Middle-East Journal of Scientific Research 24 (S1): 192-198, 2016
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2016.24.S1.39

Road Based Travel Recommendation by Mining Important Tourist Location and Travel Sequences Using Geotagged Images

C. Puvana Thanga Suba, Mrs. S. Devi Mahalakshmi and Mr. P. Vinoth

Mepco Schlenk Engineering College, Sivakasi, India

Abstract: The problem of recommendation in tourismis to provide important landmark for tourists. This proposed work predicts the user favorite location in a city.Before traveling to particular city the tourist needs information about landmark worth to visit and its order. Without this knowledge the tourist depends on the travel blogs, travel guides and public transportation sites which is time consuming process. In this paper it provides the landmark with user preference categories like education, entertainment etc. Under each categorythe important landmarks are found out. Flickr is a photo sharing website which has launched its service for adding latitude and longitude information to photos called geotagged photos. This information is extracted from flickr using photoid. Based on these geotagged photos information the users preference category get recommended and then ranked the tourist spot. To recommend the travel for the users, the traveling path is also necessary. So to recommend the path, the routing calculation is done. Then the interested travel routing between the landmarks is recommended to the users.

Key words: Recommendation · Flickr · Latitude · Longitude · Geotagged Photos · Routing · Landmarks

INTRODUCTION

The Internet also mediates tourism experiences as tourists use social media sites to relieve their trips. The application of geo spatial metadata (i.e. latitude and longitude) is geo-tag to a range of web based media (e.g., photos, videos) [1]. In recent years there is well development in tourism which delivers the prominent travel location to particular city. Existing methods provides recommendation without involving user context and other dimensions [2]. For example research by the Association of America in the travel industry found that two-thirds (64%) of online travelers use search engines for travel planning. The purpose of this paper is to satisfy the tourist's expectation to visit the important landmark and its order. So to give more priority for the user, the experienced users' knowledge must be considered. There are various types of information came through social networking sites and shared by users. In many web album services, the users not only share their photos also share their location. The most common social networking sites are Instagram, Facebook Places, Location Based Photo Sharing, Flickr etc. Location acquisition technology becoming very popular as it is the part of Internet which allows the data to share with others. This research is location based solution to tourists which involves the latitude, longitude, location related tags, title. This information is extracted through Flickr API which has resulted in huge volumes of geotagged photos and videos available on the web. Geo-tags of photosand videos provide a wealth of information about users' behaviour and their potential is increasing. It becomes ever common for images, to be associated with location specific information. In recent days, there is increasing tendency to adopt the information from these geotagged photos to find the recommendation of tourist location. Most of the recommender system suggests the place without considering the users interest or preference. When seen from tourist point of view, recommendation system is helpful in the sense that they assist user based on their taste that matched with user preference category.

To recommend the tourist location geotagged photo information extracted through FLICKR API. It provides latitude, longitude, Tags,Titleand Description.

Corresponding Author: C. Puvana Thanga Suba, Mepco Schlenk Engineering College, Sivakasi, India.

The significance of these data has been already addressed in some existing papers like landmark identification [3,4] and tag representative [5]. Flickr is a photo sharing website so tags and other forms of text are freely entered and not associated with any ontology. Hence these tags are often inaccurate, ambiguous and wrong. Each record from the photo dataset carries essential information about the surrounding area and owner of the photoid. Every photo has unique photoid.

For each location there is specific latitude and longitude points and also tags which it describes the photo may be related to location or some other events. The places where lot of people take photos present attractive area which grab attention of location specific attributes.In particular a point of interest is identified by activities of photo based on the metadata such as geographical location. To identify the landmark or to provide the semantic meaning for the tags the world geodeg database is used.In India.Tamilnadu is the fourth largest economy of the country and tourism is one of the most important sources of its revenue. India's fourth largest metropolis is Chennai and headquarter of tourism is also available in Chennai. So to recommend the important landmarks in the city, the fourth largest metropolis city is chosen as the dataset. Every tourist has different preference and taste. By considering the user preferences category, the proposed works definefive categories like Religious, Food, Tourism Attraction, Shopping, Entertainment. The basic requirement of geo visualization environment and analysis of category is Google earth. It helps to view the path of landmark and find the shortest path between the landmarks. To visualize the points of interest with category Google earth software is used.

Literature Review: In this work, the author evaluated the geotagged photos from Flickr[6].Then they proposed a framework that is used to take the associative points-of-interest patterns from geo-tagged photos. The popular data mining techniques used in this research is clustering for points-of-interest detection and association rules mining for associative points-of-interest patterns. In this paper [7], they used the combined GPS information for multiple users in order to mine a list of interesting locations and rank them. The authors [8] propose a framework that encompasses new techniques for extracting semantically meaningful geographical locations from GPS data. Then provide the ranking for these locations according to their significance. In this paper[9], the author presented a geovisual analytic approach to detect people preferences on landmarks from flickr photos. This result would help to users which landmarks are popular in particular city. An enhanced collaborative filtering method is proposed in this work [10] which recommends the restaurants to mobile users. Here the author proposed the work to identify the points of interest and ordered the locations based on their popularity score and rank them for tourists [11].Then it is linearly combined with personalized score weighted by the similarities between active users and other users. In this research work the author used k-means clustering which is used to identify tourist location based on latitude and longitude points [12]. A trip routing is calculated based on user trajectory from GPS history data. Every user historic trajectory can be considered as a historic individual routing. The recommended routing is always denoted as the average of the historic routings or the other with the highest ranking. These recommended trip routings sometimes lack convenience and also validness for users to position and plot on the digital map, since they are based on GPS trajectory in place of separate road in the road network.

System Design for Recommendation System: In the proposed system the first step starts with collection of geotagged photos through Flickr API. The data extracted from photo metadata are Photoid, Latitude, Longitude, Photo Upload Time, Photo Taken Time, Tags, Title and Owner of the photo, Description. These attributes served as input for further processing.

Feature Selection: The following attributes are extracted from xml documents

- Latitude
- Longitude.
- Location related tags.
- Photoid
- Ownerid

In this Table 1, the records of each photo id are retrieved from Flickr API. The attributes of each photo id describes the photo information and the photo taken surrounding area. The latitude and longitude specifies the location of the photo taken area.

Table 1: Sample Input

Latitude	Longitude	Photoid
12.911409	80.155728	22957859423
12.959979	80.118398	22960081934
12.901630	80.214014	22984453754

In this table 2,it shows the ownerid of the photo means user of the photo.Tags which is given by the user describes the photo activity or location event of the photo.In title it describes the photo main theme or user idea about the event.

Table 2: Sample Input for Tags

Ownerid	Tags	Title
115922745@N06	Narada Gana Sabha, art-form, bharatanatyam, Chennai, classical, dance, margazhi	Vidhya Subramanian
59225019@N05	Kilpauk, chennai, northeast monsoon, Chennai flooding	Pushing through

Clustering: There is collection of photos that are represented with location metadata (latitude and longitude). Then finding the tourist location in a city can be seen as a clustering problem in a highly photographed location. The Euclidean distance measure is used and the numeric attributes involved in the method is latitude and longitude.

The Euclidean distance measure is defined as follows

$$\text{DIST (i, j)}=\sqrt{(x_{i1}-x_{j1})^2+(x_{i2}-x_{j2})^2+...+(x_{in}-x_{jn})^2}$$

Objects within one cluster have high similarity compared with objects in other clusters. Given a set of latitude and longitude points and a cluster number k, the partitioning methods assign the n objects into k clusters. By applying k-means clustering, it forms the group between similar objects.

Semantic Annotation with Category: From the clustered photos, the proposed work needs to find the landmark. To identify the landmarks and to give semantic meaning from the clustered photos, the geodeg database is incorporated in the algorithm.

Algorithm 1: Semantic annotation with category Input-Set of locations
Output-Set of semantic locations
1. For all location l do
2. COUNT score s for each tag t belongs to photos Using TF-IDF
3. FETCH PLACES from POI DATABASE
4. SORT $T^{'}l$ based on score s.
5. CREATE list Matched List
6. for all $T^{'}_{L}$ do
7. for all place PLACES do
8. if MATCH(t, place)=true then
9. ADD place to Matched List
10. endif
11. End for
12. end for
13. endfor
14. if LENGTH(Matched List)>1 then
15. L.name<-CLOSEST (Matched List).name
16. Else if LENGTH (Matched List=1) then
17. L.name<-place.name
18. Else
19. L.name<-TOP (T)
20. Endif
21. Category<-TF(GENERALIZE(MatchedList))
22. ADD into semantic location

In the first step of the algorithm 1, the term frequency- inverse document frequency is used to find the score for each tag of the photo. Then the tags are sorted based on the maximum values. Using geodeg database the locations are matched with landmark names. In the last step the proposed system aggregate the landmark names with its category.

Ranking the Landmarks: From the semantic annotation module, the landmark names and its category are known. The trip advisor website (http://www.tripadvisor.com/) which has landmark name, address, description, review count and popularity score. To rank the landmark this website popularity score and review count is used. In this website the landmarks has reviews and comments. The trip advisor has the reviews and comments for all the seven categories. Based on the popularity score the important landmarks are recommended.The landmark of higher popularity always owns more reviews in Trip advisor. The top landmarks are generated which has high popularity score. In this figure 1,the proposed system of

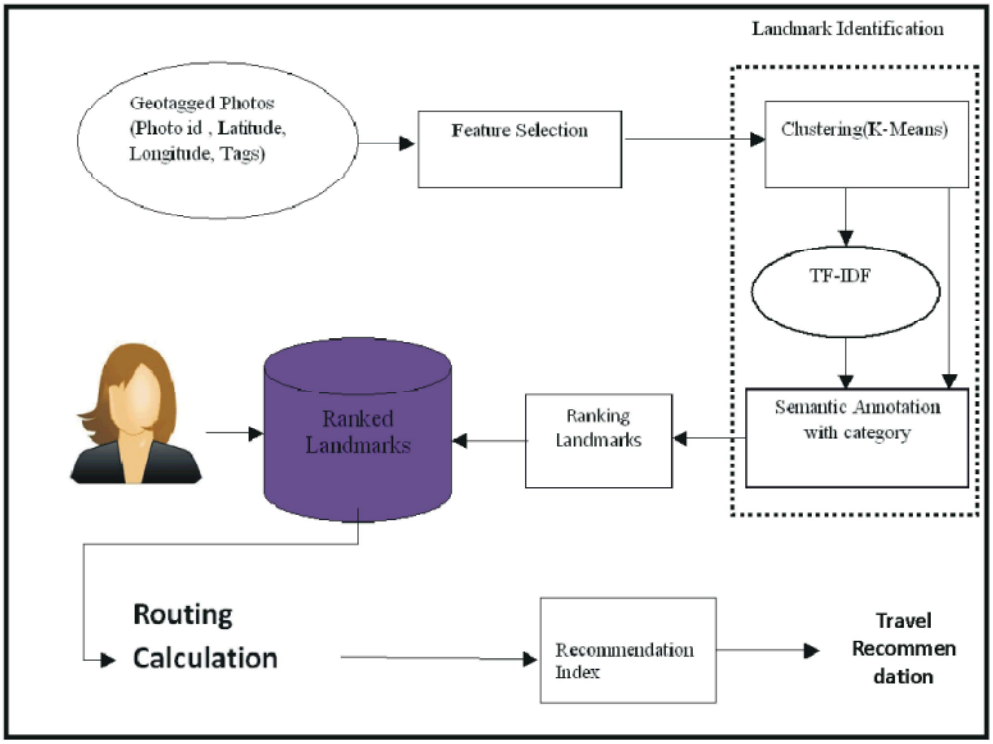


Fig. 1: System Design for Travel Recommendation

system design is shown. At first the Flickr dataset for each photoid is retrieved through API. K-means Clustering is applied to find the similarity or nearest distance between objects.TF-IDF is used to find out the tag score values. From the semantic annotation with category the landmarks names are identified. Then the landmarks are ranked with the help of trip advisor. Routing calculation is done to find the shortest path and the number of points of interest coverage.

Query Processing: The user has to choose the source and destination from the ranked landmarks. After selection of landmarks the shortest path was calculated among them and number of points of interest coverage. In the final step the tourist or users provide with the recommendation.

Routing Calculation

Shortest Path Classification: Here the shortest path algorithm was used to find the nearest road to travel. When calculating the road path there are n number of roads are generated.Although there are n number of roads, only two nearest roads are considered. The first nearest road which is less than ten meter will be represented as one. The second nearest road which is greater than ten meter will be represented as zero.This classification set {0, 1} is called as binary classification. Therefore the classification based approach used in this research.

Classification Method: In our shortest path classification the relationship between classification category and feature variables are not clearly correlated. In this classification feature, the two distances are used and then three categories are added as feature. In addition to the distances the number of points of interest is also regarded as classification feature. The subcategories of the three categories of points of interest are presented below

- Tourism Attraction-Parks, Beaches
- Eating and Drinking Sites-Food, Restaurants
- Others- Religious, Shopping, Cultural and Entertainment.

The Classification feature set is represented as (Dis_1,Dis_2,Num_1^1,Num_2^1,Num_1^2,Num_2^2,Num_1^3,Num_2^3) Where

the notations are represented asDis_1-Distance of the point to the first nearest road,Dis_2-Distance of the point to the second nearestroad,Num_1^1-Number of points of interest of category 1 on the first nearest road, Num_2^1-Number of points of interest of category 1 on the second nearest road, Num_1^2-Number of points of interest of category 2 on the first nearest road, Num_2^2-Number of points of interest of category 2 on the second nearest road, Num_1^3-Number of points of interest of category 3 on the first nearest road, Num_2^3-Number of points of interest of category 3 on the second nearest road. Binary Logistic Regression method is applied to find the relationship between classification category and feature variables. It measures the relationship between independent variables and dependent variable. It computes the probability of outputs for the different category. The Binary Logistic regression is represented as

$$P(y) = \frac{1}{1+e^{-(a_0+a_1x_1+a_2x_2+a_3x_3+a_4x_4+a_5x_5+a_6x_6+a_7x_7+a_8x_8)}} \quad (1)$$

Recommendation Index: The following features are used to find the recommendation index and it is considered to the travel recommendation. The recommendation index used to quantify the overall popularity of road.

The different categories of points of interest is described and to find the PO usability of the road as follows

$$PO(r) = W_1 Num^1_{poi}(r) + W_2 Num^2_{poi}(r) + W_3 Num^3_{poi}(r) \quad (2)$$

Where $Num^1_{poi}(r)$, $Num^2_{poi}(r)$, $Num^3_{poi}(r)$ are number of POIs of category1,2 and 3. Then W_1, W_2, W_3 are the weights of

$N^1_{poi}(r)$, $N^2_{poi}(r)$, $N^3_{poi}(r)$ and $W_1+W_2+W_3=1$.

The recommendation index is calculated as follows

$$\text{☻}_1(\text{no.of photos}) + \text{☻}_2 PO(r) - Leng(r) \quad (3)$$

Where ☻1,☻2,☻ are the weights of no.of photos,PO(R)and leng(r) separately.

Interested Routing Generation: Based on the source and destination landmarks, the interested routing is calculated. It is given below

$$Rou(L_1, L_2) = \arg\max_{rou(L_1,L_2)} \sum_{r \in Rou(L_1,L_2)} \text{Re}\, c(r) \quad (4)$$

Where Rou (L_1, L2) is the one which is of maximum value of total in recommendation index.

RESULTS AND DISCUSSIONS

Feature Selection: The following features are selected from the Flickr Photo metadata.

Table 3: Feature Selection Sample Output

Latitude	Longitude	Tags	Photoid
12.911409	80.155728	Narada Gana Sabha, art-form, bharatanatyam, chennai, classical, dance, margazhi	22960081934
12. 90163	80.214014	None	23018456003

In this table 3, it shows the attributes extracted from xml documents. The latitude and longitude where it contains location information based on this information only the proposed system begins to process further.

Preprocessed data in trip advisor Trip advisor data which is used to rank the landmarks.

Table 5: Preprocessed Data

Landmark Name	Popularity	Review Count
Kapleeshwarar Temple	294	978
Shiridi Sai Baba Temple	294	216

In this table 5, it describes the landmark name and review count.This landmark name matched with the identified landmark name from algorithm 1. Then the landmarks are ranked based on the popularity core.

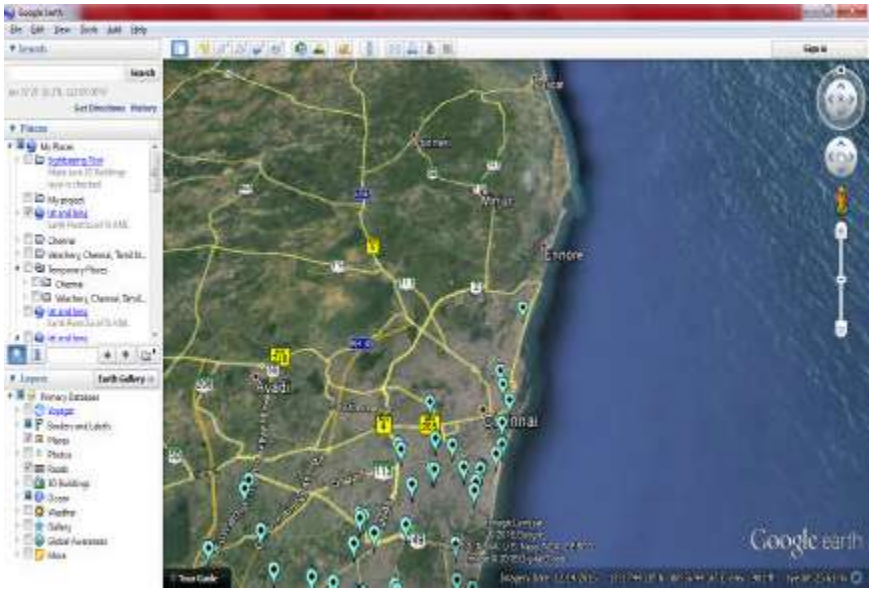
Clustered Result with Photo id: K-means cluster produce the result with photoids of each record. The cluster id are represented as cluster1, cluster 2, cluster 3.

Table 4: Sample Clusters for Photo id

CLUSTER 1	CLUSTER 2	CLUSTER 3
23065256234	23713968266	23513448016
23794070726	23692405405	23748105995
22927044314	22938888194	23426567820
23628325325	23566749551	23765098806

In this table 4,it shows the cluster formation of objects (photoid).The clusters are formed based on latitude and longitude which acts as Euclidean distance measure.

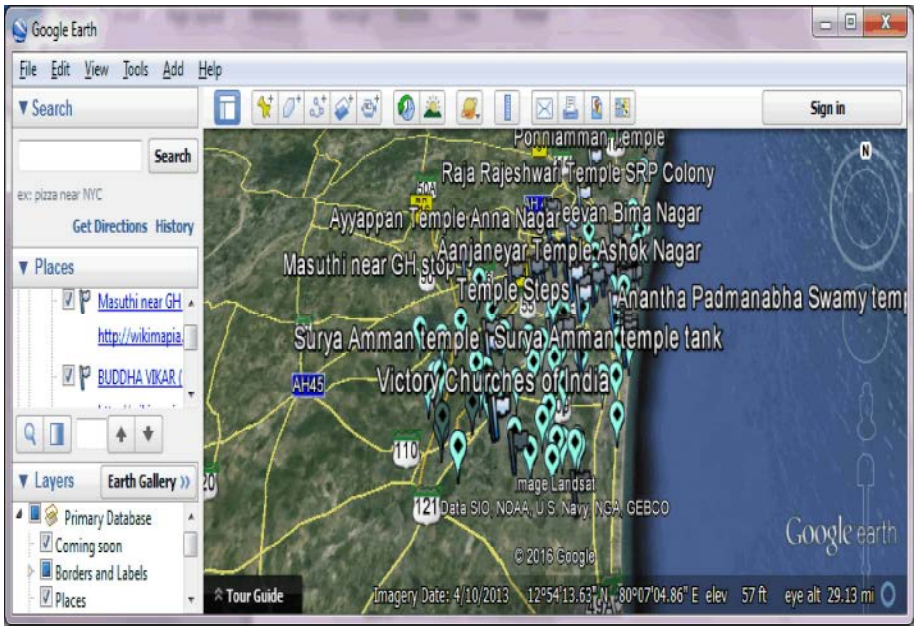
Sample Cluster in Google Earth: The latitude and longitude of each cluster is taken and stored in database. From the database it is retrieved, then shown in Google Earth.



In this figure 2, it shows the clustered latitude and longitude without location information. Google earth software is downloaded and visualized the result in it.

Category in Goggle Earth: There are seven categories involved in the proposed system. They are Tourism Attraction(Beaches,Parks), Restaurants, Religion, Education, Shopping, Entertainment and Cultural.
The given figures show the various religion category.

- Temples
- Churches
- Mosques.



CONCLUSION

This system provides solution to tourists which are useful worth to visit the landmark in particular city. The recommendation system pays more attention to the personalized user. The landmarks are recommended based on user preference rating. The routing between the landmarks and tourism popularity of the road is calculated and suggested to the users. It recommended the landmarks based on the user preferences category.

ACKNOWLEDGEMENT

The authors wish to thank the Management and Principal of Mepco Schlenk Engineering College, for the support in carrying out this research work. I am extremely thankful to my guide for sharing his expertise and for their valuable guidance and encouragement extended to me.

REFERENCES

1. Abdul Majid, Ling Chen, Hamid Turab Mirza, Ibrar Hussain and Gencai Chen, 2015. A system for mining interesting tourist locations and travel sequences from public geo-tagged photos, in Data & Knowledge Engineering, 95: 66-86. Elsevier journal.
2. De Choudhury, Munmun, Moran Feldman, Sihem Amer-Yahia, Nadav Golbandi, Ronny Lempel and Cong Yu, 2010. Automatic construction of travel itineraries using social breadcrumbs. In Proceedings of the 21st ACM conference on Hypertext and hypermedia, pp: 35-44.
3. Kennedy, L., M. Naaman, S. Ahern, R. Nair and T. Rattenbury, 2007. How Flickr helps us make sense of the world: Context and content in Community-contributed media collections. Proceedings of the 15th International conference on Multimedia, pp: 631-640.
4. Crandall, D., L. Backstrom, D. Huttenlocher and J. Kleinberg. 2009. Mapping the world's photos. InWWW,
5. Ahern, S., M. Naaman, R. Nair and J. Yang, 2007. World explorer: Visualizing aggregate data from Unstructured text in geo- referenced Collections. In Proceedings of the 7th ACM/IEEE joint Conference on Digital libraries, pp: 1{10.
6. Lee, Ickjai, Guochen Cai and Kyungmi Lee, 2014. Exploration of geo-Tagged Photos through data mining approaches. Expert Systems with Applications, 41(2): 397-405.
7. Khetarpaul, Sonia, Rashmi Chauhan, S.K. Gupt Subramaniam and Ullas Nambiar, 2011. Mining GPS Data to determine Interesting locations. In Proceedings of the 8th International Workshop on Information Integration on the Web: in conjunction with WWW, pp: 8. ACM
8. Cao, Xin, Gao Cong and Christian S. Jensen 2010. Mining Significant Semantic locations from GPS data. Proceedings of the VLDB Endowment, 3(1- 2): 1009-1020.
9. Jankowski, P., N. Andrienko, G. Andrienko and S. Kisilevich, 2010. Discovering landmark Preferences and movement patterns from photo, Postings. Transactions in GIS, 14(6): 833-852.
10. Horozov, T., N. Narasimhan and V. Vasudevan, 2006. Using location for personalized POI Recommendationsin mobile environments. In Proceedings of the International symposium on Applications on Internet Applications and the Internet23–27 January 2006. Phoenix, AZ, USA (pp: 124-129). Washington, DC: IEEE Computer Society.
11. Clements, M., P. Serdyukov, A.P. de Vries and M.J. Reinders, 2010. Using flickr Geotags to predict user travel in Information retrieval, SIGIR '10, ACM, New York NY, USA, pp: 851-852.
12. Yin, Z., L. Cao, J. Han, J. Luo and T.S. Huang, 2011. Diversified Trajectory pattern ranking in geo-tagged social media", Information retrieval, SIGIR '10, ACM, New York Proceeding of SDM'11, pp: 980-991.
13. http://www.flickrapi.com
14. http://www.tripadvisor.com
15. http://www.geodeg.com.
16. http://www.googleearth.com.