Generating Tourism Path from Trajectories and Geo-Photos

Zhixing Zeng, Richong Zhang, Xudong Liu, Xiaohui Guo, and Hailong Sun

School of Computer Science and Engineering, Beihang University,
Beijing, 100191 China
{zengzx,zhangrc,liuxd,guoxh,sunhl}@act.buaa.edu.cn

Abstract. The pervasiveness of GPS devices enables tourists recording their trajectories and uploading geo-tagged photos. Geo-related data has emerged as new source for travelers to refer to when making tourism decisions. As the increasing availability of these user-generated experiences on the social networks, there is a need to automatically discovering useful patterns for potential travelers. In this paper, we propose a tourism path by incorporating the trajectories and geo-photos. Specifically, we provide an algorithm for precisely matching user-uploaded photos to tourism sites and a density based clustering approach to identify the place of interests inside tourism sites. We then build a model that adapts the well-known HITS algorithm to detect interesting points and trajectories with high utility scores and design an algorithm for efficiently computing rational routes for visiting tourism sites. Finally, experimental results illustrate the advantage of the proposed density-based algorithm and confirm the effectiveness applicability of our tourism path discovering approach.

1 Introduction

Geo-related data, such as geo-photos and geo-spatial trajectories, is growing exponentially with the pervasive usage of GPS-enabled mobile devices. The pervasiveness of GPS-enabled devices makes it possible for travelers sharing their geo-location related information in social web communities. Travelers may discover popular visiting places, common visiting sequences and rational visiting routes from these location based social networks. Especially when tourists are planning to visit a tourism site, the prior knowledge of the attractiveness of interesting points and detailed tour paths are needed to be collected to assist travelers to make visiting plans. However, these patterns often implicitly exist in various types of data, therefore there is a need for automatically discovering patterns from the huge spatial data and unifying and matching objects across social communities.

To solve this problem, existing travel route recommendation approaches [1–3] simply extract coordinates and timestamps from uploaded photos to generate the possible visiting paths. However, these approaches are limited to as discreteness of the geo location of photos, and the extracted path from geo-photos is rough and only indicates an approximate direction of visiting sequences. In addition, some other researches [4–7] discover temporal-spatial sequences for travelers from collected GPS trajectories. The limitation of these approaches is that places of interest cannot be precisely extracted from raw trajectories and the popularity of the interesting points that trajectories pass is not taken into consideration.

Aware of these limitations, in this paper, we present a more profound approach to discover useful patterns from cross media community and build high quality tourism paths. Particularly, we take the advantages of different communities to mine interesting points. In specific, we present an algorithm for aggregating and mapping geo-photos with geo-spatial trajectory data for tourism sites. We then propose a density based clustering approach to identify the place of interest and estimate the popularity of each place. Furthermore, we build a model based on HITS algorithm to evaluate the quality of trajectories and the popularity of interesting points. Finally, we design a tour path constructing algorithm to generate rational tourism paths under a specified time limitation. The evaluation on Geolife[1] trajectory data and geo-tagged photos from Flickr confirms the efficiency and effectiveness of our proposed approaches.

The rest of the paper is organized as follows. In Sect. 2, we discuss the related works; In Sect. 3, we introduce the approach to discover interesting points and visiting path; Then we evaluate the performances of our proposed algorithms in Sect. 4; Finally in Sect. 5, we conclude the paper.

2 Related Work

Trajectory recommendation researchers have made significant advances in the recent years through trajectory mining [8, 4–6], travelogue mining [9, 10] and photo mining [11, 12, 2, 3].

Some approaches apply clustering algorithm in the photo mapping, such as in [11, 13], authors proposed an approach for clustering photo collections to find popular locations with clustering algorithm. K-means clustering [13] and mean-shift [11] are also been introduced to discover locations and photos. For the same purpose of mapping photos to points with semantic information in [1], textual tags and geo-distance are leveraged to find photos of a given point.

Authors in [3] propose a probabilistic personalized travel recommendation model using knowledge from textural and image features. Moreover, recently, photos are made use to construct visiting sequences with time attribute in order to analyze the pattern of trips in [3, 1, 11]. However, photo sequences can only indicate the order of visited interesting locations but can not provide detailed routes or walking paths between these points. Researches in [6, 7] both recommend travel sequences by mining tourist-generated GPS trajectories. In the paper [6], authors detected the classical travel sequences using location interests and users' travel experiences. An itinerary model concerning elapsed time ratio, stay time ratio, interest density ratio and classical travel sequence ratio was proposed in [7] to recommend itineraries. These two researches convert the

[1] $http://research.microsoft.com/en-us/projects/geolife/$

trajectories to locations sequence and then discard the explicit GPS points sequence. However, these GPS points sequence can show the clear path connecting two locations.

Motivated by these facts, in this paper, we proposed an improved density-based clustering algorithm to mine attractive places in tourism sites. Then we extract travel paths connecting these attractive places. At last, an intelligent tour path recommendation approach is proposed to generate a specific path within time limitation for tourists.

3 Methodology

Our work focuses on analyzing previous user-uploaded trajectories and photos, and discovering high quality visiting paths for tourism sites. In this section, we begin with introducing the definitions used throughout this paper. We then propose the algorithm for finding attractive spots from photos taken by previous travelers. With these discovered attractive spots, we build a graph for each tourism site, define utility functions for trajectories passing these spots and provide an algorithm for composing possible high utility trajectories.

3.1 Preliminary

To build a generic model for solving the tourism site visiting path discovering problem, we formulate definitions which would be used in this paper.

Definition 1 (Trajectory). *As defined in [14, 4], a trajectory, denoted by t, is a sequence of GPS points. Each GPS point, denoted by p, consists of latitude (p.lat), longitude (p.lng) and timestamp (p.time). Formally, a trajectory can be represented by:*

$$t = \{p_0,\ p_1,\ p_2,\ \ldots, p_n\},\ \forall i \in [0, n]\ \ p_i.time < p_{i+1}.time$$

As the interest of this paper is on finding movement patterns in a tourism site, we only consider tourism trajectories in the following sections.

Definition 2 (Geo-photo). *A geo-photo, denoted by gp, represents a photo that is geo-tagged and consists of a set of tags, latitude(lat) and longitude(lng) of where this photo is taken. The set of geo-photos in i_{th} tourism site can be represented by GP_i.*

Definition 3 (Interesting Point). *An interesting point, denoted by ip, means an attractive place inside a tourism site.*

In this work, we assume that travelers usually take photos at interesting points, such the interesting points can be extracted from the geo-photos set GP. We denote the set of interest points of i_{th} tourism site as IP_i.

Definition 4 (Tourism Site). *A tourism site consists of a set of interesting points, photos and trajectories. Formally, the i^{th} one can be represented by:*

$$s =< name_i, lat_i, lng_i, IP_i, GP_i, T_i >$$

where T_i denotes the trajectories set in i^{th} tourism site. The coordinate lat_i and lng_i represent roughly location on map. It will be used in mapping photos approach. It can be get from Google Geocoding API[2].

Definition 5 (Trajectory Segment). *A trajectory segment, denoted by ts, is a segment of a trajectory connecting two adjacent interesting points. The trajectory segment connecting the j^{th} and k^{th} interesting points can be denoted by $ts_{j,k}$.*

We note that there may exist several ts connecting two same interesting points, we may introduce a superscript x on ts (ts^x) to distinguish these trajectory segments.

Definition 6 (Tourism Site Graph). *In this paper, we use a graph $G =< E, V >$ where $E = \{ts_1, ts_2, \cdots, ts_n\}$ and $V = \{ip_1, ip_2, \cdots, ip_m\}$ to formal indicate the interesting points topology in a tourism site. In addition, we denote the entries and exits as sets V_s and V_e standing for entries and exits respectively.*

Definition 7 (Route). *A route, denoted by r, is a trajectory that consists of some interesting points and trajectory segments which can be formally defined as below:*

$$< ip_0 \xrightarrow{ts_{01}^p} ip_1 \xrightarrow{ts_{12}^q} ip_2 \ldots ip_n >, ip_0 \in V_s, ip_n \in V_e$$

where ts_{ij}^p denotes the p^{th} trajectory segment of ones connecting from ip_i to ip_j.

3.2 Photo Mining

Given a set of photos GP and a set of tourism sites S, the first goal of our system is to map photos to their corresponding tourism sites and extract interesting points in each site.

Mapping Photos to Tourism Sites When mapping photos to tourism sites, by photo annotations or tags, some unexpected photos may be located inside or near a specific tourism site, although, they share the same tags with expected ones. The solution proposed in [1] used both geo and tag information to map photos whose locations are close to a certain tourism site within 100 meters, as well as photos whose tags are matched with the expected ones. However, the size of tourism sites varies significantly. Photos would be mismatched merely by a fixed range.

In this approach, we propose a method which can adapt to different sizes of tourism sites for filtering photos. Given a set of geo-photos GP and a specified

[2] $https : //developers.google.com/maps/documentation/geocoding/$

tourism site s, we take the photos tagged with $s.name$ as the photo candidate set for s. To adaptively adjust various ranges of tourism sites, we exploit DBSCAN[15], a density-based cluster algorithm, to dynamically and precisely discover photos for each tourism site. We treat a tourism site and its every geo-photo as points to be clustering. After clustering, photo points sharing the same cluster with the tourism site point are mapped to the tourism site s. Empirically, two parameters of DBSCAN, Eps and $MinPts$, are set as $Eps = 25m$ and $MinPts = 2$.

Algorithm 1 Find Neighborhood of Each Point(Geo-photo)

```
Require: Input: $Points$, $K$, $Eps_D$, $Eps_S$ $MinPts$.
    Output: A link array neighborhood
for $i = 1$ to $Points.size$ do
    for $j = 1$ to $Points.size$ do
        if $dist(i, j) < kdistance[i]$ then
            $sortedInsert(knearest[i], j)$;
            $kdistance[i] \leftarrow dist(knearest[i][last], i)$;
        end if
    end for
end for
for $i = 1$ to $Points.size$ do
    for $j = 1$ to $Points.size$ do
        if $dist(i, j) < Eps_D$ and $similarity[i, j] < Eps_S$ then
            $neighborhood[i].add(j)$;
        end if
    end for
end for
return $neighborhood$;
```

Interesting Points Extraction. To extract interesting points of a tourism site s from mapped photos set GP, we use a hybrid density-based clustering algorithm based on DBSCAN [15] and Shared Nearest Neighbor (SNN) [16]. We find that DBSCAN cannot achieve a good clustering performance when data density is relatively high, while SNN is not effective for distinguishing noise data points as the data density is low. However, photo densities inside a tourism site vary significantly at different area. We take the best of both algorithms to improve the performance of clustering. We let Eps_D and Eps_S denote the original parameter Eps in DBSCAN and SNN respectively. According to the definitions in [17], we redefine the neighborhood of a point p as follows:

$$N_{Eps}(p) = \{q \in P | dist(p, q) \leq Eps_D \ and \ similarity(p, q) \geq Eps_S\} \tag{1}$$

where $dist(p, q)$ denotes the distance between p and q, and $similarity(p, q)$ denotes the SNN similarity [16] between p and q. After finding neighborhood, the subsequent algorithm are the same as that defined in the DBSCAN algorithm.

Algorithm 1 depicts the process of obtaining the neighborhood of each point that represents a geo-photo here. We let $k - distance$ denote the distance from a specific point to its k^{th} nearest point and use $kdistance[i]$ at line 3 to denote the i^{th} point's k-distance. At line 4, $knearest[i]$ is a K length array to save k nearest points of the i^{th} point in ascending order. In [15], authors define a function $regionQuery(Point, Eps)$ to represent the neighborhood of $Point$ within Eps. In this paper, we redefine this function as $regionQuery(p)$, p denoting the index of a point, which returns $neighborhood[p]$ obtained in Algorithm 1.

Each cluster generated by this process can be treated as an interesting points and the set of generated interesting points compose the vertex set, V, for the tourism graph representation of G.

3.3 Trajectory Processing

As we focus on trajectory inside tourism sites in this study, we first discover tourism trajectories inside each tourism site and further work in this paper is all on these filtered trajectories.

Trajectory Segment Extraction To discover trajectories inside a tourism site, we need to confirm the boundaries of every tourism site. We then filter out the part of a user-generated trajectory located inside every boundary. As we have discovered the cluster of photos for each tourism site in the previous subsection, the most northeast and most southwest points of these photos can be identified and we build a rectangular area which covers all photos mapped to this site.

It is obvious that we can identify the interesting points sequence of a trajectory t inside a tourism site s by finding the correlation between the discovered interesting points from the photo set GP and the trajectory t in s. Formally, we denote the interesting points sequences by $< ip_1 \rightarrow ip_2 \rightarrow \cdots ip_n >$ when it satisfies following conditions:
1.$\forall ip_i, \exists p_i \in t, \exists gp_j \in GP_i, dist(p_i, gp_j) < \varepsilon$ and $dist(gp_j, c_i) < \sigma$, where GP_i is the photos set clustered to ip_i, c_i is the central geo point of ip_i.
2.$p_1.time < p_2.time < \cdots p_m.time$, namely the interesting point sequence of t is in chronological order.

With the discovered interesting point sequences, it can be easily convert the sequences into several trajectory segments. We denote the set of trajectory segments on site t_i as TS_i and each segment ts including two interesting points it connects, the time cost and the path. The path is a geo point sequence representing the detailed movement pattern between two successive interesting points. The set of trajectory segments TS_i can also be seen as the set of edges E_i in the tourism site graph G_i.

3.4 Discovering Visiting Path

To find high quality visiting path inside a specific tourism site, interesting points that are attractive to tourists should be discovered. In this study, we exploit HITS algorithm[18] to identify the popularity of interesting points. In the context of this work, the authority scores of trajectories can be seen as the typicality of trajectories and the hub scores of interesting points can represent the popularity of interesting points. We let $auth(t)$ denotes the authority score and $hub(ip)$ denotes the hub score. A matrix **M** representing the correlation between trajectories and interesting points is introduced where rows correspond to the trajectories and columns correspond to the interesting points, and item v_{ij} in **M** represents whether t_i passes ip_j (passing denoted by 1, otherwise by 0). In this work, we use the following two formulas to calculate the authorities and hubs. First, the authority scores of a trajectory t_i is formulated as:

$$auth_n(t_i) = \frac{\alpha_i \cdot hub_{n-1}(ip)}{\sqrt{(M \cdot hub_{n-1}(ip))^T \cdot (M \cdot hub_{n-1}(ip))}} \tag{2}$$

where $\alpha_i = (v_{i1}, v_{i2}, \cdots, v_{in})$. Then, the hub score of an interesting point ip_i is formulated by:

$$hub_n(ip_j) = \frac{\beta_j \cdot auth_n(t)}{\sqrt{(M^T \cdot auth_n(t))^T \cdot (M^T \cdot auth_n(t))}} \tag{3}$$

where $\beta_j = (v_{1j}, v_{2j}, \cdots, v_{mj})$.

In (2) and (3), $auth(t) = (auth(t_1), auth(t_2), \cdots, auth(t_n))$, $hub(ip) = (hub(ip_1), hub(ip_2), \cdots, hub(ip_m))$.

We iteratively compute the authorities and hubs until algorithm converges. The hub scores and authority scores are initialized as follows:

$$hub_0(ip_i) = \frac{N(ip_i)}{\sum_{ip \in s} N(ip)}$$

$$auth_0(s_i) = 1$$

where $N(ip_i)$ denote the number of photos in the cluster of ip_i, s denotes a specified tourism site.

In practice, time constrains is often an important factor to be considered in path planning approaches. The time cost of a path, in this study, is formulated by:

$$t(r) = \sum_{i=0}^{n-1} t(ts^x_{i,i+1})$$

where $ts^x_{i\ i+1}$ denotes the x^{th} trajectory segment connecting ip_i and ip_{i+1}, and $t(ts^x_{i\ i+1})$ denotes the time cost of $ts^x_{i\ i+1}$.

The utility of the planned path is measure to represent the quality of paths. We let $u(r)$ denote the utility of the path trajectory r:

$$u(r) = \sum_{i=0}^{n} hub(ip_i) \tag{4}$$

where r contains an interesting point sequence $< ip_1 \rightarrow ip_2 \rightarrow \cdots ip_n >$, $hub(ip_i)$ denotes the hub score of ip_i generated by the HITS algorithm.

Our objective is to find a path r with the maximum utility $u(r)$ and r meets the constraint of $t(r) < \varphi$, where φ is time limitation requested by a tourist.

To improve the efficiency of our algorithm, we prepare a vector τ, which denotes minimum time that each interesting point costs to exit, based on shortest path algorithm Dijkstra [19], i.e. $\tau(ip)$ denotes the minimum time it costs to exit from the interesting point ip. In Algorithm 2, we describe a backtracking algorithm which discovers the best route. We apply τ as a pruning condition to reduce the time complexity at line 5. At the same line, $isPassed(path, l, ip)$ denotes if $path$ contains the trajectory segment from l to ip.

Such, the paths in tourism sites with the best utilities and under the user requested constraints are discovered.

Algorithm 2 The algorithm for mining route with time limitation

```
Require: Input: A graph $G =< E, V >$, a vector $\tau$ denotes minimum time each
    interesting point costs to exit, a sequence $path$ including chosen trajectory segments
    and interesting points, time limitation $t$ and current interesting point $l$. Output: A
    route $r$.
 1: if $l \in V_e$ and $u(path) > u(best)$ then
 2:   $best \leftarrow path$
 3: end if
 4: for all $ip \in V$ do
 5:   if $\tau(ip) \geq t$ and not $isPassed(path, l, ip)$ then
 6:     for all $ts$ connecting from $l$ to $ip$ and $ts$ not in $path$ do
 7:       if $t - t(ts) > \tau(ip)$ then
 8:         $newPath \leftarrow path$ add $ip, ts$;
 9:         $RouteMining(G, \tau, newPath, t - t(ts), ip)$;
10:       end if
11:     end for
12:   end if
13: end for
```

4 Experiments

In this section, we first present the dataset. Second, we introduce the cluster and path generation evaluation approaches. Last, some discussions of the results are given.

Table 1. The number of photos per considered tourism site.

Destination	Summer Palace	Temple of Heaven	Houhai	Forbidden City	Tiananmen Square
Photo Number	4250	1996	1058	3809	727

4.1 Dataset

The tourism trajectories are filtered out from Geolife dataset which is collected by 178 users in a period of over four years and contains 17621 trajectories. As the majority of the data is created in Beijing, we only consider the tourism sites in Beijing and crawled 23,649 geo-tagged photos taken at Beijing from Flickr. The tourism sites are crawled from tourism web sites like TripAdvisor[3]. Their coordinates are queried from Google Geocoding. In Table 1, we present the number of geo-photos of five considered tourism sites on which we will do experiment.

4.2 Clustering Evaluation

In this evaluation, we choose $Eps_D = 30$ and $MinPts = 5$ for DBSCAN, $KN = 9$, $Eps_S = 6$ and MinPts = 5 for SNN. These parameters are also used by our proposed Hybrid algorithm. We let σ denote the average distance of all elements in each cluster.

$$\sigma = \frac{1}{n} \sum_{p \in IP} dist(p, c_{ip}), c_{ip} =< \frac{1}{n} \sum_{p \in IP} p.lat, \frac{1}{n} \sum_{p \in IP} p.lng >$$

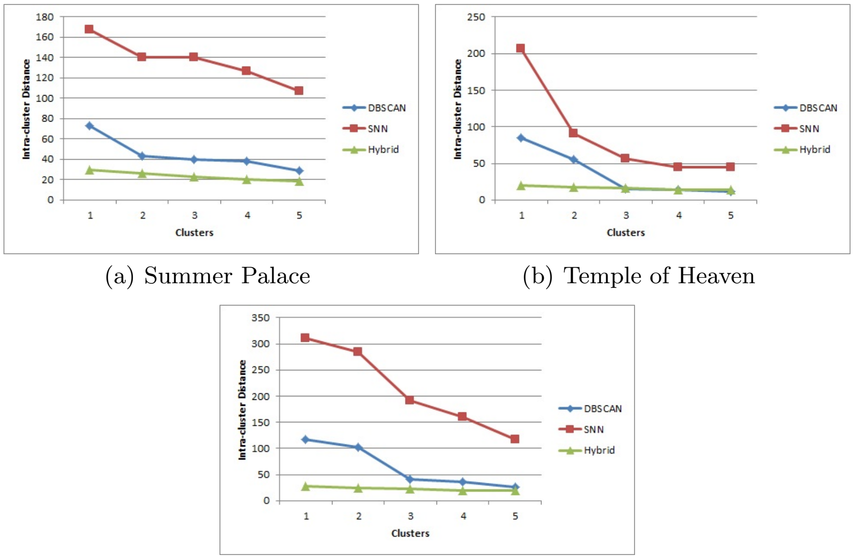
where IP denotes the photo set of a specific interesting point ip, n is the number of photos and c_{ip} denotes the center coordinate of ip.

The goal of this evaluation is to compare DBSCAN, SNN and Hybrid algorithms with respect to the clustering result. In Fig. 1, we illustrate five greatest σ of each clustering algorithm on three different tourism sites. It can be observed that the proposed Hybrid algorithm outperforms the other two compared density-based clustering algorithms in terms of the intra-cluster average distance.

As finding interesting points is the main purpose of the proposed algorithm. We evaluate the results based on precision and recall. We find ten people who are quite familiar with these five tourism sites. They can tell how many interesting points there are and where they are located in each tourism site. They also check out the clustering results that whether a cluster can represent an interesting point. We let the number of confirmed clustering interesting points divided by the number of all ones be the recall and divided by the number of clusters be the precision. Then we compare the results in terms of F-measure as shown in Fig. 2. We can observe that the Hybrid algorithm outperforms the other two.

To visually illustrate the cluster results, we plot the exacted clusters of photos on a map, as shown in Fig. 3. In Fig. 3(b), we show the photo clusters generated

[3] $http://www.tripadvisor.com$



(a) Summer Palace

(b) Temple of Heaven

(c) Forbidden City

Fig. 1. Five the greatest average intra-cluster distance by DBSACN, SNN and Hybrid on three tourism sites.

by DBSCAN. It can be observed that DBSCAN performs well at proper density area, but it regards a high density area, like area A in Fig. 3(b), as a cluster rather than divides it into clusters. So σ of the cluster covering high density area is great. In Fig. 3(c), we show the photo clusters calculated by SNN and it can be seen that it clusters well at the same area. However, as SNN only concerning nearest neighborhoods without distance limit, it mistakenly integrates some low density area as clusters like area B in Fig. 3(c). As shown in Fig. 3(d), Hybrid overcomes these two limitations and achieves a better clustering performance.

In summary, in comparison with other commonly-used density-based algorithms in terms of cluster performances, our proposed hybrid algorithm outperforms them.

4.3 Recommended Route Evaluation

In order to evaluate the tour path generation approach, we find three best original tour paths within three different periods of time at three tourism sites, the Summer Palace, the Temple of Heaven and Houhai. We evaluate the utility, as defined in E.q. 4, of the generated tourism path by the hybrid algorithm. Table 2 show the results for the Summer Palace, the Temple of Heaven and Houhai respectively. From the experimental results, we find that the generated tour paths all achieve better utility values. We note that the original tour path utility increase slower than generated. One of the possible reason is because that the longer time the tourists travel for, the more time they will take to rest.

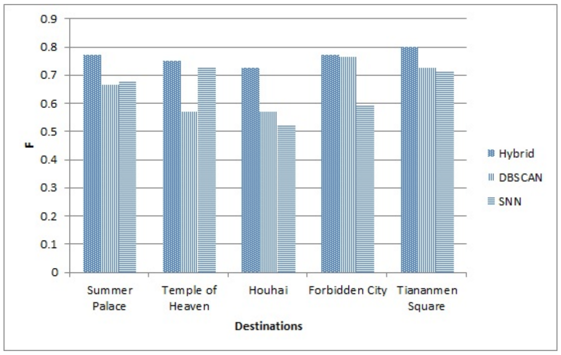


Fig. 2. Comparision of the clustering results on five tourism sites based on F-measure.

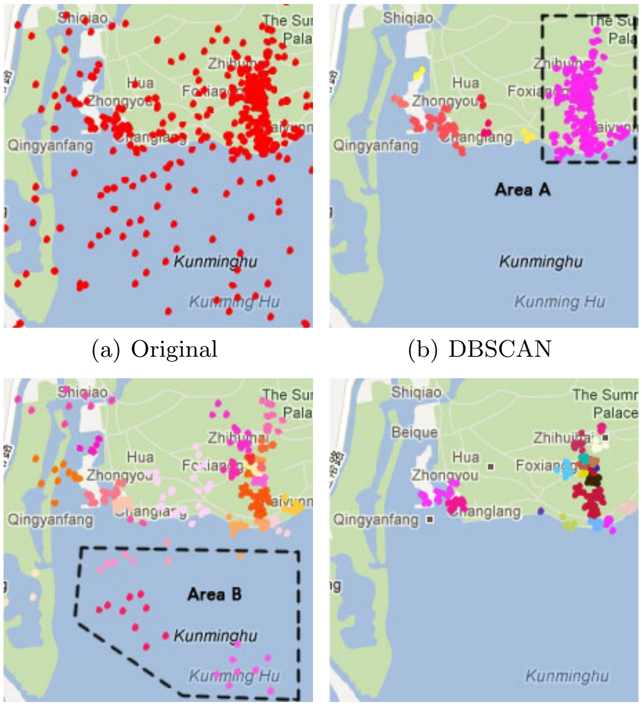
Similar as the previous experiment, we show some typical generated tourism path on the map. Fig. 4(a) and Fig. 4(b) illustrate a 4 hours and a 9 hours generated tour path in the Summer Palace. From these two figures we can observed that in the 9 hours route, our algorithm plans a boating event on the lake as some discovered interesting points are located on the lake. If the tourist only request for a short time period, the representative interesting points take priority over others, as shown in Fig. 4(c) and Fig. 4(d).

Table 2. The utility of original tour paths and the generated in the Summer Palace.

Destination	Summer Palace			Temple of Heaven			Houhai		
Time	2h	4h	9h	1h	2h	3h	1h	2h	3h
Original Utility	2.409	2.797	3.000	1.908	2.072	2.615	1.843	2.043	2.384
Generated Utility	3.797	5.915	6.756	2.143	2.615	4.190	4.987	6.415	6.978

4.4 Discussion

Traditional tourism route planning algorithms which are simply invoke APIs from other web site that provide map related services. These approaches no longer serve the increasing needs of travelers for different tourism sites. Indeed, nowadays, travelers would like to visit as much interesting points as they can. However, the quality of interesting points is not always available for route planning algorithm. With the help of user-uploaded geo-photos on the social web, this study takes the advantage of the location histories of previous travelers and presents algorithms for discovering attractive interesting points and organizing a satisfactory visiting path for potential travelers.



(a) Original (b) DBSCAN

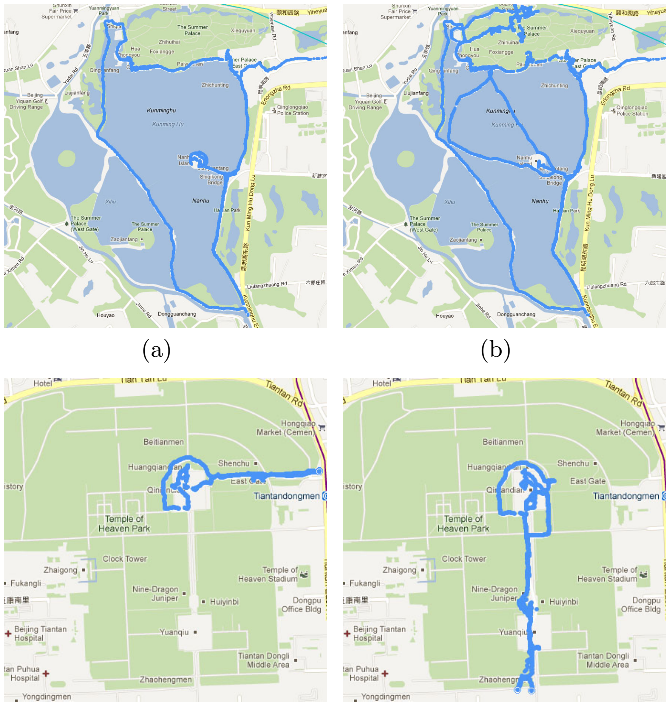
(c) SNN (d) Hybrid

Fig. 3. Part of clustering result of three algorithms in the Summer Palace. Dots with a same color are in one cluster.

5 Conclusions

In this paper, geo-tagged photos and user-generated trajectories are collected to extract travel knowledge for discovering tourism site routes. We first maped photos to tourism sites with an algorithm including annotation matching and density-based clustering. After mapping photos, a hybrid algorithm was proposed to discover interesting points in tourism sites. We then proposed a density based clustering approach to identify the places of interests and estimate the popularity of each place and trajectory. Finally we compared our density-based algorithm to two original ones in terms of average intra-cluster distance and F-measure. The experimental results of route generation have shown the intelligence of the proposed approach.

In the future, we are going to collect more geo related contents to discover the interesting points and their hub score more precisely, to build more complex topologies for tourism sites, and to evaluate the performance of the proposed approach on a much larger data set. We would like to take more aspects, e.g. season, tourism site type and travel time, into account to generate more precise paths.



(a) (b)

(c) (d)

Fig. 4. Four generated tour routes: (a) is a 4 hours tour route in the Summer Palace; (b) is a 9 hours route in Summer Palace; (c) is a 1 hour tour route in the Temple of Heaven; (d) is a 2 hours route in Temple of Heaven.

6 Acknowledgement

This work was supported by China 863 program (No. 2012AA011203), National Natural Science Foundation of China (No. 61103031), Specialized Research Fund for the Doctoral Program of Higher Education (No. 20111102120016), the State Key Lab for Software Development Environment (No. SKLSDE-2012ZX-12), and the Fundamental Research Funds for the Central Universities (No. YWF-12-LXGY-023) and (No. YWF-12-RHRS-016).

References

1. De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R., Yu, C.: Automatic construction of travel itineraries using social breadcrumbs. In: Proceedings of the 21st ACM conference on Hypertext and hypermedia, ACM (2010) 35–44
2. Lu, X., Wang, C., Yang, J., Pang, Y., Zhang, L.: Photo2trip: generating travel routes from geo-tagged photos for trip planning. In: Proceedings of the international conference on Multimedia, ACM (2010) 143–152

3. Cheng, A., Chen, Y., Huang, Y., Hsu, W., Liao, H.: Personalized travel recommendation by mining people attributes from community-contributed photos. In: Proceedings of the 19th ACM international conference on Multimedia, ACM (2011) 83–92
4. Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W., Ma, W.: Mining user similarity based on location history. In: Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems, ACM (2008) 34
5. Xiao, X., Zheng, Y., Luo, Q., Xie, X.: Finding similar users using category-based location history. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM (2010) 442–445
6. Zheng, Y., Zhang, L., Xie, X., Ma, W.: Mining interesting locations and travel sequences from gps trajectories. In: Proceedings of the 18th international conference on World wide web, ACM (2009) 791–800
7. Yoon, H., Zheng, Y., Xie, X., Woo, W.: Social itinerary recommendation from user-generated digital trails. Personal and Ubiquitous Computing (2011) 1–16
8. Farrahi, K., Gatica-Perez, D.: What did you do today?: discovering daily routines from large-scale mobile data. In: Proceedings of the 16th ACM international conference on Multimedia, ACM (2008) 849–852
9. Hao, Q., Cai, R., Wang, C., Xiao, R., Yang, J., Pang, Y., Zhang, L.: Equip tourists with knowledge mined from travelogues. In: Proceedings of the 19th international conference on World wide web, ACM (2010) 401–410
10. Ji, R., Xie, X., Yao, H., Ma, W.: Mining city landmarks from blogs by graph modeling. In: Proceedings of the 17th ACM international conference on Multimedia, ACM (2009) 105–114
11. Crandall, D., Backstrom, L., Huttenlocher, D., Kleinberg, J.: Mapping the world's photos. In: Proceedings of the 18th international conference on World wide web, ACM (2009) 761–770
12. Arase, Y., Xie, X., Hara, T., Nishio, S.: Mining people's trips from large scale geo-tagged photos. In: Proceedings of the international conference on Multimedia, ACM (2010) 133–142
13. Kennedy, L., Naaman, M.: Generating diverse and representative image search results for landmarks. In: Proceeding of the 17th international conference on World Wide Web, ACM (2008) 297–306
14. Monreale, A., Pinelli, F., Trasarti, R., Giannotti, F.: Wherenext: a location predictor on trajectory pattern mining. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM (2009) 637–646
15. Ester, M., Kriegel, H., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data mining. Volume 1996., AAAI Press (1996) 226–231
16. Jarvis, R., Patrick, E.: Clustering using a similarity measure based on shared near neighbors. Computers, IEEE Transactions on **100**(11) (1973) 1025–1034
17. Popescu, A., Grefenstette, G.: Mining social media to create personalized recommendations for tourist visits. In: Proceedings of the 2nd International Conference on Computing for Geospatial Research & Applications, ACM (2011) 37
18. Kleinberg, J.: Hubs, authorities, and communities. ACM Computing Surveys (CSUR) **31**(4es) (1999) 5
19. Dijkstra, E.: A note on two problems in connexion with graphs. Numerische mathematik **1**(1) (1959) 269–271