International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 03, Volume 4 (March 2017) Special Issue www.ijirae.com

Personalized Travel Sequence Recommendation on Multi-Source Big Social Media

O.K.Gowrishankar[1]
Assistant Professor
Computer Science and Engineering
Sengunthar College of Engineering, Tiruchengode, India

Archana.K[2]
Computer Science and Engineering
Sengunthar College of Engineering
Tiruchengode, India

Abstract— ***Big data increasingly benefit both research and industrial area such as health care, finance service and commercial recommendation. This paper presents a personalized travel sequence recommendation from both travelogues and community-contributed photos and the heterogeneous metadata (e.g., tags, geo-location, and date taken) associated with these photos. Unlike most existing travel recommendation approaches, our approach is not only personalized to user's travel interest but also able to recommend a travel sequence rather than individual Points of Interest (POIs). Topical package space including representative tags, the distributions of cost, visiting time and visiting season of each topic, is mined to bridge the vocabulary gap between user travel preference and travel routes. We take advantage of the complementary of two kinds of social media: travelogue and community-contributed photos. We map both user's and routes' textual descriptions to the topical package space to get user topical package model and route topical package model (i.e., topical interest, cost, time and season). To recommend personalized POI sequence, first, famous routes are ranked according to the similarity between user package and route package. Then top ranked routes are further optimized by social similar users' travel records. Representative images with viewpoint and seasonal diversity of POIs are shown to offer a more comprehensive impression. We evaluate our recommendation system on a collection of 7 million Flickr images uploaded by 7,387 users and 24,008 travelogues covering 864 travel POIs in 9 famous cities, and show its effectiveness. We also contribute a new dataset with more than 200K photos with heterogeneous metadata in 9 famous cities.***

Keywords: Travel recommendation, geo-tagged photos, social media and multimedia information retrieval.

I. INTRODUCTION

Automatic travel recommendation is an important problem in both research and industry. Big media especially the flourish of social media (e.g. Facebook, Flick, Twitter etc.) offers great opportunities to address many challenging problems, for instance, GPS estimation [1], [2] and travel recommendation [3]. Travelogue websites (e.g., www.igougo.com) offer rich descriptions about landmarks and traveling experience written by users. Furthermore, community-contributed photos with metadata (e.g., tags, date taken, latitude etc.) on social media record users' daily life and travel experience. These data are not only useful for reliable POIs (points of interest) [4], travel routes but give an opportunity to recommend personalized travel POIs and routes based on user 's interest. There are two main challenges for automatic travel recommendation.

First, the recommended POIs should be personalized to user interest since different users may prefer different types of POIs. Take New York City as an example. Some people may prefer cultural places like the Metropolitan Museum, while others may prefer the cityscape like the Central Park. Besides travel topical interest, other attributes including consumption capability (i.e., luxury, economy), preferred visiting season (i.e., summer, autumn) and preferred visiting time (i.e., morning, night) may also be helpful to provide personalized travel recommendation.

Second, it is important to recommend a sequential travel route (i.e., a sequence of POIs) rather than individual POI. It is far more difficult and time consuming for users to plan travel sequence than individual POIs. Because the relation- ship between the locations and opening time of different POIs should be considered. For example, it may still not be a good recommendation if all the POIs recommended for one day are in four corners of the city, even though the user may be interested in all the individual POIs.

- *Our work is a personalized travel recommendation rather than a general recommendation. We automatically mine user 's travel interest from user- contributed photo collections including consumption capability, preferred time and season which is important to route planning and difficult to get directly.*
- *We recommend personalized POI sequence rather than individual travel POIs. Famous routes are ranked according to the similarity between user package and route package, and top ranked famous routes are further optimized according to social similar users' travel records.*
- *We propose Topical Package Model (TPM) method to learn users and route's travel attributes. It bridges the gap of user interest and routes attributes. We take advantage of the complementary of two big social media to construct topical package space.*

II. RELATED WORK

In this section, we mainly introduce three aspects of related works (1) travel recommendation on various big social media; (2) personalized travel recommendation; (3) travel sequence and travel package recommendation. We also point out the differences between our work and existing works.GPS trajectory [5], check-in data [4], [6], [7] geo-tags [2], [3], [8], [9], [10] and blogs (travelogues) [11], [12] are four main social media used in recommendation. User- generated travelogues provide rich information. Kurashima et al. Extracted typical user's travel sequences according to entries, associated with multimedia information of the routes [12]. Besides travelogues, GPS and geo-tags are also widely utilized in travel recommendation. Zheng et al. conducted a series of works of travel routes mining and recommendation using GPS trajectory, and achieved promising results [5], [16], [17], [18]. However, comparing to the rich travelogues and geo-tags data on social media, GPS trajectory data are relatively difficult to obtain. Geo-tagged photos based automatic travel route planning works have attracted lot attentions [8], [9]. Recently, multi-source big social media have shown their robustness [9], [19], [20]. Liu et al. discovered Areas of Interest by analyzing geo-tagged image and check-ins data simultaneously [19].

III.SYSTEM OVERVIEW

The system we proposed is a personalized POI sequence recommendation system which could automatically mine user 's travel attributes such as topical interest, consumption capability and preferred time and season. In this section, we briefly introduce the terms used in this paper: topical package space, user package and route package. Secondly, we provide the system overview. Topic package space is a kind of space in which the four travel distributions of each topic are described by (1) representative tags mined from travelogues which describe POIs within the same topic; (2) the average consumer expenditure of the POIs within this topic, which are also mined from travelogues; (3) distribution of the visiting season of the twelve months mined by the "date taken" attached with the community-contributed photos; (4) distribution of visiting time during the day from travelogues. The usage of topic package space is to bridge the gap between user interest and the attribute of routes, since it is difficult to directly measure the similarity between user and travel sequence. From mapping both user information and route information to the same space, we get the quantitative standard to measure the similarity of user and routes.

IV. SOCIAL MEDIA MINING AND TOPICAL PACKAGE SPACE CONSTRUCTION

Our topic package space is the extension of textual descriptions of topics such as ODP [35]. We use the topical package space to measure the similarity of the user topical model package (user package) and the route topical model package (route package). In our paper, we construct the topical package space by the combination of two social media: travelogues and community-contribute photos. To construct topical package space, travelogues are used to mine representative tags, distribution of cost and visiting time of each topic, while community-contributed photos are used to mine distribution of visiting time of each topic. The reasons for using the combination of social media are

1. *Travelogues are more comprehensive to describe a location than the tags with the photos which are with so many noises [36], [37];*
2. *It is difficult to mine a user 's consumption capability and the cost of POIs directly by the photos or the tags with the photos;*
3. *To season, although both media could offer correct visiting season information of POIs, the number of photos of a POI is far larger than the number of travelogues.*
4. *The time difference between where the user lives and the "data taken" of community- contributed photos of where he or she visits make the taken time inaccurate.*

4.1 Travelogue Mining

First, we introduce the gathering and the structure of travelogues. Then we introduce how to mine representative tags, distribution of cost and time of each topic.

4.1.1 Travelogues Gathering and Structure Illustration

We downloaded 24,008 travelogues of 864 travel POIs on nine most famous cities of the world from famous travel website IgoUgo.com. These nine cities are Barcelona, Berlin, Chicago, London, Los Angeles, New York, Pairs, Rome and San Francisco [14]. A lot of travelogue related works are based on the data from IgoUgo [36], [37].

In this paper, we directly use the category definition of IgoUgo as Table 2. This category could cover most of the travel activities. The structure of data we crawled from IgoUgo is as Fig.4. The first layer is "City Layer". Under each city, there are 26 topics constructed "Topic Layer".

CATEGORY OF THE TOPICS.

No	Name	No	Name
1	Bars and Clubs	14	Public Transportation
2	Beaches	15	Scenic Drives
3	Casinos	16	Shopping
4	Cheap things to do	17	Sightseeing
5	Concerts and Shows	18	Skiing
6	Golf and Spas	19	Snorkeling and Scuba
7	Gyms and Spas	20	Spectator Sports
8	Hiking	21	State and National Parks
9	Lakes and Rivers	22	Theaters and Movies
10	Mountains	23	Theme Parks
11	Museums	24	Tours
12	Outdoor POIs and Activity	25	Wineries and Breweries
13	Parks and Gardens	26	Zoo and Aquariums

4.2.1 POI MINING

So first we introduce the way to mine POIs from crowded geo-tagged photos. POIs mining is a hot research area in recent years. First, we filter a set of photos for each city from all the users. We match city name, for example, London, with the textual tags of each photo. It cannot guarantee that all the photos matching city name definitely belong to this city, since community-contributed photos include a lot noises. We further use the geo-location restriction. If the GPS coordinate of the photo is 500km (between region level and country level) [39] away from the center of the city, we remove it. After getting a set of photos of each city, second, we extract POIs from these crowded geo-tagged photos toward each city by mean shift clustering [14], [15]. Then we choose the POIs in both the clusters and the travelogue website. Thus, these POIs have both GPS coordinates and travelogues description, which could guarantee the routes plan and routes package mining.

4.2.2 TOPIC'S SEASON MATRIX MINING

After getting POIs, to each POI, there are a set of photos with tags and "date taken" labels. To season, we use the "month" in "date taken" to get the visiting distribution during the 12 month. The season vector of a POI is defined as $\zeta(P) \in$[spring, summer, autumn, winter]. Months from March to May belong to spring and so on. According to the structure of travelogues, for each topic, we average over all the season distributions of the POIs in this topic. The season matrix $\zeta(M)$ is a $N \times 4$ matrix.

4.2.3 REPRESENTATIVE IMAGES MINING FOR POI

In order to offer vivid impression of the travel sequence, our system also provides representative images of the POIs on the route. We consider two factors of the representative images. First, we present representative viewpoints using the 4-D viewpoint vector model (i.e., horizontal, vertical, scale and orientation) [40],[41]. The diverse viewpoints could offer more comprehensive knowledge of the POI. Second, as POIs may show quite different characteristics in different seasons, we provide representative images of each season. To achieve season diversity, we extract the "date taken" information from metadata of the image, and divide the photos into four seasons. Fig.5 shows the example results of representative images.

V. USER TOPICAL PACKAGE MODEL MINING

In this section, we introduce how to extract the user package $[\alpha(U), \beta(U), \gamma(U), \zeta(U)]$, which contains user topical interest distribution $\alpha(U)$, user consumption capability distribution $\beta(U)$, preferred travel time distribution $\gamma(U)$ and preferred travel season distribution $\zeta(U)$. First we introduce user 's topical interest $\alpha(U)$ mining from mapping user 's tags to the topical package space. Then, we introduce how to get $\beta(U)$, $\gamma(U)$ and $\zeta(U)$ through topical space mapping method.

5.1 USER TOPICAL INTEREST MINING

Fig.4 illustrates user topical interest mining method. We map the textual description (tags) of user's community photos to the topical package space to present the user 's travel preference of different topics, which is defined as user topical interest distribution $\alpha(U)$. We assume that if a user 's tags appear frequently in one topic and less in others, the user has a higher interest towards this topic.

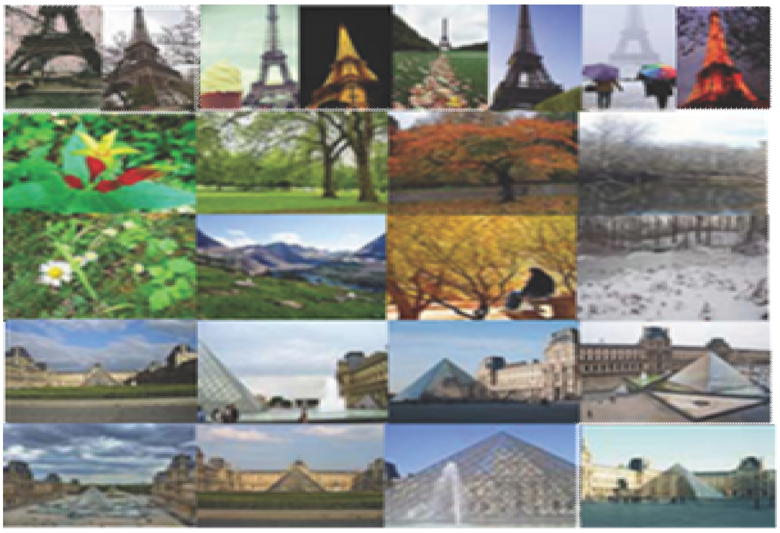

Fig.2 Examples of representative images

The figure shows representative images in three example POIs,(a) Eiffel Tower (row 1), (b) Hide Park (row 2, 3), (c) The Louvre Museum (row 4, 5). We offer two images of each season. In (b), we could see that the scenes of different season are diverse. In (c), images of different seasons are similar, but the view points are different. In (a), both season and viewpoints are diverse.

5.2 COST, TIME AND SEASON DISTRIBUTION MINING

Although the user 's photo set contains lots of information about the user, we still could not directly get the consumption capability and the time preference of the user. The easiest way to obtain the time preference seems to analyze the "date taken" of the photo. However, the time difference of the country between where the user lives and where he or she visits may cause errors. We may find the case that the "date taken" of the photo is at noon but the photo is a night scene. Here, we introduce our method to get user cost distribution $\beta(U)$, time distribution $\gamma(U)$ and season distribution $\zeta(U)$ of the user with the help of user topical interest distribution $\alpha(U)$ and cost matrix $\beta(M)$, time matrix $\gamma(M)$ and season matrix $\zeta(M)$ in topical package space. First, we introduce how to get $\beta(U)$ as:

$$\beta(U) = \alpha(U) \cdot \beta(M),$$

where $\beta(U)$ is an one-dimension vector, defined of the dot product of user topical interest distribution $\alpha(U)$ and cost matrix $\beta(M)$. We use the cost distributions of the all the topics and distribution of use's topical interest to present a user 's consumption capability.

VI. ROUTE TOPICAL PACKAGE MODEL MINING

In this section, we firstly describe travel route mining from community-contributed photos

6.1. ROUTE MINING

To save the online computing time, we mine travel routes and the attribute of the routes offline. After mining POIs, to construct travel routes, we analyze the spatiotemporal structure of the POIs among travelers' records.

6.2 ROUTE PACKAGE MINING

In this section, we describe routes' topical package model mining. We firstly mine POIs' package including POI topical interest distribution $\alpha(P)$, POI cost distribution $\beta(P)$, time distribution $\gamma(P)$ and season distribution $\zeta(P)$. Then to each route, we average all the POIs on the route to get route topical package model.

VII. TRAVEL SEQUENCE RECOMMENDATION

After mining user package $[\alpha(U), \beta(U), \gamma(U), \zeta(U)]$ and route package $[\alpha(R), \beta(R), \gamma(R), \zeta(R)]$, in this section, we introduce our travel routes recommendation module. It contains two main steps:

(1) Routes ranking according to the similarity between user package and routes packages, and

(2) Route optimizing according to similar social users' records

7.1 ROUTES RANKING

Assume $R = \{r1, r2, ..., rn\}$ is a set of n travel routes mined offline. We rank these routes according to the similarity between user package and routes packages. For user uj and route ri, we measure the similarity of each attribute among topical interest, cost, time season, denoted as $\varphi(\alpha)$, $\varphi(\beta)$, $\varphi(\gamma)$, $\varphi(\zeta)$ respectively

7.2 Route Optimizing

After POI and route ranking module, we get a set of ranked routes Rˆ. Here, we further describe the optimization of top ranked routes according to social similar users' travel records. Firstly, we introduce how to mine social similar users and their travel records. Then we introduce how to optimize the roads by social users' travel records

7.2.1 Social Similar Users based POI Ranking

The well known Location-based Collaborative Filtering (LCF) firstly mine similar users according to the high-occurrence of GPS histories [4], [5]. Then the top popular POIs among similar users' travel records would be recom mended to the user. However, if there are very few GPS records in user 's photo set, it is difficult to find location-based similar users accurately. This is called "sparsity problem".

VIII. EXPERIMENTS

In this section, we first introduce the dataset and evaluation criteria. Then we show the evaluation of the proposed framework of (1) POI recommendation, (2) route recommendation, (3) POI package mining, and (4) user package mining. We also demonstrate the visualization of the system.

8.1 Dataset

Our dataset consists of travelogues and community contributed photos. For travelogues, we downloaded almost 24,008 travelogues of 864 travel POIs on nine famous cities according to [14] over the world from the famous travel website IgoUgo. Table 3 shows the number of users, POIs, travelogues of these nine cities. For community-contributed photos, we randomly download 7,387 users' photo albums associated with heterogeneous metadata. Table 4 shows the corresponding number of users, POIs, photos in each city. There are 2,892 users, 307 POIs and 150,101 photos in total, after we remove the photos whose geo-tags or tags are missing. The community-contributed photos dataset is the same as [3]. Since CF based approaches mine similar users in one city (city1) and recommend POIs in a new city (city2) according to similar users' votings, potential similar users are required to cover travel records in both city1 and city2. Define < city1, city2 > as a city group. There are 46 city groups finally used, in which each group contains at least 20 potential similar users. The total number of potential similar users are 1156, in which each user has more than 5 photos

8.2 Ground Truth and Criteria

8.2.1 Travel sequence recommendation evaluation

In travel route evaluation, four aspects should be concerned, representativeness, diversity, rationality and the overall satisfaction [15]. In our evaluation, except these four aspects, the volunteers should also consider whether the routes meet the user 's topical interest, consumption capability and time and season preference. In our work, we invited 12 volunteers with different ages and domains to conduct human evaluation as Argo [35]. Before evaluation, the volunteers have to learn users' interest by reviewing their traveling records including photo albums and textual travelogues. We use the average precision (AP) and weighted average precision (WAP) as the performance metrics [35]:

$$AP = (p + r)/(p + r + i),$$
$$WAP = (p + 0.5r)/(p + r + i),$$

Where p denotes the number of recommended routes that volunteers are very satisfied with r denotes the number of recommended routes which are relatively related to user 's preference, but still need to be improved. i denotes the recommended routes are not relevant to user 's preference. In our work, we evaluate 100 Flickr users and recommend 10 routes for each user.

8.2.2 POI recommendation evaluation

We follow the same criteria, mean average precision (M AP @n) [14], used in state-of-art POI recommendation work [3] to evaluate the POI recommendation module in our framework. Denote n as the number of recommended POIs, the average precision (APk @n) for user k is calculated as

$$APk\ @n = (Xn\ i=1(Xj=1relj,k\ )/i)/n$$

Where the relevance value relj,k = 1 if uk has visited the POI j according to ground truth. Otherwise, relj,k = 0. For a test user, the textual tags are used as inputs of the system, and the original geo-tags are used as ground truth

8.3 Evaluation of POI Recommendation

To evaluate the performance of POI recommendation by TPM, we compare TPM with recommendation by popularity (PO), collaborative filtering (CF) and Latent Dirichlet Allocation (LDA) based method under MAP@n.

Recommendation by POPULARITY (PO): It is non- personalized recommendation. Only the popularity of the POIs is considered as the criterion of ranking. We measure the popularity according to the number of users who upload photos related to this POI.

Recommendation by Collaborative Filtering (CF): Location-based collaborative filtering is a widely method in recommendation system and it can be easily implemented [4], [5].

First of all, user-POI matrix is constructed from users' location records. Then similar users are detected through this user-POI matrix. Finally POIs are recommended based on similar users' travel records

RECOMMENDATION BY LATENT DIRICHLET ALLOCATION (LDA): To test the impact of the combination of travelogue and community-contributed photos, we compare our TPM with Latent Dirichlet Allocation (LDA) [42] based travel recommendation, in which only the community-contributed photos are used

8.4 DISCUSSION OF IMPACTS OF ATTRIBUTES IN TPM

In order to discuss the impacts of four attributes in TPM (i.e., user topical interest, cost, time and season) in Eq. (9), in this section, we conduct two sets of experiments. First, we discuss the performance if only one of the cost, time and season is combined with user topical interest in Table 6. Second, we discuss the impact under different setting of weights in Eq. (9) in Fig.7. The meanings of notations in Table 6 are shown as follows: α means that we only use the similarity of user topical interest to detect similar users. Attributes of cost, time and season are not considered. $\alpha + \beta$ means that when detecting similar users, we use the attributes of both user topical interest and cost. $\alpha + \gamma$ means that when detecting similar users, we use the attributes of both user topical interest and time. $\alpha + \zeta$ means that when detecting similar users, we use the attributes of both user topical interest and season. Table 6 shows the recommendation results of TPM (ours) on MAP@1, 5, 10, 20 and 30 in comparison with α, $\alpha + \beta$, $\alpha + \gamma$ and $\alpha + \zeta$. Note that in Table 6 all the weights in TPM are set as 1. This is not the best performance of TPM when comparing with other settings of weights, shown in Fig Impact of the weight of topical interest, cost, time and season.

8.5 EVALUATION ON ROUTE RECOMMENDATION

We compare our personalized routes recommendation with the other methods of routes planning as follows: Random Routes Planning (RAM): This method constructs travel route by randomly selecting 5 POIs. Famous Routes Planning (FAM): This method recommends the famous routes but without ranking and optimizing. The recommended POIs for all the users are the same.

8.6 EVALUATION ON POI PACKAGE MINING

We compare the results of POI package automatically mined with the information referred to the official website. POI's topical interest, cost, time mined by our method are denoted as IP, CP, TP, while results from the official website are denoted as IO, CO, TO. We also show four examples of visiting distributions of POIs during 12 month

8.7 EVALUATION ON USER PACKAGE MINING

In this section, we evaluate the performance of our user package mining method with AP and WAP on four attributes. Meanwhile, in order to evaluate the effectiveness of IgoUgo (IU), we compare IgoUgo category to Trip Advisor (TA) category as (a) beached sun, (b) casinos, (c) family fun, (d) history & culture and (e) skiing in the relevance-oriented recommendation [36]. Table 10 shows the AP and WAP permanence using IU and TA categories on topical interest, cost, time and season. IIU, CIU, DIU and SIU respectively represent performance on the topic, cost, time and season by IU category, and ITA, CTA, DTA and STA respectively represent the topical interest, cost, time and season by TA category.

IX. CONCLUSION

In this paper, we proposed a personalized travel sequence recommendation system by learning topical package model from big multi-source social media: travelogues and community-contributed photos. The advantages of our work are 1) the system automatically mined user 's and routes' travel topical preferences including the topical interest, cost, time and season, 2) we recommended not only POIs but also travel sequence, considering both the popularity and user 's travel preferences at the same time. We mined and ranked famous routes based on the similarity between user package and route package. And then optimized the top ranked famous routes according to social similar users' travel records. However, there are still some limitations of the current system. Firstly, the visiting time of POI mainly presented the open time through travelogues, and it was hard to get more precise distributions of visiting time only through travelogues. Secondly, the current system only focused on POI sequence recommendation and did not include transportation and hotel information, which may further provide convenience for travel planning. In the future, we plan to enlarge the dataset, and thus we could do the recommendation for some non-famous cities. We plan to utilize more kinds of social media (e.g., check-in data, transportation data, weather forecast etc.) to provide more precise distributions of visiting time of POIs and the context- aware recommendation.

REFERENCES

[1]. H. Liu, T. Mei, J. Luo, H. Li, and S. Li, "Finding perfect rendezvous on the go: accurate mobile visual localization and its applications to routing," in Proceedings of the 20th ACM international conference on Multimedia. ACM, 2012, pp. 9–18.

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

[3]. S. Jiang, X. Qian, J. Shen, Y. Fu, and T. Mei, "Author topic model based collaborative filtering for personalized poi recommendation," IEEE Transactions on Multimedia, vol. 17, no. 6, pp. 907–918,2015.
[4]. J. Sang, T. Mei, and C. Sun, J.T.and Xu, "Probabilistic sequential pois recommendation via check-in data," in Proceedings of ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2012.
[5]. Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W. Ma, "Recommending friends and locations based on individual location history," ACM Transactions on the Web, vol. 5, no. 1, p. 5, 2011.
[6]. H.Gao, J. Tang, X.Hu,and H. Liu, "Content-aware point of interest recommendation on location-based social networks," in Proceedings of 29th International Conference on AAAI. AAAI, 2015.
[7]. Q.Yuan, G.Cong, and A.Sun, "Graph-based point-of-interest recommendation with geographical and temporal influences," in Proceedings of the 23rd ACM International Conference on Information and Knowledge Management. ACM, 2014, pp. 659–668.
[8]. H. Yin, C. Wang, N. Yu, and L. Zhang, "Trip mining and recommendation from geo-tagged photos," in IEEE International Conference on Multimedia and Expo Workshops. IEEE, 2012, pp. 540–545.
[9]. Y. Gao, J. Tang, R. Hong, Q. Dai, T. Chua, and R. Jain, "W2go: a travel guidance system by automatic landmark ranking," in Proceedings of the international conference on Multimedia. ACM,2010, pp. 123–132.
[10]. X. Qian, Y. Zhao, and J. Han, "Image location estimation by salient region matching," IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 4348–4358, 2015.