Scientific Journal of Impact Factor (SJIF): 4.72
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 5, May -2017

# Travel Recommendation Based On Data From Social Media

Mandar Raskar[1], Abhishek Raut[2], Ashwin Dongare[3]

[1]*Department of Computer Engineering, Indira College of Engineering, Pune*
[2]*Department of Computer Engineering, Indira College of Engineering, Pune*
[3]*Department of Computer Engineering, Indira College of Engineering, Pune*

**Abstract** — *Big data increasingly benefit every analysis and industrial area like health care, finance service and industrial recommendation. This paper presents a customized travel sequence recommendation from every travelogues and community-contributed photos and additionally the heterogeneous data (e.g., tags, geo-location, and date taken) related to these photos. in contrast to most existing travel recommendation approaches, our approach isn't only customised to user's travel interest but to boot prepared to recommend a travel sequence rather than individual Points of Interest (POIs). Topical package house together with representative tags, the distributions of value, visiting time and visiting season of every topic, is well-mined to bridge the vocabulary gap between user travel preference and travel routes. we tend to profit of the complementary of a pair of kinds of social media: attraction and community-contributed photos. we tend to map every user's and routes' matter descriptions to the topical package house to induce user topical package model and route topical package model (i.e., topical interest, cost, time and season). To suggest customised dish sequence, first, notable routes are stratified consistent with the similarity between user package and route package. Then prime stratified routes are any optimized by social similar users' travel records. Representative pictures with viewpoint and seasonal diversity of POIs are shown to provide a extra comprehensive impression. we tend to value our recommendation system on a set of seven million Flickr pictures uploaded by 7,387 users and twenty four,008 travelogues covering 864 travel POIs in nine notable cities, and show its effectiveness. we tend to additionally contribute a brand new dataset with quite 200K photos with heterogeneous information in nine notable cities.*

***Keywordst;*** *Point of Intrest, Crowdsourcing, Recommendation.*

## I. INTRODUCTION

Automatic travel recommendation is an vital downside in each analysis and business. Big media, especially the flourish of social media (e.g., Facebook, Flick, Twitter etc.) offers great opportunities to address several difficult issues, for instance, GPS estimation [1], [2] and travel recommendation [3]. Travelogue websites (e.g., www.igougo.com) offer wealthy descriptions regarding landmarks and traveling expertise 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 solely helpful for reliable POIs (points of interest) Ming dynasty [4], travel routes ming, but provide Associate in Nursing chance to suggest customized travel POIs and routes supported user's interest.

There are two main challenges for automatic travel recommendation. First, the recommended POIs ought to be customized to user interest since completely different users might like different varieties of POIs. Take New York town as Associate in Nursing example. Some people might like cultural places like the Metropolitan deposit, while others might like the cityscape like the park. Besides travel topical interest, other attributes as well as consumption capability (i.e., luxury, economy), preferred visiting season (i.e., summer, autumn) and preferred visiting time (i.e., morning, night) may conjointly be useful to produce customized travel recommendation. Second, it is important to suggest a consecutive travel route (i.e., a sequence of POIs) rather than individual POI. It is much more difficult and time consuming for users to set up travel sequence than individual POIs. Because the relationship between the locations and gap time of various POIs ought to be thought of. For example, it may still not be a decent recommendation if all the POIs suggested for in the future ar in four corners of the town, even though the user is also curious about all the individual POIs. Existing studies on travel recommendation mining famous travel POIs and routes ar in the main from four sorts of huge social media, GPS trajectory [5], check-in information [4], [6], [7] geo-tags [2], [3], [8], [9], [10] and blogs (travelogues). However, general travel route planning cannot well meet users' personal necessities. Personalized travel recommendation recommends the POIs and routes by mining user's travel records. The most famous methodology is location-based cooperative filtering (LCF). To LCF, similar social users are measured based mostly on the situation co-occurence of antecedently visited POIs. Then POIs are stratified based mostly on similar users' visiting records. However, existing studies haven't well solved the 2 challenges. For the first challenge, most of the travel recommendation works only targeted on user topical interest mining however while not considering alternative attributes like consumption capability. For the second challenge, existing studies focused a lot of on famed route mining however while not mechanically mining user travel interest. It still remains a challenge for most existing works to produce both "personalized" and "sequential" travel package recommendation. To address the challenges mentioned above, we propose a Topical Package Model (TPM) learning methodology to mechanically mine user travel interest from 2 social media, community-contributed photos and travelogues. To address the first challenge, we take into account not solely user's topical interest

@IJAERD-2017, All rights Reserved 

however conjointly 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 house, and map both user's and route's matter descriptions to the topical package house to get user topical package model (user package) and route topical package model (route package) underneath topical package house.

## II. LITERATURE SURVEY

In this section, we principally introduce 3 aspects of connected works (1) travel recommendation on numerous massive social media; (2) personalised travel recommendation; (3) travel sequence and travel package recommendation. We additionally purpose out the variations between our work and existing works. GPS trajectory [5], check-in knowledge [4], [6], [7] geo-tags [2], [3], [8], [9], [10] and blogs (travelogues) are four main social media used in recommendation. User generated travelogues provide wealthy data. Kurashima et al. extracted typical user's travel sequences according to entries, associated with multimedia data of the routes [12]. Besides travelogues, GPS and geo-tags are additionally wide used in travel recommendation. Zheng et al. conducted a series of works of travel routes mining and recommendation using GPS flight, and achieved promising results[5] .However, comparing to the wealthy travelogues and geo-tags knowledge on social media, GPS trajectory knowledge are comparatively difficult to get. Geo-tagged photos based automatic travel route coming up with works have attracted a heap attentions [8], [9].

Recently, multi-source big social media have shown their lustiness [9], However, general travel recommendations only thought of the quality of POIs or routes. Recently, personalized travel recommendations have attracted a lot of attentions. The three main approaches of personalised recommendation are cooperative Filtering (CF) mathematician Chains and matrix resolution Location based mostly CF firstly strip-mined similar users according to location co-occurrence. For example, Clements et al. modeled the co-occurrence with mathematician density estimation. Second, POIs are counseled according to similar users' balloting. However, location based CF could face 2 issues. First of all, the computational complexness will increase dramatically with massive quantity of users and locations, which is particularly serious in massive knowledge situation. Second, if the user has very few location records or most of those records belong to non-famous places, it would be very laborious to mine correct similar users. To solve these challenges, Jiang et al. proposed the Author Topic Model based mostly cooperative Filtering [3].

They mined the class of travel topics and user topical interest at the same time through Author Topic Model. Personalized travel sequence recommendation is a lot of convenient for users than the individual POIs recommendation. The system enabled user to input personal performance in an interactive manner. How ever it did not very mechanically mineuser's interest. What's more, in recent years, studies of the travel package recommendation which contained a lot of attributes (e.g. time, cost, season) have shown more effective performance than works that solely thought of topical interest [7]. Yuan et al. proposed a Geographical-Temporal influences Aware Graph for time aware dish recommendation [7]. Ge et al. developed a cost aware model, and they analyzed the relation between price and keep days. However, although these studies thought of user's travel attributes, few of them really mechanically strip-mined these attributes. The existing studies associated with travel sequence recommendation failed to well consider the recognition and personalization of travel routes at constant time. What's more, the multi-attributes of users and routes (e.g., consumption capability, preferred season, etc.) have not been mined mechanically. 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 contemplate the quality of routes and user's personal preference along by the thought of ranking famed travel routes supported user's travel interest, and optimizing top hierarchal routes by social similar users' travel records.

## III. PROPOSED SYSTEM

The system we have an inclination to planned is also a personal dish sequence recommendation system which will automatically mine user's travel attributes like topical interest, consumption capability and most popular time and season. during this section, we have an inclination to briefly introduce the terms utilized during this paper: topical package space, user package and route package. Secondly, we provide the system outline. Topic package space could be a fairly area during which the four travel distributions of each topic are drawn by (1) representative tags mined from travelogues which describe POIs at intervals constant topic;(2) the common shopper expenditure of the POIs at intervals this subject, that are mined from travelogues; (3) distribution of the visiting season of the twelve months mined by the "date taken" hooked up with the community-contributed photos; (4) distribution of visiting time throughout the day from travelogues. The usage of topic package area is to bridge the gap between user interest and thus the attribute of routes, since it's difficult to directly live the similarity between user and travel sequence. From mapping each user knowledge and route knowledge to constant space, we have an tendency to urge the quantitative customary to measure the similarity of user and routes. User topical package model (user package) is learnt from mapping the tags of user's photos to topical package area. It contains user topical interest distribution $\alpha(U)$, user consumption capability $\beta(U)$, preferred period distribution $\gamma(U)$ and most popular travel season distribution $\zeta(U)$. Route topical package model (route package) is learnt from mapping the travelogues associated with the POIs on the route to topical package area. It contains route topical interest $\alpha(R)$, route's value distribution $\beta(R)$, route's time distribution $\gamma(R)$ and season distribution $\zeta(R)$. Fig.2 illustrates the system framework, that consists of offline and on-line module. The offline module aims at getting prepared topical

@IJAERD-2017, All rights Reserved

package space and mining dish and noted route and their topical package emodels. It consists of the subsequent two parts: 1) social media mining and topical package space construction, 2) routes package mining. the web module is concentrating on mining user's travel interest and recommending travel routes that contains two steps: 1) routes ranking, 2) routes optimizing.

Figure briefly illustrates whether one attribute is mined directly from travelog or community contributed photos or indirectly from topical area mapping methodology.

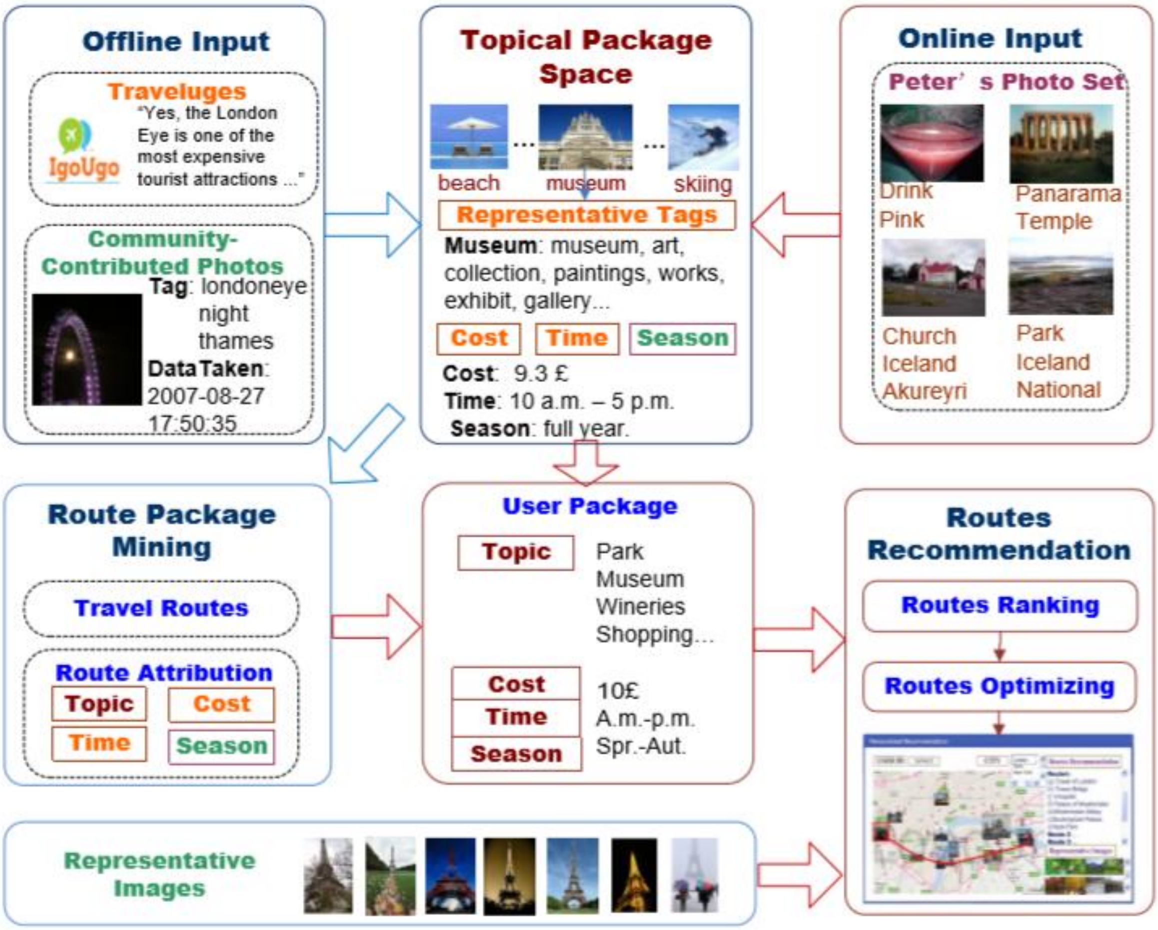



Figure: Planned System Design

## IV. METHODOLOGIES

Our topic package space is the extension of matter descriptions of topics like ODP [35]. We use the topical package house to live the similarity of the user topical model package (user package) and also the route topical model package (route package). In our paper, we construct the topical package house by the combination of 2 social media: travelogues and community-contribute photos. To construct topical package space, travelogues are used to mine representative tags, distribution of cost and visiting time of every topic, while community-contributed photos area unit used to mine distribution of visiting time of every topic.

**Travelogue Mining :**

**1. Travelogues Gathering and Structure Illustration:**

We downloaded twenty four,008 travelogues of 864 travel POIs on nine most famed cities of the world from famed travel web site IgoUgo.com. These nine cities area unit urban center, Berlin, Chicago, London, Los Angeles, New York, Pairs, Rome and San Francisco [14]. A lot of travelog connected works area unit supported the information from IgoUgo. In this paper, we directly use the class definition of IgoUgo as Table a pair of. This category might cowl most of the travel activities. The structure of data we have a tendency to crawled from IgoUgo is as Fig.4. The first layer is "City Layer". Under every town, there are twenty six topics created "Topic Layer". We denote C = as the category of topics. N is the number of topics that is twenty six in our work. Under every topic ck is "POI Layer".

@IJAERD-2017, All rights Reserved 

**2. Topic's Representative Tags Mining**

Representative tags are tags that not solely have high frequency in one topic, but additionally might distinguish this topic from others. We use travelogs to mine representative tags since previous works indicated that travelogue area unit additional appropriate than Flick information set once mining comprehensive tags of locations [36], [37]. To mine representative tags, first, we take away mindless symbols and stop words [3]. We then use Term Frequency Inverse Document Frequency (TF-IDF) technique to get the score of every tag [38]. Tag score is used to reflect the importance of a tag to the subject. We define the i-th tag's score of the k-th topic as $\chi_{i,k}$. The TF part reflects the frequency the i-th tag seem in the k-th topic. The IDF half reflects however abundant classes contain the i-th tag.

**3. Topic's Cost and Time Matrix Mining**

After mining representative tags, in this section, we mine value and time attributes for all the topics from travelogues. They are defined as value matrix $\beta(M)$ and time matrix $\gamma(M)$.

However, not all the sentences contain numbers related to value or time. So we have a tendency to ought to mechanically choose sentences associated with value or time. The processes of cost and time area unit similar, so we have a tendency to take value as associate degree example. First, we choose sentences containing numbers. And then we utilize tongue process (NLP) to find out the feature of every sentence. We use "faridani matlab NLP" to method every sentence. The basic idea is as follows. For each sentence, first, we pass it through "comment Sanitizer". Then we initialize world hash map. For each word within the ensuing 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 price. Do the same thing for the world hash map. Then we take the prices in your hash map that have value >= n, for example, n=3, and store those keys as a header. For each hash map of a sentence, we gift it mistreatment headers. For example if header is 'book', 'note', 'I', a comment like "I love my book" should be [1,0,1]. The second step is to train a text classifier with positive samples and negative samples. We manually mark five hundred positive samples and five hundred negative samples. We define the sentences, which contain value info (e.g., ticket fee) as positive samples. Sentences, which contain numbers not connected to value, are defined as negative samples. After coaching the classifier, we place the sentences containing numbers into the classifier to take a look at whether or not a sentence is said to value. We use the sentences that area unit each connected to value and from the travelogues of the subject to mine the price info regarding this subject. We calculated the mean price of the numbers appeared in these sentences. It mainly consists of the price ticket value of the dish and typically additionally together with occupation value and transportation value. Notice that although the tickets fee for adults and youngsters area unit totally different, the mean cost might still distinguish low cost and dearly-won things. After obtaining the value vector of every topic, we apply a value matrix to gift the price vector of all the N topics. Cost matrix $\beta(M)$ is a $N \times 4$ matrix. Each row of the matrix is the value distribution for one topic. It is calculated by the typical of cost distribution for all the POIs during this topic. Similarly, we get the time matrix $\gamma(M)$ for all the topics. $\gamma(M)$ is also $N \times 4$ matrix.

**2. Community-Contributed Photos Mining**

In this section, we tend to introduce POIs mining, season attribute mining for every topic, and representative pictures mining for POIs from community-contributed photos.

**1. POI Mining**

So first we have a tendency to tend to introduce the thanks to mine POIs from packed geo-tagged photos. POIs mining could be a hot analysis house in recent years. First, we tend to filter a cluster of photos for each city from all the users. we tend to match city name, for instance, London, with the matter tags of every image. It can't guarantee that each one the photos matching city name definitely belong to this city, since community-contributed photos embody lots noises. we tend to further use the geo-location restriction. If the GPS coordinate of the picture is 500km (between region level and country level) off from the center of city, we tend to take away it. when getting a cluster of photos of each city, second, we tend to extract POIs from these packed geo-tagged photos toward each city by mean shift agglomeration. Then we elect the POIs in every the clusters and thus the attraction information processing system. Thus, these POIs have each GPS coordinates and travelogues description, that may guarantee the routes got wind of and routes package mining.

**2 Topic's Season Matrix Mining**

After getting POIs, to every dish, there square live a bunch of photos with tags and "date taken" labels. To season, we tend to use the "month" in "date taken" to induce the visiting distribution throughout the twelve month. The season vector of a dish is defined as $\zeta(P) \in$ [spring, summer, autumn, winter]. Months from March to could belong to spring and then on. consistent with the structure of travelogues, for every topic, we tend to average over all the season distributions of the POIs throughout this topic. The season matrix $\zeta(M)$ could be a $N \times 4$ matrix. every row of the matrix is that the season distribution for one topic.

**3 Representative footage Mining for dish**

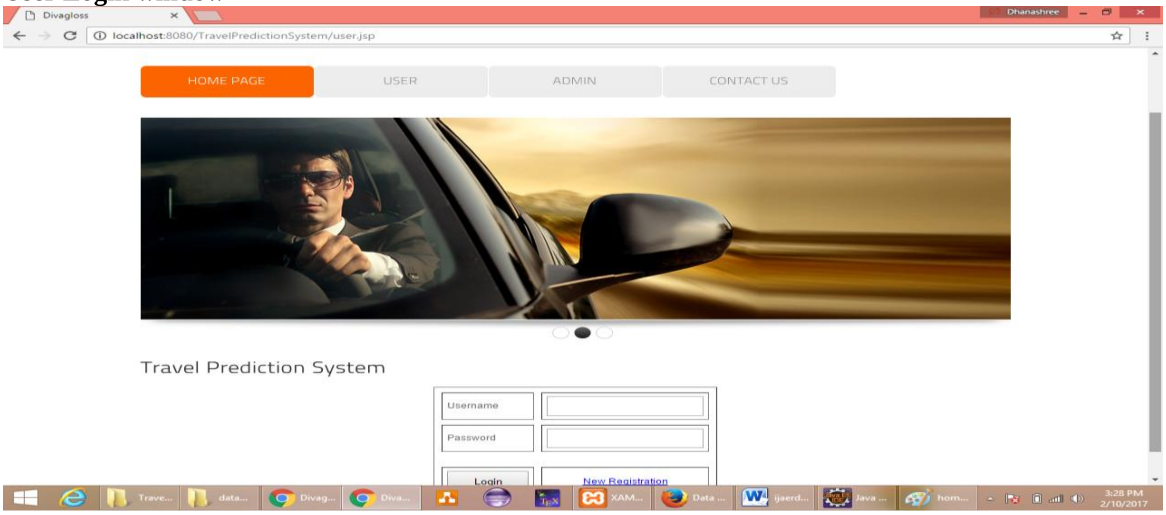
In order to produce vivid impression of the travel sequence, our system conjointly provides representative footage of the POIs on the route. we tend to ponder a pair of factors of the representative footage. First, we tend to gift representative viewpoints victimization the 4-D viewpoint vector model (i.e., horizontal, vertical, scale and orientation) . the varied

@IJAERD-2017, All rights Reserved

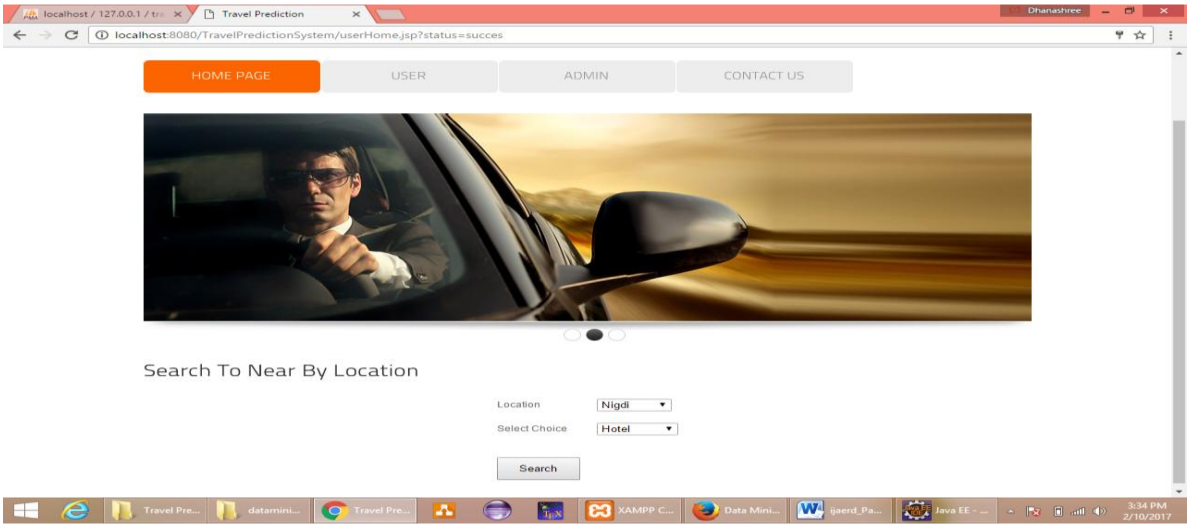
viewpoints may provide lots of comprehensive data of the dish. Second, as POIs could show quite all totally different characteristics in varied seasons, we provide representative footage of each season. to realize season diversity, we tend to extract the "date taken" information from info of the image, and divide the photos into four seasons.

## V. RESULT AND DISCUSSIONS

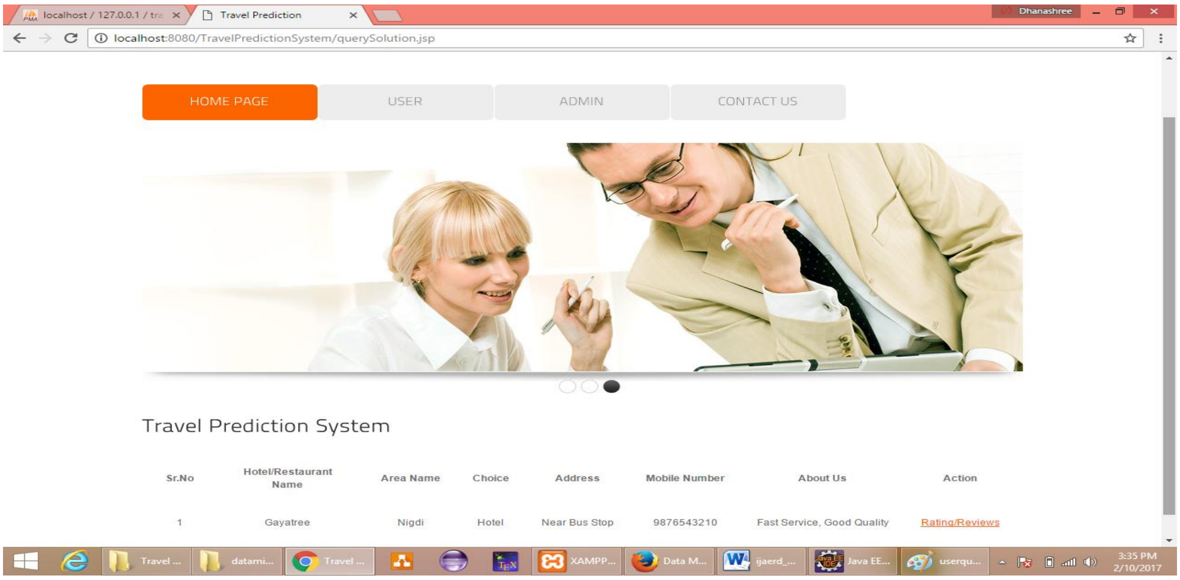
### 1. User Login window




### 2. Home window

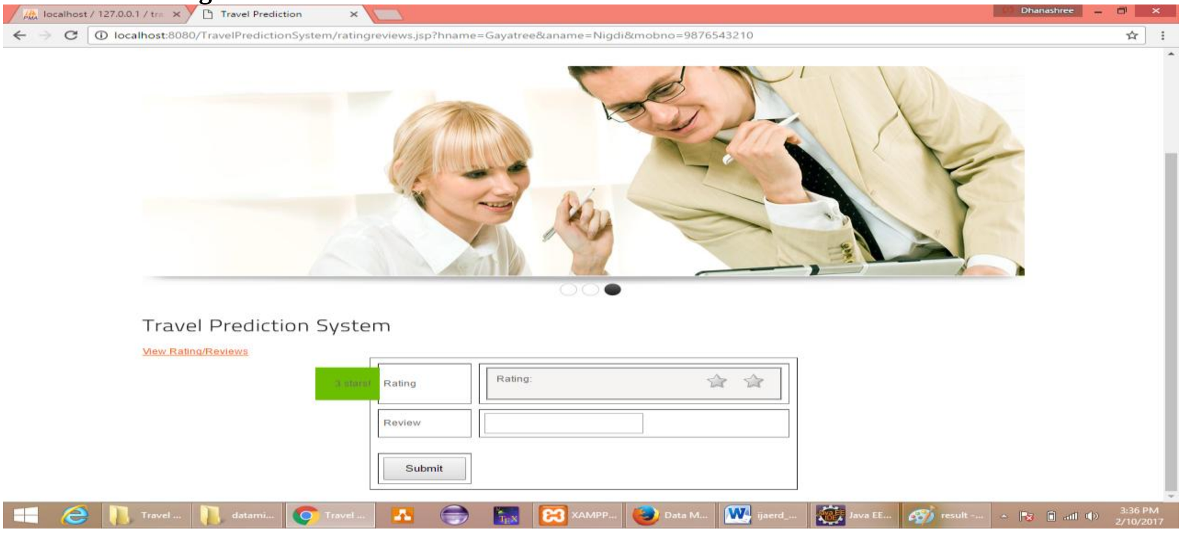



### 3. Prediction Window

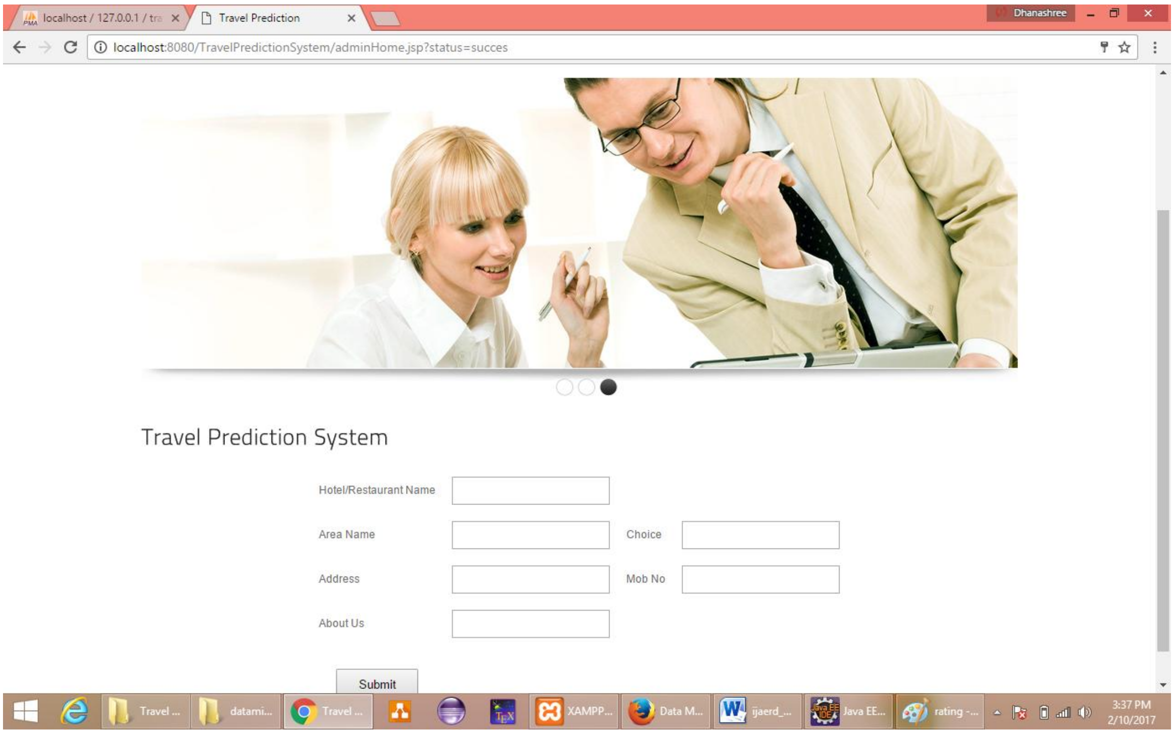



@IJAERD-2017, All rights Reserved 

**4. Review and rating Window:**




**5. Admin Home Window:**




## VI. . CONCLUSION

In this paper, we projected a customized travel sequence recommendation system by learning topical package model from massive multi-source social media: travelogues and community-contributed photos. The advantages of our work ar 1) the system mechanically deep-mined user's and routes' travel topical preferences as well as the topical interest, cost, time and season, 2) we suggested not solely POIs however conjointly travel sequence, considering both the quality and user's travel preferences at an equivalent time. We deep-mined and hierarchical famed routes based mostly on the similarity between user package and route package. And then optimized the highest ranked famed routes consistent with social similar users' travel records. However, there are still some limitations of the current system. Firstly, the visiting time of POI principally bestowed the open time through travelogues, and it was hard to urge additional precise distributions of visiting time solely through travelogues. Secondly, the current system only targeted on dish sequence recommendation and failed to embrace transportation and building data, which could more give convenience for travel coming up with. In the future, we set up to enlarge the dataset, and thus we have a tendency to might do the recommendation for a few non-famous cities. We set up to utilize additional types of social media (e.g., check-in knowledge, transportation data, weather forecast etc.) to provide additional precise distributions of visiting time of POIs and therefore the context aware recommendation.

@IJAERD-2017, All rights Reserved 

## ACKNOWLEDGMENT

We might want to thank the analysts and also distributers for making their assets accessible. We additionally appreciative to commentator for their significant recommendations furthermore thank the school powers for giving the obliged base and backing.

## REFRENCES

[1] H. Liu, T. Mei, J. Luo, H. Li, and S. Li, “Finding perfect rendezvous on 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 ACM SIGSPATIAL International Conference on Advances in Geographic
[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] Analytics: An Intelligent Approach in Clinical Trail Management Ankit Lodha* Analytics Operations Lead, Amgen, Thousand Oaks, California, USA.
[12] Agile: Open Innovation to Revolutionize Pharmaceutical Strategy Ankit Lodha University of Redlands, 333 N Glenoaks Blvd #630, Burbank, CA 91502.
[13] Clinical Analytics – Transforming Clinical Development through Big Data Ankit Lodha University of Redlands, 333 N Glenoaks Blvd #630, Burbank, CA 91502.

@IJAERD-2017, All rights Reserved 