USER QUERY BASED TRAVEL PATH ADVISORY SYSTEM BY USING DATA OBJECT MODELING (DOM) PARSER.

[1]Shital N. Raul, [2]Nitin N. Patil

[1] P. G. Student , [2] Head & Associate Professor
[1] Department of Computer Engineering
R. C. Patel Institute of Technology, Shirpur, India

Abstract : Big data has significantly benefitted both Research and Industrial area such as health care, finance service and commercial recommendation. This paper presents a user Query based travel advisory system from both travelogues and community-contributed photos and the heterogeneous metadata (e.g., tags, geo-location, and date taken) associated with the photos on social media. Unlike most existing travel recommendation approaches, our approach is not only personalized to user's travel interest and 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 using DOM tree parser algorithm to the topical package space to get user topical package model and route topical package model (i.e., topical interest, cost, time and season, Medical). To propose personalized POI order, initially, famous routes are ordered according to the similarity between user package and route package. Then apex ranked routes are more optimized by social similar users' travel records.

Index Terms - **Travel recommendation, Geo-tagged photos, DOM Parser, Multimedia Information Retrieval.**

I. INTRODUCTION

The assists of big data collects large volume of data, it is great computational challenge for the big data Hadoop which uses map reduce to maintain and process this data and also helps to extract useful information in an efficient manner . It helps people to take decision in advance for their any outdoor events.

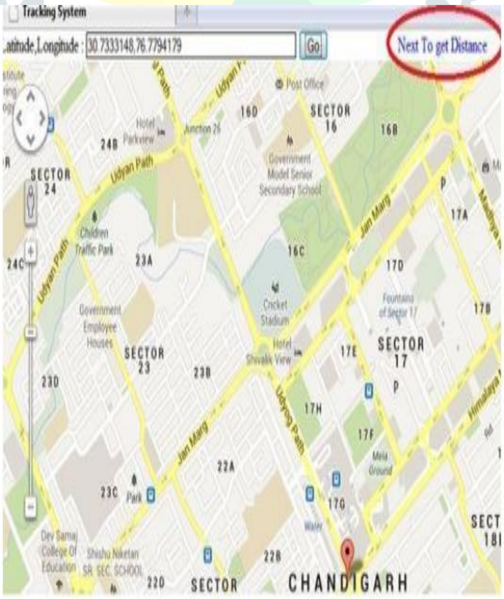


Figure1:Location based route tracking.

Gives the exact location of the target. The applications also alerts about the distance travelled by the target and also the¬ routes that are possible to reach to the target. Automatic travel recommendation is an important problem in both research and industry. Big media, especially the flourish of social media (e.g., Facebook, Flickr, 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) mining [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 relationship 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. Existing studies on travel recommendation mining famous travel POIs and routes are mainly from four kinds of big social media, GPS trajectory [5], check-in data [4].and blogs (travelogues)[11], [12]. However, general travel route planning cannot well meet users' personal requirements. Personalized travel recommendation recommends the POIs and routes by mining user's travel records [13].

The most famous method is location-based collaborative filtering (LCF). To LCF, similar social users are measured based on the location co-occurrence of previously visited POIs. Then POIs are ranked based on similar users' visiting records. However, existing studies haven't well solved the two challenges.

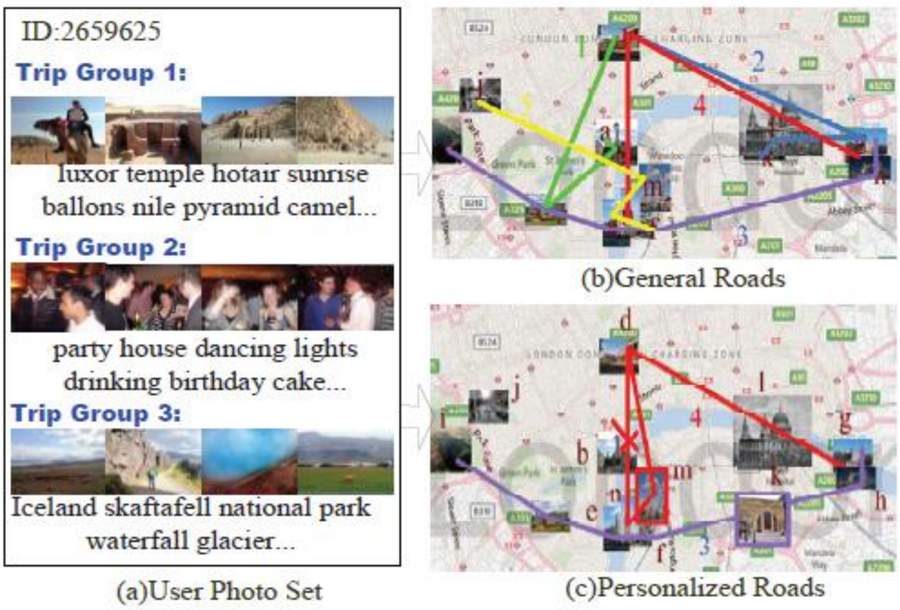


Figure.2 Example of our proposed PTSR.

(a) Example photo set of a real Flicker user. First, the photo share divided to trip groups according to "Data Taken". Three example groups are shown with the photos and the tags.(b) Results by general recommendation. five famous routes of London we extracted are shown in the map.

We know from user package mining that this user prefers park, museum, wineries and sightseeing topical interest, could afford normal and luxury event, and prefers afternoon and evening visiting time and spring visiting season. The famous routes are ranked as 3,4,1,2,5 according to user package and route package. We analyze that the park distribution (i) raises the ranking of route 3 and the museum (d) and sightseeing (f, g) distributions raise the ranking of route4. We offer the optimized route 3 and route 4 as example. In route 3, the system optimizes the route by adding the "Vinopolis", which is a London commercial wine attraction, and it is on the way, In route 4 optimization, the system adds London Eye which meets user's travel preference and also on the way. The landmark names are shown as follows: a) Trafalgar Square, b) National Gallery, c) Buckingham Palace, d) British Museum, e) Westminster Abbey, f) Palace of Westminster, g) Tower of London, h) Tower Bridge, i) Hyde Park, j) Oxford street, k) Globe Theatre, l) Staple's Cathedral, m) London Eye, n) Big Ben. mediation works only focused6 on user topical interest mining but without considering other attributes like consumption capability. For the second challenge, existing studies focused more on famous route mining but without automatically mining user travel interest. It still remains a challenge for most existing works to provide both "**personalized**" and "**sequential**" travel package recommendation. To address the challenges mentioned above, we propose a Topical Package Model (TPM) learning method to automatically mine user travel interest from two social media, community-contributed photos and travelogues. To address the first challenge, we consider not only user's topical interest but also the consumption capability and preference of visiting time and season. As it is difficult to directly measure the similarity between user and route, we build a topical package space, and map both user's and route's textual descriptions to the topical package space to get user topical package model (user package) and route topical package model (route package) under 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 [7] geo-tags [10] and blogs (travelogues) 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 [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 a 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].However, general travel recommendations only considered the popularity of POIs or routes. Recently, personalized travel recommendations have attracted more attentions[13], [14], [21]. The three main approaches of personalized recommendation are Collaborative Filtering (CF) [14].

Location based CF firstly mined similar users according to location co-occurrence. For example, Clements et al. modeled the co-occurrence with Gaussian density estimation [14]. Second, POIs are recommended according to similar users' voting. However, location based CF may face two problems. First of all, the computational complexity increases dramatically with large amount of users and locations, which is especially serious in big data scenario. Second, if the user has very few location records or most of these records belong to non-famous places, it would be very hard to mine accurate similar users. To solve these challenges, Jiang et al. proposed the Author Topic Model based Collaborative Filtering [3]. They mined the category of travel topics and user topical interest simultaneously through Author Topic Model. Personalized travel sequence recommendation is more convenient for users than the individual POIs recommendation[15], [26], [27], [29].

The system enabled user to input personal performance in an interactive manner [15]. However it did not really automatically mine user's interest. What's more, in recent years, studies of the travel package recommendation which contained more attributes(e.g. time, cost, season) have shown more effective performance than works which only considered topical interest[7].Yuan et al. proposed a Geographical-Temporal influences Aware Graph for time aware POI recommendation [7]. Ge et al. developed a cost aware model, and they analyzed the relation between cost and stay days [31]. However, although these studies considered user's travel attributes, few of them really automatically mined these attributes .The existing studies related to travel sequence recommendation did not well consider the popularity and personalization of travel routes at the same time. What's more, the multi-attributes of users and routes (e.g., consumption capability, preferred season, etc.) have not been mined automatically.

To solve these problems, in this paper, first, topical package model is learnt to get users' and routes' multi-attributes (i.e., topical interest, cost, time, season preference).Second, we take the advantage of the complementation of travelogues and community-contributed photos. Third, we consider the popularity of routes and user's personal preference together by the idea of ranking famous travel routes based on user's travel interest, and optimizing top ranked routes by social similar users' travel

Global Positioning System:

Recent advances in position localization techniques have fundamentally enhanced social networking services, allowing users to share their locations and location-related content, such as geo-tagged photos and notes. We refer to these social networks as location-based social networks (LBSNs). Location data both bridges the gap between the physical and digital worlds and enables a deeper understanding of user preferences and behavior. This addition of vast geospatial datasets has stimulated research into novel recommender systems that seek to facilitate users' travels and social interactions. In this paper, we offer a systematic review of this research, summarizing the contributions of individual efforts and exploring their relations. We discuss the new properties and challenges that location brings to recommendation systems for LBSNs. We present a comprehensive survey of recommender systems for LBSNs, analyzing 1) the data source used, 2) the methodology employed to generate a recommendation, and 3) the objective of the recommendation. We propose three taxonomies that partition the recommender systems according to the properties listed above. First, we categorize the recommender systems by the objective of the recommendation, which can include locations, users, activities, or social media. Second, we categorize the recommender systems by the methodologies employed, including content-based, link analysis-based, and collaborative filtering-based methodologies. Third, we categorize the systems by the data sources used, including user profiles, user online histories, and user location histories. For each category, we summarize the goals and contributions of each system and highlight one representative research effort. Further, we provide comparative analysis of the recommendation systems within each category. Finally, we discuss methods of evaluation for these recommender systems and point out promising research topics for future work. This article presents a panorama of the recommendation systems in location-based social networks with a balanced depth, facilitating research into this important research theme.

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.

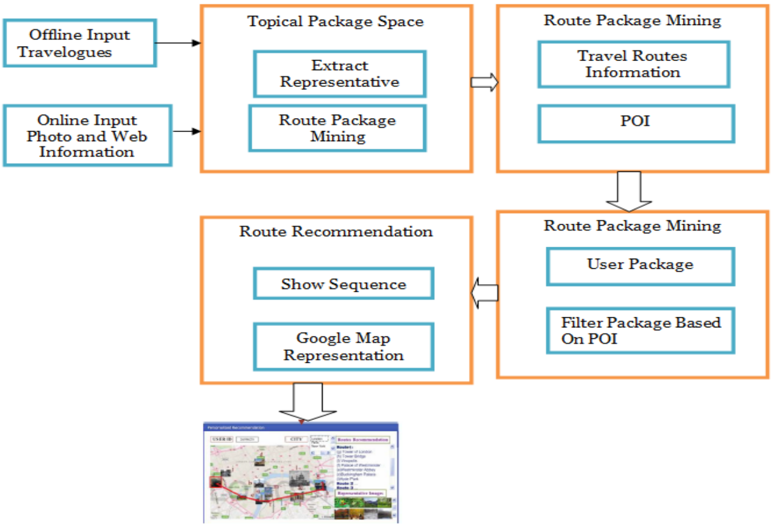


figure 3:System Architecture

3.1 Proposed Methodology:

3.1.1Travelogue Mining:

The structure of data we crawled from Flicker. The first layer is “City Layer”. Under each city, there are 26 topics constructed “Topic Layer”. We denote C = {c1, c2, ...,ck, ..., c N } as the category of topics. N is the number of topics which is 26 in our work. Under each topic ck is “POI Layer”. For example, under the topic“ Museum” in NYC, there would be POIs like “Modern Art” and “Natural History”. To each POI, there are a lot of user-generated travelogues. A POI may appear in more than one topics. So the topical interest of each POI is described by the distribution of N dimensions.

3.2 Data Preprocessing:

Data preprocessing prepares raw data for further processing. The traditional data preprocessing method is reacting as it starts with data that is assumed ready for analysis and there is no feedback and impart for the way of data collection. The data inconsistency between data sets is the main difficulty for the data preprocessing.

- **Data Integration:** Integration of multiple databases, data cubes, or files.
- **Data Transformation:** Data transformation is the task of data normalization and aggregation.
- **Data Reduction:** Process of reduced representation in volume but produces the same or similar analytical results.
- **Data Discretization:** Part of data reduction but with particular importance, especially for numerical data .The proposed model and task for preprocessing is described in the following sections.

3.3 Community-Contributed Photos Mining:

We randomly collected 7 million images worldwide up-loaded by 7,387 users from Flicker open API. For each user, there is a photo album consists of the photos shared by the user, associated with the heterogeneous metadata including user ID (e.g.,19181920@N03), textual tags (e.g., room, roams, colossal), date taken (or timestamp) (e.g., 2004-09-2500:00:01), latitude and longitude (e.g., 41.890585, 12.493171)provided by the user, or recorded by camera or smart phone. In this section, we introduce POIs mining, season attribute mining for each topic, and representative images mining for POIs from community-contributed photos.

3.4 DOM Parser:

According to our page generation model, data instances of the same type have the same path from the root in the DOM trees of the input pages. Thus, our algorithm does not need to merge similar sub trees from different levels and the task to merge multiple trees can be broken down from a tree level to a string level. Starting from root nodes of all input DOM trees, which belong to some type constructor we want to discover, our algorithm applies a new multiple string alignment algorithm to their first-level child nodes. There are at least two advantages in this design.

First, as the number of child nodes under a parent node is much smaller than the number of nodes in the whole DOM tree or the number of HTML tags in a Webpage, thus, the effort for multiple string alignment here is less than that of two complete page alignments in Road Runner [4]. Second, nodes with the same tag name (but with different functions) can be better differentiated by the sub trees they represent, which is an important feature not used in EXALG [1]. Instead, our algorithm will recognize such nodes as peer nodes and denote the same symbol for those child nodes to facilitate the following string alignment. After the string alignment step, we conduct pattern mining on the aligned string S to discover all possible repeats (set type data) from length 1 to length jSj=2. After removing extra occurrences of the discovered pattern (as that in DeLa [11]), we can then decide whether data are an option or not based on their occurrence vector, an idea similar to that in EXALG [1]. The four steps, peer node recognition, string alignment, pattern mining, and optional node detection, involve typical ideas that are used in current research on Web data extraction. However, they are redesigned or applied in a different sequence and scenario to solve key issues in page-level data extraction.

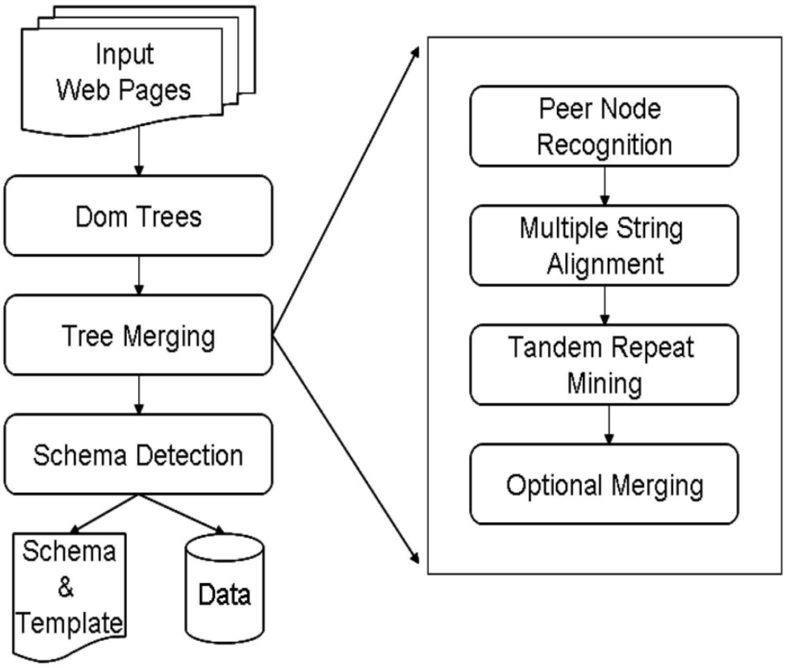


figure 4:Working of DOM Parser.

```
Algorithm MultipleTreeMerge(T, P)
// T is a set of DOM trees of the same type;
// P is the tag for the roots of T.
Initialize M; i = 0;
for each tree t in T
    j = 0;
    for each child c in t
        M[j++][i] = c;
    endfor
    i++;
endfor
recognizePeerNode(M);
childList = matrixAlignment(M);
childList = repeatMining(childList, 1);
mergeOptional(childList);
for each node c in childList
    if (c is a tree) then
        C = multipleTreeMerg(peerNode(c, M), tag(c));
    else // c is a leaf node
        C = c;
    endif
    Insert C as a child of P;
endfor
return pattern tree P;
```

Figure 5:Tree Merge algorithm:

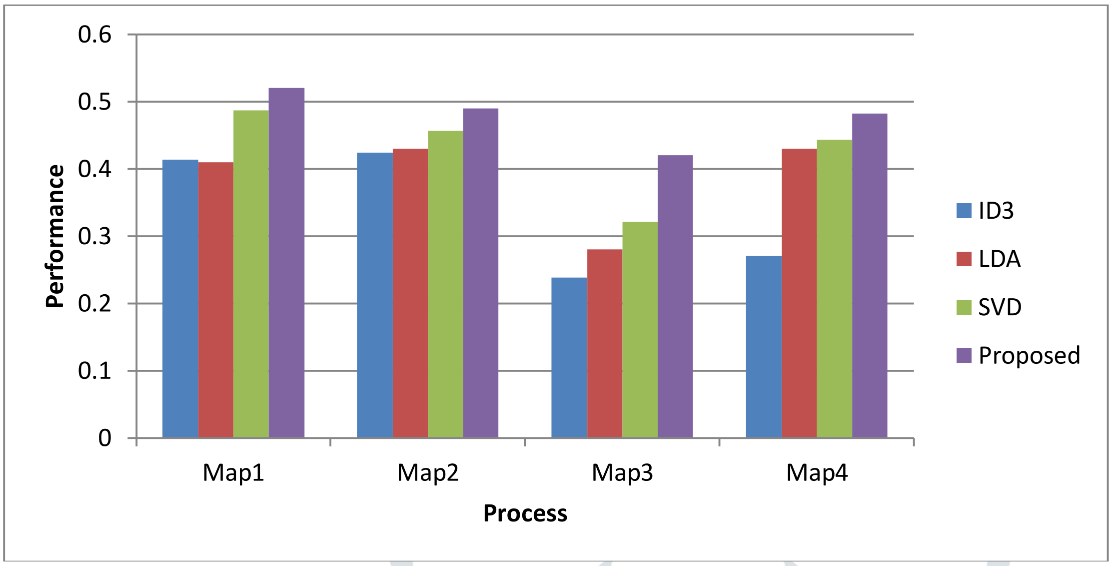
IV. EXPERIMENTAL RESULTS:

We randomly collected 7 million images worldwide up-loaded by 7,387 users from Flicker open API. For each user, there is a photo album consists of the photos shared by the user, associated with the heterogeneous metadata including user ID (e.g.,19181920@N03), textual tags (e.g., room, roams, colossal), date taken (or timestamp) (e.g., 2004-09-2500:00:01), latitude and longitude (e.g., 41.890585, 12.493171)provided by the user, or recorded by camera or smart phone.

	ID3	LDA	SVD	Proposed
Map1	0.4138	0.41	0.4871	0.52
Map2	0.4239	0.43	0.4567	0.49
Map3	0.2389	0.28	0.321	0.42
Map4	0.271	0.43	0.443	0.4823

Table 1.

Table shows the MAP and POI based on the user rating. Accuracy calculated for the maps with the comparison of the recommendations for user based on the search query. The Proposed approach seems to give more precise output for the search query.



V. Conclusions:

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 the system automatically mined user's and routes' travel topical preferences including the topical interest, cost, time and season. 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 different 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 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 ACMSIGSPATIAL 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.
[11] H. Kori, S. Hattori, T. Tezuka, and K. Tanaka, "Automatic generation of multimedia tour guide from local blogs," Advances in Multimedia Modeling, pp. 690–699, 2006.
[12] T. Kurashima, T. Tezuka, and K. Tanaka, "Mining and visualizing local experiences from blog entries," in Database and Expert Systems

Applications. Springer, 2006, pp. 213–222.
[13] Y. Shi, P. Serdyukov, A. Hanjalic, and M. Larson, "Personalized landmark recommendation based on geo-tags from photo sharing sites," ICWSM, vol. 11, pp. 622–625, 2011.
[14] M. Clements, P. Serdyukov, A. de Vries, and M. Reinders, "Personalized travel recommendation based on location co-occurrence, "are Xiv preprint arXiv:1106.5213, 2011.
[15] X. Lu, C. Wang, J. Yang, Y. Pang, and L. Zhang, "Photo2trip: generating travel routes from geo-tagged photos for trip planning," in Proceedings of the international conference on Multimedia. ACM 2010, pp. 143–152.
[16] Y. Zheng, L. Zhang, X. Xie, and W. Ma, "Mining interesting locations and travel sequences from gps trajectories," in Proceedings of the 18th international conference on World wide web. ACM, 2009, pp.791–800.
[17] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang, "Collaborative location and activity recommendations with gps history data," in Proceedings of the 19th international conference on World wide web.ACM, 2010, pp. 1029–1038.
[18] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering urban functional zones using latent activity trajectories," IEEE Trans. Knowles. Data Eng.vol. 27, no. 3, pp. 712–725, 2015.[Online]. Available: http://dx.doi.org/10.1109/TKDE.2014.2345405
[19] J. Liu, Z. Huang, L. Chen, H. T. Shen, and Z. Yan, "Discovering areas of interest with geo-tagged images and check-ins," in Proceedings of the 20th ACM international conference on Multimedia.ACM, 2012, pp. 589–598.
[20] Y. Pang, Q. Hao, Y. Yuan, T. Hu, R. Cai, and L. Zhang, "Summarizing tourist destinations by mining user-generated travelogues and photos," Computer Vision and Image Understanding, vol. 115, no. 3,pp. 352–363, 2011.
[21] L. Cao, J. Luo, A. Gallagher, X. Jin, J. Han, and T. Huang, "A world wide tourism recommendation system based on geo tagged web photos," in IEEE International Conference on Acoustics Speech and Signal Processing. IEEE, 2010, pp. 2274–2277.
[22] H. Huang and G. Gartner, "Using trajectories for collaborative filtering–based poi recommendation," International Journal of Data Mining, Modelling and Management, vol. 6, no. 4, pp. 333–346, 2014.
[23] C. Zhang and K. Wang, "Poi recommendation through cross region collaborative filtering," Knowledge and Information Systems ,pp. 1–19, 2015.
[24] A. Majid, L. Chen, G. Chen, H. Mirza, and I. Hussain, " travel suggestions using geo tagged photos," in Proceedings of the21st international conference companion on World Wide Web. ACM, 2012, pp. 577–578.
[25] C. Cheng, H. Yang, M. R. Lyu, and I. King, "Where you like to go next: Successive point-of-interest recommendation," in IJCAI,2013.