

Hybrid Approach for Location Based Customized POI Travel Recommendation System

Chitra P S¹, Dr. Girijamma H A²

¹Department of Computer Science and Engineering
RNS Institute of Technology, Bengaluru- 560 098
chitra.arunkrishnan@gmail.com

² Professor, Department of Computer Science and Engineering
RNS Institute of Technology, Bengaluru- 560 098
girijakasal@email.com

Abstract: *The creation of world of internet lead to many advancements in the technical as well as non-technical fields. Many thriving industries gained enormous advantages through the development of internet. Tourism is one among the flourishing industries of this era. Earlier the approach to visit a place is far beyond what is available today. Ranging from pamphlet to maps were made use to visit a place. Routing of places was done using maps and travel guides. Unlike before, the tourism industry has grown to great heights and handles the tours and travel without any manual intercession. The places are endorsed to users automatically using the recommendation system. Existing recommendation systems suffer from a problem of sparsity. Sparsity problem deals with inadequate data related to a user or any travel region. The proposed approach concentrate on eliminating this snag in the existing approach and generating a travel recommendation structure which focus on customized suggestions for users. The tags as well as geotags of images uploaded by the users on social sites like Flickr is made use in order to recommend the points of interest (POI) to the users. A hybrid approach with use of tags, geotags and the owner (author) of the image is proposed in this paper. User based collaborative filtering is used to create suggestions for the users. The benefits of the proposed approach is manifested by experimental analysis on a large cluster of data.*

Keywords: Collaborative filtering, hybrid approach, POI, sparsity problem

1. Introduction

Many travel guiding websites are providing collection of information to facilitate the tour and travel of the end users. The recommendation of travel places is done for the users which helps them to take decisions regarding which place to visit. Although this method is found to be very helpful for travelers but at same time it suffers from many problems. Existing systems do not provide appropriate suggestions for places if any user has visited very less places or even if very little information about any place is known. This disadvantage of the existing recommendations systems is known as sparsity problem. Sparse data is not suited for recommendation systems. To resolve the problem, a hybrid approach to travel recommendation system is proposed in this paper which concentrates on using the contents of the images up-loaded on social sites like Flickr in order to make appropriate endorsements of places to the users. Collaborative filtering is a popular approach to travel recommendation and provides appropriate results to the end users [1], [4]. The proposed method makes use of tags, geo-tags and the author information for various purposes in this paper. A customized suggestion of places is carried out in order to propose places according to interest level of users [2].

The hybrid approach is implemented using tags of images as well as geotags in order to create a domain object which would be used for further developments. Existing systems uses either tags or geotags for this work [3], [5] and [6]. The locations are mined using the geotags extracted from the photos uploaded by the users and shared publically [7], [8]. These are called as points of interest or POI's and are saved for future integration purpose. Then the tags are separated and used to create a concept or a topic along with the information of its creator.

This method is known as topic model-ling [11]. The topics here refers to points of interest like beaches, hill stations, historical sites etc. These topics are mined from the useful tags of the user uploaded photographs [12]. Sites like Flickr contributes millions of publically shared photographs uploaded by the users and taken during their travel to the respective places. Finally the similarity of the users is calculated to generate the similarity values of all the users to the current end user for whom recommendation has to be done. The users who are similar to end user would be extracted and the places travelled by them is suggested to the end user. The ranking of places is done according to the interest level of users to that particular place [10].

The major contributions of this work is to make use of tags as well as geotags for the recommendation purpose thus leading to a hybrid approach. Also utilizing tags for recommendation solves sparsity problem to enormous extent. And since using geotags instead of tags there helps to create an appropriate poi list rather consisting of locations only. Using tags to create poi list gives all tags extracted including location names along with other unnecessary tags. Ranking of places according to the number of travelled users is a further advantage which helps users in decision making.

The remaining part of the paper includes: Section II focusing on system overview. In part III, hybrid approach based filtering for personalized recommendation is elaborated. Experimental analysis and its results are shown in section IV. Conclusions are strained in part V.

2. System Overview

The basic intention behind this work is to recommend travel places according to the interest level of particular user. This work focus on proposing a hybrid approach utilizing tags as

well as geotags for creating personalized suggestions and hence eliminating the existing problem of sparsity to a large extent. Availing the information from the existing approaches of either using tags or using geotags for recommendation, the new hybrid model is proposed which is a combination of both. The earlier methods consisted of two parts: offline module and online module. Offline part focus on mining the information and online part generates recommendations. The approach is time consuming and uses all the locations information from the beginning till the end. This drawback has been overcome by avoiding the split into offline and online. Entire system is considered as a single unit and leads to creation of POI list after which the tags are put to use for topic modelling. Similarity calculation is the next succeeding step of topic modelling after which the locations visited by similar users are suggested to end user. The location names are extracted at the end of the process in order to make the process more efficient and less time consuming.

2.1 Data Collection

In this paper, the work has been carried out using cluster of real world data. Flickr.com is a rich image sharing site which lets to download the publically shared photographs through its open API [8], [13]. The data set for my paper is downloaded using this API of Flickr for two cities of Los Angeles and Berlin. Using the data set directly would lead to many errors. Thus pre-processing of data is done to eliminate any noise in the data set. The file downloaded for the images consist of raw information consisting of user id, length and width, tags, date and time and the geo tags consisting of latitude and longitude. The extraction of geotags and tags is done with respect to further process of location identification and similarity calculation and topic modelling. The parser written to extract useful information from raw data helps in creating a domain object which gives content for POI location mining as well as user history mining. Natural language processing tool is used to extract the useful tags from the whole data set.

2.1 Architecture

Fig.1, explains the architecture of the poi recommender system. The approach followed is to use geotags to mine the locations from the dataset and tags for topic modelling along with the user information in the photo dataset. In the absence of geotags for any image or improper geotags, in that case the tags is likely to be utilized for mining the points of interest from the dataset. . The similarity measure calculates the resemblance of the given user with all the other existing users. And the most similar N users are extracted for further processing. Ranking of places is also done for accuracy in recommendation.

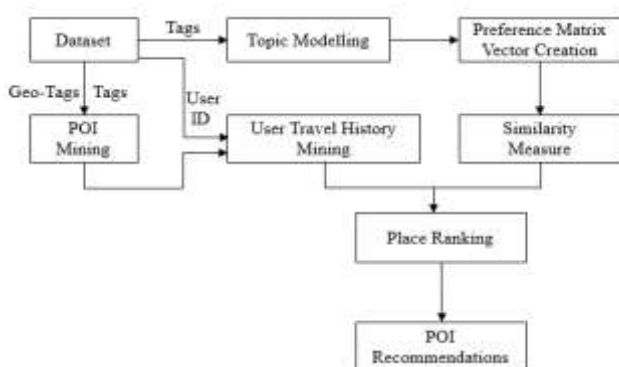


Figure 1: Architectural representation of Customized Travel Recommender

3. Hybrid Approach

The works carried out in previous papers shows the existence of problem of sparseness in data. The works also portrays the difficulty to recommend places to users who have not travelled frequently. To overcome these prevailing hurdles, this paper suggest to combine the tags and geo-tags in order to acquire effective recommendations. The hybrid approach introduces the usage of both the tags as well as geotags for the customization of places for particular users.

3.1 Tags for Topic Modelling

The domain object created after parsing of the raw dataset contains the separated user id, tags and geotags of images. Let N_c denote the number of photos that is usable as per the requirements. And the geo tags, tags be denoted by G and T_g respectively. An image is given by I and possess two variables given by G and T_g . The initial process is to determine the relation between topics, tags and user. Similar kind of words in tags related to travel makes up the topics denoted by T_o . The generation of preference matrix vector V leads to a $N_u \times N_t$ matrix where the number of users denoted by N_u and the number of topics be given by N_t . The output of the matrix $V(i,j)$ shows the tags of the users assigned to the topics. This is further normalized to extract the topic distribution of particular users.

3.2 Geo tags and tags for POI Mining

The geo tags are available with the images in the domain objects. These geotags are made use for the purpose of mining the poi's. For the basis of making the whole work less complicated, the geotags are retrieved as locations at the time of recommending the locations to the users. The conversion of all geotags to location names at the beginning will create unnecessary load making the process of recommendation very slow. And there exist a limit on conversion of geo tags of pictures into names per day. To solve these problems the location names are extracted only at the time of suggesting the particular place to any users. The absence of geotags in any of the images leads to use of tags for the same purpose. The location names are extracted from the tags of the images and this leads to creation of a poi list which is used for the purpose of user travel history mining.

3.3 Customized POI Recommendations

The customized poi recommendation system aims to put forward a list of places to visit in a city according to his or her interest level. The system is called customized due to the fact that the suggestions given for the places are as per the interest level of the particular traveler. If a person has travelled more of any places like beach, historical places etc. then that sort of recommendations will be provided to the user first along with other places of interest. The kind of recommendation approach consists of two sub processes: 1) Similarity measure, and 2) Place ranking.

1) Similarity measure: Let u_1 and u_2 be two users who have

travelled l_1 and l_2 . The similarity measure between these two users will be calculated on the basis of the preference matrix vectors created in previous steps. The similarity calculations would be based on cosine similarity measure. $V(i)$ be topic distributions of u_i , $i \in \{1, 2, \dots, N_u\}$ in l_j . Other similarity measures are also explained in the experimental analysis. The estimation of cosine measure is as

$$\hat{C}(u, i) = \frac{V(u) \times V(i)}{\|V(u)\| \cdot \|V(i)\|} \quad (1)$$

2) Place ranking: For any particular location l_i , if the number of visitors for that particular place is n times more than the number of visitors for another location l_j , then ranking for l_i will be more than ranking of l_j . The places of more rank is suggested to the users on the top of the recommendation list. Let $L_2 = \{l_1, l_2, \dots, l_N\}$ be set of poi locations. Ranking of L_2 for all the users according to the user travel history Q_{Nu} is calculated. The ranking of places helps the users to decide on which place to travel and is of most importance rather than deciding on any place without any knowledge. The rank of poi in L_2 is arranged in decreasing order. The top most poi in the list is served as recommendations for the users. And is given by

$$C(j, k) = \sum_{i=1}^{N_u} q_i^j \quad (2)$$

Ranking is done in decreasing order of user interest. The most travelled places by the users are given highest ranks and with least number of travelers are given lesser ranks. For example, if Eiffel tower in Paris has most number of visitors and tags associated with it, then it is given rank 1.

4. Experimental Analysis

The proposed work have been proved through experimental analysis on dataset taken from the famous site Flickr.com. The real world data was retrieved using the open API of Flickr. The raw data consist of all information related to an image like its author name, latitude and longitude information, tags given by the users, dates on which the photographs were taken etc. For this work, useful information related to an image is extracted which includes tags and geotags. More information about the dataset used in the work is discussed in next section.

4.1 Test Data

The test data from Flickr consists of images shared by the users. The information related to these images are the input for the proposed work done in this paper. Parsing these data leads to a domain object which includes the place name also called as the PoI, author details, geotags and tags. The test dataset of the poi list and user id after the mining process is shown in table 1. The places of interest visited by various users is fetched from the dataset after the mining of places of interest of the travelers along with their user ids. The tags and geotags are also obtained after this process which is used in topic modelling. The topics are modelled in order to ease the process of recommendation. Tags of same category are put together into same set. For example all the water related places like beaches, waterbodies, etc. are put under Topic 52. In same way all the

other topics are modelled. This topics are used to create a matrix vector which helps in calculating similarity of users.

Table 1: The Poi List along with User Information

No	POI	User No.
1	Wilhelm-Wolff-Straße 20, 13156 Berlin	User414
2	Marinesteig 6A, 14129 Berlin, Germany	User522
3	Großer Stern, 10557 Berlin, Germany	User112
4	3950 Los Feliz Blvd, Los Angeles	User909

4.2 Results

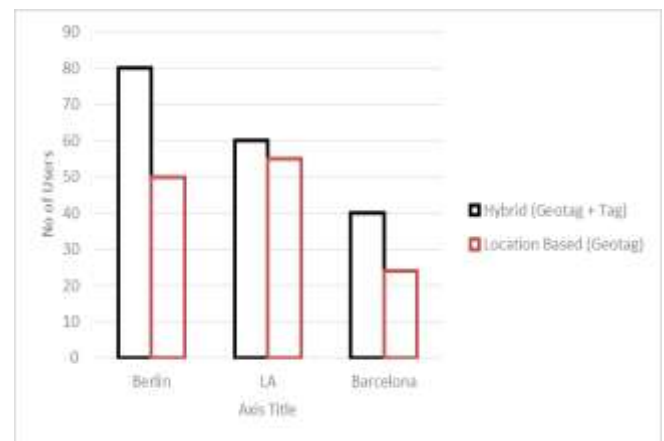


Figure. 2. Number of users with correct recommendations of places in different cities.

The final result provided by the customized recommendation system is showing the names of travel places in a city which are suited according to the personal interest level of the users. The performance of the proposed approach shows an accuracy in recommendation which outperforms the previous approaches of just using either tags or just geotags for suggesting the places of travel. The problem of sparsity is also solved to a much greater extent than any other previously suggested methods. Using the combination of tags and geotags have enhanced the accuracy level to a greater extent. According to the study of literature survey of previous approaches which utilized only geotags, the proposed hybrid approach recommends appropriate places to more number of users as shown in fig. 2. Thus increasing the efficiency of the work and reducing the prime complication of sparse data.

5. Conclusion

In this paper, the proposed approach of hybrid method for recommendation of travel sites to people is solving the existing problems of sparsity. Data from Flickr website has been used to

do the experimentation and to provide customized recommendations. If only geotags are used to mine the locations the users will be matched on the basis of their visit to any particular place. And then the recommendations of each other which are not visited by them would be shown. This creates a problem where a user who travelled to less places won't be given any sort of suggestions for a new place. To overcome these hurdles, the proposed method in this paper considers geotags along with tags to suggest the places of interest to the users. The PoI's are also given ranks according to its popularity among the people. The locations are extracted from the latitude and longitude and recommended to users based on their similarity with other users. Topic modelling based on tags and the authors is done to find the preference matrix which is the input for the similarity calculation. Finally the most similar users are found and the places are recommended to them which is not visited by them. The method is more accurate in comparison to the previously adopted approaches and is proved experimentally.

References

- [1] Shuhui Jiang, Xueming Quian, Jialie Shen, Yu Fu and Tao Mei, " Author Topic based Collaborative Filtering for Personalized POI Recommendations," IEEE Transaction Multimedia, vol.17, no.6, June.2015 (IEEE Transaction).
- [2] Qi Lui, Yong Ge, Zhongmou Li, Enhong Chen, Hui Xiong, " Personalized travel package recommendation," IEEE Int. Conf. Data mining, pp 407-416,2011 (IEEE Transaction).
- [3] S. Jiang, X. Qian, J. Shen, and T. Mei, "Travel recommendation via author topic model based collaborative filtering," in Proc. 21st Int. Conf., MMM 2015, Sydney, N.S.W., Australia, pp. 392–402, 2015 (Conference Proceedings).
- [4] H. Huang and G. Gartner, "Using trajectories for collaborative filtering-based poi recommendation," Int. J. Data Mining, Model.,Manage., vol. 6, no. 4, pp. 333–346, 2014 (International Journal Citation).
- [5] M. Clements, P. Serdyukov, A. de Vries, and M. Reinders, "Using flickr geo-tags to predict user travel behaviour," in Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, pp. 851–852, 2010 (Conference Proceedings).
- [6] T. Kurashima, T. Iwata, G. Irie, and K. Fujimura, "Travel route recommendation using geo-tags in photo sharing sites," in Proc. 19th ACM Int. Conf. Inf. Knowl. Manage, pp. 579–588., 2010 (Conference Proceedings).
- [7] J. Bao, Y. Zheng, and M. F. Mokbel, "Location-based and preference aware recommendation using sparse geo-social networking data," in Proc. 20th Int. Conf. Adv. Geographic Inf. Syst, pp. 199–208, 2012 (Conference Proceedings).
- [8] J. Li, X. Qian, Y. Y. Tang, L. Yang, and T. Mei, "GPS estimation for places of interest from social users' uploaded photos," IEEE Trans. Multimedia, vol. 15, no. 8, pp. 2058–2071, Dec. 2013 (IEEE Transaction).
- [9] Qi Liu, Enhong Chen, H. Xiong, Yong Ge, Zhongmou Li and Xiang Wu," A cocktail approach for travel package recommendation,"IEEE transaction Knowledge and data engineering, vol.26, no.2, Feb. 2014 (IEEE Transaction).
- [10] K.Venkatesh and J.Jabej, "Recommendation and ranking of travel package to tourist," in Contemporary engineering sciences, vol. 8, no.10, pp.401-407, 2015 (Journal or magazine citation).
- [11] H. Feng and X. Qian, "Mining user-contributed photos for personalized product recommendation," Neurocomput., vol. 129, pp. 409–420, 2014 (Journal or magazine citation).
- [12] X. Liu, X. Qian, D. Lu, X. Hou, and L. Wang, "Personalized tag recommendation for Flickr users," IEEE Int. Conf. Multimedia Expo, pp. 1–6, Jul. 2014 (IEEE Conference citation).
- [13] X. Qian, X. Liu, C. Zheng, Y. Du, and X. Hou, "Tagging photos using users' vocabularies," Neurocomput., vol. 111, pp. 144–153, 2013 (Journal or magazine citation).

Author Profile

1. **Chitra P S** is pursuing MTech in Computer Science and Engineering from RNS Institute of Technology, Bengaluru under Visvesvaraya Technological University, India.

2. **Dr. Girijamma H A** is currently working with RNS Institute of Technology, Bengaluru as professor in Department of Computer Science and Engineering.