

공간패널모형을 이용한 인스타그램 콘텐츠 특성과 유동인구의 관계 분석

: 용산동2가, 이태원동, 한남동을 사례로*

장지원**, 김민준***, 최진무****

A Study on the Relationship between Foot Traffic and Instagram Contents
Using Spacial Panel Model

: the Case of Yongsan-dong 2-ga, Itaewon-dong, and Hannam-dong

Jiwon Jang, Minjun Kim, Jinmu Choi

요약 인스타그램, 페이스북 등 소셜미디어는 모바일 시대의 보편적인 서비스로 사용되고 있으며, 소셜미디어 사용자가 콘텐츠를 제작하고 소비하는 것은 자연스러운 현상이 되었다. 본 연구는 인스타그램 사용자가 지역을 방문한 이후 지역과 관련된 인스타그램 콘텐츠를 생성하는 현상에 주목하여, 인스타그램 콘텐츠 특성이 유동인구에 미치는 영향을 분석하였다. 분석에 사용한 인스타그램 콘텐츠 자료와 유동인구 자료는 인접한 지역에서 집계된 패널자료이기 때문에 두 자료 사이의 관계를 분석하기 위해 패널공간더빈모형을 사용하였다. 분석결과 40대, 50대, 60대 세대 유동인구보다 10대, 20대, 30대 세대 유동인구에 인스타그램 콘텐츠가 더 많이 영향을 미치며, 남성 유동인구보다 여성 유동인구에 인스타그램 콘텐츠가 더 많은 영향을 미치는 것을 확인할 수 있었다.

핵심용어 공간패널모형, 인스타그램, 빅데이터, 유동인구, 사용자 생성 콘텐츠

Abstract Social media such as Instagram and Facebook are used as universal services in the mobile era, and it has become a natural phenomenon for social media users to produce and consume the contents. This study analyzed the impact of Instagram content characteristics on the foot traffic by paying attention to the phenomenon of creating Instagram contents related to a region after Instagram users visited the region. Since the Instagram content data and foot traffic data used in the analysis are panel data aggregated in the adjacent areas, the spatial panel model was used to analyze the relationship between the two data. The analysis showed that Instagram content affects the foot traffic of teenagers, 20s and 30s more than the foot traffic of the 40s, 50s and 60s, and that Instagram contents affect the foot traffic of women more than that of men.

Keywords Spatial panel model, Instagram, Big Data, Foot Traffic, User generated Contents

* 본 연구는 석사학위논문 내용 중 일부를 수정, 보완하여 작성되었음.

** 경희대학교 대학원 소셜네트워크과학과 석사과정(Master's Student, Department of Social Network Science, Kyung Hee University, jangjiwon@khu.ac.kr)

*** 경희대학교 대학원 지리학과 박사과정(Ph. D. Student, Department of Geography, Kyung Hee University, mjkim@khu.ac.kr)

**** 교신저자, 정회원, 경희대학교 지리학과 교수(Corresponding Author, Member, Professor, Department of Geography, Kyung Hee University, cjm89@khu.ac.kr)

1. 서론

인스타그램(Instagram), 페이스북(Facebook), 트위터(Twitter)와 같은 소셜미디어(Social media) 사용자가 증가함에 따라 정보를 찾기 위해 소셜미디어를 이용하는 것은 보편적인 현상이 되었다. 소셜미디어는 개인 사용자 및 기업 등 다양한 주체가 참여하여 콘텐츠를 생산하고 소비할 수 있는 플랫폼 역할을 하고 있으며, 소셜미디어에서 생성되는 콘텐츠는 상품소비(Kwak and Kim, 2016), 신규점포 선정(Mckenzie and Slind, 2019) 등 의사결정에 영향을 미치고 있다. 수억 명의 사용자를 가지고 있는 소셜미디어에서 방대한 양의 콘텐츠가 매일 생성되고 있기 때문에, 소셜미디어 콘텐츠가 빅데이터로서 가지는 가치는 다양한 관점에서 주목받고 있다.

온라인에서 발생하는 빅데이터가 주목받는 것은 단순히 그 양의 방대함 때문이 아니라 소셜미디어 콘텐츠와 같은 온라인에서 발생하는 자료를 이용하여 현실 문제를 설명할 수 있기 때문이다. 소셜미디어 콘텐츠를 활용하여 현실 문제에 적용하여 식물 종, 철새의 분포, 교통계획 등을 설명하고자 하는 시도가 이루어지고 있다. ElQadi *et al.*(2017)은 곤충과 식물 종(species)의 분포 및 범위에 대한 데이터를 구축하는데 소셜미디어 콘텐츠를 활용하는 방안을 연구하였고 Kim *et al.*(2016)은 교통계획분야에서 수요 추정을 위해 소셜미디어 콘텐츠를 활용하는 방안을 제시하였다. 또한, 소셜미디어 사용자는 장소를 명시하기 위해 소셜미디어 콘텐츠에 지오태그(geotag)와 같은 공간정보를 추가할 수 있기 때문에 공간 및 지역 연구에 사용할 수 있는 새로운 자료로 주목받고 있다(Park and Yu, 2015). Sasaki *et al.*(2010)은 트위터 데이터를 사용하여 사용자가 생성하는 게시물을 통해 지진의 영향 범위를 추정하고 사용자의 트윗과 지진감지시스템을 연동하여 지진모니터링 체계에 대한 연구를 진행하였다. 또한 Dashti *et al.*(2014)는 홍수 후 피

해 복구 단계에서 소셜미디어 콘텐츠를 활용하여 피해범위를 시각화하고 현장 복구 지원을 위한 경로정보 등 위치 정보를 제공하는 방안에 대한 연구를 지원하였다.

인스타그램 사용자는 사진 콘텐츠를 활용하여 다른 사용자와 소통하며, 소통하는 과정에서 방문한 카페, 식당, 여행지 등 공간에 대한 공감대를 형성한다. 인스타그램 콘텐츠를 중심으로 인기가 많고 주목받은 지역은 그렇지 않은 지역보다 더 많은 콘텐츠를 발생시키기 때문에, 지역을 방문하고자 하는 예비 방문자에게 인스타그램 콘텐츠 수는 지역의 유명한 정도를 판단하는 중요한 지표가 되고 있다. 인스타그램 콘텐츠의 유명도와 같이 온라인에서 생성되는 정보에 기반하여 지역 방문에 대한 의사결정을 내리는 과정은 모바일 시대의 보편적인 현상이 되었다(Cox *et al.*, 2009).

이와 같은 현상을 검증하기 위해 지역 상권에서 인스타그램 콘텐츠가 유동인구에 미치는 영향을 분석할 필요가 있을 것이다. 따라서 본 연구에서는 이전 연구에서 주목받지 못한 인스타그램의 콘텐츠를 활용하여 이태원 상권 사례로 이 지역의 성별, 세대별 유동인구를 설명하고자 하였다. 분석에 사용한 인스타그램 콘텐츠와 유동인구 자료는 인접한 집계구를 중심으로 집계된 자료이며 시간에 따라 반복적으로 수집한 패널자료이기 때문에 공간패널모형을 활용하여 분석을 진행하였다. 공간적 인접성과 시간적 연속성을 포함하는 자료의 특성으로 인해 두 자료의 관계를 온전히 분석하기 위해 공간요인을 반영할 수 있는 공간모형이 요구된다(Anselin *et al.*, 2008). 따라서 본 연구는 공간패널모형인 패널공간더빈모형을 이용하여 유동인구와 인스타그램 콘텐츠 사이의 관계를 분석하였으며, 두 변수 사이의 영향관계와 독립변수의 공간전이효과를 분석하였다. 또한 패널공간더빈모형의 결과를 공간요인을 반영하지 않은 패널모형의 결과와 비교하여 변수별 영향력 차이를 보고자 하였다.

Table 1. List of variables

	Variables	Details	Source
Foot Traffic	Teenagers, twenties, thirties. male foot traffic (YOUNGMAN)	Monthly average male foot traffic in 10s, 20s and 30s between 5 p.m. and 11 p.m.	Seoul City
	Teenagers, twenties, thirties. female foot traffic (YOUNGYOUNG)	Monthly average female foot traffic in 10s, 20s and 30s between 5 p.m. and 11 p.m.	Seoul City
	Forties, fifties, sixties and older male foot traffic (OLDMAN)	Monthly average male foot traffic in 40s, 50s, 60s and older between 5 p.m. and 11 p.m.	Seoul City
	Forties, fifties, sixties and older female foot traffic (OLDWOMAN)	Monthly average female foot traffic in 40s, 50s, 60s and older between 5 p.m. and 11 p.m.	Seoul City
Instagram Content	Number of Content Upload Users (UPLOADER)	Monthly average number of Instagram content uploaders referring to the stores	Self collecting
	Number of 'likes' of content (LIKES)	Monthly average number of 'likes' for Instagram content referring to shops	Self collecting
	Total number of content (POSTS)	Monthly average number of Instagram content referred to stores	Self collecting
Resident Population	Resident registration demographics (POPULATION)	Resident Registration Demographics in Administrative district	Ministry of the Interior and Safety.

본 연구의 구성은 다음과 같다. 제2장에서는 연구지역과 분석에 사용한 변수에 대한 기초통계를 제시하고 제3장에서 분석에 사용한 패널모형, 패널공간더빈모형에 대해 소개하였다. 제4장은 패널모형과 패널공간더빈모형을 이용하여 분석한 결과를 논의하며, 제5장에서는 결론을 제시하였다.

2. 연구지역 및 변수 설정

연구지역으로 선정한 이태원 상권은 이태원역을 중심으로 형성되어 있으며 인근의 해방촌(용산동2가), 경리단길(이태원동), 한남동으로 상권이 확산되었다. 상권이 확산되는 과정에서 대형 쇼핑몰 유치 등 대형 자본을 통해 상권이 확장된 것이 아니라, 개성 있는 식당과 카페를 중심으로 소셜미디어 입소문을 통해 상권이 확장되는 것을 확인할 수 있

었다(Kyung and Jeong, 2019). 상권이 확산되는 과정에서 소셜미디어 콘텐츠가 젊은 세대에 영향을 미친 것으로 판단되었기 때문에 본 연구에서는 이태원 상권을 연구지역으로 선정하였다.

인스타그램 콘텐츠와 유동인구 사이의 관계를 분석하기 위해 종속변수로 서울시에서 제공하는 생활인구 자료¹⁾를 사용하였다. 생활인구 자료는 이태원 상권을 포괄하는 65개 집계구에서 측정된 현존인구에 대한 자료이며 이를 기반으로 성별, 세대별 유동인구를 집계하여 사용하였다. 연구지역인 이태원

1) 생활인구 자료란 서울시의 대중교통 이용통계, 건물 DB 등 공공 빅데이터와 통신사 KT의 통신 데이터를 이용하여 추계한 서울의 특정지역, 특정시점에 존재하는 인구에 대한 자료이다. KT의 통신 데이터는 LTE 기지국과 통신하는 이력을 수집한 자료이다.

Table 2. Descriptive Statistics

	Variable	Mean	Standard Deviation	Minimum value	Maximum value
Dependent Variable (Obs. : 1300, N : 65, T : 20)	YOUNGMAN	67.713	90.103	0.066	575.369
	YOUNGWOMAN	76.620	106.174	0.010	652.987
	OLDMAN	54.413	58.234	0.016	329.396
	OLDWOMAN	60.485	59.447	0.046	366.964
Independent Variable (Obs. : 1300, N : 65, T : 20)	LIKES	26.891	33.584	0	391.458
	POSTS	8.009	17.265	0	273
	UPLOADER	5.336	13.042	0	218
	POPULATION	500.241	291.908	65.812	2023.749

은 상권이 발달한 지역이므로 지역 방문객이 주로 활동할 수 있는 시간대인 오후 5시부터 11시까지를 집계 시간으로 설정하였다. KISDI (2019)의 한국미디어패널조사에 따르면 인스타그램을 사용하는 비중은 남성보다 여성이 많으며, 소셜미디어를 사용하는 연령이 40대부터 급격하여 낮아지기 때문에 성별 유동인구와 연령별 유동인구를 구분하여 분석할 필요가 있다. 따라서 인스타그램 콘텐츠의 영향이 성별, 연령별로 다를 것이 예상되기 때문에 10대, 20대 30대 세대의 남성(YOUNGMAN)과 여성 유동인구(YOUNGWOMAN)와 40대 50대 60대 이상 세대의 남성(OLDMAN)과 여성 유동인구(OLDWOMAN)로 변수를 구분하여 종속변수로 사용하였다.

분석에 사용한 인스타그램 콘텐츠를 수집하기 위해서 23대의 서버를 이용한 분산형 웹 크롤러(Web Crawler)를 제작하여 사용하였다. 각 집계구에 위치한 가게이름을 인스타그램의 검색 쿼리(Query)로 사용하여 콘텐츠를 수집하였고 광고 등 해당 가게와 무관한 콘텐츠를 분류하기 위해 수집된 콘텐츠의 해쉬태그(hashtag), 콘텐츠 내용, 댓글 내용을 분석하여 수집된 자료를 전처리하였다.

수집된 인스타그램 콘텐츠를 집계하여 ‘총 콘텐츠 수(POSTS)’를 변수로 사용하였고 또한 콘텐츠를 게시한 인스타그램 사용자 수를 집계하여 ‘콘텐츠 업로더 수(UPLOADER)’ 변수를 사용하였다. ‘콘텐츠 업로더 수’ 변수는 인스타그램 콘텐츠를 올리는 사용자가 방

문에 대한 기록으로 콘텐츠를 올리는 것이 아니라 해당 가게의 사업자가 홍보를 위해 콘텐츠를 반복적으로 게시하는 현상을 통제하기 위해 사용하였다. 마지막으로 콘텐츠의 유명도를 측정하기 위해 콘텐츠에 대한 ‘좋아요’ 수를 집계하여 ‘콘텐츠의 좋아요 수(LIKES)’ 변수를 사용하였고, 콘텐츠에 대한 ‘좋아요’ 수 변수는 콘텐츠의 영향력과 같은 질적인 측면을 유동인구를 설명하는 변인으로 사용하고자 하였다. 2018년 1월부터 2019년 8월까지 20개월 동안 65개의 집계구에서 수집된 자료를 분석에 사용하였고 사용한 변수는 <Table 1>과 같다. 본 연구에서 사용한 종속변수와 독립변수에 대한 기초통계량은 <Table 2>와 같다. 분석에 사용한 자료는 65개의 집계구에서 20개월 동안 반복하여 집계된 패널자료이며 총 1300개의 관측치가 집계되었다.

종속변수인 성별, 세대별 유동인구 자료가 가지고 있는 공간적 자기상관을 파악하기 위해 기간별 각각의 종속변수에 대해 global Moran's I 검정을 진행하였다(Moran, 1948). <Table 3>은 모든 종속변수에 대한 global Moran's I 검정 결과이다. 2018년 1월에서 2019년 8월 사이에 집계구에서 집계된 모든 종속변수의 Moran's I 값이 양의 계수를 가지면서 유의미한 것을 확인할 수 있다. 이는 연구지역의 종속변수가 양의 공간적 자기상관을 가지고 있는 것으로, 유동인구가 성별 및 세대를 막론하고 인접한 지역의 영향을 받고 있기 때문이다(Tobler,

Table 3. Global Moran's I statistics of Dependent Variable

Period	YOUNGMAN	YOUNGWOMAN	OLDMAN	OLDWOMAN
2018. 1	0.422 ***	0.439 ***	0.315 ***	0.275 ***
2018. 2	0.356 ***	0.377 ***	0.240 ***	0.195 ***
2018. 3	0.437 ***	0.446 ***	0.336 ***	0.301 ***
2018. 4	0.436 ***	0.436 ***	0.331 ***	0.309 ***
2018. 5	0.446 ***	0.444 ***	0.364 ***	0.337 ***
2018. 6	0.474 ***	0.483 ***	0.384 ***	0.352 ***
2018. 7	0.468 ***	0.486 ***	0.350 ***	0.307 ***
2018. 8	0.418 ***	0.404 ***	0.307 ***	0.246 ***
2018. 9	0.425 ***	0.389 ***	0.313 ***	0.257 ***
2018. 10	0.421 ***	0.396 ***	0.321 ***	0.259 ***
2018. 11.	0.421 ***	0.394 ***	0.319 ***	0.266 ***
2018. 12.	0.376 ***	0.401 ***	0.262 ***	0.221 ***
2019. 1	0.401 ***	0.398 ***	0.275 ***	0.190 ***
2019. 2	0.418 ***	0.389 ***	0.311 ***	0.245 ***
2019. 3	0.397 ***	0.372 ***	0.299 ***	0.238 ***
2019. 4	0.386 ***	0.358 ***	0.297 ***	0.242 ***
2019. 5	0.403 ***	0.399 ***	0.310 ***	0.262 ***
2019. 6	0.385 ***	0.385 ***	0.288 ***	0.246 ***
2019. 7	0.378 ***	0.373 ***	0.286 ***	0.246 ***
2019. 8	0.390 ***	0.395 ***	0.274 ***	0.239 ***

- Note 1 : * p < 0.1, ** p < 0.05, *** p < 0.01

W. R., 1970). 자료가 가지고 있는 공간적 자기상관으로 인해 본 연구는 공간패널모형을 사용하였다.

3. 시공간 데이터 분석모형

자료의 공간적 자기상관으로 인해 개체 사이의 독립성 가정이 위반되기 때문에, 공간요인을 통제하여 성별, 세대별 유동인구와 인스타그램 콘텐츠 사이의 관계를 분석해야 하는 것이 요구된다. 따라서 공간계량분석기법인 패널공간더빈모형을 사용하였고, 공간요인을 반영하지 않은 패널모형의 결과와 비교하여 추정된 변수별 영향력 차이를 보고자 하였다. 성별, 세대별 유동인구(YOUNGMAN, YOUNGWOMAN, OLDMAN, OLDWOMAN)에 대한 패널모형은 Eq.(1) ~ Eq.(4)와 같다. Eq.(1) ~ Eq.(4)의 i 는 개별 개체를 나타내며 t 는 집계된 시점을 나타낸다. 시간불변항인 u_i 를 고정효과 모형에서 개체별 모수로 간주하지만 확률효과 모형은 u_i 를 오차항으로 간주한다. 고정효과 모형과 확률효과 모형 중에서 적합한 모형을 선택하

기 위해 하우스만 검정을 사용하여 적합한 모형을 결정하였다(Hausman, 1978). u_i 가 설명변수와 상관관계를 가지는 경우 확률효과 모형을 선택하며 상관관계를 가지지 않는 경우 고정효과 모형을 선택한다(Baltagi, 2005). 확률효과 모형을 사용하는 본 연구는 u_i 를 오차항으로 간주하고 분석을 진행하였다(Hills *et al.*, 2011). ε_{it} 는 순수 오차항으로 $\varepsilon_{it} \sim N(0, \sigma_e^2)$ 을 가정한다(Min and Choi, 2012).

$$\begin{aligned} YOUNGMAN_{it} = & \beta_1 UPLOADER_{it} + \beta_2 LIKES_{it} \\ & + \beta_3 POSTS_{it} \\ & + \beta_4 POPULATION_{it} + \\ & u_i + \varepsilon_{it} \end{aligned} \quad (1)$$

$$\begin{aligned} YOUNGWOMAN_{it} = & \beta_1 UPLOADER_{it} + \\ & \beta_2 LIKES_{it} + \beta_3 POSTS_{it} + \\ & \beta_4 POPULATION_{it} + \\ & u_i + \varepsilon_{it} \end{aligned} \quad (2)$$

$$\begin{aligned}
OLDMAN_{it} = & \beta_1 UPLOADER_{it} + \beta_2 LIKES_{it} \\
& + \beta_3 POSTS_{it} \\
& + \beta_4 POPULATION_{it} \\
& + u_i + \epsilon_{it}
\end{aligned} \quad (3)$$

$$\begin{aligned}
OLDWOMAN_{it} = & \beta_1 UPLOADER_{it} + \\
& \beta_2 LIKES_{it} + \beta_3 POSTS_{it} + \\
& \beta_4 POPULATION_{it} + \\
& u_i + \epsilon_{it}
\end{aligned} \quad (4)$$

개체 사이의 공간적 의존성이 존재하는 패널자료를 사용하는 모형을 최소자승법 (ordinary least square)을 이용하여 모형의 계수를 추정하는 경우, 추정된 계수는 편향(biased)되고 비효율적(inefficient)인 추정치가 추정된다(Anselin, 2010; Anselin et al., 2008; LeSage and Pace, 2009). 이와 같은 문제를 해결하기 위해 공간자기상관 모형, 공간오차 모형, 공간더빈모형 등 공간모형을 사용할 필요가 있으며, 본 연구는 독립변수가 인접한 지역에 미치는 공간전이효과 (spillover)를 확인하기 위해 패널공간더빈모형을 사용하였다.

성별, 세대별 유동인구 종속변수에 대한 패널공간더빈모형은 Eq.(5) ~ Eq.(8)과 같다. 패널공간더빈모형은 개별 개체(i)가 시차(t)를 가지는 패널 자료를 사용하며 시간불변항(μ_i)을 가지는 것은 패널모형과 같지만, 공간요인을 통제하기 위해 공간자기상관함수(ρ)와 공간오차함수(λ)를 추가하였다(Anselin, 2008). 또한 독립변수의 인접지역에 대한 영향력을 모형에 추가하기 위해 각각의 독립변수와 공간가중행렬(W)의 곱을 모형에 반영하였다. 인접한 개체사이의 관계를 반영하기 위해 사용한 공간가중행렬은 rook 방식(contiguity weight matrix)으로 산정하였다. 산정된 공간가중행렬은 가중치를 표준화하기 위해 행정규화 가중치(row normalized weight)로 변형하여 사용하였다. 또한 고정효과 모형과 확률효과 모형 중에서 적합한 모형을 선택하기 위해 하우스만 검정(Hausman's Test)을

사용하여 적합한 모형을 결정해야하지만 패널모형과는 달리 공간패널모형의 경우 공간요인을 반영할 수 있는 하우스만 검정을 적용하여 분석하였다(Elhorst, 2003).

$$\begin{aligned}
YOUNGMAN_{it} = & \rho W * YOUNGMAN_{it} \\
& + \beta_1 UPLOADER_{it} \\
& + \beta_2 LIKES_{it} + \beta_3 POSTS_{it} \\
& + \beta_4 POPULATION_{it} \\
& + W * UPLOADER_{it} \\
& + W * LIKES_{it} \\
& + W * POSTS_{it} \\
& + W * POPULATION_{it} \\
& + \mu_i + u_{it}
\end{aligned} \quad (5)$$

$$u_{it} = \lambda W u + \epsilon_{it}$$

$$\begin{aligned}
YOUNGWOMAN_{it} = & \rho W * YOUNGWOMAN_{it} \\
& + \beta_1 UPLOADER_{it} \\
& + \beta_2 LIKES_{it} \\
& + \beta_3 POSTS_{it} \\
& + \beta_4 POPULATION_{it} \\
& + W * UPLOADER_{it} \\
& + W * LIKES_{it} \\
& + W * POSTS_{it} \\
& + W * POPULATION_{it} \\
& + \mu_i + u_{it}
\end{aligned} \quad (6)$$

$$u_{it} = \lambda W u + \epsilon_{it}$$

$$\begin{aligned}
OLDMAN_{it} = & \rho W * OLDMAN_{it} \\
& + \beta_1 UPLOADER_{it} \\
& + \beta_2 LIKES_{it} \\
& + \beta_3 POSTS_{it} \\
& + \beta_4 POPULATION_{it} \\
& + W * UPLOADER_{it} \\
& + W * LIKES_{it} \\
& + W * POSTS_{it} \\
& + W * POPULATION_{it} \\
& + \mu_i + u_{it}
\end{aligned} \quad (7)$$

$$u_{it} = \lambda W u + \epsilon_{it}$$

$$\begin{aligned}
OLDWOMAN_{it} = & \rho W * OLDWOMAN_{it} \\
& + \beta_1 UPLOADER_{it} \\
& + \beta_2 LIKES_{it} \\
& + \beta_3 POSTS_{it} \\
& + \beta_4 POPULATION_{it} \\
& + W * UPLOADER_{it} \\
& + W * LIKES_{it} \\
& + W * POSTS_{it} \\
& + W * POPULATION_{it} \\
& + \mu_i + u_{it}
\end{aligned} \quad (8)$$

$$u_{it} = \lambda W u + \epsilon_{it}$$

4. 인스타그램 콘텐츠 특성과 유동인구 관계 분석

패널모형과 패널공간더빈모형을 이용한 분석결과를 모형별로 제시하면 다음과 같다. 먼저 패널모형을 이용하여 분석한 결과는 <Table 4>와 같다. 고정효과 모형과 확률효과 모형 사이에서 적합한 모형을 찾기 위해, 패널 모형에 대한 하우스만 검정을 시행한 결과 모든 종속변수에 대하여 확률효과 모형

Table 4. Panel Model Results

	YOUNGMAN	YOUNGWOMAN	OLDMAN	OLDWOMAN
LIKES	0.014 (0.512)	0.021 (0.449)	0.024 (0.130)	0.022 (0.238)
POSTS	-0.326 ** (0.034)	-0.450 ** (0.025)	-0.223 ** (0.044)	-0.220 ** (0.090)
UPLOADER	0.426 ** (0.033)	0.593 ** (0.023)	0.293 ** (0.042)	0.301 * (0.075)
POPULATION	0.176 *** (0.000)	0.209 *** (0.000)	0.101 *** (0.000)	0.101 *** (0.000)
Hausman	2.0124 (0.847)	2.838 (0.725)	9.293 (0.098)	5.389 (0.370)
R-square	within = 0.011	within = 0.007	within = 0.009	within = 0.008
	between = 0.277	between = 0.338	between = 0.233	between = 0.223
	overall = 0.266	overall = 0.321	overall = 0.221	overall = 0.209

- Note 1 : Parenthesis numbers mean p-value

- Note 2 : * p < 0.1, ** p < 0.05, *** p < 0.01

(random effect)을 사용하는 것이 적합한 것을 확인할 수 있었다.

모든 성별, 세대별 종속변수에 인스타그램 콘텐츠 변수인 POSTS와 UPLOADER가 유의미한 것을 확인할 수 있으며 통제변수로 사용한 POPULATION 또한 유의미한 것을 확인할 수 있었다. LIKES 변수는 성별, 세대별 유동인구에 유의한 영향을 미치지 않는 것으로 보아, 인스타그램 콘텐츠 자체에 대한 ‘좋아요’ 수는 유동인구에 영향을 미치지 않는 것을 확인할 수 있었다. 이를 통해 콘텐츠의 질적인 특성이 유동인구에 유의미한 영향을 미치지 않는 것으로 나타나는 것을 확인할 수 있으며, 이는 기업의 매출에 소셜 미디어 포스트의 ‘좋아요’ 수가 유의한 관계를 보이지 않는 것과 같은 결과로 볼 수 있다(Park *et al.*, 2017).

POSTS 변수는 모든 그룹에 음의 영향을 미치는 것으로 나타났다. POSTS 변수가 유동인구에 부정적인 영향을 미치는 것은 지역을 방문하는 것과 방문한 이후에 인스타그램 콘텐츠를 생성하는 것 사이에 일정한 시간 간격(time lag)이 존재하기 때문에 발생하는 것으로 예상된다. 또한 해당 지역에서 식당, 카페를 운영하는 이해관계자가 가게 홍보를 위해 지속적으로 콘텐츠를 생성하기 때문에,

인스타그램 콘텐츠는 해당 가게와 관련하여 반복적으로 발생한다. 이와 같이 홍보를 위해 반복적으로 게시되는 콘텐츠는 유동인구에 부정적인 영향을 주는 것으로 예상된다. 이를 통해 단순히 소셜미디어의 콘텐츠 수가 많은 것이 유동인구에 긍정적인 영향을 주는 것이 아닌 것을 알 수 있다.

반면 UPLOADER 변수는 모든 그룹에 양의 영향을 미치는 것으로 나타났다. 이태원 상권 이내의 특정한 지역을 방문한 인스타그램 유저가 많을수록 성별, 세대별 유동인구에 긍정적인 영향을 미치는 것으로 나타났다. 소셜미디어에서 콘텐츠를 활용한 구전행위가 실제 유동인구에도 영향을 주는 것을 나타내며, 이는 소셜미디어 구전량이 기업성과에 긍정적인 영향을 주는 것과 같은 결과로 볼 수 있다(Lee and Park, 2019).

유동인구에 대한 인스타그램 콘텐츠의 영향력은 세대에 따른 차이가 나타나는 것을 확인할 수 있었다. 유동인구에 대한 POSTS 변수의 영향력은 10대, 20대, 30대 남성과 여성 유동인구(-0.326, -0.450)에 미치는 영향이 40대, 50대 60대 이상 남성과 여성 유동인구(-0.223, -0.220)에 미치는 영향보다 더 큰 것을 확인할 수 있었다. 이 같은 세대 차이는 UPLOADER 변수에서도 관찰할 수 있었다. 뿐만 아니라 유동인구에 대한 인스타그램 콘텐츠의 영향력은 성별에 따른 차이

Table 5. Panel Spatial Durbin Model Results

	YOUNGMAN	YOUNGWOMAN	OLDMAN	OLDWOMAN
LIKES	0.178 (0.409)	0.027 (0.340)	0.025 (0.112)	0.024 (0.187)
POSTS	-0.340 ** (0.025)	-0.479 ** (0.015)	-0.228 ** (0.040)	-0.234 * (0.072)
UPLOADER	0.454 ** (0.021)	0.644 ** (0.012)	0.303 ** (0.036)	0.325 * (0.054)
POPULATION	0.154 *** (0.000)	0.186 *** (0.000)	0.091 *** (0.000)	0.090 *** (0.000)
W x LIKES	-0.026 (0.485)	-0.041 (0.397)	-0.005 (0.864)	-0.0278 (0.380)
W x POSTS	0.449 (0.139)	0.432 (0.274)	0.194 (0.383)	0.273 (0.291)
W x UPLOADER	-0.680 * (0.088)	-0.696 (0.180)	-0.341 (0.243)	-0.454 (0.182)
W x POPULATION	-0.008 (0.865)	-0.012 (0.807)	0.012 (0.679)	0.013 (0.680)
ρ	0.199 *** (0.000)	0.219 *** (0.000)	0.053 *** (0.000)	0.075 *** (0.000)
Hausman	4.71 (0.859)	5.39 (0.799)	4.44 (0.880)	5.15 (0.821)
LM test (lag)	184.8 *** (0.000)	149.89 *** (0.000)	63.347 *** (0.000)	27.67 *** (0.000)
LM test (Error)	215.41 *** (0.000)	163.15 *** (0.000)	63.347 *** (0.000)	29.933 *** (0.000)
R-square	within = 0.0127	within = 0.009	within = 0.010	within = 0.011
	between = 0.2834	between = 0.348	between = 0.239	between = 0.229
	overall = 0.2726	overall = 0.331	overall = 0.227	overall = 0.215

- Note 1 : Parenthesis numbers mean p-value

- Note 2 : * p < 0.1, ** p < 0.05, *** p < 0.01

가 나타나는 것을 확인할 수 있었다. UPLOADER 변수의 10대, 20대, 30대 세대의 유동인구에 미치는 영향력은 남성 유동인구보다 여성 유동인구에 더 많은 영향을 미치는 것을 확인할 수 있었고, 이는 40대, 50대, 60대 세대의 유동인구에서도 똑같이 확인할 수 있었다. UPLOADER 변수는 모든 성별, 세대별 유동인구 중에서 10대, 20대, 30대 여성 유동인구에 가장 큰 영향을 미치는 것을 알 수 있었다. 이는 인스타그램을 사용하는 10대, 20대, 30대 여성 사용자의 비중이 높기 때문인 것으로 판단된다.

패널공간더빈모형을 이용하여 분석한 결과는 <Table 5>와 같다. 패널공간더빈모형을 사용하여 분석하기 앞서, 모형선정의 타당성

을 검증하기 위해 LM 검증을 이용하여 Spatial lag(ρ)와 Spatial Error(λ)에 대한 검증을 진행하였다(Elhorst, 2010). LM(lag) 검증을 통해 공간자기상관함인 $\rho=0$ 이라는 귀무가설을 기각하고 LM(error) 검증을 통해 $\lambda=0$ 이라는 귀무가설을 기각한다. 모든 종속변수에 대해 LM(lag), LM(error) 검증 결과가 통계적으로 유의미한 것을 <Table 5>에서 확인할 수 있었다. LM 검증결과 spatial lag, spatial error를 충분히 고려하는 모형인 패널공간더빈모형을 이용하는 것이 적합한 것을 알 수 있었다. 또한 공간자기상관함(ρ)의 추정 계수값이 모든 종속변수에 대해 유의미하기 때문에 성별, 세대별 유동인구는 인접한 지역의 영향을 상당히 받는

Table 6. Spillover effect

	YOUNGMAN	YOUNGWOMAN	OLDMAN	OLDWOMAN
Direct Effect				
LIKES	0.162 (0.452)	0.024 (0.388)	0.025 (0.114)	0.023 (0.200)
POSTS	-0.322 ** (0.037)	-0.461 ** (0.022)	-0.226 ** (0.044)	-0.229 * (0.080)
UPLOADER	0.428 ** (0.032)	0.617 ** (0.018)	0.300 ** (0.038)	0.319 * (0.060)
POPULATION	0.154 *** (0.000)	0.186 *** (0.000)	0.090 *** (0.000)	0.089 *** (0.000)
Indirect Effect				
LIKES	-0.026 (0.540)	-0.044 (0.463)	-0.004 (0.892)	-0.028 (0.405)
POSTS	0.475 (0.181)	0.425 (-0.501)	0.200 (0.375)	0.284 (0.289)
UPLOADER	-0.735 (0.116)	-0.716 (0.249)	-0.356 (0.231)	-0.477 (0.177)
POPULATION	0.026 (0.606)	0.033 (0.573)	0.017 (0.584)	0.020 (0.544)
Total Effect				
LIKES	-0.011 (0.832)	-0.200 (0.782)	0.021 (0.528)	-0.005 (0.905)
POSTS	0.154 (0.703)	-0.036 (0.946)	-0.255 (0.918)	0.055 (0.853)
UPLOADER	-0.307 (0.559)	-0.993 (0.887)	-0.055 (0.865)	-0.158 (0.684)
POPULATION	0.180 *** (0.000)	0.219 *** (0.000)	0.108 *** (0.000)	0.110 *** (0.000)

- Note 1 : Parenthesis numbers mean p-value

- Note 2 : * p < 0.1, ** p < 0.05, *** p < 0.01

것으로 해석된다. 따라서 공간요인을 반영하는 공간모형을 사용하여 분석하는 것이 적합한 것을 알 수 있었다.

패널공간더빈모형을 이용하여 분석한 결과 POSTS, UPLOADER, POPULATION 변수가 성별, 세대별 유동인구에 유의미한 영향을 주는 것을 확인할 수 있었고 각 독립변수의 종속변수에 대한 영향력의 경향은 패널모형의 분석결과와 같았다. 패널공간더빈모형의 추정결과와 비교하였을 때 패널모형의 추정 결과는 모든 종속변수에 대한 독립변수의 계수가 비슷한 경향을 보이지만 다소 차이가 나는 것으로 나타난다.

상기의 분석을 통해 공간적으로 인접한 개체에 대한 분석하는 경우, 자료 및 오차항의

공간적 자기상관을 충분히 반영한 모형을 사용하여야 편향되지 않고 효율적인 계수를 추정할 수 있는 것을 확인할 수 있었다.

공간계량모형을 사용하는 경우 비공간모형과 달리, 독립변수의 개체에 대한 직접효과와 인접한 지역에 미치는 영향력인 간접효과를 확인할 수 있는 장점이 있다. 개별 독립변수의 인접지역에 대한 영향력을 공간 전이효과(spillover effect)를 통해 분석할 수 있으며(Lesage, 2009), 이를 통해 인접한 개체 사이의 상호작용을 분석할 수 있다.

각 독립변수의 공간 전이효과는 <Table 6>과 같다. 직접효과와 간접효과 중에서 인스타그램 콘텐츠 변수인 POSTS, UPLOADER는 종속변수에 직접효과만 영향을 미치는 것으로 나타난다. 이와 같은 결과

는 인스타그램 콘텐츠 변수가 집계된 특정 집계구에만 영향을 미치며 인접한 집계구의 성별, 세대별 유동인구에는 영향을 미치지 않는 것을 의미한다. 인스타그램 콘텐츠는 공간전이효과가 유의미하게 나타나지 않는 것으로 확인되는데, 이는 특정 지역의 콘텐츠가 인접한 지역에 영향을 미치지 않는 것을 나타낸다. 특정 지역의 온라인 유명도는 특정 지역에만 영향을 미치며, 인접한 지역에는 영향을 주지 않기 때문에 온라인 유명도로 인한 집적효과는 나타나지 않는 것으로 볼 수 있다.

이상의 분석결과 다음과 같은 함의를 얻을 수 있었다. 온라인에서 생성되는 인스타그램 콘텐츠 변수가 유동인구를 설명하는 유의미한 변수가 될 수 있음을 확인하였다. 집계구와 관련된 인스타그램 콘텐츠를 생성하는 사용자 수(UPLOADER)가 유동인구에 긍정적인 영향을 미치는 것을 확인할 수 있으며 이태원 상권과 관련된 인스타그램 콘텐츠를 생성하는 사용자가 많을수록 상권의 유동인구에 긍정적인 영향을 미치는 것을 확인할 수 있었다. 또한 인스타그램 콘텐츠는 40대, 50대, 60대 이상 세대보다 10대, 20대, 30대 세대에게 더 많은 영향을 미치고, 남성보다 여성에게 더 많은 영향을 주는 것을 확인할 수 있었다.

5. 결론

본 연구는 인스타그램 콘텐츠 특성과 유동인구 사이의 관계를 분석하기 위해 2018년 1월부터 2019년 8월 사이에 이태원 상권의 65개 집계구에서 수집된 자료를 패널모형과 패널공간더빈모형을 이용하여 분석하였다. 종속변수인 성별, 세대별 유동인구에 대해 Global Moran's I 검증결과 및 각 종속변수의 패널모형에 대한 LM(lag), LM(error)의 검증결과 공간모형을 사용하는 것이 적합하였기 때문에 패널공간더빈모형을 사용하였다. 또한 패널공간더빈모형을 이용하여 분석한 결과와 비공간 모형인 패널모형을 이용하

여 분석한 결과를 비교하였다.

분석결과 패널모형과 패널공간더빈모형 모두 성별, 세대별 유동인구에 인스타그램 콘텐츠 변수인 POSTS(인스타그램 콘텐츠 수)와 UPLOADER(인스타그램 콘텐츠 업로더 수)가 유의미한 영향을 미치는 것을 확인하였다. 인스타그램 콘텐츠는 40대, 50대, 60대 이상 세대의 유동인구보다 10대, 20대, 30대 세대의 유동인구에 더 많은 영향을 주는 것을 확인할 수 있으며 남성 유동인구보다 여성 유동인구에 더 많은 영향을 주는 것을 확인할 수 있었다. 또한 개별 독립변수의 종속변수에 대한 공간 전이효과를 분석한 결과, POSTS 변수와 UPLOADER 변수는 직접효과만 존재하며 인접한 집계구에 대한 간접효과는 존재하지 않기 때문에 인스타그램 콘텐츠는 인접한 지역의 유동인구에 영향을 주지 않는 것으로 확인된다. 이와 같은 연구결과를 바탕으로 이태원 상권의 유동인구는 인스타그램 콘텐츠의 영향을 받으며, 특히 인스타그램 콘텐츠를 생성하는 사용자 수가 많은 것이 상권의 유동인구에 긍정적인 영향을 미치는 것을 확인할 수 있었다.

본 연구는 인스타그램 콘텐츠 특성과 유동인구 사이의 관계를 분석한 의의가 있으며, 온라인에서 발생하는 빅데이터가 현실에 미치는 영향에 대한 분석의 가능성을 제시하였다. 본 연구에서 제시한 방법을 이용하면 향후 재난 상황 등에서 소셜미디어 자료를 이용하여 특정 지역의 유동인구를 추정하는데 도움이 될 수 있을 것으로 예상된다.

다음과 같은 연구한계도 존재하였다. 인스타그램 콘텐츠 수 변수(POSTS)가 유동인구에 부정적인 영향을 미치는 것에 대한 추가적인 연구가 요구되며, 인스타그램 콘텐츠 수의 증가가 유동인구 미치는 정확한 효과를 분석하기 위해 향후 연구에서 시차를 고려한 동적인 모형을 사용하여 분석할 필요가 있다고 판단된다. 또한 인스타그램 자료를 수집하는 과정에서 발생하는 자료의 삭제 및 비공개 등의 문제로 인해 자료가 누락되는 문

제가 발생하기 때문에 자료의 정확도 문제를 보완하기 위한 자료수집 방법이 요구된다.

또한 향후 분석에서는 본 연구에서 다루지 못한 다양한 변수를 사용할 필요가 있다. 인스타그램 콘텐츠 변수에서 콘텐츠의 질적인 요소를 면밀히 분석에 반영하기 위해 소셜미디어 인플루언서의 영향력을 모형에 반영할 필요가 있다. 또한 전체 유동인구 모수에 비해 소셜미디어의 사용자 수가 충분히 크지 않기 때문에 유동인구를 설명할 수 있는 추가적인 인구통계 변수를 사용하는 것이 요구된다. 주중과 주말의 유입 및 유출인구, 주중 산업 인구, 요식업 종사자 수 등 유동인구에 영향을 미칠 수 있는 다양한 요인들을 모형에 추가할 필요가 있다. 마지막으로 공간특성에 따라서 공간방문에 대한 목적이 달라지기 때문에 상권의 공간특성을 변수로 반영할 필요가 있다. 상권은 중심 상권, 주변 상권, 주거지, 관광 명소 등 다양한 특성으로 이루어져 있다. 따라서 상권 내부의 다양한 공간적 특성을 추후 모형에 반영할 필요가 있다. 이 같은 연구한계에 대해서는 향후 연구를 통해 논의를 지속할 수 있기를 기대한다.

감사의 글

본 연구는 행정안전부 재난안전 산업육성 지원 사업의 지원을 받아 수행된 연구임 (No.20009742).

참고문헌

- [1] Anselin, L., 2010, Thirty years of spatial econometrics, *Regional science*, 89(1), 3-25
- [2] Anselin, L., Le Gallo, J. and Jayet, H., 2008, *Spatial panel econometrics*, pp. 625-660.
- [3] Baltagi, B. H., 2005, *Econometric analysis of panel data*, 3rd Ed., John Wiley & Sons. Chichester, UK, p. 302.
- [4] Cox, C., Burgess, S., Sellitto, C. and Buultjens, J., 2009, The Role of User-Generated Content in Tourists' Travel Planning Behavior, *Journal of Hospitality Marketing & Management*, Vol. 18, No. 8, pp. 743-764.
- [5] Dashti, S., Palen, L., Heris, M., Anderson, K. M., Anderson, J. and Anderson, S., 2014, Supporting Disaster Reconnaissance with Social Media Data: A Design-Oriented Case Study of the 2013 Colorado Floods, *Proc. of 11th International ISCRAM Conference*, University Park, Pennsylvania.
- [6] Elhorst, J. P., 2003, Specification and Estimation of Spatial Panel Data Models. *International Regional Science Review*, INT REG SCI REV, Vol. 26, pp. 244-268.
- [7] Elhorst, J. P., 2010, Spatial Panel Data Models, *Handbook of Applied Spatial Analysis*, Manfred M. Fischer and Arthur Getis (Eds), Springer, pp. 377-407.
- [8] ElQadi, M. M., Dorin, A., Dyer, A. G., Burd, M., Bukovac, Z. and Shrestha, M., 2017, Mapping species distributions with social media geo-tagged images: Case studies of bees and flowering plants in Australia. *Ecological Informatics*. 39.
- [9] Hausman, J. A., 1978, Specification Tests in Econometrics, *Econometrica*, Vol. 46, No. 6,

- pp. 1251-1271.
- [10] Hill, R. C., Griffiths, W. E., Lim, G. C., 2011, Principles of econometrics, 4th Ed., Wiley, p. 792.
 - [11] Kim, S. H., Kim, J. Y. and Lee, S. J., 2016, A Study on the Application of Spatial Big Data from Social Networking Service for the Operation of Activity - Based Traffic Model, The Journal of The Korea Institute of Intelligent Transportation Systems, Vol. 15, No. 4, pp. 44-53.
 - [12] Korea Information Society Development Institute (KISDI), 2019, 2019 Korea Media Panel Survey.
 - [13] Kwak, S. W., Kim, H. S., 2016, The Effect of Use Motives of Social Media on Dining-Out Purchase Decision Making, Korean Journal of Hospitality and Tourism, Vol. 25, No. 4, pp. 57-76.
 - [14] Kyung, S. W., Jeong, K. R., 2019, Who are the people leading the gentrification process in Itaewon?, Seoul Studies
 - [15] Lee, J. W., Park, C., 2019, The Effects of eWOM volume of Social Media on Corporate Performance Mediating Brand Equity : A Difference between Product and Service, korean management review, 48(3), 653-681.
 - [16] LeSage, J. P. and Pace R. K., 2009, Introduction to Spatial econometrics, CRC Press Taylor & Francis Group, Boca Raton, FL, USA.
 - [17] Mckenzie, G. and Slind, R. T., 2019, A user-generated data based approach to enhancing location prediction of financial services in sub-Saharan Africa, Applied Geography, Vol. 105, pp. 25-36.
 - [18] Min, I. S., Choi, P. S., 2019, STATA Panel data analysis, Jiphil media.
 - [19] Moran, P. AP., 1948, the interpretation of statistical maps, Journal of the royal statistical society. Series B (methodological), Vol. 10, No. 2, pp. 243-251.
 - [20] Park, S.,W., Kwon, O., B., Na, H., J., 2017, The Effect of Social Media Related to Informal Data of Big Data on Sales, Accounting information review, 35(2), pp.321-342
 - [21] Park, W. and Yu, K., 2015, Spatial Clustering Analysis based on Text Mining of Location-Based Social Media Data. Journal of Korean Society for Geospatial Information Science, Vol. 23, No. 2, pp. 89-96.
 - [22] Sakaki, T., Okazaki, M. and Matsuo, Y., 2010, Earthquake shakes Twitter users: real-time event detection by social sensors. In Proceedings of the 19th international conference on World wide web (WWW '10). Association for Computing Machinery, New York, NY, USA, pp. 851-860.
 - [23] Tobler, W. R., 1970, A computer

movie simulating urban Growth
in the Detroit Region, Economic
Geography, Vol. 46, pp. 234-240.