

# 비정형 데이터를 이용한 여가통행 행태 추정: 서울시 한양도성을 중심으로

조준식<sup>1</sup> · 광주현<sup>2</sup> · 구동균<sup>3</sup> · 유연택<sup>1</sup> · 이승재<sup>4\*</sup>

<sup>1</sup>서울시립대학교 교통공학과/스마트시티학과 석사과정, <sup>2</sup>서울시립대학교 교통공학과/스마트시티학과 석박사통합과정, <sup>3</sup>서울시립대학교 교통공학과 연구교수, <sup>4</sup>서울시립대학교 교통공학과 교수

## Estimation of Leisure Behavior Using Unstructured Data: Focusing on Seoul City Wall

JO, Joonsik<sup>1</sup> · KWAK, Juhyeon<sup>2</sup> · KU, Donggyun<sup>3</sup> · YOO, Yeontaek<sup>1</sup> · LEE, Seungjae<sup>4\*</sup>

<sup>1</sup>Master Course, Department of Transportation Engineering/Department of Smart Cities, University of Seoul, Seoul 02504, Korea

<sup>2</sup>PhD Course, Department of Transportation Engineering/Department of Smart Cities, University of Seoul, Seoul 02504, Korea

<sup>3</sup>Research Professor, Department of Transportation Engineering, University of Seoul, Seoul 02504, Korea

<sup>4</sup>Professor, Department of Transportation Engineering, University of Seoul, Seoul 02504, Korea

\*Corresponding author: sjlee@uos.ac.kr

### Abstract

Recently, as the types of leisure activities have diversified and the leisure traffic has increased, individual travel patterns have become more complex. It is impossible to precisely track non-daily travel with the existing survey method. Therefore, this study attempted to estimate leisure travel behavior by collecting social media platform information that is easy to grasp leisure activities and using text mining techniques. First, Instagram posts were collected for Hanyang City Wall in Seoul. The collected text data was extracted into a total of four leisure types: social, exercise, tourism, and rest by applying the Latent Dirichlet Allocation (LDA). After that, hot spot analysis using Kernel Density Estimation (KDE) was conducted to understand how each leisure activity is spatially distributed in Hanyang City Wall. Social-type activities occurred relatively much in the eastern region, exercise-type activities in the north-west, tourism-type activities in the central region, and rest-type activities in the southern region. In addition, a Geographically Weighted Regression (GWR) was conducted to find out how the characteristics of leisure activities by region affect the mode choice behavior. The more social/tourism-type activities were observed, the more traffic increased, and the more exercise/rest-type activities were observed, the more traffic decreased. By region, social/tourism-type activities greatly increase the use of transportation in the northern region, and exercise/rest-type activities greatly reduce the use of each means in the central region.

**Keywords:** GWR, KDE, LDA, leisure travel, social media data

### 초록

여가활동의 종류가 다양해지고 여가 통행의 수가 증가하면서, 개인의 통행 패턴은 더욱 복잡해졌다. 이에 따라 기존의 설문조사 방식으로는 비일상 통행 데이터를 정확하게 구축하는 것이

J. Korean Soc. Transp.  
Vol.41, No.1, pp.135-147, February 2023  
<https://doi.org/10.7470/jkst.2023.41.1.135>

pISSN : 1229-1366  
eISSN : 2234-4217

#### ARTICLE HISTORY

Received: 28 November 2022

Revised: 5 December 2022

Accepted: 3 January 2023

Copyright ©  
Korean Society of Transportation

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

불가능해졌다. 따라서 본 연구에서는 여가활동 파악이 용이한 소셜 미디어 플랫폼 정보를 수집하고, 텍스트마이닝 기법을 활용함으로써 여가통행 행태를 추정하고자 하였다. 먼저, 서울시 한양도성을 대상으로 인스타그램의 게시물을 수집하였다. 수집된 텍스트 데이터는 Latent Dirichlet Allocation(LDA) 모델을 적용함으로써, 사교, 운동, 관광, 휴양의 총 4개의 여가 유형으로 추출되었다. 이후 각각의 여가활동들이 한양도성 내에서 공간적으로 어떻게 분포하고 있는지 파악하기 위해, Kernel Density Estimation(KDE)를 활용한 핫스팟 분석을 시행하였다. 시행 결과, 상대적으로 사교형 활동은 동부지역, 운동형 활동은 서북부지역, 관광형 활동은 중부지역, 휴양형 활동은 남부지역에서 많이 발생한 것으로 확인되었다. 또한, 지역별 여가활동 특성들이 통행수단 선택행위에 어떠한 영향을 미치는지 알아보기 위해 Geographically Weighted Regression(GWR) 분석을 실시하였다. 분석 결과, 사교형 및 관광형 활동이 많이 관측될수록 교통량이 증가하였고, 운동형 및 휴양형 활동은 많이 관측될수록 교통량이 감소하였다. 지역별로는 북부지역의 경우 사교형 및 관광형 여가활동이 교통수단 이용량을 크게 증가시키며, 중부지역에서는 운동형 및 휴양형 여가활동이 각 수단별 이용량을 크게 감소시키는 것으로 분석되었다.

**주요어:** GWR, KDE, LDA, 여가통행, 소셜 미디어 데이터

## 서론

경제성장과 함께 자동차의 급속한 보급과 더불어 대중교통수단의 첨단화는 출퇴근 등의 일상적인 통근통행뿐 아니라 여가활동과 같은 비일상통행도 급격히 증가시키고 있다. 주 5일제와 Work life Balance를 추구하면서 노동시간은 지속적인 감소추세를 보이는 반면 여가통행은 증가하고 있다(Koo et al., 2021).

교통분야에서는 통행행태를 분석하기 위한 다양한 연구들이 진행되었다. Kim et al.(2018)는 가구통행실태조사를 통해 서울시 연계수단 통행행태 영향요인을 분석하였다. Park et al.(2022)는 개인의 10분 단위 행동을 기록한 생활시간조사 자료를 활용하여 통행보행과 여가보행의 통행패턴을 분석하였다. 기존의 통행행태 분석은 가구통행실태조사, 여객기종점통행량조사, 국민여행실태조사, 국민여가활동조사 등 단순한 설문조사 자료를 기반으로 수행되었다. 하지만 설문조사는 통행목적과 수단을 명확히 알 수 있다는 장점과 동시에 실시간적으로 변화하는 도시의 환경을 반영하지 못한다는 단점이 존재한다. 또한, 여가 및 관광활동의 수요 증가에 따라 기존의 여가통행자료 수집을 위한 전통적인 방법으로는 불규칙적인 통행발생과 다수의 목적지 경로 선택 등을 반영하지 못한다는 한계점을 지닌다(Jang, 2015).

제4차 산업혁명 이후 첨단기술의 도입을 통하여 다양한 수집원으로부터 통행정보를 포함한 교통 빅데이터들이 수집되기 시작하면서 표본자료가 아닌 전수 조사에 가까운 정보를 분석이 가능해졌다. 이를 통해, 사람, 차량 등 이동 주체의 움직임을 연속적인 형태로 파악함으로써 전국적인 통행행태 및 특성을 분석하는 것이 가능해졌다. Shin et al.(2019)은 교통카드 빅데이터를 사용하여 수도권에서의 버스요금과 지하철 수입금의 변화를 분석하였다. Jang et al.(2021)은 모바일 데이터와 GPS 기반 기지국 정보를 활용함으로써 보행자 탐지 알고리즘을 개발하였다. 한편, SNS(Social network service)는 최근 전세계적으로 많은 사람들이 이용하고 있으며, 실시간으로 생성되는 빅데이터 자료이다. SNS 이용률은 꾸준히 증가하여 2019년 47.7%, 2020년 52.4%, 2021년 55.1%에 이르렀다(Kim, 2022). 사람들은 인스타그램에 시간과 장소를 포함한 일상을 텍스트, 이미지 등 다양한 형태로 기록하여 지인들과 공유한다(Lee et al., 2020a). SNS는 사람들이 자발적으로 게시물을 공유하고 실시간으로 업데이트되기 때문에 기존의 구조화된 설문 방식을 보완할 매체로 각광 받고 있다(Im, 2017).

여가통행은 계속해서 증가하고 있어, 기존 직접조사 방식으로 여가통행을 추정하는 것은 한계가 있다. 이에 본 연구는 여가통행 추정을 위한 새로운 접근방식을 제안한다. 연구의 내용은 SNS를 통한 비정형데이터의 수집과 전처리 과정을 포함하며, 여가통행 유형을 구분하고 공간적 특성 및 교통수단 선택의 특성을 발견하는 일련의 과정을 포함한다. 연구의 공간적 범위는 서울의 녹색교통진흥지역인 한양도성 내부를 대상으로 하며, 시간적 범위는 2020년-2021

년까지로 설정하여 데이터를 수집하였다. 본 연구를 통해 여가통행 행태의 분석을 수행하고 공간적 특성 및 교통수단선택과의 관계를 발견할 수 있다. 특히, 본 연구는 새로운 형태의 데이터인 SNS데이터를 이용하여 통행행태를 추정하는 새로운 접근방식을 제안하는 것에 목적이 있다.

## 선행연구 검토

### 1. 여가통행행태 연구

비일상 통행에 대한 연구의 필요성과 함께 여가통행에 대한 연구가 진행되었다. 비일상통행 중 하나인 여가통행은 다양한 목적으로 분류될 수 있다. Jang and Lee(2016)는 20-30대 1인 가구를 대상으로 여가통행목적지 선택에 미치는 영향을 쇼핑복합형, 관람행락 취미활동과 사교형으로 나누어 공간 선택과 선호패턴을 분석하고 시각화하였다. Jang(2014)은 여가활동 유형을 자연·관광·휴양형, 관람·행락·스포츠형, 사교형, 쇼핑·복합형으로 구분하였다. 쇼핑·복합형 통행은 주로 거주지 주변에서 단거리 통행으로 발생하고 자연·관광형 통행은 주로 장거리 통행인 것으로 나타났다. Magdolen et al.(2022)은 LCA(Latent Class Analysis)를 사용하여 잠재적인 여가 통행유형을 장거리 통행자, 연간 휴가 통행자, 여행중독자, 지역 통행자의 4가지로 구분하였다. 연구에서 연간 휴가 통행자와 지역 통행자는 유사한 사회인구학적 특성과 일상적 이동성을 보였으며 장거리 통행자들은 비행기 위주의 통행을, 여행중독자는 자가용 차량을, 지역 통행자는 대중교통을 이용하는 경향을 보였다. Mahdi et al.(2022)은 여가통행자와 교통수단의 관계를 이산선택모형을 통해 평가하고 여가통행자가 선택할 수 있는 교통수단 대안을 MNL(Multinomial Logit) 모형을 통하여 분석하였다.

경제성장을 통해 사람들은 이전에 비해 더 많은 여가활동이 가능해졌다. 연구자들은 여가통행을 다양한 유형을 통해 분류하고 유형별 이용 통행행태 및 통행특성을 주로 분석하는 경향을 나타냈다. 또한, 지역별, 여가 유형별 특성에 따라 이용하는 수단이 다른 것으로 확인되었다. 따라서 본 연구에서는 SNS를 통해 다양한 여가활동에 대한 정보를 수집하고, 여가 유형을 분류하여 유형에 따른 통행행태를 파악하고자 한다.

### 2. 소셜미디어/비정형데이터 활용 연구

스마트폰 보급이 보편화되면서 전세계적으로 소셜미디어의 사용이 증가함에 따라 소셜미디어 데이터를 통해 얻을 수 있는 정보는 다양해지고 있다(Toivonen et al., 2019). 하지만 소셜미디어 데이터는 대부분 비정형화된 데이터로 기존 정형데이터 분석과 다른 데이터마이닝 기법이 필요하다(Injadat et al., 2016). 다양한 방법 중 SNS 데이터로부터 유용한 정보를 분석하기 위한 도구로 텍스트마이닝 기법이 적합하다(Go and Hong, 2022). Lee and Kang(2020)은 SNS 중 하나인 플리커(Flickr)의 데이터를 의미연결망 분석을 통해 텍스트를 분석하여 부산지역에서 외국인 관광객의 선호관광지를 문화권, 관광지역별로 분류했다.

LDA는 비정형데이터 분석에 주로 사용되는 텍스트마이닝 기술 중 하나로 잠재 데이터 발견과 토픽 모델링으로 활용된다(Jelodar et al., 2019). Chen et al.(2019)은 잠재적인 운전행태와 구조화된 패턴을 파악하기 위해, 도로실험 데이터에 LDA 모델과 k-means 클러스터링을 포함한 비지도 학습 방법을 적용하였다. Sa et al.(2021)는 온라인 기사와 설문조사를 통해 수집된 데이터에 LDA 기법을 적용하여 코로나 블루와 여가활동의 관계를 설명하였다.

특히, Kwak et al.(2022a)은 소셜 미디어 데이터에 LDA 모델을 적용하여 지역별 대표 토픽을 추출하고, 지역별 친환경 교통수단 이용량과 비교하여 여가 목적과 친환경 교통수단의 관계성을 검토하였다. 따라서 본 연구에서는 소셜 미디어에서 다양한 정형데이터와 비정형데이터를 수집하고, LDA를 활용하여 여가 유형과 관련된 토픽들을 추출하고자 한다. 또한, 토픽들이 공간적으로 어떻게 분포하는지 세부적으로 검토함으로써, 지역별로 다르게 나타나는 여가 관련 특성들이 사람들의 교통수단 선택행위에 어떠한 영향을 미치는지 파악해보고자 한다.

### 3. 교통분야에서의 공간분석

통행자는 개인의 목적과 편의에 따라 서로 다른 교통수단을 이용하며 시공간적으로 분리된다. 연구자들은 다양한 교통수단의 통행행태와 분포에 대한 공간분석을 통해 통행자들의 특성을 설명하고자 하였다. Kwak et al.(2022b)은 공유자전거의 이용률 변화를 코로나로 인한 거리두기 단계에 따라 분석하고자 Moran Index와 Hot-spot 분석을 사용했다. Cho et al.(2019)은 대중교통 접근성 분석을 위해 공간가중행렬을 구성하여 Moran Index를 계산하고, 보행 접근성과 교통시설공간을 분석하였다. Shafabakhsh et al.(2017)는 교통사고 특징을 분석하기 위해 Mashhad의 교통사고 자료를 전통적인 KDE와 Hot-spot 클러스터링으로 표현하였다. Son and Park(2022)는 교차로에서의 사고심각도 정보를 GIS 환경에 맵핑하고, 해당 변수에 KDE를 적용함으로써 도심 내 주요 교통인프라에서의 세부 위험구역을 분석하였다. Kwon et al.(2014)은 서울의 행정동별 지역적 요소가 통행수단으로써 보행을 결정하는 행위에 미치는 영향의 정도를 파악하기 위해, 지역별 보행특성을 기반으로 공간적 이질성을 고려하는 GWR 모델을 사용하였다. Bin et al.(2013)은 교통인프라와 통행행태를 통합적으로 고려하여 지역간 교통행평성을 분석하기 위해 GWR 모델을 활용하였다.

교통정보는 시공간적으로 다양하게 분포하며 정보를 분류 및 분석하기 위해 효과적인 기준이 필요하다. 따라서 본 연구에서는 인스타그램 게시물의 공간적 분포를 확인하기 위해 KDE를 활용하여 Hot-spot 클러스터링을 진행하였고, 교통수단 이용량의 일정한 공간적 연관성을 증명하기 위해 Moran's Index 검증을 수행하였다. 이후 GWR 모델을 사용하여 여가 행위와 각 교통수단 선택행위 간의 상관성을 검토하였다.

### 연구 방법론

본 연구의 프레임워크는 Figure 1과 같다. 먼저, 웹 크롤링 기법을 사용하여 인스타그램의 게시물을 수집하였다. 게시물의 텍스트 데이터는 텍스트 마이닝을 적용하였고, 게시물의 위치정보는 지오코딩을 통해 좌표 형태로 변환되었다. 전처리된 텍스트 데이터는 LDA를 적용하여 각 게시물에 대한 토픽을 할당하였다. 이후 게시물의 토픽별 공간분포를 확인하기 위해 KDE 모델을 사용하여 지역별 토픽 등장 횟수를 시각화하였다. 각 게시물의 대표 토픽은 통행자들의 여가 목적으로 간주함으로써, GWR 모델을 적용하여 여가 목적에 따른 지역별 교통수단 이용량을 분석하였다.

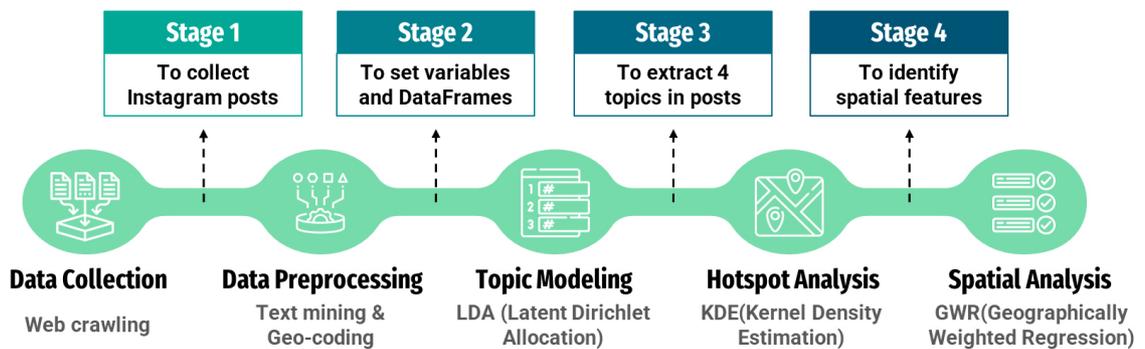


Figure 1. Framework of study

### 1. 데이터 수집 및 전처리

소셜 미디어 데이터를 효율적으로 수집하기 위해 인스타그램의 해시태그 기능을 이용하여 서울의 녹색교통진흥지역인 '한양도성'을 키워드로 4,039개의 게시물을 크롤링하였다. 각 게시물은 생성된 위치의 위도와 경도를 추출하

기 위해 카카오 API 키를 호출하여 위치명 변수를 좌표 형태로 변환하였다. 이후 한양도성에 업로드된 게시물만 추출한 결과 2020년부터 2021년까지 1,973개의 데이터 지점이 특정되었다. 텍스트 분석을 효과적으로 하기 위해서는 전처리과정이 중요하며, 이를 위해 텍스트 데이터를 단어 단위로 토큰화함으로써 문서 단위를 단어 단위로 나누었다. 이러한 방식으로 추출된 단어들은 규칙 기반 통합, 불용어 제거 및 형태소 분석을 통해 정규화하였다. 수집된 소셜 미디어 데이터의 변수 및 형태는 Table 1과 같다.

**Table 1. Components of social media data**

Components	Description	Property
Upload id	ID of user who uploaded the post	TEXT
Location name	The location where the post was created or artificially specified by the user	TEXT
Date	The date when user uploaded the post (year-month-day)	DATE
Time	The time when user uploaded the post (hour-minute-seconds)	TIME
Content	The content the user has written in the body	TEXT
Comments	The content of the comments in the post	TEXT
Likes	The number of people who clicked "Like" on the post	NUMBER

## 2. Latent Dirichlet Allocation(LDA)

LDA는 문서 집합에서 토픽을 식별하는 프로세스인 토픽모델링 기법의 대표적인 알고리즘이다. LDA의 그래픽 모델 표현은 Figure 2와 같다. LDA는 토픽  $K$ 가 모든 문서  $D$ 에 걸쳐 분포되어 있다는 가정 하에 토픽  $k$ 의 개수를 결정한다. 따라서 사용자는 하이퍼파라미터인 토픽  $K$ 의 개수를 할당해야 한다. LDA는 모든 문서의 모든 단어에 대해 토픽 중 하나인  $k$ 를 무작위로 할당한다. 결과적으로 각 문서에는 토픽이 있고 토픽에 맞는 단어 분포가 있다. 이 결과는 무작위로 할당되기 때문에 정확하지는 않을 것이다. LDA가 각 단어에 대해 토픽을 할당할 때 문서의 각 단어  $w$ 는 잘못된 토픽에 할당되지만 다른 모든 단어는 올바른 토픽에 할당된다고 가정한다. 위의 가정하에 LDA는 문서  $D$ 의 단어 중 토픽  $k$ 에 해당하는 단어의 비율과 단어  $w$ 가 있는 모든 문서의 비율이라는 두 가지 기준에 따라 단어의 토픽을 재할당한다. 여기까지의 과정을 반복하면 모든 단어에 대한 토픽 할당이 점차 수렴 상태로 통합된다.

LDA는 다음과 같은 생성 프로세스를 거쳐 문서의 토픽 분류를 수행한다 : (1)  $\theta_d \sim Dir(\alpha)$ ,  $d \in \{1, \dots, D\}$  (2)  $\phi_k \sim Dir(\beta)$ ,  $k \in \{1, 2, \dots, K\}$  (3)  $z_{d,n} \sim Mult(\theta_d)$ ,  $d \in \{1, \dots, D\}$ ,  $n \in \{1, \dots, N\}$  (4)  $w_{d,n} \sim Mult(\phi_{z_{d,n}})$ ,  $d \in \{1, \dots, D\}$ ,  $n \in \{1, \dots, N\}$  (Blei et al, 2003). 이에 대한  $d$ 번째 문서의  $i$ 번째 단어의 토픽  $z_{d,i}$ 가  $j$ 번째에 할당될 확률은 Equation 1로 제시된다.

$$P(z_{d,i} = j \mid z_{-i}, w) = \frac{n_{d,k} + \alpha_j}{\sum_{i=1}^K (n_{d,i} + \alpha_i)} \times \frac{v_{k,w_{d,n}} + \beta_{w_{d,n}}}{\sum_{j=1}^V (v_{k,j} + \beta_j)} = AB \quad (1)$$

여기서,  $\theta$ 는 문서 내 토픽 분포,  $\phi$ 는 토픽 내 단어 분포,  $w_{d,n}$ 은  $d$ 번째 문서의  $n$ 번째 단어,  $n_{d,k}$ 는  $k$ 번째 토픽에 할당된  $d$ 번째 문서의 단어 빈도,  $V_{k,w_{d,n}}$ 은 전체 코퍼스에서  $k$ 번째 토픽에 할당된  $w_{d,n}$ 의 빈도,  $\alpha$ 는 문서의 토픽 분포를 생성하기 위한 Dirichlet 분포 매개변수,  $\beta$ 는 토픽의 단어 분포를 생성하기 위한 Dirichlet 분포 매개변수,  $A$ 는  $d$ 번째 문서와  $k$ 번째 토픽의 연관 정도,  $B$ 는  $w_{d,n}$ 과  $k$ 번째 토픽의 연관 정도를 의미한다.

본 연구에서는 LDA의 토픽 분포가 고정되어 있지 않고 Dirichlet 분포에서 랜덤하게 생성된다는 가정 하에 토픽 모델링을 수행하였다. 인스타그램 각각의 게시물을 문서 단위로 간주하여 토픽의 단어 분포와 문서의 토픽 분포를 추정 한 후, 주어진 게시물들에 대하여 어떤 토픽들이 존재하는지 살펴보았다.

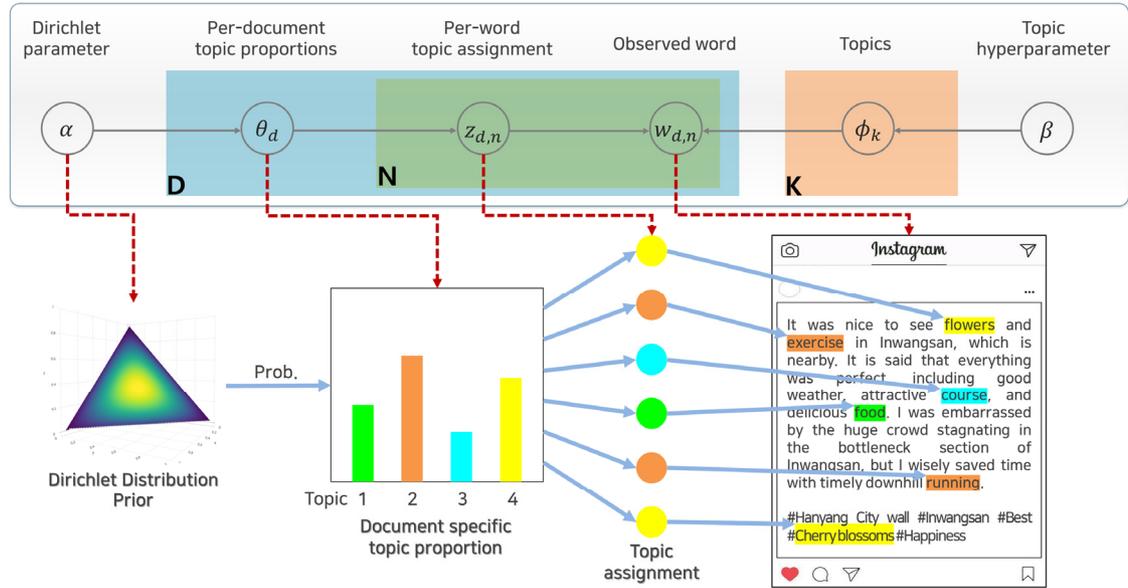


Figure 2. Graphical model representation of LDA

### 3. 공간 분석 방법론

본 연구에서는 한양도성 내 지역별 토픽의 분포를 확인하기 위해 KDE 모델을 사용하여 히트맵을 분석하고자 하였다. KDE를 활용한 핫스팟 분석은 좌표 간 거리를 바탕으로 모든 지역의 사건 밀도를 추정하는 방법으로 다른 핫스팟 분석들에 비해 더 사실적이고 정밀한 분포 모양을 나타낸다(Paulsen and Robinson, 2004). KDE는 자료의 분포를 비모수적(Non-Parametric)으로 추정하는 방법으로 관측된 데이터 각각에 대하여 해당 데이터의 값을 중심으로 갖는 커널 함수를 생성한 뒤 만들어진 커널 함수들을 모두 더한 후에 전체 데이터 개수로 나눈 값이다(Lee et al., 2020b). KDE의 공식은 Equation 2와 같다. 커널 함수는 다양하게 정의될 수 있으며, 커널 함수의 종류에는 Quartic, Triangular, Uniform, Triweight, Epanechnikov 등이 있다. 본 연구에서는 커널 함수로 Quartic을 사용하였으며, bandwidth 폭은 500m로 설정하였다.

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \tag{2}$$

여기서,  $x$ 는 확률변수,  $x_i$ 는 관측된 하나의 데이터 포인트,  $h$ 는 bandwidth를 결정하는 파라미터,  $K$ 는 커널 함수를 의미한다.

대다수의 사회 현상은 서로 연관되어 있고 지역적으로 각 현상에 서로 영향을 미친다. 이는 ‘모든 것은 모든 것들과 연관되어 있지만, 그러한 연관성은 멀리 떨어진 것들보다 인접한 것들 사이에 더 강하다’라는 지리학 제1법칙에 개념적 바탕을 두며, 이를 공간적 종속성이라고 한다(Tobler, 1970). GWR 모델을 시행하기 위해서는 우선적으로 공간적 종속성을 먼저 검토해야 하며, 공간적 종속성은 전역적(global) 관점과 국지적(local) 관점에서 분석할 수 있다. 전역적 관점에서 공간적 종속성을 측정하기 위한 대표적인 지표 중 하나는 Moran’s I이다(Cliff et al., 1981). Moran’s I의 산출식은 Equation 3과 같다. Moran’s I는 -1에서 +1 사이의 값을 갖는 일종의 상관계수로 양의 값은 군집화되는 경향, 음의 값은 흩어지는 경향, 0의 값은 랜덤한 공간적 패턴을 의미한다. 본 연구에서는 각 지점의 교통량이 인접한 지점의 교통량의 영향을 받는지 분석하기 위해 각 교통수단의 정류장별 교통량을 종속변수로 Moran’s I 검정을 시행하였다.

$$Moran's\ I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \hat{y}_i)(y_j - \hat{y}_j)}{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{3}$$

여기서,  $y_i$ 는  $i$ 번째 지역의 실제 종속변수 값,  $\hat{y}_i$ 는  $i$ 번째 지역의 모형에 의해 추정된 종속변수 값,  $w_{ij}$ 는  $i$ 번째 지역과  $j$ 번째 지역 간의 공간가중치,  $n$ 은 전체 지역의 수를 의미한다.

GWR은 공간상에서 서로 다른 위치에 존재하는 데이터들에 대하여 각각 다른 회귀모형을 적용하여 공간적 비정상성(non-stationarity)을 탐색하는 방법이다(Kim et al., 2011). 일반적인 선형회귀분석은 분석대상지의 총값 혹은 평균값을 변수로 사용하여 회귀계수가 전 지역에서 같은 것으로 간주한다. 반면, GWR 모델은 회귀계수를 상수로 인식하는 것이 아니라 위치의 함수로 인식함으로써, 지역별로 각각 다른 회귀계수를 추정함으로써 변수들의 국지적인 영향력을 파악할 수 있는 결과를 제시한다(Chun et al., 2021). 따라서 본연구에서는 GWR 모델을 활용해서 지역별 토픽 개수에 따른 교통수단별 이용량의 상관관계를 분석하고자 하였다. GWR 모델의 독립변수로는 한양도성 내 동별 각 토픽의 개수로 설정하였으며, 종속변수로는 한양도성 내 동별 자전거 대여 수, 버스 승차자 수, 지하철 승차자 수로 설정하여 총 3개의 GWR 모델을 생성하였다. Equation 4는 GWR 모델을 나타낸다.

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \tag{4}$$

여기서,  $y_i$ 는 지점  $i$ 에서의 종속변수( $i = 1, 2, \dots, n$ ),  $x_{ik}$ 는 지점  $i$ 에서  $k$ 번째 독립변수,  $\beta_k$ 는 모형에서 추정된  $k$ 번째 파라미터,  $\varepsilon_i$ 는 지점  $i$ 에서의 오차항,  $p$ 는 파라미터의 개수,  $(u_i, v_i)$ 는  $i$ 번째 회귀중심점의 좌표,  $\beta_k(u_i, v_i)$ 는 회귀 분석의 중심점에서  $k$ 번째 독립변수의 회귀계수를 의미한다.

## 분석 결과

이 절에서는 토픽모델링 및 공간분석 결과를 통해 여가활동의 유형과 녹색교통수단의 이용 간의 지역별 상관관계를 분석한 결과를 다룬다.

### 1. Latent Dirichlet Allocation(LDA)

본 연구에서는 Gibbs Sampling 기반 LDA 알고리즘을 총 1,973개의 전처리된 인스타그램 게시물에 적용하여 토픽을 추출하고 게시물을 재분류하였다. 이를 통해 4개의 토픽을 설정하고, 각 토픽별로 상위 30개의 키워드를 도출하였다. 주요 키워드와 문서 수는 Table 2와 같다. Topic 1의 주요 키워드는 사람, 친구, 사랑, 카페, 레스토랑이었다. Topic 2의 주요 키워드는 달리기, 하이킹, 산, 트레일, 걷기였다. Topic 3의 주요 키워드는 관광, 여행, 문화, 가이드, 박물관이었으며 마지막 Topic 4의 주요 키워드는 꽃, 산책, 나들이, 휴식, 예술이었다. 따라서 Topic 1-4는 사회, 운동, 관광, 휴양으로 분류하였다.

Table 2. Result of LDA

Topic	Subject	Primary keywords	Num. of Documents	Perc. of documents
1	Social	people, friends, love, cafe, restaurant	517	0.26
2	Exercise	running, hiking, mountain, trail, walking	444	0.23
3	Tourism	tour, travel, culture, guide, museum	433	0.22
4	Rest	flower, walks, outings, relaxation, art	579	0.29

토픽 모델링 결과는 LDAvis 기법을 이용하여 시각화하였다(Sievert and Shirley, 2014). LDAvis는 Principal Component Analysis(PCA)를 사용하여 토픽과 단어들을 2차원으로 축소하여 토픽별 토픽 유사도 및 단어 분포를 보여준다. 원의 크기는 토픽과 관련된 단어의 수에 비례한다. 원 사이의 거리는 토픽 간의 유사성을 의미한다. 즉, 가까울수록 토픽이 더 비슷하다는 것을 말한다. 막대는 토픽을 구성하는 주요 키워드와 해당 키워드와 관련된 토픽을 보여준다. 키워드 추출은 Saliency와 Discriminative power 두 가지를 기준으로 한다. Saliency는 다수의 문수에서 자주 등장하며 단어의 중요성이 높은 것을 의미한다. 반대로 Discriminative power는 문서 전체에 걸쳐 단어가 반복적으로 나타나므로 중요성이 낮은 것을 의미한다. 즉, Saliency와 Discriminative power 사이에는 음의 상관관계가 존재한다. Saliency와 Discriminative power의 가중치는  $\lambda$  값을 조정하여 설정할 수 있으며, 본 연구에서는  $\lambda$ 를 0.5로 설정하였다. Figure 3은 LDAvis의 결과를 보여준다. Intertopic distance map에서 원의 크기가 모두 유사하고 겹치는 부분이 없으므로, 각 토픽의 포함된 단어들의 수가 모두 유사하고 토픽들이 서로 유사하지 않게 잘 분류된 것으로 확인된다.

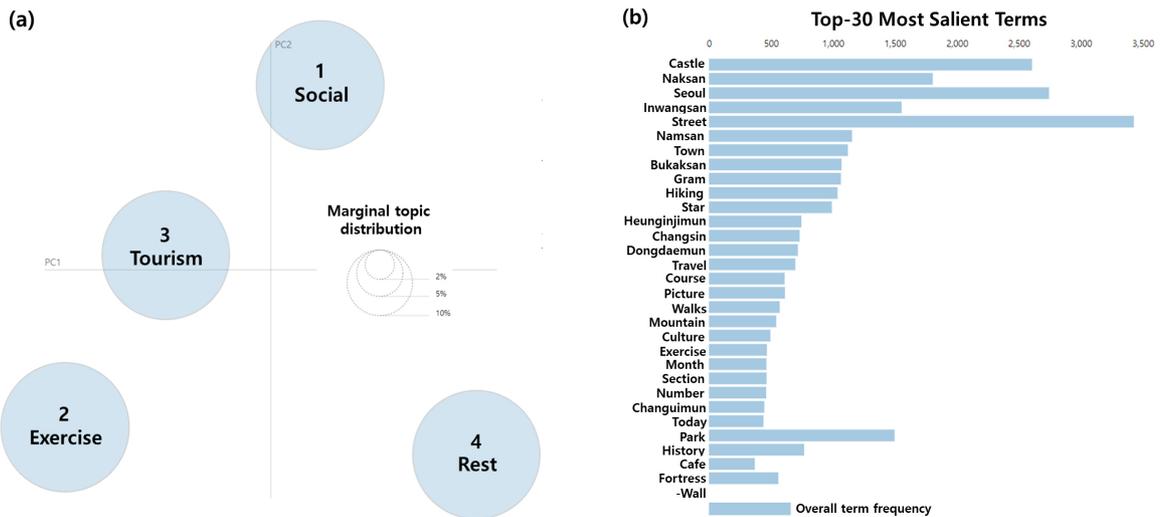


Figure 3. (a) Intertopic distance map (via multidimensional scaling) (b) Top-30 most salient terms

## 2. 공간 분석

먼저, 공간 분석을 진행하기 위해 인스타그램 게시물 데이터를 경도와 위도 좌표에 따라 GIS 맵 상 한양도성 내부에 마킹하였다. 이후 게시물을 토픽별로 분류하고, KDE를 적용하여 한양도성내 지역별로 각 토픽의 빈도수를 나타내는 히트맵을 제작하였다. 히트맵에 따른 핫스팟 분류 결과는 Figure 4와 같다. 토픽별 밀도 분석에 의하면 각 토픽들은 주로 한양도성의 중심보다 외곽지역에 많이 분포된 것으로 나타났다. 토픽별로 핫스팟 지점의 차이를 확인해 봤을 때, 사교형 토픽은 외곽지역에 전반적으로 고루 분포되어 다양한 지역에서 사교형 여가활동이 이뤄진다는 것을 알 수 있었고, 특히 다른 토픽들에 비해 을지로, 대학로 같은 핫플레이스들이 많이 위치한 동부지역에 많이 분포되었다. 운동형 토픽은 5단계의 핫스팟 지역이 인왕산, 북악산 등 산악 지역으로 형성된 서북부지역에 형성되어, 등산을 목적으로 많은 통행이 이루어진다는 것을 확인할 수 있었다. 관광형 토픽은 다른 토픽들과 달리 중부지역에 많이 분포되었으며, 이는 중부지역에 창덕궁, 종묘 같은 다양한 문화재가 존재하여 관광수요가 많이 발생하는 것으로 확인되었다. 마지막으로 휴양형 토픽의 경우 남부지역에 5단계의 핫스팟 지역이 밀집되어 있었는데, 이는 남부지역에 남산공원, 남산벚꽃길 등의 휴양지가 존재하여 사람들이 남산에서 휴양을 많이 즐기는 것으로 판단된다.

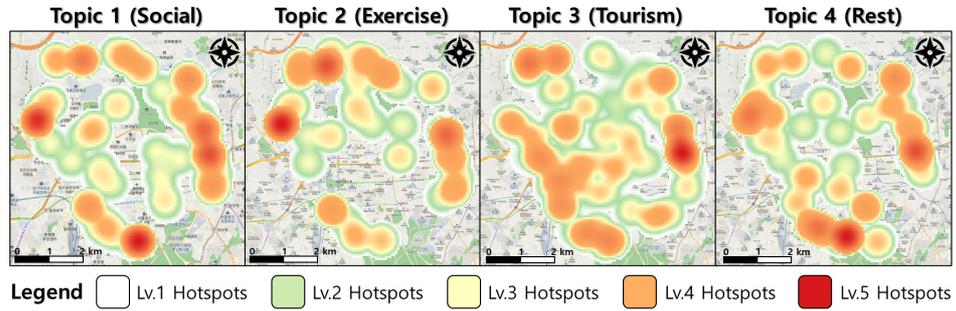


Figure 4. Results of hotspots visualization by KDE

본 연구에서는 대상지의 각 교통수단 이용량의 공간적 자기상관성을 확인하기 위해 Moran's I를 검토하였다. 입력 변수로는 한양도성 내에 존재하는 대여소별 자전거 대여 수, 정류장별 버스 승차자 수, 역별 지하철 승차자 수를 사용하였다. Moran's I 산출 결과는 Table 3과 같다. 자전거 대여 수, 버스 승차자 수, 지하철 승차자 수 모두 Moran's I가 양수로 산출되어 특정 지점과 인접한 지점의 교통량이 양의 상관관계에 있음을 확인하였다. 또한, 세 변수의 p-value가 모두 0.05 미만으로 산출되어, 각 교통수단별 Moran's I 통계량이 유의수준 0.05에서 통계적으로 유의미한 것으로 확인되었다. Figure 5는 각 교통수단 이용량 별 Global Moran's I 산점도이다.

Table 3. Global Moran's index summary

Variables	Moran's index	Expected index	StDev	z-score	p-value
Bicycle	0.1551	-0.0175	0.1010	1.7284	0.04
Bus	0.2207	-0.0032	0.0175	12.8054	0.001
Subway	0.3946	-0.0714	0.3052	1.8034	0.05

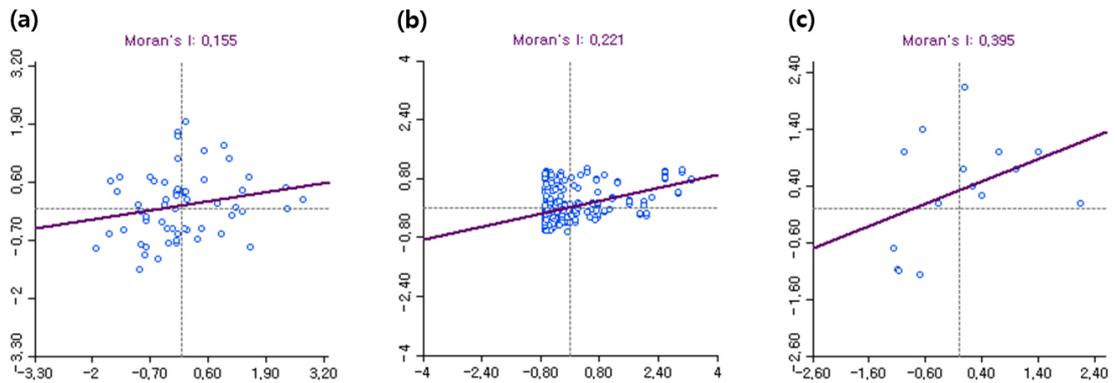


Figure 5. (a) Moran's I scatter of bicycle rides (b) Moran's I scatter of bus rides (c) Moran's I scatter of subway rides

한양도성 내에서 여가 목적별로 각 교통수단별 이용량에 미치는 영향을 분석하기 위해 GWR 분석을 수행하였다. 분석단위는 동을 기준으로 토픽별 등장 횟수와 자전거, 버스, 지하철 이용량 간의 상관관계를 확인하고자 하였다. GWR 모델별로 독립변수의 회귀계수를 시각화하였을 때 Figure 6과 같이 나타났다. 각 행은 GWR 모델별로 구분되어 있으며, 각 열은 독립변수인 4개의 토픽별 등장 횟수를 의미한다. 분석 결과 자전거 대여 수 GWR 모델, 버스 승차자 수 GWR 모델, 지하철 승차자 수 GWR 모델의 R2 값은 각각 0.428, 0.355, 0.661로 산출되어, 지하철 승차자 수 GWR 모델의 설명력이 가장 높은 것으로 분석되었다.

각 독립변수별 회귀계수 확인 결과, 토픽1과 토픽3은 모든 GWR 모델의 회귀계수가 양수로 산출되었고 토픽 2와

토픽4는 모든 GWR 모델의 회귀계수가 음수로 산출되었다. 이는 사교형 및 관광형 여가활동이 많이 관측될수록 교통량이 증가하는 것을 의미하며, 운동형 및 휴양형 여가활동은 많이 관측될수록 교통량이 감소하는 것을 의미한다. 지역별로 회귀계수 값의 차이를 확인했을 때, 토픽1은 대체적으로 북부 및 동부지역에서, 토픽3은 북부 및 중부지역에서 각 교통수단 이용량에 미치는 영향력이 정(+)의 방향으로 큰 것으로 나타났다. 반면 토픽2와 토픽4는 중부지역에서 각 교통수단 이용량에 미치는 영향력이 부(-)의 방향으로 큰 것으로 나타났다. 즉, 북부지역에서는 사교형 및 관광형 여가활동이 각 수단별 이용량을 크게 증가시키며, 중부지역에서는 운동형 및 휴양형 여가활동이 각 수단별 이용량을 크게 감소시키는 것으로 분석된다.

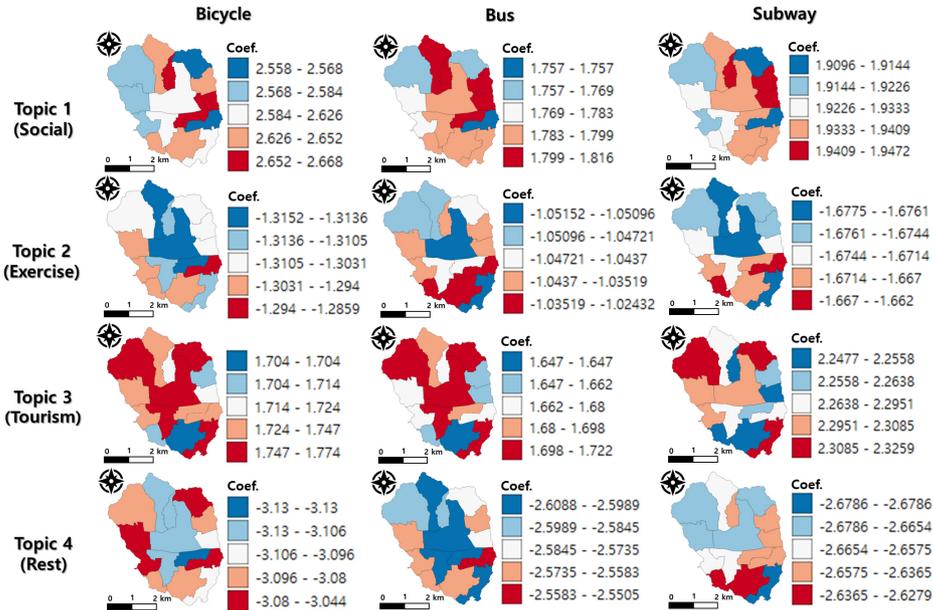


Figure 6. Coefficients of GWR models

## 결론

본 연구에서는 인스타그램 게시물을 텍스트마이닝하여 4가지 여가 유형인 사회, 운동, 관광, 휴양으로 분류된 토픽들을 추출하고, 각 토픽의 공간적 분포를 확인하기 위해 KDE를 적용하여 핫스팟 분석을 시행하였다. 토픽별로 핫스팟을 확인한 결과 지역별 특성에 따라 상이한 결과를 확인할 수 있었다. 동부지역은 핫플레이스가 많이 존재하여 사교 목적의 활동이, 서북부지역은 등산지가 많이 위치하여 운동 목적의 활동이, 중부지역은 다양한 문화 관광지가 존재하여 관광 목적의 활동이, 남부지역에는 남산 휴양지가 있어 휴양 목적의 활동이 많이 이루어진 것으로 확인되었다. 또한 지역별 여가활동에 따른 교통수단 이용량을 분석하여 통행목적과 교통수단 간의 상관관계를 규명하고자 하였다. GWR 모델을 사용하여 각 토픽의 분포가 수단별 교통량에 미치는 영향을 확인하였다. GWR 모델의 회귀계수 값을 토대로 결과를 해석하였을 때, 사교형 및 관광형 여가활동을 목적으로 이동하는 사람들은 대중교통수단과 자전거를 많이 이용하였으며, 반면 운동형 및 휴양형 여가활동이 목적인 사람들은 대중교통수단과 자전거를 적게 이용하였다. 이는 사교형 및 관광형 여가활동을 목적으로 통행하는 사람들과 달리 운동에 이미 너무 많은 에너지를 소모한 사람들과 휴식을 취하기 위해 이동하는 사람들은 통행에 많은 에너지를 소모하고 싶지 않아 하기 때문으로 판단된다. 지역별로 각 토픽이 교통수단 이용량에 미치는 영향력의 크기를 살펴봤을 때, 사교형 및 관광형 여가활동은 각각 북동부, 북중부 지역에서 대중교통 이용량을 크게 증가시켰으며, 운동형 및 휴양형 여가활동은 중부지역에서 대중교통 이용량을 크게 감소시켰다.

본 연구를 통해 지역별로 행해지는 주요 여가활동이 상이하다는 점과 이러한 지역적 특성이 교통수단 선택행위에 있어 유의미한 영향을 미친다는 점을 확인하였다. 따라서 본 연구의 결과를 활용하여 신촌 ‘차없는 거리’, 을지로 ‘세운상가’ 등의 지역에 적용한다면, 지역별로 주된 여가활동 유형과 그에 따른 통행패턴을 파악함으로써 상권 분석 및 개발에 도움이 될 수 있다. 상권이 포함된 지역을 대상으로 본 연구의 프로세스를 적용한다면 지역별 주요 여가활동을 분류할 수 있다. 이를 통해 상업지의 방문목적은 파악함으로써 각 상권에 특화된 마케팅 전략을 수립할 수 있다. 또한, 공간별로 나타나는 주요 키워드와 교통수단별 이용량을 비교함으로써 지역별 대중교통 정책에 활용될 수 있다. 예를 들어, 대중교통 이용이 활발한 지역과 대중교통 이용에 많은 영향을 미치는 활동유형을 도출함으로써, 해당 수단에 대한 시설 개선 및 여가활동과의 연계 등의 교통 정책을 통해 대중교통 활성화에 기여할 수 있다.

본 연구는 기존 설문조사 방식의 정보 수집이 아닌 개인이 일상생활에서 자유롭게 업로드하는 소셜 미디어 데이터를 수집하여, 지역별 여가활동 목적과 교통수단 선택 행위의 관계성을 확인하였다. 그러나 소셜 미디어는 특정 젊은 연령대가 주로 이용하며 고른 표본집단을 추출할 수 없다는 한계점이 있다. 그럼에도 본 연구는 지역별 주요 여가활동에 초점을 맞춰 교통수단 이용량의 통계적 유의성을 확인하였으며, 특히 젊은 세대들이 여가 목적으로 많이 방문하는 지역에서 더 정확한 통행행태 파악이 가능하다는 장점이 있다. 또한 향후에는 편향된 연령대를 보완하기 위한 여러 연구들을 수행할 수 있다. 첫 번째로, 본 연구에서 사용한 인스타그램 외에 다양한 SNS 플랫폼을 활용하고, 각 플랫폼의 이용자 특성을 비교함으로써 다양한 연령층, SNS 플랫폼별 시사점을 도출할 수 있다. 두 번째로, 모바일 데이터, 신용카드 사용실적 데이터 등 다양한 빅데이터와 연계한다면, 본 연구의 결과로 도출된 주요 활동지역의 연령별 통행비율과 비교 및 검증함으로써 여가 목적별 통행 횟수를 보정하고 전 연령대에 대한 비일상 통행패턴을 구축할 수 있다. 이외에도 생활인구 데이터, 승용차 교통량, 보행량 등 다양한 변수를 추가함과 동시에 해당 지역에서 발생하는 모든 게시물들을 위치 검색 기능을 통해 수집할 수 있도록 개선된다면, 개별 통행행태를 정밀하게 추적하여 지역별 교통수요 예측 역시 가능할 것으로 판단된다. 또한 소셜 미디어로부터 실시간으로 게시물들의 토픽 추이 변화를 파악하는 데이터베이스관리시스템을 구축한다면 대규모 군중 밀집 상황 등 다양한 교통 돌발상황 예측 및 대응에도 활용될 수 있다.

## Funding

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(2019K1A4A7A03112460, 2022R1A6A3A01087328). This work was also supported by Korea Ministry of Land, Infrastructure and Transport(MOLIT) as 「Innovative Talent Education Program for Smart City」.

## 알림

본 논문은 대한교통학회 제87회 학술발표회(2022.9.30)에서 발표된 내용을 수정·보완하여 작성된 것입니다.

## ORCID

JO, Joonsik  <http://orcid.org/0000-0002-6916-5335>

KWAK, Juhyeon  <http://orcid.org/0000-0003-0308-1777>

KU, Donggyun  <http://orcid.org/0000-0002-8106-393X>

YOO, Yeontaek  <https://orcid.org/0000-0002-2941-4309>

LEE, Seungjae  <http://orcid.org/0000-0001-9081-2835>

## REFERENCES

- Bin M.Y., Lee W.D., Moon J.B., Joh C.H. (2013), Integrated Equity Analysis Based on Travel Behavior and Transportation Infrastructure: In Gyeonggi-Do Case, *J. Korean Soc. Transp.*, 31(4), Korean Society of Transportation, 47-57.
- Blei D.M., Ng A.Y., Jordan M.I. (2003), Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 3(Jan), 993-1022.
- Chen Z., Zhang Y., Wu C., Ran B. (2019), Understanding Individualization Driving States Via Latent Dirichlet Allocation Model, *IEEE Intelligent Transportation Systems Magazine*, 11(2), 41-53.
- Cho S.J., Lee K.I., Yoon S.H., Moon S.G., Kim H.Y. (2019), Modifiable Areal Unit Problem in Transit Accessibility Analysis, *J. Korean Soc. Transp.*, 37(6), Korean Society of Transportation, 499-513.
- Chun H.J., Kim S.A., Yang G.P. (2021), The Effect of Residential Environment on Housing Price Using GWR Model, *Korea Association of Real Estate Law*, 25(1), 135-151.
- Cliff A. D., Ord J. K. (1981), *Spatial processes: models & applications*, Taylor & Francis.
- Go W.Y., Hong J.G. (2022), Framework for Key Performance Index of Use Cases in Smart Transportation Using Text Mining: Focusing on Daegu Metropolitan City, *J. Korean Soc. Transp.*, 40(4), Korean Society of Transportation, 449-461.
- Im K. H. (2017), SNS Bigdata Analysis Technology Trend and Development Direction, *Review of Korea Contents Association*, 15(1), 38-43.
- Injadat M., Salo F., Nassif A.B. (2016), Data Mining Techniques in Social Media: A Survey, *Neurocomputing*, 214, 654-670.
- Jang Y.J. (2014), Research Trends and Issues Relating to Leisure and Tourist Trips, *International Journal of Tourism and Hospitality Research*, 29(3), 139-149.
- Jang Y.J. (2015), A Methodological Study for Collection of Leisure Travel Data: Focused on ICT Technologies, *International Journal of Tourism and Hospitality Research*, 28(9), 65-84.
- Jang Y.J., Ku D.G., Lee S.J. (2021), Pedestrian Mode Identification, Classification and Characterization by Tracking Mobile Data, *Transportmetrica A: Transport science*, 1-29.
- Jang Y.J., Lee C.H. (2016), A Study of Behavioural Characteristics of 20-30's Single Households on Destination-Choices and Preferences for Leisure Travel, *Seoul Urban Research*, 17(2), 77-96.
- Jelodar H., Wang Y., Yuan C., Feng X., Jiang X., Li Y., Zhao L. (2019), Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, A Survey, *Multimedia Tools and Applications*, 78(11), 15169-15211.
- Kim J.H., Park I.S., Jeong J.H. (2011), Trip Generation Model Based on Geographically Weighted Regression, *J. Korean Soc. Transp.*, 29(2), Korean Society of Transportation, 101-109.
- Kim S.J., Choo S.H., Kim Y.J., Han Y.J. (2018), Analyzing Factors to Affect Trip Mode Chaining Behavior Using Travel Diary Survey Data in Seoul, *The Journal of The Korea Institute of Intelligent Transport Systems*, 17(1), 55-70.
- Kim Y.H. (2022), Social Network Service(SNS) Usage Behavior by Generation, *ICT KISDI STAT Report*, 22(11), 1-6.
- Koo J.H., Choo S.H. (2021), Analysis of Factors Affecting Travel Time Considering Travel Purpose, *J. Korean Soc.*

- Transp., 39(5), Korean Society of Transportation, 697-709.
- Kwak J.H., Jo J.S., Ku D.G., Lee S.J. (2022a), The Relationship between Green Transportation and Leisure Travel Based on Social Media Data Chemical Engineering Transactions, 97, 115-120.
- Kwak J.H., Ku D.G., Oh H.R., Jeong I.H., Shin S.H., Lee S.J. (2022b), Impact Analysis in Spatial Use Changes by COVID: Focusing on Shared Bicycles, J. Korean Soc. Transp., 40(2), Korean Society of Transportation, 178-189.
- Kwon D.Y., Suh T.J., Kim S.Y., Kim H.S. (2014), Effects of Urban Environments on Pedestrian Behaviors: a Case of the Seoul Central Area, J. Korean Soc. Transp., 32(6), Korean Society of Transportation, 638-650.
- Lee B.K., Moon, Y.S., Yun T.W., Lee J.K., Jeong J.N., Noh H.H., Rim H.B. (2020a), An Exploratory Study on Instagram Usage Behavior: Based on Structural Relationships between User Motivation and Big 5 Personality Traits, Journal of the HCI Society of Korea, 15(3), 13-22.
- Lee H.J., Kang Y.O. (2020), Understanding Tourist's Region of Attraction and Image of City Through Social Network Data Analysis, Journal of the Korean Urban Geographical Society, 23(1), 101-114.
- Lee J.S., Son W.j., Lee H.T., Cho I.S. (2020b), A Study on the Factors Affecting Optimal Site of Offshore Wind Farm from the Perspective of Maritime Traffic using Spatial Analysis, Journal of Coastal Disaster Prevention, 7(2), 85-96.
- Magdolen M., von Behren S., Chlond B., Vortisch P. (2022), Long-distance Travel in Tension with Everyday Mobility of Urbanites: A Classification of Leisure Travellers, Travel Behaviour and Society, 26, 290-300.
- Mahdi A., Hamadneh J., Esztergár-Kiss D. (2022), Modeling of Travel Behavior in Budapest: Leisure Travelers, Transportation Research Procedia, 62, 310-317.
- Park J.H., Ku, D.G., Jeong I.H., Lee S.J. (2022), A Study of Walking Activity Time Characteristics Based on the Time Use Survey, Journal of the Korea Institute of Intelligent Transportation Systems, 21(3), 53-61.
- Paulsen D. J., Robinson M. B. (2004), Spatial Aspects of Crime: Theory and Practice, Allyn & Bacon.
- Sa H.J., Lee W.S., Lee B.G. (2021), Corona Blue and Leisure Activities: Focusing on Korean Case, Journal of Internet Computing and Services, 22(2), 109-121.
- Shafabakhsh G.A., Famili A., Bahadori M.S. (2017), GIS-based Spatial Analysis of Urban Traffic Accidents: Case Study in Mashhad, Iran, Journal of Traffic and Transportation Engineering (English edition), 4(3), 290-299.
- Shin S.G., Lee C.H., Cheon S.H. (2019), Analyzing Changes in Revenue of Subway Organizations according to Bus Fare Increase in Seoul Metropolitan Area, J. Korean Soc. Transp., 37(6), Korean Society of Transportation, 486-498.
- Sievert C., Shirley K. (2014), LDAvis: A Method for Visualizing and Interpreting Topics, In Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, 63-70.
- Son S.O., Park J.Y. (2022), Identifying Micro-Level Hotspots of Urban Network Infrastructures Using GIS Data, J. Korean Soc. Transp., 40(4), Korean Society of Transportation, 585-598.
- Tobler W.R. (1970), A Computer Movie Simulating Urban Growth in the Detroit Region, Economic Geography, 46(sup1), 234-240.
- Toivonen T., Heikinheimo V., Fink C., Hausmann A., Hiippala T., Järvi O., et al. (2019), Social Media Data for Conservation Science: A Methodological Overview, Biological Conservation, 233, 298-315.