International Journal of Advanced Trends in Computer Applications
*www.ijatca.com*

# Personalized Travel Recommendation Using K-Means Clustering

**Pranita Sathe[1], Archana Augustine[2]**
[1,2] Department of Information Technology,
Pillai HOC College of Engineering & Technology, Rasayani,
Raigad, 410207, India.
[1]*pranusathe8@gmail.com*, [2]*aan.archana@email.com*

**Abstract:** *Travel route planning is one of the most important steps for a tourist to prepare his/her trip. Before traveling to an unknown location, most people have questions about how to plan their trips. Although users can take help of travel guide or ask questions to web based communities, the process is generally not efficient and the results may not be customized. An automatic and interactive travel route planning service is highly desired to plan a customized trip according to users' preferences. This paper provides personalized travel sequence recommendation with the help of both travelogue and community contributed photos. Travelogue websites offers 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, travel routes mining, but give an opportunity to recommend personalized travel POIs and routes based on user's interest. Compared with general routes recommendation, our recommended personalized travel sequential POIs are more relevant to user's interest and more convenient for travel planning. We propose Topical Package Model (TPM) method to learn users and route's travel attributes. It bridges the gap of user interest and routes attributes. We map both user's and routes' textual descriptions to the topical package space to get user topical package model and route topical package model (i.e., topical interest, cost, time and season).[9]To recommend personalized POI sequence, first, famous routes are ranked according to the similarity between user package and route package. Then top ranked routes are further optimized by social similar users' travel records.*

**Keywords:** *travel recommendation, topical package model, photo collection and information retrieval.*

## I. INTRODUCTION

Travelogue websites provides lots of information regarding locations and user's travel experiences. Community contributed photos with metadata on social media also records the user's daily life and travel experience. These data recommends personalized travel POI and routes according user's interest. In modern days, the rapid growth of cities has paved way for the development of a huge number of points of interest (POIs), which enliven and entertain the people, providing us with more choices of living experience than before. Nowadays people decided where to go based on their personal interests and the various choices of POIs. The relationship between the locations and opening time of different POI [12] it is difficult and time consuming for the users to plan their travel sequence. Therefore it is required to recommend travel sequence. To solve the problems of existing travel recommendation systems we propose topical package model (TPM) which provides travelogue and community contributed photo collection [8].TPM is also used to obtain the attributes of user topical package and route topical package. To mine user's topical interest we mapped user's tagged photos with topical package model. After user package mining we rank famous travel routes by finding the similarity between user package and route package. Then top ranked routes are further optimized using similar users travel records. To find similar users we mapped user package with social user's package.

The remainder of this paper is organized as follows. Related studies are reviewed in section 2. Section 3 introduces overview of our system. Section 4 describes how to build topical package model.

# II. RELATED WORK

In this section we provide existing work of travel recommendation system also we point out the differences between proposed and existing system.GPS trajectory [5][6][11],geo-tags [3][7], blogs (travelogues) are main sources used in travel recommendation. User generated travelogue provides more information.GPS trajectory and geo-tags are also widely used to recommend the travel plan. While comparing user generated travelogue and geo-tags data on social media, GPS data are relatively difficult to obtain. However general travel recommendations only considered the popularity of POI or routes. Therefore personalized travel recommendation is important nowadays. The main approach of personalized recommendation is collaborative filtering (CF)[6][2]. Location based collaborative filtering first mined similar users according to location co-occurrence and then POIs are recommended according to similar users voting. However CF may face problem that is computational complexity increases with large amount of users and locations. To solve this problem author topic model based collaborative filtering [4] is used. In this technique travel topics and user topical interest mined simultaneously. 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. The existing studies did not well consider the popularity and personalization of travel routes at the same time. The attributes like consumption capability, preferred season have not been mined automatically. This paper provides the solution with the help of building TPM. Topical package model is learnt to get users and routes attributes (topical interest, cost, time, season).For ranking famous routes we consider the user topical package and route package. After this we optimized top rank routes by social similar users travel records in TPM

.

# III.SYSTEM OVERVIEW

The system we proposed is personalized POI sequence recommendation. By using this system we automatically mine users travel attributes such as topical interest, preferred time, season, consumption capability. In this section, we briefly introduce the term used in this paper: topical package space, user package, route package. Topical package space consists of (1) representative tags mined from travelogue (2) average consumer expenditure (3) distribution of visiting season (4) visiting time. User topical model is learnt from mapping the tags of user's photos to topical space. It includes user topical interest $\alpha^{(U)}$, user's consumption capability $\beta^{(U)}$, preferred visiting time $\gamma^{(U)}$ and season $\zeta^{(U)}$. Route topical model is learnt from mapping travelogues related to the POIs on the route to topical package space. It includes routes interest $\alpha^{(R)}$, routes cost $\beta^{(R)}$, time $\gamma^{(R)}$ and season distribution $\zeta^{(R)}$.Figure1 illustrates system framework of proposed system which consists of two input modules. Offline module aims at preparing topical package space, POI mining, famous route and their topical package. Online module is mainly concentrated on mining users travel interest and recommending travel routes.

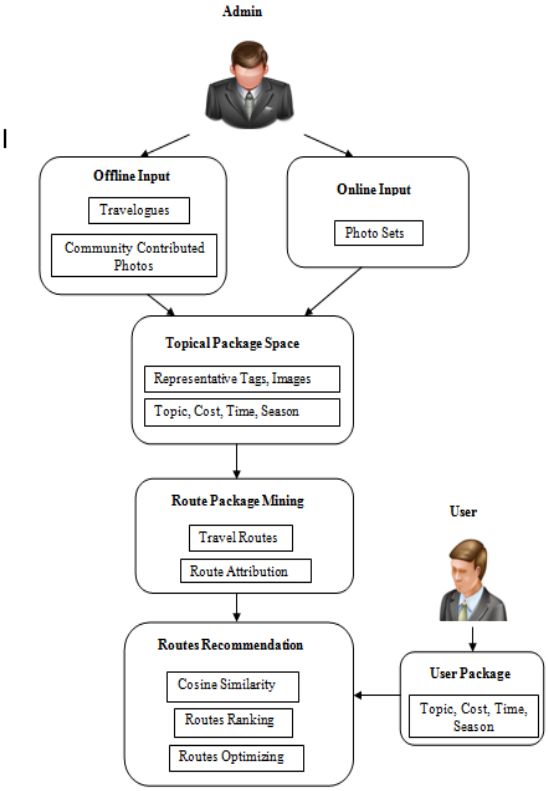


**Figure 1:** System Architecture

### 3.1 Topical package space construction

In our paper, we construct the topical package space by the combination of two social media: travelogues and community-contribute photos.

We introduce the gathering and structure of travelogues and then how to mine representative tags, distribution of cost and time of each topic. We downloaded most famous cities travelogues from famous travel websites. Representative tags are tags which not only have high frequency in one topic but also could distinguish topic from others. To mine representative tags, first, we remove meaningless symbols and stop words. We then use Term Frequency Inverse Document Frequency (TF-IDF) method to get the score of each tag.

Tag score is used to reflect the importance of a tag to the topic. We define $i^{th}$ tag's score of the $^{k\text{-}th}$ topic as $\chi$ i, k. The TF part reflects the frequency $^{i\text{-}th}$ tag appear in the $^{k\text{-}th}$ topic. The IDF part reflects how much categories contain i-th tag. After mining representative tags, we mine cost and time attributes for all the topics from

travelogues. We select sentences containing numbers and then we utilize natural language processing (NLP) to learn the feature of each sentence. First, we pass sentence through "comment Sanitizer". Then we initialize global hash map. For each word in the resulting string, first, we pass the word to "porter Stemmer". Then if the word is not in your hash map, add it. If it is, just add one to its value. The second step is to train a text classifier with positive samples and negative samples. After training the classifier, we put the sentences containing numbers into the classifier to test whether a sentence is related to cost. We use the sentences which are both related to cost and from the travelogues of the topic to mine the cost information about this topic. We calculated the mean value of the numbers appeared in these sentences. Community contributed photos are collected by randomly downloaded images worldwide. There is a photo album which includes no of photos shared by the user, associated with metadata including user id, textual tags, date taken, latitude, longitude etc. By using community contributed photo collection we can mine POI and season attribute. After getting POIs, there are a set of photos with tags and "date taken" labels. To season, we use the "month" in "date taken" to get the visiting distribution during the 12 month.

### 3.2 User topical package model mining

In this we introduce how to mine user package. User package consists of α (U) - distribution of user's topical interest β (U) - distribution of user's consumption capability γ (U) - distribution of user's preferred travel time ζ (U) - distribution of user's preferred travel season.

### 3.3 Route topical package model mining

In this we introduce how to build route package. Route package consists of distribution of route's topics, route's cost, route's visiting time and route's visiting season. We can define this as α(R) - distribution of route's topics β(R) - distribution of route's cost γ(R) - distribution of route's visiting time ζ(R) - distribution of route's visiting season.

**K-means clustering:** Clustering is the process of partitioning a group of data points into a small number of clusters. K-means is a clustering method that aims to find the positions of the clusters that minimize the distance from the data points to the cluster. K-means clustering algorithm is used to cluster the attributes of user topical package and route topical package.

**K-means clustering algorithm steps:**

1. Place k points into space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closet centroid.
3. When all objects have been assigned, recalculate the positions of the k centroids.
4. Repeat steps 2 and 3 until the centroids no longer move.

### 3.4 Travel recommendation

For travel recommendation first routes are ranked according to similarity between user and route package. Second routes are optimized.

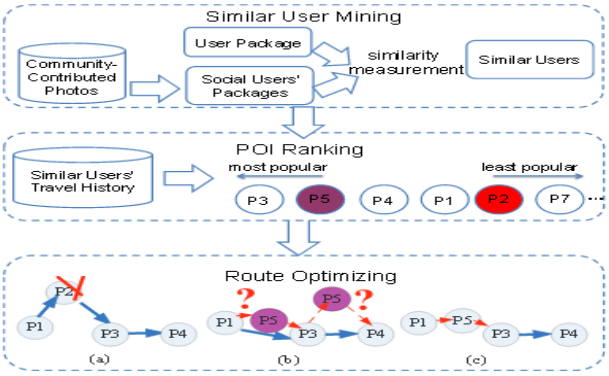


**Figure 2:** Route Optimization

## IV. CONCLUSION

We proposed a personalized travel sequence recommendation system by learning topical package model from big multi-source social media: travelogues and community-contributed photos. The advantages of our work are 1) the system automatically mined user's and routes' travel topical preferences including the topical interest, cost, time and season, 2) we recommended not only POIs but also travel sequence, considering both the popularity and user's travel preferences at the same time.

## REFERENCES

[1] Shuhui Jiang, Xueming Qian "Pesonalized travel sequence recommendation on multisource big social media", IEEE transaction on big data, vol.x.x, 2016

[2] 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

[3] 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.

[4] 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.

[5] 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.

**[6]** H. Huang and G. Gartner, "Using trajectories for collaborativefiltering based poirecommendation," International Journal of Data Mining, Modelling and Management, vol. 6, no. 4, pp. 333–346, 2014**.**

[7] 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.

[8] A. Cheng, Y. Chen, Y. Huang, W. Hsu, and H. Liao, "Personalized travel recommendation by mining people attributes from community-contributed photos," in Proceedings of the 19th ACM international conference on Multimedia. ACM, 2011, pp. 83-92.

[9] Y. Ge, Q. Liu, H. Xiong, A. Tuzhilin, and J. Chen,"Cost-aware travel tour recommendation," in Proceedings of the 17th ACMinter- national conference on Knowledge discovery and data mining. ACM, 2011, pp. 983–991.

[10] 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, 2015.

[11] 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.

[12] 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.