. Figure 1) [16]. It is promising to leverage such freely available user-contributed photos and further detect important people attributes. However, the previous works only focus on the statistics of the travel logs such as the popularity of locations but neglected the important and rich dimension of people attributes.

Meanwhile, in [11] Kumar et al. exploited the community-contributed photos for learning facial attributes (e.g., gender, age and race) which contribute to mining people attributes in large-scale media. Such rich attributes are shown having reasonable accuracy (>80%) for further applications. Therefore, we propose an approach to enable personalized travel recommendation by mining the parameterizing factors directly and automatically from the community-contributed photos – especially emphasizing on the (automatically) detected people attributes from the photos.

3.2 Correlation between Travel Patterns and People Attributes

By intuition, we know that some landmarks are female-favored, and some are male-favored. So are by other attributes such as race, age. Here we want to use entropy and mutual information [6] to measure the correlation between predicting next travel location and people attributes. The mutual information $I(X; Y)$ between two random variables X and Y is defined as follows:

$$I(X; Y) = H(X) - H(X|Y), \quad (1)$$

where entropy $H(X)$ measures the uncertainty with a random variable X . The entropy reduction as introducing another random variable (i.e., Y) is the mutual information $I(X; Y)$ between the two random variables. Intuitively, the reduction means the help that random variable Y brings for predicting X .

Measured by entropy and mutual information and illustrated in Table 1, we demonstrate the relationship of people attributes and travel patterns in 5 locations sampled from 5 major worldwide cities. Taking the nearest five locations

(indexed by j) into considerations as starting from location i , the entropy for the uniform choice is $H(L_{i \rightarrow j}) = 2.3219$. We will show how the entropy (uncertainties in choosing next visiting location) will be reduced as introducing mined (from community-contributed photos) knowledges. Here F denotes the calculated visiting frequencies between location i and j from the photo trips (cf. Figure 4(d)); A denotes the (detected) people attributes along with the photo trips; we use gender (i.e., male and female) only in this measurement. Thus, $H(L_{i \rightarrow j}|F)^2$ denotes the entropy of predicting the next location as given the mined travel frequencies (F). It is apparent that, from $H(L_{i \rightarrow j}|F)$, the mined travel frequency (F) helps choosing (recommending) the next travel location from the current one. Further introducing the people attributes (A), the recommendation will be more accurate; that is, salient entropy reduction is shown in $H(L_{i \rightarrow j}|A, F)$. Note that $I(L_{i \rightarrow j}; A|F) = H(L_{i \rightarrow j}|F) - H(L_{i \rightarrow j}|A, F)$, as depicted in the last column.

Taking Madison Square in Manhattan as an example (cf. the first row in Table 1), the mutual information $I(L_{i \rightarrow j}; A|F)$ is 0.5329 (bits), the reduction is about 25% of the entropy when given the people attributes (A). To interpret that intuitively, we can illustrate like this, if there are 4 random choices for the next destination, after knowing the people attribute (e.g., for the male only), the number of choice is down to 3. Therefore, we can see that the people attributes (A) have quite correlations with predicting travel locations $L_{i \rightarrow j}$. In this paper, we will show how to leverage such automatically mined people attributes for personalized travel recommendation.

4. SYSTEM OVERVIEW

Figure 2 shows the architecture of our system, which is divided into four main parts. At first, in order to mine the travel information within each city, we crawl the photos from the on-line photo-sharing websites (i.e., Flickr). We then use a mean-shift based method on geo-locations of these photos to generate the important locations on each city (Section 6.2) for the following user trip mining process. For the trip attribute pattern mining, the faces in the photos are firstly detected. The attributes are generated by applying people attribute detectors (Section 5).

Travel paths and locations mined from the community-contributed photos also contain rich people attributes from the detected people. For instance, if many faces in this location are detected as male, this destination is probably favored by male. This phenomenon can also be found in travel paths. We can further identify the demographic information (via people attributes) within travel paths by analyzing the associated photos. By mining the travel patterns users' day trips (Section 6.3), we propose two personalized travel recommendation applications – mobile travel recommendation (Section 7) and route planning (Section 8), which are further entailed by a probability Bayesian model and dynamic programming techniques.

Concept Definition

We first define some concepts used in the following sections:

Photo collections All the geo-tagged photos in the database. (i.e., all the points in Figure 4(a)).

Destination *Destinations* are the popular or important locations in a city, e.g., Brooklyn bridge in Manhattan, Colos-

²The detailed deviations can be referred to Section 7.

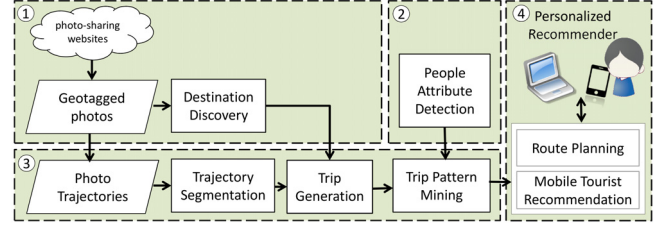


Figure 2: System Diagram: There are four main components. The first part contains data crawling process and destination generation (Section 6.2). After that, people attribute (i.e., gender, age, race) detection is applied on each photo (Section 5). The next step is the trip generation process and the attribute information of each trip is inferred by the detected faces in the trip (Section 6.3). Combined with the people attributes and the travel logs, we can build an personalized tourist recommender, such as mobile travel recommendation (Section 7) and route planning (Section 8).

seum in Rome, and Brandenburg Gate in Berlin. (i.e., the red point in Figure 4(c)).

Trajectory A collection of time series geo-tagged photos for a specific photographer.

Path & Route *Path* refers to a sequence of time series geo-tagged photos traversed by a user in one specific day (i.e., Figure 4(b)). After quantizing the geo-tagged photos to *destinations*, the results are defined as *route* (i.e., Figure 4(c)).

Trip An individual user's *trip* is the *route* with attribute information (i.e., Figure 4(d)).

Trip Segment A *trip segment* is a tuple with (start location, end location, attributes) and is a part of the *trip*. We will take the trip segments as major knowledge sources (or observed data) for mobile travel recommendation and route planning.

5. PEOPLE ATTRIBUTE DETECTION

In this work, we utilize nine people attributes including two gender attributes (female, male), four age attributes (kid, teen, middle-aged, elder) and three race attributes (Caucasian, African, Asian) to profile the travel preference of users. Currently, we only focus on the facial attributes since they represent rich information of people and can be learned through an adaptive framework [11]. Considering the scalability to deal with the increasing number of facial attributes, [11] proposes a generic framework for learning varying facial attributes adaptively. To begin with, we crawl the user-contributed photos from Flickr and extract the facial region by face detector. Further, the face images are decomposed into face components for learning and constructing a mid-level feature bank (by SVM) to describe varying facial attributes. Finally, Adaboost is conducted to select the most important mid-level feature set for the designated facial attribute.

5.1 Preprocessing

To construct a training dataset naturally with rich and diverse images close to real life, we crawl training images from social media (Flickr) containing abundant user-contributed photos. Moreover, the facial region is extracted from the

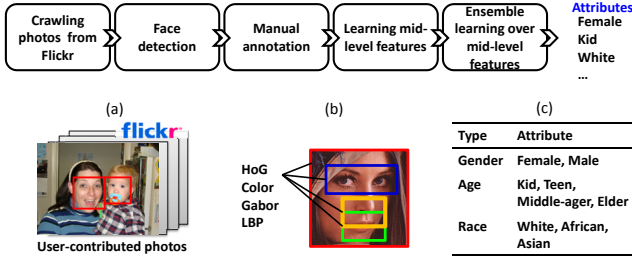


Figure 3: The framework of facial attribute detection: (a) crawling the user-contributed photos from social media and extracting the facial region by face detection to prepare the training images; (b) manually annotating the facial images for learning mid-level features by SVM; (c) utilizing Adaboost to learn the best combination of mid-level features for describing different facial attributes respectively.

whole image by face detection (cf. Figure 3(a)) and further applied facial feature detection [17] to locate important parts of a face such as eyes, philtrum and mouth. Features (e.g., edge, color, texture) extracted from different facial components pose different discriminating abilities across facial attributes. Therefore, we construct a mid-level feature bank based on those facial components for providing better generalization capability to deal with various facial attributes.

5.2 Mid-level Features

The preprocessed facial images are annotated manually as the training data to learn the mid-level features. A mid-level feature (Figure 3(b)) is a SVM classifier [3] with varying low-level features (e.g., Gabor filter, HoG, Color, Local Binary Patterns) extracted from different face components (e.g., whole face, eyes, philtrum, mouth). Those combinations constitute a feature bank to provide possible mid-level features required by facial attributes. Rather than using low-level features, mid-level features have better semantic meanings for describing facial attributes through multiple modalities. Note that, unlike image object classification, randomly selecting the images from other categories (attributes) as negative training data is impractical since attributes are often partially correlated such as female and Asian. Hence we simply group the attributes into several attribute type, for example, the attribute type “age” includes kid, teen, middle-ager and elder. Since the attributes with the same attribute type are mutually exclusive, thus providing negative training images with better discriminating capability. For example, we acquire the negative training images of kid attribute from the rest of the attributes belonging to the same attribute type (e.g., teen, middle-ager and elder).

5.3 Ensemble Learning for Different Attributes

For a facial attribute, the best combination of mid-level features is learned by Adaboost (Figure 3(c)). Here, we treat each mid-level feature as a weak classifier. The combined strong classifier represents the most important parts of that attribute, for example, <whole face, Gabor> is most effective for female attribute while <whole face, color> is most effective for African attribute. The important mid-level feature set for the designed facial attribute is adaptively selected and weighted through the boosting scheme.

Experimenting in the benchmark data [11], the approach can effectively detect facial attributes and achieve more than 80% accuracy on the average. Meanwhile, the framework is generic for various facial attributes thus providing better scalability to precisely profile users for more attributes.

6. PHOTO TRIP GENERATION

The trip generation is a challenging problem because of the various trip types (e.g., itinerary durations) and different travel behaviors across demographic groups. We adopt the common frameworks, e.g., [14, 2], to mine the trips from the (noisy) user-contributed geo-tagged photos and further consider the people attributes in the photos. We conduct three main procedures: path extraction, destination discovery, and trip attribute pattern mining, to extract the observations (in trip segments between destinations) to be further incorporated in the next recommendation model.

We first use the time information in the photos to segment the *trajectories* from each user. These segments are with the limitation in one-day that can be seen as the one-day *paths* in the given city. The next processing is about the destination discovering because we try to map these uninformative *paths* into a higher level, i.e., *routes* which are based on the detected *destinations*. The next challenge is to give the attribute knowledge into the *routes*, resulting in the informative *trips*. These *trips* can be separated into *trips segments* and then will become the inputs in our proposed travel recommendation.

We emphasize here that, although people may visit places as a couple or group and usually not alone, in this beginning stage for proving the novel concept by leveraging people attributes in the photos for personalized recommendation, we focus on the individual travelers in this study first and we believe it can be extended to group travelers by further investigations; for example, conducting face clustering in the original travel patterns.

6.1 Path Extraction

We sort photos by the captured time for each user. The one-day paths are then detected by separating the sorted photos according to their date information. Also, we adopt the soft assignment to the photos which are at the boundary between two days, that is, we make each boundary photo belong to two paths, one is its original date path, another is the previous date path in order to handle the case that if a user’s trip is across the midnight. The result paths are shown in Figure 4 (b).

6.2 Destination Discovering and Route Mapping

Because the generated paths are too noisy and complicated to mine the information, the next task is to refine these paths to a more informative level, *routes*, by quantizing the photos into the destinations. In the beginning, we apply mean-shift clustering procedure on the geo-tagged photos for each city, where each photo is represented by latitude and longitude coordinates, to discover the locations with high photo densities. Empirically, for mean-shift method, we use a flat kernel function and set the bandwidth parameter as 0.001, which is roughly the radius of a landmark, as mentioned in [12, 5, 7]. We use clusters containing at least 0.1% photos in the city and no less than 100 photos as destina-

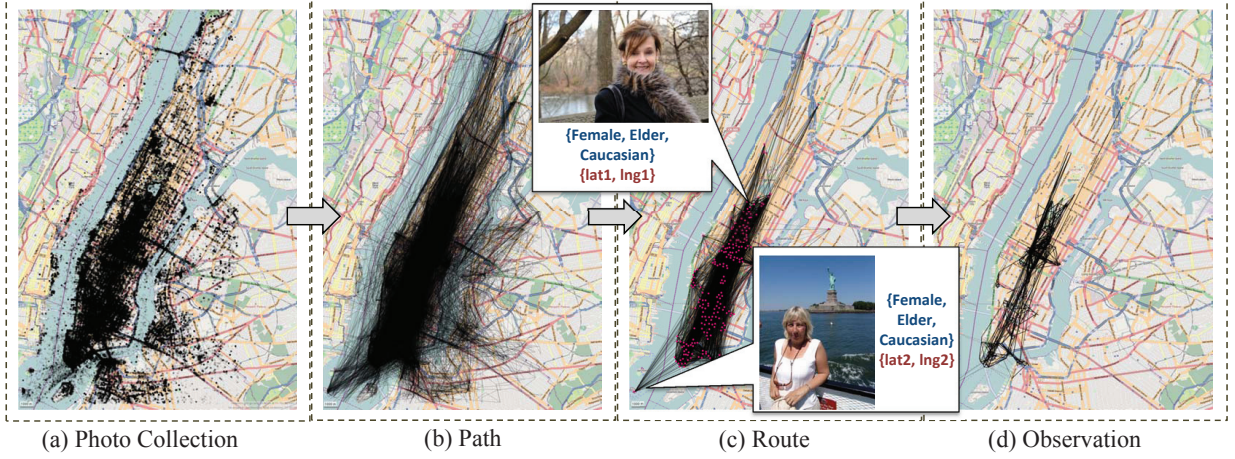


Figure 4: The photo trip generation process illustrated in Manhattan collection. (a) The collection of geo-tagged photos; each dot is the place where the photo is taken. After separating by date information on the photos for each user, the resulted travel *path*, which connects its photo sequences of the same user, is shown in (b). However, these noisy path trajectories are not informative. In (c), we first cluster these geo-tagged photos as *destinations*, representing locations of interests in the city and also marked as red points; we then quantize the paths into these discovered destinations as *routes*. The two photos are the examples for geo-tagged photos along with the trips, which contain the people attributes (i.e., gender, age, race) and the geo-location information (i.e., Lat for latitude, Lng for longitude) and can be further utilized to mine the relationship between attributes and travel behaviors. The *trips* with people attributes are shown in (d) which will finally become the learning data in our recommendation model. (Best seen in color)

tions. The red points in Figure 4 (c) illustrate the detected destinations on Manhattan via this procedure.

The path sequences of all users now can be translated into destination sequences, i.e., *routes*, by assigning the photo to the nearest destination if the distance between them is under a given threshold.

6.3 Trip Attribute Pattern Mining

A photographer can be seen as a human-sensor [16]; hence each detected face in the photos reveals not only the trip but also an observed people attributes. Besides, the detected face can be regarded as an individual sharing the same movements with this photographer. That is, we propagate the same movements from photographer to all the faces in the taken photos. For instance, given that there is a route sequence ($A \rightarrow B \rightarrow C$) which is generated by an user. A , B , and C are the traversed destinations sequentially. All the detected faces in this route will all own the same route sequences and then result in many trips according to the face number. Although these trips have the same movement sequence, they might have different attributes based on the detection on the faces. Besides, all the trips generated from this route sequence will contain trip segments (e.g., ($A \rightarrow B$) and ($B \rightarrow C$) in this example). These trip segments are treated as the atomic references as the observation data for learning and recommendation later.

Formally, we define the trip T containing n destinations for a specific person (i.e., detected face) $u \in \mathbf{U}$ (i.e., \mathbf{U} is the set of users) as $T_u = (L_{1_u} \rightarrow L_{2_u} \rightarrow \dots \rightarrow L_{n_u}, \mathbf{A}_u)$ where $L_{i_u} \in \mathbf{L}$ (i.e., \mathbf{L} is the set of destinations) is the sequentially traversed destinations of user u and \mathbf{A}_u is a tuple containing the detected attributes such as gender and age. For example, in Figure 4 (c), the two photos have the same (detected) attributes (female, elder, Caucasian), where each dimension represents the gender, age, and race accordingly.

7. TOURIST RECOMMENDATION MODEL

For personalized recommendation, we tackle the following scenario: “I am a male, and I am at Central Park now. What is the next suggested destination for me?” Obviously, the inputs of this framework include the target user u ’s profile A_u , composed of known people attributes, u ’s present location L_i , and the output is the recommended next destination L_j .

To recommend the optimal next destination L^* , this problem can be formulated as follows:

$$L^* = \operatorname{argmax}_{L_j} P(L_j | A_u, L_i, \Theta) \quad \Theta = \{\mathbf{F}, \mathbf{A}\}, \quad (2)$$

where Θ is our learned tourist model composed of the travel frequencies between the destinations (\mathbf{F}) and detected attributes in photos (\mathbf{A}). Intuitively, we will recommend the proper destination that most tourists visit and matches the target user’s profile A_u . That is why we need to mine both \mathbf{F} and \mathbf{A} , the existing travel preferences among the destinations, from the large-scale community-contributed photos.

Until now, we have the trips with attribute patterns $T_{u \in \mathbf{U}} = (L_{1_u} \rightarrow L_{2_u} \rightarrow \dots \rightarrow L_{n_u}, \mathbf{A}_u)$. And each trip $T_u \in \mathbf{U}$ can be regarded as a sequence of trip segments as described above and which then become the observation dataset of our recommendation model.

This section describes a way of constructing the learning tourist knowledge that takes two factors into account: (1) the sequence correlation between destinations ($P(L_{i \rightarrow j})$ or denoted as $P(L_{ij})$), and (2) the people attribute patterns on the movements ($P(A|L_{ij})$).

7.1 Bayesian Learning Model

We adopt Bayesian learning model as our recommendation model because Bayesian has shown effective in recommendation systems and, most importantly, it can be applied for real-time mobile recommendation and immediate route

planning service. It is the first attempt for us and can be replaced by any other recommendation models. By Bayes' theorem, the probability that the location L_j is the suggested destination given a start location L_i and the attribute value \mathbf{A}_u of a specific user u is:

$$P(L_{i \rightarrow j} | \mathbf{A}_u) = \frac{P(L_{i \rightarrow j}, \mathbf{A}_u)}{P(\mathbf{A}_u)} = \frac{P(L_{i \rightarrow j})P(\mathbf{A}_u | L_{i \rightarrow j})}{P(\mathbf{A}_u)}, \quad (3)$$

where $P(L_{i \rightarrow j})$ is the probability that moving from location L_i to location L_j . Intuitively, it is equivalent to that we choose L_i as a start location (with the probability $P(L_i)$) (i.e., given L_i) and then transit to L_j (with the probability $P(L_j | L_i)$) and can be jointly formulated as $P(L_i)P(L_j | L_i)$. Therefore, the Equation 3 will become

$$P(L_{i \rightarrow j} | \mathbf{A}_u) = \frac{P(L_i)P(L_j | L_i)P(\mathbf{A}_u | L_{i \rightarrow j})}{P(\mathbf{A}_u)} \quad (4)$$

In other words, based on the target user profile and his current location and given the observation data $\Theta = \{\mathbf{F}, \mathbf{A}\}$, the optimal prediction of next location is the destination L^* such that Equation 4 is maximized,

$$L^* = \operatorname{argmax}_L \{P(L_j | L_i)P(\mathbf{A}_u | L_{i \rightarrow j})\} \quad (5)$$

We assume the independence between different types of attributes (i.e., gender and race are independent), such that the joint probability $P(\mathbf{A}_u | L_{i \rightarrow j})$ can be further expressed as the product of the marginals,

$$L^* = \operatorname{argmax}_L \{P(L_j | L_i) \prod_{a_u \in \mathbf{A}_u} P(a_u | L_{i \rightarrow j})\} \quad (6)$$

The a_u is one of the tuple of \mathbf{A}_u (i.e., gender). $P(a_u | L_{i \rightarrow j})$ and $P(L_j | L_i)$ in the product above are easy to be estimated from the frequencies of the training corpus (i.e., observation data):

$$P(a_u | L_{i \rightarrow j}) = \frac{\text{count}(L_{i \rightarrow j} \wedge \mathbf{A}_u = a_u)}{\text{count}(L_{i \rightarrow j})} \quad (7)$$

$$P(L_j | L_i) = \frac{\text{count}(L_{i \rightarrow j})}{\sum_{j \in \mathbf{L}} \text{count}(L_{i \rightarrow j})} \quad (8)$$

The $\text{count}(L_{i \rightarrow j})$ is the total number of observations that start at L_i and end at L_j . Similarly, The $\text{count}(L_{i \rightarrow j} \wedge \mathbf{A}_u = a_u)$ is the total number of observations that start at L_i and end at L_j with the attribute a_u .

7.2 Background Smoothing

We cannot ignore the possibility that the user will be in the location or want to go to the location that beyond our observed travel logs. For example, if a male is now in L_i , and there is no male which starts from L_i to any other locations in the historical data, we cannot predict the next location for this user due to the zero probability. Therefore, in order to deal with the unobserved events, we need to add a smoothing factor in the probabilistic model. Intuitively, we take destination popularity into consideration for the location correlation. This information is obtained from the destination generation process for each location L_j . The background probability is defined by,

$$P_{C.L}(L_j) = \frac{\text{number of photos in the destination } L_j}{\text{number of photos in the city } C} \quad (9)$$

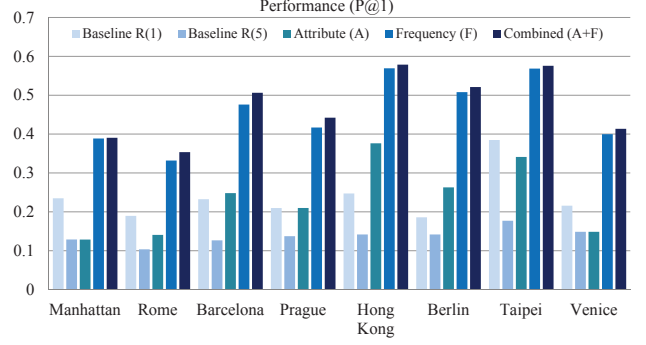


Figure 5: Mobile recommendation performance of different methods on eight worldwide cities. The knowledge-based methods, i.e., frequency (F), attribute (A), and combined (F+A), perform better than naively choose the first near location (R(1)) or the top 5 nearest (R(5)). Best seen in color.

For attribute probability, we use the city C 's background attribute information to smooth the attribute effect.

$$P_{C.A}(a_k) = \frac{\text{number of people with the attribute } a_k}{\text{number of people in this city } C} \quad (10)$$

It must be noted that the background information of attributes is based on the detected faces, which may include the trips staying in the same location (i.e., it generates no trip segment) or the faces in the photos which does not quantizing into destinations.

We adopt a linear interpolation of the maximum likelihood model with the collection background information, using the coefficient α and β to control the influence of background model.

When we substitute Equation (9), (10) into Equation (6), we get the following final estimate for the joint probability:

$$L^* = \operatorname{argmax}_L \{((1 - \alpha)P(L_j | L_i) + \alpha P_{C.L}(L_j)) \prod_{a_k \in \mathbf{A}_u} ((1 - \beta)P(a_k | L_{i \rightarrow j}) + \beta P_{C.A}(a_k))\}. \quad (11)$$

To avoid floating-point underflow when computing products of probabilities, all of the computations are in log-space.

Because the term $((1 - \alpha)P(L_j | L_i) + \alpha P_{C.L}(L_j))$ considers the "frequency" of movements between locations, the recommendation using this probability is regarded as *frequency-based method* (F). Similarly, using the term $\prod_{a_k \in \mathbf{A}_u} ((1 - \beta)P(a_k | L_{i \rightarrow j}) + \beta P_{C.A}(a_k))$ in recommendation is the *attribute-based method* (A). The combination of the F and A (i.e., Equation 11) is denoted as "F+A".

8. ROUTE PLANNING

The *route planning* problem can be treated like this: starting from location L_s and ending at location L_e , we want to go through N interesting destinations, and we have designated profile (or attributes) A_u (e.g., male).

We can take the graph-based method to solve the problem. A graph is generally composed of nodes, edges, and the weights on the edges. **Nodes**: each destination in the city is a node in the graph. **Edges**: The edge connects two nodes. From our learning model, we can calculate score of edges from the paths connecting the two destinations, however, parameterized by the user profile A_u . Because we want

Table 2: The 4 million geo-referenced photos collected from eight major cities in the world.

City	Images	Trips with face	Face in trips	Trip Segments
Manhattan	718,194	1,621	5,871	26,734
Rome	494,483	523	1,684	12,289
Barcelona	538,027	467	1,523	6,108
Prague	289,304	304	857	6,326
Hong Kong	446,813	379	1,393	4,316
Berlin	530,298	294	778	3,639
Taipei	579,033	585	3,242	3,658
Venice	334,513	199	606	3,140
Total	3,930,665	4,372	15,954	66,210

to go through popular landmarks rather than obscure ones, $P(L_i)$ should be placed into consideration. The other factor is the distance between candidate destinations; users do not prefer a devious journey, so we need a penalty function to diminish the overall route distance. Besides, we do not want to visit a destination twice, so we add the constraint that if the destination has been visited, the weight connecting the destination (from the current destination) is 0. Thus the weights for the edges can be formulated as:

$$s_{ij} = \begin{cases} 0 & , \text{if } j \text{ was visited} \\ P(L_i)P(L_{ij}|A_u) + \frac{\omega}{1+\text{dist}(L_i, L_j)}, & \text{otherwise} \end{cases} \quad (12)$$

where $\text{dist}(L_i, L_j)$ is the geo-distance between destination L_i and L_j . The former term can be derived from Section 7. The latter term, $\frac{\omega}{1+\text{dist}(L_i, L_j)}$, is a penalty function of distance, and ω is the weight for this penalty function.

Because of the smoothing technique mentioned in Section 7.2, it is a fully connected and directed graph. The route planning problem is to recommend best travel route from L_s to L_e via N other destinations. We define the route R as:

$$R = \{r_0, r_1, \dots, r_N, r_{N+1}\} \quad (13)$$

where r_i is the i -th destination of the route R . By definition, $r_0 = L_s$, and $r_{N+1} = L_e$. In addition, s_{ij} is the score from destination i to j . We will locate a route that connecting the starting and ending destinations that maximize the scores among the traversed edges. Thus, the route we want to find is:

$$R^* = \underset{R}{\operatorname{argmax}} \left\{ \sum_{k=0}^N s_{r_k r_{k+1}} \right\} \quad (14)$$

Therefore, the route problem can be treated as the shortest path problem. In our constructed graph, the path with the highest score is the route we expect for recommendation. As mentioned in Section 7.2, we can get L^* , which is the maximum score between current and the next destination given A_u (designated user profile). Hence we can solve this problem by dynamic programming to derive the best route.

9. EXPERIMENTS

Our dataset was collected by downloading images and photo metadata using Flickr API for the major cities such as Manhattan, Rome and Barcelona. Each city's boundary can be determined by the Wikimapia³, which is an on-line collaboratively map service containing the user-contributed geographic coordinates. We first issue a roughly rectangle boundary query and then use the precise boundary information to remove the photos outside the city. Table 2 summa-

³www.wikimapia.org

rizes the dataset statistics. These real travel logs are practical for interpreting the performance (by accuracy) of our recommendation algorithm. Although there are total about 4 millions images, we only experiment on the trips which contain detected attributes, i.e., those trips with face detected. In total, there are about 16K faces in the 4,372 trips as shown in the third and fourth columns in Table 2. The observations (L_i, L_j, \mathbf{A}) are generated by separating each trip into segments and duplicating these segments according to the detected face number in the trip. That is, faces belonging to the same trip share the same observations but have different attribute information (Section 6). Without loss of generality, we discard the trip segments not including in at least 2 different people.

For the processing time of our framework, the backend processes, including people attribute detection, trip generation, and learning prediction model are all in the off-line procedures which can be separated into different data splits and tasks to be processed in the distributed computation framework (e.g., Hadoop/MapReduce). The efficiency of the on-line recommendation scenarios can be easily achieved by the proposed Bayesian probability model. For on-line time cost, we measured on four randomly selected cities using the non-optimized MATLAB codes. On the average, it takes 0.0032 (sec) for personalized mobile recommendation. For route planning, we issued 5 queries per city, with randomly selected start/end locations and fixed 3 destinations in the midway; the averaged computation time is 3.5076 (sec).

9.1 Mobile Travel Recommendation

The first experiment is designed to verify the performance of our probability model on travel recommendation. We adopt a 5-fold cross-validation approach for the experiment. That is, the observations (i.e., the combination of (L_i, L_j, \mathbf{A})) are first split into 5 folds; 5 iterations of training and testing are performed; the training set contain 4 folds iteratively. Instances in the testing fold are excluded if their initial locations are not seen in the training set because they can not benefit from the recommendation model. For the background smoothing factors, α and β , we conduct a sensitivity test which varies the parameters by increments of 0.1 between 0 and 1. The best performance is measured by the average cross-validation accuracy, which is defined by *precision at one* ($P@1$) for the recommended destination. We set the parameters as those performed the best in this experiment.

In our experiment, we consider three kinds of people attributes, i.e., gender, age, and race, where gender $\in \{\text{male, female}\}$, age $\in \{\text{elder, middle-aged, teenager, kid}\}$, and race $\in \{\text{Caucasian, African, Asian}\}$. For each kind of attribute, we apply the detectors for an input face and assign this face to the class with the highest score. Take gender as an example; if the score from male detector is higher than female's, this face will be classified as a male.

For measuring the frequency of attributes in Equation 7, it just needs to count on how many peoples in the same movement sequence having the same attribute value a_u (e.g., $a_u = \text{male}$).

9.1.1 Impact of Background Smoothing

The trade-off parameters α and β in the proposed algorithm influence the relative contribution from the background destination popularity and the people's attribute dis-

Table 3: Improvements in P@1 for mobile travel recommendation across eight worldwide cities by exploiting the mined travel patterns and people attributes. R(5) is the baseline that recommends randomly choosing the top 5 nearest destinations. F represents the method which considers the visiting frequency between locations only. F+A combines the people attributes mined from photos and the F-based method. The mined knowledges, either travel frequencies and people attributes, both are helpful for mobile travel recommendation. We further categorize the city locations into different classes determined by prediction entropy $H(L_{ij}) = H(P(L_{i \rightarrow j}))$, measured in the collection and predicting by the travel frequencies only. The entropy $H(L_{ij})$ is higher as more diverse and hard choices for recommending the next travel destination. As observed, the relative gain between F and F+A (cf. the sixth and the last columns) increase for the difficult predictions.

City	Entropy	R(5)	F	F+A	F vs. F+A (%)	City	Entropy	R(5)	F	F+A	F vs. F+A (%)
Manhattan	$H(L_{ij}) \geq 0$	0.13	0.39	0.39	0.5	Hong Kong	$H(L_{ij}) \geq 0$	0.14	0.57	0.58	1.6
	$H(L_{ij}) \geq 1$	0.13	0.35	0.35	0.6		$H(L_{ij}) \geq 1$	0.14	0.41	0.43	4.6
	$H(L_{ij}) \geq 2$	0.11	0.28	0.28	0.03		$H(L_{ij}) \geq 2$	-	-	-	-
Rome	$H(L_{ij}) \geq 0$	0.10	0.33	0.35	6.5	Berlin	$H(L_{ij}) \geq 0$	0.14	0.51	0.52	2.6
	$H(L_{ij}) \geq 1$	0.10	0.28	0.30	8.5		$H(L_{ij}) \geq 1$	0.12	0.39	0.41	6.0
	$H(L_{ij}) \geq 2$	0.09	0.23	0.25	8.5		$H(L_{ij}) \geq 2$	0.08	0.28	0.32	13.5
Barcelona	$H(L_{ij}) \geq 0$	0.13	0.48	0.51	6.4	Taipei	$H(L_{ij}) \geq 0$	0.18	0.57	0.58	1.3
	$H(L_{ij}) \geq 1$	0.12	0.38	0.41	9.6		$H(L_{ij}) \geq 1$	0.16	0.43	0.44	1.7
	$H(L_{ij}) \geq 2$	0.12	0.26	0.30	14.7		$H(L_{ij}) \geq 2$	-	-	-	-
Prague	$H(L_{ij}) \geq 0$	0.14	0.42	0.44	6.1	Venice	$H(L_{ij}) \geq 0$	0.15	0.40	0.41	3.5
	$H(L_{ij}) \geq 1$	0.14	0.38	0.41	6.4		$H(L_{ij}) \geq 1$	0.14	0.37	0.38	4.5
	$H(L_{ij}) \geq 2$	0.10	0.30	0.30	-0.2		$H(L_{ij}) \geq 2$	0.11	0.29	0.30	3.5

tributions in each city. We find that for most of the cities, i.e., Manhattan, Rome, Barcelona, the optimal α is below 0.4, except for Hong Kong that has a higher α value (0.8). This phenomenon probably is caused by the sparse distribution of the observation data of Hong Kong such that most of the destinations have few faces detected. For other cities, the trips spread out well over the destinations, and can accurately predict using the training corpus.

9.1.2 The Recommendation Performance

Figure 5 shows the performance of the destination recommendation. We compare our knowledge-based methods with two baselines which randomly recommend the next destination by the top 1 (i.e., R(1)) or top 5 (i.e., R(5)) nearest destinations. The baselines are intuitive and proper because usually people will choose the nearby location as their next destination. As observed, the knowledge-based methods (i.e., frequency, combined) outperform the two baselines over 20% with respect to P@1. The method with attribute only sometimes performs worse than top 1 random due to the dense location distribution in these cities, incorporating the attribute into frequency factors generally increases the performance, which indicates that the attributes indeed bring up additional benefits besides those travel frequencies between destinations.

9.2 The Impact of People Attributes

In Figure 5, we can see that with the help of the location frequency mined from the historical travel logs, the performance is much better than the two baseline methods. The method with attribute only also improves but not so much as the method with frequency only. The phenomenon is reasonable since for major landmarks the visiting sequences between them are stable and people of different attributes might all likely travel between them. Therefore, we look at a more difficult testing location subset where the uncertainty of the number of choices for the next destination is large, i.e., where the conditional entropy $H(L_{ij})$ is above a certain threshold. A location with a higher movement (or prediction) entropy $H(L_{ij})$ means that the choices of the next stop are more diverse for the people who leave from this location.

For these cases, the attribute information is more useful to aid the performance of recommendation.

Table 3 presents the performance gain between different methods. We compare the random choice among the five nearest R(5) baseline with the two knowledge-based methods. As shown in Table 3, the improvement comparing with baseline degrades as the entropy increases. The gain between the frequency (F) and the combined method (F+A) usually increases when the testing subset becomes more harder to predict ($H(L_{ij}) \geq 0$ to $H(L_{ij}) \geq 1$). There are some missing entries in Table 3 due to the lack of testing sets.

9.3 Route Planning

In the route planning framework, we take the graph-based method to solve this problem and the scores (over edges) are all based on the above destination recommendation steps which have been shown effective through the experiment (Section 9.1).

Initially we recruit eight testers to use the designed route planning system for travel planning on the cities they are familiar with. The ratings are satisfactory and consistent with their expectations. Here we list 3 cases from 3 major cities and try to analyze the results⁴.

In the case of the routes from Central Park to Battery Park in Manhattan (see Figure 6(a)), we can see that Metropolitan Museum of Art is famous that it becomes a must-see landmark for people no matter what gender they have; Intrepid Sea, Air & Space Museum, a military and maritime history museum, is apparently male-tended. Note that the route planning is suggested according to the user profiles (i.e., male and female) and the learned travel preferences between destinations, automatically mined from the community-contributed photos.

In the case of the routes from Pincio to Piramide di Caio Cestio in Rome (see Figure 6(b)), it shows that females prefer going to Fontana Dei Quattro Flumi, while males like visiting Vatican Museum. In this case, we hypothesize that

⁴Note that our system assumes the higher score between two destinations reflects the fact that the traffic between them is accessible. We do not directly concern whether there are some barriers (e.g. river) or traffic barriers via the route.

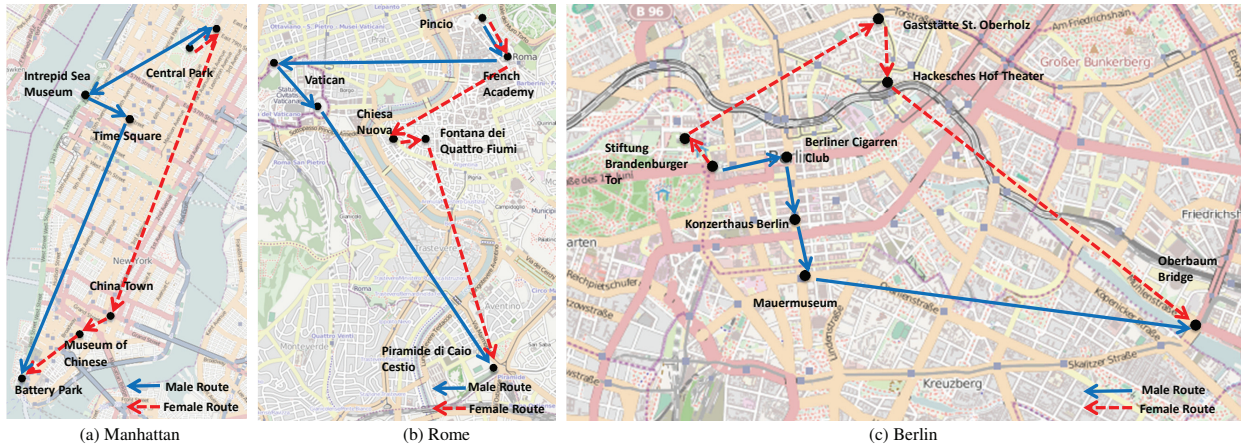


Figure 6: The route planning samples, with designated profiles – male vs. female, in 3 cities: (a) Manhattan, (b) Rome, and (c) Berlin. The blue route is for male, while the red one for female. Both routes in the same city are fixed at the same starting and end locations (or destinations). In addition, they must go through 3 other destinations in midway. Our route planning system could recommend different routes considering the given user profiles such as gender, age. From these results, Intrepid Sea, Air & Space Museum (i.e., Intrepid Sea Museum in (a)) and Berliner Cigarren Club tend to be suggested to males. (Best seen in color.)

the total distance might matter; that is the route through Vatican is longer.

In the case of the path from Stiftung Brandenburger Tor to Oberbaum Bridge in Berlin (see Figure 6(c)), Berliner Cigarren Club seems male-favored, while Hackesches Hof Theater is suggested for the female route. A leisurely route with theater visiting is more attractive than the cigarette museum to females. Besides, Mauermuseum in the route of male is the famous crossing point on the Berlin Wall.

Generally speaking, there are some landmarks very popular with both males and females, but it is without doubt that there are still some landmarks favored by only males or females according to the recommended routes of the same starting and ending destinations. By such attribute knowledge, our system is able to provide personalized route planning given the users profiles (i.e., gender, race, age)

10. CONCLUSIONS

In this work, we propose a probabilistic personalized travel recommendation model which adopts the automatically mined knowledges from the travel photo logs and the automatically detected people attributes in the photo contents. By information-theoretic measures and experiments in eight major cities, we confirm that people attributes are effective for mining demographics for travel landmarks and paths and very helpful for personalized travel recommendation and route planning. Meanwhile, people attributes are orthogonal to the travel logs alone and can further yield more satisfactory results especially in more challenging recommendations. For the future work, we are conducting experiments on more people attributes and adjust the probabilistic model for such diverse attributes to address more capabilities in personalization. Besides, the more competitive recommender models need to be investigated as well. We also want to expand our model with more contexts such as travel durations, traveling seasons. We believe such location- and individual-aware models are promising for further applications such as advertisement.

11. REFERENCES

- [1] M. Ames et al. Why we tag: motivations for annotation in mobile and online media. In *ACM CHI*, 2007.
- [2] Y. Arase et al. Mining people’s trips from large scale geo-tagged photos. In *ACM MM*, 2010.
- [3] C.-C. Chang et al. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [4] A.-J. Cheng et al. Gps, compass, or camera?: investigating effective mobile sensors for automatic search-based image annotation. In *ACM MM*, 2010.
- [5] M. Clements et al. Using flickr geotags to predict user travel behaviour. In *ACM SIGIR*, 2010.
- [6] T. M. Cover et al. *Elements of Information Theory*. Wiley, 1991.
- [7] D. J. Crandall et al. Mapping the world’s photos. In *WWW*, 2009.
- [8] Y. Gao et al. W2go: a travel guidance system by automatic landmark ranking. In *ACM MM*, 2010.
- [9] Q. Hao et al. Equip tourists with knowledge mined from travelogues. In *WWW*, 2010.
- [10] R. Ji et al. Mining city landmarks from blogs by graph modeling. In *ACM MM*, 2009.
- [11] N. Kumar et al. Facetracer: A search engine for large collections of images with faces. In *ECCV*, 2008.
- [12] Y. Li et al. Landmark Classification in Large-scale Image Collections. In *IEEE ICCV*, 2009.
- [13] D. Liu et al. Location sensitive indexing for image-based advertising. In *ACM MM*, 2009.
- [14] X. Lu et al. Photo2trip: generating travel routes from geo-tagged photos for trip planning. In *ACM MM*, 2010.
- [15] T. Mei et al. Knowledge discovery from community-contributed multimedia. *IEEE MultiMedia*, 17, October 2010.
- [16] V. K. Singh et al. Social pixels: genesis and evaluation. In *ACM MM*, 2010.
- [17] P. Viola et al. Rapid object detection using a boosted cascade of simple features. In *IEEE CVPR*, 2001.
- [18] Y. Zheng et al. Mining interesting locations and travel sequences from gps trajectories. In *WWW*, 2009.
- [19] Y. Zheng et al. Learning travel recommendations from user-generated gps traces. *ACM Trans. Intell. Syst. Technol.*, 2, January 2011.
- [20] Y. Zheng et al. Recommending friends and locations based on individual location history. *ACM Trans. Web*, 2011.