소셜 미디어 상에서 개인화된 여행 경로 추천 기법
Personalized Travel Path Recommendation Scheme on Social Media

Paul Aniruddha, 임종태, 복경수, 유재수
충북대학교 정보통신공학과

Paul Aniruddha(libra16101983@gmail.com), Jongtae Lim(jtlim@chungbuk.ac.kr), Kyoungsoo Bok(ksbok@chungbuk.ac.kr), Jaesoo Yoo(yjs@chungbuk.ac.kr)

요약
소셜 미디어 환경에서 여행과 커뮤니티에서 기고한 사진과 관련된 메타 데이터 (태그, 지리적 위치 및 찍은 날짜)에 기반한 개인화 된 여행 경로 추천 기법이 연구되고 있다. 사용자는 소셜 미디어를 사용하고 자신의 위치 기록을 여행 경로의 형태로 기록한다. 이러한 여행 경로 정보는 미래의 여행자들에게 새로운 추천 정보를 제공하기 위한 유용한 정보로 활용될 수 있다. 본 논문에서는 라이프 로그를 기반으로 한 개인화 된 여행 경로 추천 기법을 제안한다. 제안하는 기법은 여행자 및 지역 사회가 제공한 라이프 로그 및 사진 정보를 활용하여 사용자에게 개인화된 추천 서비스를 제공할 수 있을 뿐만 아니라 개별 관심 장소가 아닌 대중적인 여행 경로도 추천 할 수 있다 (POI). 제안하는 개인화된 여행 경로 추천 기법은 POI 가치기 단계와 여행 경로 생성 단계로 구성된다. POI 가치기 단계에서는 POI 전체 데이터로부터 사용자에게 추천할 경로를 생성하는데 필요한 POI만을 남기고 가치기를 수행한다. 여행 경로 생성 단계에서는 POI 가치기 단계를 통해 도출된 POI 사용자 관심도, 비용, 시간, 이벤트 등을 고려하여 후보 경로를 생성한다.

■ 중심어 : 여행추천 경로 | POI | 라이프 로그 | 소셜미디어 | 추천 서비스 |

Abstract
In the recent times, a personalized travel path recommendation based on both travelogues and community contributed photos and the heterogeneous meta-data (tags, geographical locations, and date taken) which are associated with photos have been studied. The travellers using social media leave their location history, in the form of paths. These paths can be bridged for acquiring information, required, for future recommendation, for the future travellers, who are new to that location, providing all sort of information. In this paper, we propose a personalized travel path recommendation scheme, based on social life log. By taking advantage, of two kinds of social media, such as travelogue and community contributed photos, the proposed scheme, can not only be personalized to user's travel interest, but also be able to recommend, a travel path rather than individual Points of Interest (POIs). The proposed personalized travel route recommendation method consists of two steps, which are: pruning POI pruning step and creating travel path step. In the POI pruning step, candidate paths are created by the POI derived. In the creating travel path step, the proposed scheme creates the paths considering the user’s interest, cost, time, season of the topic for more meaningful recommendation.

■ keyword : Travel Path Recommendation | POI | Life Log | Social Media | Recommendation Service |
1. Introduction

Recently, travel path recommendations have been studied[1-21]. Whenever tourists visit a tourist destination, there is always a necessity, regarding how the sightseeing can be efficient and what the best local dish with budget compliance depending on every season is. It can be during the festive season in other words travel season or can be during the off-season. Generally, in India, tourism is very much likely to be dependent on the season due to a mixture of tropical and alpine climate in one nation[4].

Typically, people would desire to know which locations are the most interesting places in a geo-spatial region. In this paper, interesting location is the culturally important places such as Victoria Memorial overlooking the mighty Ganges River in the city of Kolkata (formerly known as Calcutta), India or commonly frequented public areas such as shopping malls/streets, restaurants, movie theatres, coffeehouses etc. Further, given these interesting locations in a geo-spatial region like a city, users might also wonder what the most classical travel sequences are among them. For example, an individual would be more likely to go to a coffeehouses for resting out after visiting a cultural landmark than they would before, making landmark-to-coffeehouses a classical travel sequence. With the information mentioned above, an individual can understand an unfamiliar city in a very short period and plan their journeys with minimal effort. Meanwhile, such information would enable mobile guides given the recommendation of the interesting places and travel sequences around them, mobile users are more likely to enjoy a high quality travel experience while saving lots of time for location finding and trip planning.

The recommended POIs should be personalized to user interest since different users may prefer different types of POIs[1-3]. POI is a place in which users are interested such as a restaurant, a tourist destination, and so on. some people may prefer cultural places like the Victoria Memorial, Martyr’s Tower or Shaheed Minar in the local language, Indian National Museum (situated in the heart of the city), while others may prefer the cityscape like the South City (situated in the southern part of Kolkata) or Eden Gardens and Park Street (situated near the heart of Kolkata). Besides travel topical interest, other attributes including consumption capability (i.e., luxury, economy), preferred visiting season (i.e., spring, autumn, winter which are the festive seasons in India) and preferred visiting time (i.e., morning, evening, afternoon, night).

It is important to recommend a sequential travel path (i.e., a sequence of POIs) rather than individual POI[4][14]. It is more difficult that users plan travel sequence than choosing individual POIs because users need to consider the relationship between the locations, and opening time of POIs. For example, it may still not be a good recommendation if all the POIs recommended for one day are in four corners of the city, even though the user may be interested in all the individual POIs.

Therefore, how can a tourist who is new to an Indian city travel without having any problem? Nowadays tourists mostly use social media especially during their travel and thus leave their location history in the form of paths. These paths bridge the gap of information between the travelers and locations. Currently, social media applications like Instagram, QQ, Expedia, Foursquare, Yelp, Zagat, Facebook, Flick, Zomato, and MySpace are offering great opportunities to address many challenging problems, like GPS estimation, travel recommendation. Furthermore, the photos contributed from the communities with metadata (e.g. longitude
position and the latitude position, the date took etc.) on social media record user’s daily and travel experiences. These data are not only useful for reliable (points of interest) POIs travel paths but they also provide an opportunity to recommend personalized travel POIs and paths based on user’s interest. To recommend the personalized POI path, first, famous paths surrounding the landmark areas or top rated POIs are ranked, by considering the user preference and path package. However, the challenges are as follows. There are thousands of POI’s and at which time that POI is best to visit and how to match with the visitor’s time and not to forget the price.

In this paper, we propose a personalized travel path recommendation system on social media. In recent years, new information has been constantly created and existing information changes through social and citation networks and the Internet of Things (IoT). There are challenge in any recommendation application. It is “what time is the best to recommend a particular POI by considering the factor of price”. After considering time, the challenge is “when the tourists visits the city, which path is the best to recommend considering timing and rating”. Here the time zones are divided in to mainly three parts, mainly in Morning, Afternoon and Evening and sometimes night as well making it 4 sections, incase a tourist arrives at night. For example a POI known as X has free entry so it can fit into any POI category in terms of price, since it is free. And let us say the X has a high rating in the morning recommendation compared to the other timings. Therefore based on the timing the tourist stays, and the timing is the morning the path will include that POI.

II. Related Works

The existing recommendation systems currently used as an application are introduced in this section. Many studies on techniques that store on recommendation systems have been conducted in recent years. Most of applications are widely used as they give out path recommendation in India, UK, Brazil, Australia, Germany, US, Spain, Portugal, Bangladesh, Sri Lanka, Canada and Japan.

1. FOURSQUARE

One of the top widely used recommendation system in the form of an application. This application is free to Subscribe and gives out the information without even logging into site. It provides the directions and easy to search. These applications always links the homepage of restaurant or hotels by providing the contact details. This application provides ratings and feedbacks of POIs. However, this application Shows no timings of Opening and Closing of POIs. The directions are not embedded but rather the directions are linked with the Google maps only and the directions are given out in a pop-up window rather than given out in a new tab. And the path is not personalized rather linked on Google map, and it does not give out the which path is best according to a time. like for an example a POI in a particular city which is closed at night is also recommended for a tourist who visits at night in that very particular city. Moreover, this service does not cater to any preference. For example if no alcohol preference clicked for an Indian city, then also it will show the bars and pubs within the recommended path. Price is not considered as part of recommendation.

2. YELP

One of the top widely used recommendation system in the form of an application mainly for hospitals, or clinics, dentists, pharmacy store and even laundry service as well required by a tourist. This application
is free to Subscribe and gives out the information without even logging into site. The directions are embedded and they give out the directions based on the preference. It takes out the tourist location through GPS and gives nearby POIs as a bouquet of recommended locations. For more information one has to get logged in for more information like nearby POIs. This application Shows no timings of Opening and Closing of POIs. Therefore just like FOURSQUARE does not give out best path according to a time. Moreover this service does not accept any preference. Price is not considered as part of recommendation.

3. ZAGAT

One of the top widely used recommendation system mainly in USA, India, China and Japan. This application is a paid service and provides detailed information about the closing and opening time of POIs and provides a good path based in the embedded directional map based on the timings and provides a detailed ratings. It provides rating and feedback very accurately. However, this application gives recommendation according to ratings and gives out the information which are among top 15 but unlike other services, this one does not cater to user’s preference. For contact information of POIs, they provide physical format of directory books rather than providing contact details online and the price is not considered.

4. ZOMATO

One of the most widely used recommendation system mainly in South Asia, South East Asia, Europe, Middle East, Oceania and Americas for dining. It is free to subscribe provides rating and feedback very accurately based on the user’s real time location. The directions are provided in the embedded map within the service. Especially if any POIs area having special event due to festive season, the information is updated especially the discounts or price reductions offers. Also due to weather or local political based strikes, the POIs which are difficult to reach are also updated in the service. It provides the opening and closing time of POIs with live updates. This service also provides contact details of POIs and a feature to connect to POIs directly as well. It provides to exclusive service as a paid service as a social drinking membership, where the tourists can meet the locals and dine together. Zomato's data can be utilized to merge into one application. This application is confined to information about restaurants only and does not provide a good path.

5. EXPEDIA

One of the most widely used recommendation system used for travelling in world-wide for hotels and flight bookings. Just like Zomato it is free to subscribe provides rating and feedback very accurately based on the user’s real time location. The directions are provided in the embedded map within the service. Especially if any POIs area having special event due to festive season, the information is updated especially the discounts or price reductions offers. Also due to weather or strike, the POIs are difficult to reach are also updated in the service. It provides the opening and closing time of POIs with live updates. This service also provides contact details of POIs and a feature to connect to POIs directly as well. This application is confined to information about hotels and flights only and does not provide a good path.

However, existing schemes haven’t well solved the two challenges. For the first challenge, most of the travel recommendation works only focused on user topical interest mining without considering other
attributes like consumption capability. For the second challenge, existing schemes focused more on famous path mining without automatically mining user travel interest. It still remains a challenge for most existing works to provide both “personalized” and “sequential” travel package recommendation. In addition, the features and problems of each existing scheme are explained.

III. Proposed Personalized Travel Path Recommendation Scheme

In this paper, we propose a personalized travel path recommendation scheme, based on social life log. Especially, the proposed scheme considers the rating according to the time of POIs. So, the proposed scheme improves the accuracy of travel path.

1. Characteristics

[Fig. 1] shows the overall structure of the proposed system. We can get the geographical location data of users as well from the social network services Instagram and Expedia, QQ. And the location rating from the Zomato and Foursquare including the traffic as well budget restaurants. To collect the events during festive season such as Holiday Season or the period from Christmas Eve to New Years, we can be retrieved from Zomato. Incase of QQ, we can get the ratings of hour wise to certain level of segregation, especially as popular rating during daytime hours and non-daytime hours around which part of day a POI is more popular and has more higher rating.

The proposed system collects and stores the POI data from social media. In order to perform path recommendation, the proposed system consists of mainly two phases. Firstly, POI data set is acquired in the format of location, type, ratings, hours, price and events. So, in the Pruning POI phase, the proposed system prunes the POI data by location, hour, price, and rating of POI. Second, in the Making path phase, the proposed system makes all cases of paths (candidates) and processes the top-k query with candidates. In a top-k query, the proposed system considers hour, price, and rating of POIs.

2. Pruning of POIs

In here we actually prune out the POIs. [Fig. 2] shows how to prune of POIs. For example, we have around 20 POIs, which are: A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, and T. Therefore, we start pruning by considering the factors of hour and price. The reason is that we need to match the time, when a tourist is visiting and whether the opening and closing time matches with the period the tourist is present, which is the utmost importance followed by the budget that is why price is considered as well with the cheapest ones to the dearer ones as per the tourist’s wishes. By considering the factor of user or here, tourist’s preferences, we can prune 20 POIs to 14 POIs. Then these 14 pruned POIs are further pruned down by the factor of ratings given by the previous users who have already visited the POIs to retrieve the final pruned POIs such as 8 POIs, which comes down to B, F, J, K, L, Q, S, and T.
3. Pruning of Paths

Now the paths of the 8 POIs through the pruning process of POIs are therefore considered for retrieving the best top k paths. By considering the 8 POIs following paths can be obtained, like $F \rightarrow B \rightarrow K \rightarrow J \rightarrow L \rightarrow Q \rightarrow T \rightarrow S$, then $B \rightarrow J \rightarrow F \rightarrow K \rightarrow Q \rightarrow L \rightarrow S \rightarrow T$ ..... path 1, $T \rightarrow F \rightarrow Q \rightarrow L \rightarrow K \rightarrow S \rightarrow J \rightarrow B$ ..... path 2, $Q \rightarrow S \rightarrow L \rightarrow F \rightarrow T \rightarrow J \rightarrow B \rightarrow K$ ..... path 3 and many more can be obtained. Here, in our case, since 20 POIs are pruned down to 8 POIs, therefore $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 40,320$ paths can be formed. Therefore, the challenge is to obtain the best top k paths and recommending to the user is the main aim. [Fig. 3] shows the path paths obtained from POIs.

So how can we obtain the top k path out of 40,320 path paths. Therefore, just like POIs, the paths also need to be pruned down in order to obtain the top k paths. We have considered omega $\Omega$ as the weight of every POIs for every particular path. Similar to the POIs, we would prune the 40320 path paths to top k path paths in the factors of price, rating and finally for hours. So among the 40,320 paths let us consider the first path $F \rightarrow B \rightarrow K \rightarrow J \rightarrow L \rightarrow Q \rightarrow T \rightarrow S$

We need to find a score for every path in order to compare with other paths to find the top k path. So, we need to get three scores without which path scoring cannot take place. The categories of scores are as follows:- Price, Rating and Hour.

We need to find a factor for every path in order to calculate the score of every categories. Let us consider $\Omega$ as weight of the paths. Therefore, we divide the $\Omega$ into $\Omega_1$, $\Omega_2$, $\Omega_3$ for the following factors, which goes by $\Omega_1$ for price, $\Omega_2$ for rating & $\Omega_3$ for hour respectively for a particular path. We have initialized the weights for easier calculation. To reduce the candidates, we calculate the factor which is calculated easily first. We obtain chi $\chi$ as weight per price for a POI of a particular path by multiplying weight $\Omega_1$ with price score, where Price is the cost of every POI for budget friendly trip for tourist. The Price Score is calculated by (2). In formula 2, $Price(pi)$ means the cost such as admission fees when you visit $pi$. $Price(Pi, Pi+1)$ means the cost that you spend moving from $Pi$ to $Pi+1$. $n$ means the number of POI.

$$\Omega_1 \times \text{Price Score} = \chi \quad (1)$$

$$\sum_{i=1}^{n} (\text{Price}(P_i) + \text{Price}(P_i, P_{i+1})) \quad (2)$$
Just like chi $\chi$, we obtain kappa $\varphi$ as weight per rating for a POI of a particular path by multiplying weight $\Omega_2$ with rating score. Rating Score is calculated by (4), where, rating score is the rating of every POI for in that path for a time of a day, which can day, night, or evening or afternoon.

$$\Omega_2 \times \text{Price Score} = \varphi$$  \hspace{1cm} (3)

$$\sum_{i=1}^{n} (\text{Rating}_{(P_i)})$$  \hspace{1cm} (4)

The Final category, we obtained is theta $\Theta$ as weight per rating for a POI of a particular path by multiplying weight $\Omega_3$ with rating score. The Hour Score is calculated by (6), where where hour score is the time taken among every POI for in a particular path, in order to make the path path very comfortable for a user. Say for example if we look at the considered path $F \rightarrow B \rightarrow K \rightarrow J \rightarrow L \rightarrow Q \rightarrow T \rightarrow S$, so, how long the time takes from F POI to reach B POI. In formula 6, Hour($p_i$) means the time when you visit $p_i$. Hour($P_i$, $P_{i+1}$) means the time that you spend moving from $P_i$ to $P_{i+1}$.

$$\Omega_3 \times \text{Hour Score} = \Theta$$  \hspace{1cm} (5)

$$\sum_{i=1}^{n} (\text{Hour}_{(P_i)} + \text{Hour}_{(P_i, P_{i+1})})$$  \hspace{1cm} (6)

We decided to have the path score as gamma $\Gamma$, where $\Gamma = \chi + \varphi + \Theta$. The path score can be divided into $\Gamma_1$, $\Gamma_2$, $\Gamma_3$, $\Gamma_4$, $\Gamma_5$, $\Gamma_6$, ...... and $\Gamma_n$. By obtaining $\chi$, $\varphi$ and $\Theta$ for every POI, in each path score gamma $\Gamma_1$, $\Gamma_2$, $\Gamma_3$, $\Gamma_4$, $\Gamma_5$, $\Gamma_6$, ...... and $\Gamma_n$ can be obtained. Now, these path scores are compared among each other to get the top k path. Finally, top k paths with top scores can be obtained and can be recommended.

IV. Performance Evaluation

The performance of the proposed scheme was verified by comparing the memory usage and query response times with those of existing schemes. Furthermore, threshold values were also evaluated to determine which threshold showed the best efficiency. The performance valuation environment is presented in [Table 1]. The experiment was conducted in an
environment of Intel Core i5-4440 CPU 3.10 GHz with 8 GB memory. The data used in the performance evaluation are presented in [Table 2].

Table 1. Environment of the Performance Evaluation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>Intel (R) Core (TM) i5 - 4440 CPU 3.10GHz</td>
</tr>
<tr>
<td>Memory</td>
<td>8 GB</td>
</tr>
<tr>
<td>Language</td>
<td>JAVA</td>
</tr>
<tr>
<td>Operating System</td>
<td>Microsoft Windows 7</td>
</tr>
</tbody>
</table>

Table 2. Data of the Performance Evaluation

<table>
<thead>
<tr>
<th>Item</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of POI</td>
<td>06 ~ 14</td>
</tr>
<tr>
<td>Number of POI in Path</td>
<td>03 ~ 06</td>
</tr>
<tr>
<td>Number of POI in Path</td>
<td>60 ~ 140</td>
</tr>
</tbody>
</table>

We considered the number of POIs or Point of Interest for around 06 ~ 14. And considering the number of POIs in a single path around 03 ~ 06. Out of the POIs and making the path paths, the top k path paths are around 60 ~ 140.

We have considered to compare the existing system with our proposed method. Therefore, firstly, we will consider the scores using the existing system.

We considered 10 POIs after the pruning of POIs therefore, static int num_of_POI = 10. While for every path paths 3 POIs would exist. Henceforth, static int num_of_POI_in_path = 3. So, that makes around 300 path paths. And out of 300 path paths we took 100 top k paths. Therefore, static int top k = 100

In the existing method, since the score of is almost similar, it is very difficult to differentiate the paths, especially when we want to segregate through the score of time because it has an average score. Let us consider the score for the path “p1, p5, p9” has a score of = 65.7341638281733. If we look at the next path “p1, p4, p5” has a score of = 65.6791058637173, which are almost more or less similar. Therefore by this score or time, generated from the existing system, the differences of path cannot be segregated. Therefore the comparison is almost impossible to perform, due to the fact of generation of the similar scores which average score for the factor of time.

[Fig. 7] shows the calculation of POI to generate the score for POI \( \rightarrow \) p1, p4, p5
In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to the query. In Precision, the graph shows a decline phase, because as more POIs are added the paths are supposed to increase, however, to narrow down to exact matching, the data the graph shows a decline phase. [Fig. 8] shows evaluating POIs in a path through precision.

In information retrieval, recall is the fraction of the relevant documents that are successfully retrieved. [Figure 9] shows evaluating POIs in a path through recall. Initially, we aimed for 3 POI, this time we like to see what happens if we add more POIs in a path, so we tried the result by 3 POIs, 4 POIs, 5 POIs and 6 POIs in a single path. Therefore number of paths will increase.

**V. Conclusion**

In this paper, we proposed our research about the personalized travel path recommendation. Every POI according to their geographical position has ratings. And these ratings are very important to provide the more effective recommendation. The existing system which provides recommendation of POIs is based on the average score, which does not gives a good fruitful recommendation. However our proposed scheme can provide time segregation for every POI and the highest score of a POI for a particular time, having the ratings will be formed within the path to provide the bouquets of top k paths to the user. The experiment results showed that the proposed scheme having the punning steps for POIs as well as for paths provides efficiency by up to 80% in respect to recommendation to existing schemes. However, it
cannot guarantee better performance for a long range travel for example, a tour for 5 days. But by this system the POIs of long term travel can benefit. In our future research, we will use this method for long term travel where the best POIs can be visited and gives a new edge in defining the recommendation. As a result, the tourism industry can benefit if the method is implied into their applications meant for recommending the tourists the POIs in the perfect time.

참고 문헌


저자 소개

Paul Aniruddha
준회원
• 2012년 2월 : Graduated Degree of Bachelors, in Computer Sciences from Sam Higginbottom Institute of Agriculture, Technology & Sciences (공학사)
• 2011년 10월 ~ 2015년 2월 : Worked as an Software Engineer in Nomura Research Institute, Financial Technologies
• 2015년 3월 ~ 2016년 2월 : 충북대학교 한국어 연수생
• 2016년 3월 ~ 현재 : 충북대학교 정보통신공학과 석사과정
<관심분야> : 빅데이터, 소설 미디어, 추천 정보경로, 소셜 네트워크

임 종태 (Jongtae Lim)
정회원
• 2009년 2월 : 충북대학교 정보통신공학과 (공학사)
• 2011년 2월 : 충북대학교 정보통신공학과 (공학석사)
• 2015년 8월 : 충북대학교 정보통신공학과 (공학박사)
• 2015년 9월 ~ 현재 : 충북대학교 정보통신공학과 Postdoc
<관심분야> : 소셜 네트워크, 빅데이터, 텍스트 마이닝
소셜 미디어 상에서 개인화된 여행 경로 추천 기법 295

복 경 수(Kyoungsoo Bok) 종신회원

- 2009년 2월 : 충북대학교 수학과 (이학사)
- 2000년 2월 : 충북대학교 정보통신공학과(공학석사)
- 2005년 8월 : 충북대학교 정보통신공학과(공학박사)
- 2005년 3월 ∼ 2008년 2월 : 한국과학기술원 정보전산공학부 초빙교수 Postdoc
- 2008년 3월 ∼ 2011년 2월 : 가인정보기술 연구소 연구원
- 2011년 3월 ∼ 현재 : 충북대학교 전자정보대학 정보통신공학부 초빙교수
  <관심분야> : 데이터베이스 시스템, 이동 객체 데이터베이스, 이동 P2P 네트워크, 소셜 네트워크 서비스, 빅데이터 등

유 재 수(Jaesoo Yoo) 종신회원

- 1989년 2월 : 전북대학교 컴퓨터공학과(공학사)
- 1991년 2월 : 한국과학기술원 전산공학과(공학석사)
- 1995년 2월 : 한국과학기술원 전산공학과(공학박사)
- 1995년 2월 ∼ 1996년 8월 : 목포대학교 전산통계학과 전임강사
- 1996년 8월 ∼ 현재 : 충북대학교 전자정보대학 정교수
  <관심분야> : 데이터베이스 시스템, 멀티미디어 데이터베이스, 센서 네트워크, 바이오인포매틱스, 빅데이터 등