

# A Study on Deriving an Optimal Route for Foreign Tourists through the Analysis of Big Data

Seong-Taek Park<sup>1</sup>, Young-Ki Kim<sup>2\*</sup>

<sup>1</sup>Professor, Department of Bigdata, KASOM

<sup>2</sup>Professor, Dept. of Management Information Systems, Chungbuk National University

## 빅데이터 분석을 통한 외국인 관광객을 위한 최적 경로 도출

박성택<sup>1</sup>, 김영기<sup>2\*</sup>

<sup>1</sup>한국소프트웨어기술인협회 빅데이터 교수, <sup>2</sup>충북대학교 경영정보학과 교수

**Abstract** The purpose of this paper is to derive an optimal route for foreign tourists in Korea. To that end, the data gained from domestic tourist portal sites was analyzed with a big data analytics tool R. The destinations most visited by inbound foreign tourists, the shortest route and the most economical route were derived from the analysis results. The findings suggest original Korean culture is the factor for successful tourist destinations and relevant products, and will serve as some reference data conducive to planning the tourist products in practice.

**Key Words** : Bigdata, Foreign, Tourist, Optimal route, R, Bigdata analysis

**요약** 본 연구는 외국인의 국내 관광의 최적의 경로 도출하는데 그 목적이 있다. 이를 위해 국내관광 관련 포털 사이트 데이터를 참고하여 빅데이터 분석 툴인 R을 활용하여 분석을 시도하였다. 분석결과 국내 방문 외국인의 가장 방문 빈도수가 높은 곳과 최단시간 경로, 최저 비용의 경로를 도출하였다. 이를 통해 관광상품의 성공요인은 한국만의 특유한 문화가 보이는 곳으로 상품을 구성해야 한다는 결론을 도출하게 되었다. 본 연구의 결과는 국내 관광 상품을 기획하는 단계에서 실무적인 지침으로 활용이 가능할 것으로 보인다.

**주제어** : 빅데이터, 외국인, 여행, 최적 경로, R, 빅데이터 분석

### 1. Introduction

Tourism contributes to the country's economic growth in addition to semiconductor and car exports, as evidenced by the ever-increasing number of inbound tourists, with as many as 17.24 million foreigners visiting Korea in 2016. That said, Korea's tourism industry is in its infancy compared with that of the advanced countries including Europe. With Korea's national competitiveness improving, the number of inbound tourists has

been increasing every year. Therefore, it is necessary to recommend some optimal routes and destinations to foreign tourists.

Quite a few studies and reports have proposed a range of ideas for boosting tourism in Korea. In the same vein, such events as Korea Big Sale parallel to Black Friday have been planned and held to attract foreign tourists, but not been well-received as much as expected.

The sensational Korean Wave has engendered

\*Corresponding Author : Young-Ki Kim(ykkim@cbnu.ac.kr)

Received August 29, 2019

Accepted October 20, 2019

Revised September 27, 2019

Published October 28, 2019

culture evangelists in many fields, arousing much interest in Korea around the world. Yet, the lack of information about tourist destinations or recommended attractions in Korea, price gouging by some transport, lodging and catering businesses, and inconvenient transport services need be addressed. Undeniably, extensive efforts are exerted to attract tourists, which still leave much to be desired for foreigners.

Currently, investment in tourism industry is concentrated on widely known tourist attractions such as Myungdong, Itaewon and Gangnam, whereas the War Memorial Museum of Korea was ranked top in a survey of foreign tourists' hot spots in 2016, which suggests tourist needs should be identified [1].

It is certainly necessary to introduce diverse cultural traditions, things to enjoy and foods. However, it is crucial to boost tourism by meeting the needs of foreign tourists. As shown on a TV show, 'Welcome, First Time in Korea?', foreign tourists get information on SNS, blogs and Twitter or rely on the recommendations from those who have travelled to or lived in Korea [2].

Importantly, as people have different tastes, it is essential to develop measures that meet the taste of foreigners. For example, suppose you go on a trip to 'Sapporo'. You will search information about the destination on Google and Naver. Not only individuals but also tourist agents post relevant information. Also, there are many other online tourist information sources such as those offered by Sapporo City Council, Hokkaido Provincial Office and the Incorporated Association of Sapporo Tourism. You build your own itinerary based on the information searched on the internet and visit the tourist attractions in person.

Meanwhile, foreign tourists are likely to find it difficult to get the information about Korea. As Korea's tourism industry is still comparatively in its infancy, it is necessary to develop some approaches

to attract tourists not from certain countries but from around the world.

Hence, this paper analyzes Chinese tourists, who outnumber those from the other countries, as the first step to formulate some measures to attract foreign tourists from around the world. The present analysis is intended to propose some attractions for Chinese tourists. Indeed, the proposed attractions will help Chinese tourists to build their own itineraries in Korea.

## 2. Theoretical Background

### 2.1 Tourist Trend

The number of foreign tourists reached record 17.24 million in 2016, increasing by approximately 11 million compared to 6.15 million in 2006. The figure increased to 13.35 million and 15.46 million in 2017 and 2018, respectively. The uptrend in inbound tourist population is projected to continue, coupled with the globally sensational Korean Wave and the country's successful improvement of its national competitiveness [3].

Chinese tourists outnumber those from all the other countries. Yet, in 2017, the number of Chinese tourists decreased due to the controversial THAAD deployment in Korea. The consequence of the controversies over the THAAD missile shield system still lingers in 2019. In that respect, it is crucial to diversity the country's tourism policy toward attracting tourists from around the world instead of those from a few countries. At the same time, identifying tourist needs is a requisite to minimize the current crisis of the country's tourism industry.

### 2.2 Bigdata

Big data collection consists of three steps, i.e. target selection, planning and data gathering. In the target selection step, the availability and usability of the target data, the inclusion of specific items

meeting the purpose of such data, the infiltration of privacy or personal information and the cost for data collection need be considered first [4, 5].

Crawling engines, or crawlers, are largely used to collect external data. The crawlers, or (software) robots, visit internet sites and generate copies of all pages to collect data. Crawlers are applicable to collecting huge volumes of data relevant to various activities of internet users including SNS, UCC (User Created Contents), shopping and searching.

### 2.3 Analysis Method and Content

Based on tourism-related reports, arrivals and stays of foreign tourists, data about foreign tourists by country and the 2016 survey data about foreigners available from the 'Tourism Knowledge Information System,' this paper identifies the existing tourist destinations and attractions and proposes new destinations and attractions based on the crawled data on Twitter and TripAdvisor (international tourist information site), with intent

to help foreign tourists determine their destinations on arrival at the airport and set up the most efficient route [5-7].

The present analysis is limited to Seoul for the city's well-designed public transport system, multiple cultural assets, and accommodations as well as the ready availability of tourist information.

## 3. Research Methods and Procedure

By gathering the opinions of foreign tourists concerning the destinations in Korea by means of crawling, this paper derives an optimal tourist route. Also, following the planning of the methods of using and analyzing information and data needed to derive the optimal tourist routes, this paper elicits the results from the analysis and visualization using the open source tool R [8-11]. The research procedure is shown in Fig. 1.

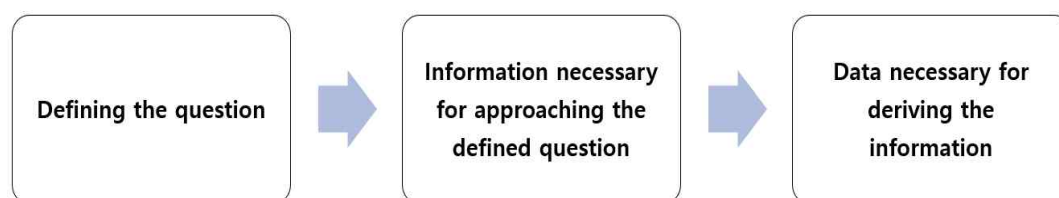


Fig. 1. Research Methods and Procedure

### 3.1 Defining the question

At present, undifferentiated programs and imitative products are prevalent throughout the tourism industry in Korea. Also, the industry is inclined to quantitative, not qualitative, development. In addition, the absence of local originality and cultural traits has caused the redundant products and thus the growing dissatisfaction of tourists. Hence, this paper defines the poor performance in contrast to tourist arrivals and the low tourist satisfaction as the research question to propose a 'successful tourist route' [12-13].

### 3.2 Information necessary for approaching the defined question

The information about the foreign tourist visits to the country's tourist destinations and the tourist satisfaction is necessary. This paper refers to TripAdvisor's 2016 survey data about tourist satisfaction to analyze the relevance to the tourist visits per destination. Ultimately, the analysis was intended to derive some success factors and propose a desirable model for a successful product, as well as to derive failure factors and propose some ideas for improvement.

### 3.3 Data necessary for deriving the information

First, this paper visualizes the information to put into perspective the foreign tourists visiting the tourist destinations in Korea. Second, it analyzes the most and second-most visited destinations to illuminate the association. Finally, it examines real-time responses of foreigners via the data crawled on Twitter and TripAdvisor.

## 4. Data Collect and Analysis

### 4.1 Data Collect and pre-processing

Inbound tourist data up to 2016 is as follows [8]. 17,241,823 foreigners visited Korea for sightseeing, with \$17,087,600 earned from tourism as shown in Table 1.

Foreigners visiting Korea in 2016 were from 67 countries excluding unknown ones and overseas Koreans. Therefore, this paper classifies the nationalities into Asian states, Middle Eastern states, the US states and others. The average stays were 6.7 days (Asian states), 18.5 days (Middle Eastern states), 13.4 days (the US states) and 11.4 days (others) [8].

**Table 1. Tourists data**

	number of people
Jan-16	1,077,431
Feb-16	1,126,250
Mar-16	1,389,399
Apr-16	1,469,674
May-16	1,492,680
Jun-16	1,554,413
Jul-16	1,703,495
Aug-16	1,664,303
Sep-16	1,523,928
Oct-16	1,587,797
Nov-16	1,309,055
Dec-16	1,343,395

### 4.2 Analyzed subjects and period

The subjects and period analyzed were determined based on the gathered data. Then, as

the elementary statistics data, percentages of inbound arrivals, monthly arrivals from top-ranked source countries, and monthly tourism revenues were derived to determine the major subjects and period (Given the overwhelmingly larger visitor population, Chinese and Japanese visitors are included as the subjects for analysis).

Thus, this paper defines the foreign tourists who arrived in Korea from Asian states from June 6th to August and stayed for 6.7 days on average as the subject of analysis. The data was mainly collected on the 'Tourism Knowledge Information System.

### 4.3 Association and time series analyses

Based on the 2016 foreigner survey data offered on the 'Tourism Knowledge Information System', the association was analyzed. The purpose of the present analysis was to derive the most-visited destinations. The survey data was derived from the percentages of inbound arrivals, 1,000 visitors yearly and 12,000 respondents in total. The details of the survey data are shown in Table 1 below as shown in Table 2.

To analyze the association[14-15], the response data about the question 'Write the numbers of the destinations you have visited' was extracted, so as to derive related destinations (\*Lift  $\geq 1.5$ ).

**Table 2. Tourists data analysis results**

		BB	DO	GG
1	Jan-16	6.2483939	7.6131230	1207.409
2	Feb-16	6.5320820	8.1246050	1232.675
3	Mar-16	8.0583070	8.7993630	1082.195
4	Apr-16	8.5238900	8.0532080	936.3301
5	May-16	8.6573210	10.015450	1146.528
6	Jun-16	9.0153630	8.7788810	965.0588
7	Jul-16	9.8800170	7.5352890	755.8578
8	Aug-16	9.6527090	8.8965100	913.4154
9	Sep-16	8.8385550	7.7079290	864.2797
10	Oct-16	9.2108986	8.5207990	916.9938
11	Nov-16	7.5923240	8.1316280	1061.453
12	Dec-16	7.7915080	7.8232170	995.0886



Related information was extracted from the 40page reviews of No. 1~5 destinations based on the crawled data on the global tourist information site TripAdvisor.

In short, Gyeongbokgung, Changdeokgung, Myungdong, Insadong and Namsan Tower were ranked in the top five in the order named. The crawling returned the following results.

The crawling extracted many nouns related to things 'original to Korea' and 'history'. Thus, it is desirable to recommend the tourist destinations that may represent some historicity of Korea as shown in Table 4.

**Table 4. TripAdvisor Crawling Review Results**

7	Gyeongbokgung	45	Changdeokgung
186	Joseon dynasty	23	Gyeongbokgung
168	Republic of Korea	86	Joseon dynasty

121	Myungdong	188	Korean style
300	Korean	184	Korean
118	Korean Food	120	small & cute

Also, the daily spending of the subjects was less than the average, which should be considered in recommending the tourist destinations. Moreover, the result seemed to account for the high interest in the 'War Memorial Museum of Korea' as mentioned in the Introduction.

## 5. Recommendation of routes by destination

The route was developed based on the shortest distance to each destination. Also, the route involved the contents related to social issues.

Given the interest in something original to Korean as the crawling results showed, the subway and bus were selected as the more accessible transports than cars. Given foreigners moved to Seoul Station from the airport, and considering the highest Lift as found in the association analysis, Myungdong was chosen as the start point.



**Fig. 5. Recommendation of optimal route for foreign visitors**

Hence, the proposed route includes the following destinations as shown in Fig. 5. : Myungdong → Namdaemun Market → Gwanghwamun → Insadong → Dongdaemun Market → Namsan Tower → Itaewon or National Museum of Korea → Shinchon/Hongdae or National Folk Museum of Korea → Bukchon Hanok Village → History Museum → Seoul Station.

In addition to the subways and buses, tourists are recommended to go from Myungdong to Namdaemun using Seoul Metro 7017 and then walk while overlooking the city. They may move from Insadong to Dongdaemun Market using the No.1 subway line, and go from Dongdaemun to Namsan Tower using the buses (Dongdaemun Blue Bus No. 105 and transfer to Blue Bus No. 02 at Toigyero 5ga). Itaewon was separated from the National Museum of Korea, where tourists spent less money, so that the destination generating more visitor spending came last.

In replacement of this, tourists are recommended to visit the National Museum of Korea from the Namsan Tower using the cable car and bus as the crawling results showed. For the same reason, tourists are recommended to visit the destinations that can be accessed from the Seoul Station with

Shinchon/Hongdae proposed as an option. Also, based on the crawling result, tourists are recommended to visit the National Folk Museum of Korea. Then, they may walk to the Bukchon Hanok Village and finally the History Museum History, where they may take the Blue Bus No. 703 to the Seoul Station before they fly back home.

## 6. Conclusion

Noting the ever-increasing foreign tourists, this paper proposed a customized route of destinations for foreign tourists by analyzing the big data including those gathered on Twitter and TripAdvisor. To that end, association analysis and time series analysis were used. Specifically, a wide range of texts on the web were crawled and refined, as well as analyzed with the text mining technique.

Then, the optimal route for foreign visitors was proposed as follows: Myungdong → Namdaemun Market → Gwanghwamun → Insadong → Dongdaemun Market → Namsan Tower → Itaewon or National Museum of Korea → Shinchon/Hongdae or National Folk Museum of Korea → Bukchon Hanok Village → History Museum → Seoul Station.

The present study was intended to inform foreign tourists of the proposed route in advance, and help them to experience the history of the country and cities. However, the limited time and data precluded more diversified analysis and variables.

Hence, further studies need to consider other variables, review relevant literature published in Korea and overseas countries, establish those variables and earlier findings into a system, and consider a wider range of analysis methods.

## REFERENCES

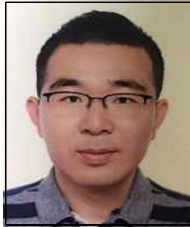
- [1] TripAdvisor. <https://www.tripadvisor.com>
- [2] Twitter. <https://twitter.com>
- [3] Korea Tourism Organization. <https://kto.visitkorea.or.kr/eng.kto>
- [4] D. Y. KIM. S. T. Park & M. H. Ko. (2018). A Study on the Analysis of IT-related Occupational Cluster using Big Data. *IAENG International Journal of Computer Science*, 45(1), 7-11.
- [5] S. T. Park. S. W. Lee & M. H. Ko. (2018). Industry 4.0 on Keyword Network Analysis. *Journal of Engineering and Applied Sciences*, 13, 2442-2446.
- [6] Tourism Knowledge & Information System. <https://know.tour.go.kr/customer/engboard/engboardlist.do>
- [7] M. J. Choi & K. S. Noh. (2016). Exploratory Study on Crime Prevention based on Bigdata Convergence-Through Case Studies of Seongnam City. *Journal of Digital Convergence*, 14(11), 125-133.  
DOI : 10.14400/jdc.2016.14.11.125
- [8] K. W. Kim. D. W. Kim. K. S. Noh & J. Y. Lee. (2015). An exploratory study on improvement method of the subway congestion based big data convergence. *Journal of Digital Convergence*, 13(2), 35-42.  
DOI : 10.14400/jdc.2015.13.2.35
- [9] K. W. Kim. W. J. Park & S. T. Park. (2015). A study on plan to improve illegal parking using big data. *Indian Journal of Science and Technology*, 8(21), 1.  
DOI : 10.17485/ijst/2015/v8i21/78274
- [10] T. G. Kang. Y. R. Kim. D. W. Kim & S. T. Park. (2016). A Study on Improving Comparative Analysis and Providing Information on Bicycle Roads through Big Data Analysis. *Indian Journal of Science and Technology*, 9(24), 1-7.  
DOI : 10.17485/ijst/2016/v9i24/96152
- [11] D. W. Kim. T. G. Kang. G. Li & S. T. Park. (2015). Analysis of User's Behaviors and Growth Factors of Shopping Mall using Bigdata. *Indian Journal of Science and Technology*, 8(25), 1-7.  
DOI : 10.17485/ijst/2015/v8i25/80979
- [12] B. S. Kim. K. W. Kim & S. T. Park. (2015). The Improvement Plan for Fire Response Time using Big Data. *Indian Journal of Science and Technology*, 8(23), 1.  
DOI : 10.17485/ijst/2015/v8i23/79198
- [13] E. M. Park. J. H. Seo & M. H. Ko. (2016). The effects of leadership by types of soccer instruction on big data analysis. *Cluster computing*, 19(3), 1647-1658.

DOI : 10.1007/s10586-016-0609-2

- [14] K. S. Noh. J. H. Kim. S. T. Park & B. S. Kim. (2018). *Big data analysis for R*, Seoul: Wowpass.
- [15] J. H. Kim. S. T. Park & S. W. Lee. (2018). *Big data R point*, Seoul: Kwangmoonkag.

박 성 택(Seong-Taek Park)

[종신회원]



- 2001년 2월 : 한국교통대학교 문학사
- 2003년 8월 : 충북대학교 경영대학원 (경영학석사)
- 2010년 2월 : 충북대학교 경영정보학과 (경영학박사)
- 2014년 5월 ~ 현재 : 한국소프트웨어 기술인협회 교수

· 관심분야 : 빅데이터, 빅데이터분석, 경영정보, 기술혁신, 특허 가치평가

· E-Mail : solpherd@nate.com

김 영 기(Young-Ki Kim)

[정회원]



- 1978년 2월 : 서울대학교 학사
- 1986년 12월 : Univ. of Iowa MBA (경영학 석사)
- 1992년 5월 : Univ. of Iowa MBA(경영정보학 박사)
- 1993년 3월 ~ 현재 : 충북대학교 경영 대학 경영정보학과 교수

· 관심분야 : 정보통신, 특허가치평가, 특허경영전략, 이러닝, 소프트웨어 방법론 등

· E-Mail : ykim@cbnu.ac.kr