

# Analysis of Travel Patterns of Seoul Tourists by Trajectory Data Mining<sup>a</sup>

Juyoon Lee<sup>b</sup>, Youngok Kang<sup>b\*</sup>, Nayeon Kim<sup>b</sup>, Dongeun Kim<sup>b</sup>, Yearim Park<sup>b</sup>

<sup>a</sup> This Research was supported by the Technology Advancement Research Program funded by Ministry of Land, Infrastructure and Transport of Korean government (Grant No: 19CTAP-C151886-01)

<sup>b</sup> Ewha Womans University, leedew12@gmail.com, ykang@ewha.ac.kr, rlaskdus993@naver.com, kan00100@naver.com, yeahrim023@naver.com

\* Corresponding author

**Keywords:** Social Network Service, Flickr, Trajectory Data Mining, Markov Chain, Apriori

## Abstract:

It is necessary to identify the preferences and characteristics of tourists for vitalizing the tourism industry, as the tourism industry is of the fast-growing industries in the economic sector. A trajectory of tourists, a movement of tourists over time, is a very valuable information since it shows tourism characteristics, such as the length of tourists' visit, tourists' preferred attraction, the time of arrival to specific tourist attractions, and the movements in between different tourists' destinations. Earlier studies regarding the movements of tourists were conducted by surveys, or by analyzing the data derived from GPS devices that were handed out to the research areas. However, these approaches using surveys or GPS devices are not only time consuming and requiring a lot of time for the analysis, but also difficult to detect the actual travel patterns. Recently, advances in mobile technologies and multimedia have allowed large amounts of user-generated data, such as travel photos, to be created and shared. This expands the use of such data on tourism industry since it enables to extract and analyze the trajectory of users by using geotagged data that were uploaded on Social Network Services (SNS)(Vu *et al.* 2015; Zheng *et al.* 2012).

The purpose of this research is to analyze the movement pattern of tourists who visited Seoul by using Flickr data, which is one type of SNS. For this, we went through the following process. First, we collected the Flickr data and preprocessed. Second, we separated the visitors and residents from the collected data. Third, we selected the main tourist attractions. Fourth, we produced trajectory data targeting tourists. Lastly, we analyzed the movement characteristics of tourists using trajectories.

First, by using Flickr API, we gathered the data which encompasses the parameter of Seoul consisting of latitude of 37.4°~37.8°, and longitude of 126.8°~127° from January 1, 2015 to December 31, 2017. The collected Flickr data includes metadata (user ID, information of users' residence, the time of the photo taken, the longitude and latitude, the url and titles of the photos, and tagging information). The data collected for this period were 86,304 and 1,974 users in total. Second, we distinguished the tourists from residents. The users who specified their residence information were about 44 percent which sums up to 868 people. From these users, we excluded those who self-identified their residence area to Seoul or South Korea and distinguished 689 people as tourists. In order to distinguish tourists from the users who did not write their residence information, we developed a classification model using the data (the data of one's visiting period—the gap between the initial and the last photo that users uploaded) from the users who provided their residence information. We selected the threshold of 30 days which had the highest prediction accuracy of tourists, and identified 787 people out of 1,106 as tourists. As a result, 1,476 people were identified as tourists from 1,974 users. Third, we extracted the Regions of Attraction (RoA) from the 39,157 geotagged Flickr data uploaded by 1,476 tourist users. In order to extract RoA, we used DBSCAN (Density Based Spatial Clustering of Application with Noise) algorithm. We extracted the cluster based on the density of data, then defined each clusters' Convex Hull as RoA. 11 RoA in total were extracted, which included Jongno·Namsan, Shinchon·Hongdae, War Memorial of Korea, Samsung Station·Bongeunsa Temple·Cœx, Jamsil, Itaewon, Yeouido, Garosu-gil Road, Apgujeong, etc. Fourth, we constructed trajectory data of 1,476 people who were classified as tourists in order to analyze the movement characteristics of tourists. The spatio-temporal characteristics of geotagged data, which were uploaded by tourists, were defined as the visits of according locations. Then a series of data over time was defined as a trajectory. Since the constructed trajectories include all the geotagged data of photos that were taken in Seoul from 2015 to 2017, there is a possibility that the visiting period of re-visitors will be overestimated and mislead the results when transition probability calculation is used. In order to prevent this, if the time gap between data [i] and data [i+1] within trajectory is more than 30 days, we considered it as a re-visit and separated it as a visit trajectory. We extracted 1,747 visit trajectories from the trajectories of 1,476 people. Finally, we analyzed the movement pattern of Seoul tourists by using trajectory data of each users. For this, we first defined the number and length of visit trajectory as the number of visits and visiting time of Seoul tourists then analyzed the visiting characteristics of Seoul. We analyzed the visiting characteristics and movement patterns of RoA by applying a Markov chain and apriori algorithm to RoA visits within the visit trajectories. A Markov chain is a stochastic model describing a

sequence of possible events in which the probability of each event depends only on the state attained in the previous event. It is assumed that RoA of  $[n+1]$  visit depends entirely on RoA of  $[n]$  visit over time. With this assumption, we derived the probabilities of visits among each RoA. By applying apriori algorithm we extracted the pattern of concurrent appearance RoA within the visit trajectories and derived the tourists' visiting rules. We confirmed the significance of the rules by having the support higher than or equal to 0.03, confidence higher than or equal to 0.3, and lift higher than or equal to 1.

Accordingly, the analyzation of movement characteristics of the tourists who visit Seoul are as follows. The tourists stayed 5.13 nights on average on their first visit and re-visited Seoul about 1.27 times. The first tourist attraction they visited in Seoul was mainly Jongno and Namsan(68%), followed by Shinchon and Hongdae(8%), and Itawon(6%). By applying Markov chain, we studied the probabilities of sequential movements among RoA. As a result, the probability of moving to Jongno·Namsan from every RoA (except Samsung Station·Bongeunsa Temple·Coex and Apgujeong) came out to be more than or equal to 50%. Also, it is found that the movement probabilities were higher among adjacent RoA. We extracted eight visiting rules by using apriori algorithm. Tourists tended to visit Garosu-gil Road and Jongno·Namsan together, and Jamsil and Jongno·Namsan together. Also, showed the rules of concurrence appearance among Itaewon, Jongno·Namsan, and Shinchon·Hongdae RoA.

This study has its significance for defining the movement trajectories through geotagged SNS data and analyzing the tourism characteristic through existing social research. However, it has its limitations by having big differences in sizes and numbers among RoA and having the majority of data focused on Jongno·Namsan RoA. It also has its limitation that this study did not handle the movement within each RoA. It is anticipated to have more meaningful results if the study analyzes the characteristic of movement trajectories with more various variables, such as weather, dates, time, and tourist attractions, by extending the temporal and spatial range.

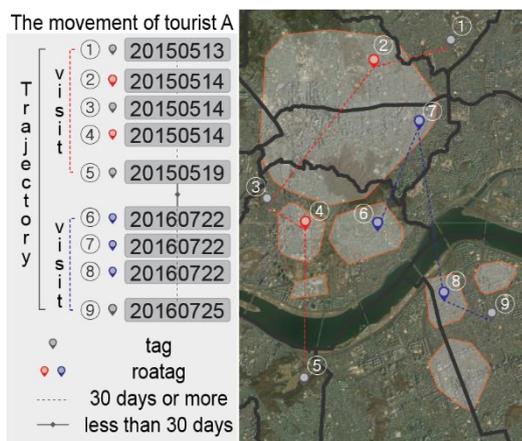


Figure 1 Example of tourist's trajectory.

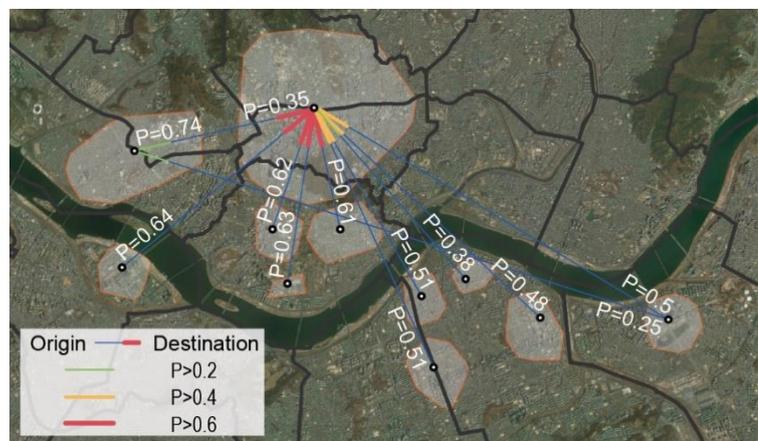


Figure 2 The probability of movement( $p>0.2$ )

#### Reference

- Vu, H. Q., Li, G., Law, R. and Ye, B. H., 2015, "Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos". *Tourism Management*, 46, 222~232.
- Zheng, Y. T., Zha, Z. J. and Chua, T. S. 2012, "Mining travel patterns from geotagged photos". *Transactions on Intelligent Systems and Technology*, 3(3), 56-73