

도시 지역 트윗 데이터의 시간대별 공간분포 특성*

- 부산광역시를 사례로 -

A Study on the Spatial Patterns of Tweet Data for Urban Areas by Time

- A Case of Busan City -

구자용**

Ku, Cha Yong

Abstract

The process of spatial big data, such as social media, is being paid more attention in the field of spatial information in recent years. This study, as an example of spatial big data analysis, analyzed the spatial and temporal distribution of Tweet data based on the location and time information. In addition, the characteristics of its spatial pattern by times were identified. Tweet data in Busan city are collected, processed, and analyzed to identify the characteristics of the temporal and spatial pattern. Then, the results of Tweet data analysis were compared with the characteristics of the land type. This study found that spatial pattern of tweeting in the city was associated with given time periods such as daytime and nighttime in both weekdays and weekends. The spatial distribution patterns of individual time periods were compared with the characteristics of the land for the spatially concentrated area. The results of this study showed that tweeted data would be related to different spatial distribution depending on the time, which potentially reflects the daily pattern and characteristics of the land type of urban area to some extent. This study presented the possible incorporation of social media data, e. g. Tweet data, into the field of spatial information. It is expected that there will be more advantage to use a variety of social media data in areas such as land planning and urban planning.

Keywords: Spatial big data, Social media, Spatial pattern by time, Tweet data

1. 서 론

인터넷과 스마트폰으로 대표되는 현대의 정보 통신 기술은 우리의 일상생활에 필수적인 요소로 자리 잡고 있다. 우리는 매일 인터넷과 스마트폰을 통하여 매

일 엄청난 정보를 접하고 있으며 이러한 정보의 처리를 위한 정보통신 기술은 지속적으로 발달하고 있다. 최근 SNS(social networking service)로 대표되는 소셜 미디어(social media)의 등장으로 매일 엄청난 양의 정보가 생성되고 있으며, 이러한 대용량 자료의 처

* 본 연구는 2015학년도 상명대학교 교내연구비를 지원받아 수행하였음.

** 상명대학교 지리학과 교수 Geography, Sangmyung University (first author: koostar@smu.ac.kr)

리를 위하여 빅 데이터 분석이라는 새로운 정보처리 기술이 발달하기에 이르렀다. 특히 소셜 미디어의 등장으로 기존에 정보의 수요자였던 일반 대중이 이제는 정보의 제공자로 활동하기에 이르렀으며, 매일 수많은 사람들이 생성하는 데이터 역시 급증하게 되었다. 따라서 빅 데이터의 분석을 통하여 정보 제공자로서의 대중이 가지는 기호나 관심 분야를 분석할 수 있게 되었다.

이러한 빅 데이터 분석 기법의 등장은 공간 정보 분야의 연구에도 많은 영향을 끼치게 되었다. 공간 정보는 기존의 지리 정보뿐만 아니라 위성 영상, RFID 센서, 토지 정보 등 매일 엄청난 양의 정보가 생성되기 때문에 빅 데이터의 대표적인 사례라 할 수 있다. 따라서 공간 정보 분야에서도 이러한 공간 빅 데이터의 분석과 처리를 위하여 많은 연구가 수행되고 있다. 특히 정부 3.0이 추진되면서 국가 정책적인 차원에서 공간 빅 데이터의 구축과 활용을 위한 연구가 활발히 추진되어 왔다(김대중 2014).

위치 정보를 가진 소셜 미디어 자료는 정보를 제공하는 대중들의 공간적인 위치를 포함하고 있기 때문에 대표적인 공간 빅 데이터라 할 수 있다. 위치 정보를 가진 소셜 미디어 자료는 공간 정보 분석 기법을 통하여 공간적인 분포 패턴과 특성을 분석할 수 있으며, 이를 통하여 대중들의 특성을 공간적 차원에서 파악할 수 있다. 또한 소셜 미디어 자료는 게시 시간을 제공하고 있어, 시간적인 차원에서도 분석이 가능하다. 즉 게시물을 게재한 시간과 공간을 동시에 분석하여 대중의 특성을 시간과 공간적으로 파악할 수 있는 것이다. 특히 스마트폰과 SNS에 익숙해 있는 도시 지역의 경우 소셜 미디어 자료의 시간대별 공간분포를 통하여 도시 지역에 거주하는 대중의 일상생활의 특성을 시간과 공간적으로 파악할 수 있다. 즉, 개인이 머무르는 공간이 어느 시간대인가에 따라 그 공간의 기능적 특성을 파악할 수 있다. 소셜 미디어 자료의 특성을 시간과 공간적인 차원에서 분석함으로써, 도시

지역 공간의 기능적 특성을 파악할 수 있는 것이다.

이러한 시간대별 공간분포 분석을 통하여 대중의 일상생활 주기를 파악할 수 있으며, 이를 토대로 각 도시가 가지고 있는 특성(산업 도시 또는 관광 도시 등)도 파악할 수도 있다(Lee et al. 2013). 따라서 도시 지역의 소셜 미디어 자료를 이용하여 도시 내부의 토지이용 특성과 같은 지역적 특성을 파악할 수 있다.

본 연구에서는 소셜 미디어 자료의 공간 정보와 시간 정보를 바탕으로 시간대별 공간분포 특성을 분석하고 도시내부 토지이용 특성과 비교함으로써, 도시 및 토지 계획 분야에서 소셜 미디어 자료의 활용 가능성을 제시하고자 한다. 이를 위하여 도시 지역의 소셜 미디어 자료를 추출하고, 시간대별 공간분포 분석을 통하여 소셜 미디어 자료의 분포 특성을 파악하여, 그 지역의 토지이용 특성과 비교·평가하고자 한다.

본 연구에서는 소셜 미디어 자료로 트윗(Tweet) 데이터를 사용하였다. 트위터는 대표적인 소셜 미디어로 누구나 간단하게 자신의 의견이나 상태를 게시할 수 있어, 많은 사용자를 확보하고 있다. 특히 트윗 데이터를 수집하고 처리할 있는 API를 제공하고 있어, 소셜 미디어 자료의 마이닝과 분석에 주로 활용되고 있다. 본 연구에서는 트윗 데이터의 수집과 분석을 위하여 R 소프트웨어를 이용하였다. 오픈 소스 통계 소프트웨어인 R 소프트웨어는 통계 기능뿐만 아니라 다양한 비정형 데이터 처리 기능과 각종 API를 갖추고 있어 빅 데이터의 처리와 분석에 이용되는 대표적인 소프트웨어이다. 이에 R 소프트웨어를 이용하여 트윗 자료의 추출과 트윗 내용 분석을 수행하였다. 또한 위치 정보의 공간분포 분석을 위하여 공간정보 처리 소프트웨어인 ArcGIS를 이용하였다.

본 연구의 사례 지역은 부산광역시로 하였으며, 시간적으로는 2016년 9월 1일부터 30일까지의 자료를 활용하였다. 부산광역시는 우리나라 제2의 도시로 트윗과 같은 소셜 네트워크의 게시가 활발한 지역이며, 남포동을 중심으로 한 구도심과 서면과 해운대를 중

심으로 한 신도심 등과 같이 도시 내 다양한 토지이용 특성이 나타나고 있어 트윗 자료의 시간대별 공간분포 특성을 분석하기 용이하다는 특징을 가지고 있다. 따라서 본 연구에서는 부산광역시 지역에서 게시한 2016년 9월의 트윗 데이터를 수집하고 분석하여 시간대별 공간분포 특성을 파악하였다. 부산광역시 지역에서 게시한 트윗 데이터를 시간적으로 분석하기 위하여 트윗에 게시한 시간과 날짜를 기준으로 평일 주간(평일 오전 6시~오후 6시), 평일 야간(평일 자정~오전 6시, 오후 6시~자정), 휴일 주간(주말 및 추석 공휴일 오전 6시~오후 6시), 휴일 야간(주말 및 추석 공휴일 자정~오전 6시, 오후 6시~자정)으로 구분한 후, 각 시간대별 트윗 데이터의 공간적 분포 특성을 파악하였다.

2. 관련 연구

최근 정보통신 기술과 공간정보 처리 기술이 급격히 발달하면서 위치를 가진 수많은 공간 정보를 처리하기 위한 연구가 수행되어 왔다. 특히 공간 빅 데이터와 소셜 미디어의 개념이 등장할 때부터 공간 정보 분야에 미치는 영향에 대한 연구가 활발히 수행되었다. Sui and Goodchild(2011)는 공간정보 분야에서 소셜 미디어를 접목하여 나타날 수 있는 연구 주제와 사회적 영향에 대하여 예측하고, 앞으로의 연구 방향을 제시하였다. 그들은 위치 기반적인 소셜 미디어가 증가하면서 기존의 공간적인 개념을 넘어선 새로운 형태의 공간정보 처리 기술이 필요하며, 사회적으로도 데이터와 관계된 집단으로 구분되어 다양한 연구 주제가 등장할 것으로 예측하였다. Li et al.(2013)은 위치 기반 소셜 미디어인 트윗과 플리커 자료의 위치를 지도화하고 공간분석을 통하여 미국 지역의 사회경제적 특성을 파악하였다.

우리나라에서는 2012년 이후 정부 3.0을 표방한 정책들이 수립되고 추진되면서 국가적인 차원에서 빅

데이터, 특히 공간 빅 데이터에 대한 관심이 크게 증가하게 되었다(김대중 2014). 이를 위하여 기존의 국가 공간정보 통합체계에 공간 빅 데이터의 개념을 통합하여 국가적 차원에서 공간 빅 데이터를 구축하고 활용하기 위한 기초 연구가 수행되었다(안종욱 외 2013). 이러한 맥락에서 공간 빅 데이터의 대표적인 사례인 소셜 미디어 데이터의 위치 정보를 수집하고 지도화하는 연구가 활발히 진행되어 왔다.

박재희·강영옥(2014)은 서울시 지역의 트윗 데이터를 분석하여 주거 환경에 대한 만족도를 행정구역 단위로 파악하였으며, 신정엽(2014)은 미국 지역을 사례로 트윗 데이터의 공간적 분포를 지도화하고 공간적 분포 특성을 파악하였다. 또한 홍일영(2014)은 우리나라 전국의 트윗 데이터를 추출하고 지도화하여 시군구별로 분석하였고, 구자용(2015)은 서울시 지역의 트윗 데이터를 지도화하고 공간적 분포 특성을 파악하였다. 홍일영(2015)은 위치 정보를 가진 소셜 미디어인 포스퀘어를 이용하여 공간 분포를 지도화하고 분석하였다. 박우진·유기윤(2015)은 트윗 데이터의 공간 분석을 통하여 군집지역을 추출하고 각 군집지역별 트윗 내용을 분석하여 그 특성을 파악하였다. 이와 같이 소셜 미디어 자료를 이용하여 지역의 특성을 파악하고, 위치 정보를 이용하여 공간적 분포 특성을 분석하는 연구가 활발히 진행되어 왔다.

최근에는 소셜 미디어 자료를 분석하여 개인의 지리적 인식을 파악하거나 도시 지역의 기능적 특성을 파악하는 연구도 활발히 진행되어 왔다. Xu et al.(2013)은 소셜 미디어에 게시된 문장을 분석하여 지명과 같은 공간적 내용을 담고 있는 데이터와 그 공간적 분포를 비교함으로써 개인이 가지고 있는 지리적 인식을 파악하였다. Lee et al.(2013)은 소셜 미디어 자료의 위치 정보와 시간 정보를 이용하여 대중들의 생활 패턴을 유형화하고, 이를 이용하여 도시 지역의 기능적 특성을 파악하였다. 그들은 Figure 1와 같이 소셜 미디어에 나타난 대중들의 생활을 시간에 따

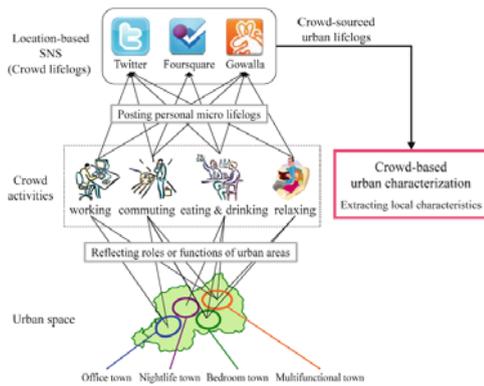


Figure 1. The model to characterize urban area using crowd behavior from social media data

Source: Lee et al. 2013

라 업무, 통근, 식사, 휴식 등으로 유형화하고, 이들 유형이 공간적으로 어느 지역에서 집약적으로 분포하고 있는가를 분석함으로써 도시 지역의 특성을 파악하였다. 그 결과 트윗 데이터의 시간대별 공간분포 분석을 통해 도시의 기능 지역을 파악할 수 있었다.

이와 같이 공간 정보 분야에서는 공간 빅 데이터와 소셜 미디어에 대한 개념이 등장한 이후, 소셜 미디어 자료를 활용한 연구가 크게 증가하고 있으며, 앞으로 도 소셜 미디어 데이터를 주축으로 한 공간 빅 데이터의 처리와 분석에 대한 연구가 활발히 진행될 것으로 전망된다.

3. 트윗 데이터의 수집과 분석

3.1. 트윗 데이터의 수집

트위터(Twitter)는 미국의 트위터 사가 운영하고 있는 소셜 네트워크 서비스로, 개인 사용자가 자신의 상태나 느낌 등을 간단한 문장으로 표현하여 공유하는 마이크로 블로그 서비스이다. 트위터는 누구나 자신의 내용을 표현하고 인터넷을 통해 공유할 수 있기

때문에 사용자 수는 폭발적으로 증가하였고, 최근에는 스마트폰의 등장으로 어디서나 간단히 정보를 게시하고 공유할 수 있게 되었다. 특히 트위터는 개방적으로 정보를 공유하기 때문에 정보의 전달이나 확산 측면에서 다른 소셜 네트워크에 비하여 더욱 영향력 이 큰 미디어라 할 수 있다.

트위터는 자체적으로 API를 제공하고 있어, 일반 사용자가 간단한 프로그램으로 트위터에 게시된 트윗 데이터를 검색하거나 분석할 수 있다. 이러한 프로그램을 이용하면 트윗 데이터를 수집하고, 빅 데이터 분석기법을 통하여 분석하여 필요한 정보를 추출할 수 있다. 트위터 API는 R 소프트웨어나 Python 등 오픈 소스 프로그램에서 활용할 수 있도록 제공하고 있다.

본 연구에서는 R 소프트웨어에서 트위터 API를 이용하여 트윗 데이터를 수집하였다. R 소프트웨어에는 트위터용 API인 “twitteR”과 “streamR” 등의 트윗 데이터 마이닝 패키지를 제공하고 있다. “twitteR”은 트윗에 게시된 데이터를 검색하는 기능을 가지고 있으며, “streamR”은 실시간으로 트윗에 게시되는 데이터를 검색하는 기능을 가지고 있다. 본 연구에서는 “twitteR” 패키지를 이용하여 부산시 지역의 트윗 데이터를 수집하였다.

R 소프트웨어에서 트위터 API를 사용하기 위해서는 트위터 애플리케이션 개발자 웹사이트(<http://apps.twitter.com>)에서 새로운 앱을 등록하고 사용에 필요한 Key와 Access Token 값을 받아야 한다. 트위터 API를 사용하기 위해서는 Consumer Key, Consumer Secret, Access Token, Access Token Secret 등의 값이 필요하다. R 소프트웨어에서는 이들 키 값을 변수로 지정한 후 “setup_twitter_oauth” 명령어를 통하여 트위터 API를 사용할 수 있는 권한을 부여받는다.

R 소프트웨어에서 트윗 데이터를 검색하는 기능은 “searchTwitter” 명령어를 통하여 수행된다. 검색할 단어를 키워드로 지정하면, 지정된 내용을 담고 있는 트윗 데이터를 검색할 수 있다. 또한 검색 옵션인

(location sparsity)이라 하며, 이러한 문제를 해결하기 위하여 사용자의 특성을 파악하여 거주 지역을 유추하는 연구가 진행되고 있다(강애띠·강영욱 2015).

본 연구에서 수집한 위치 정보를 가진 트윗 데이터의 분포는 Figure 4와 같다.

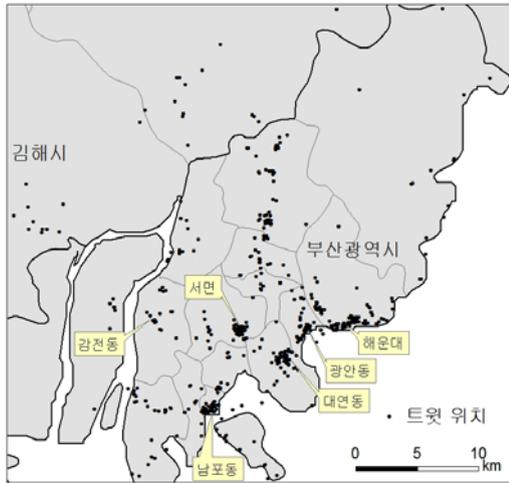


Figure 4. Location of Tweet data in Busan area

그림에서 나타나고 있는 바와 같이 부산광역시 지역에서 트윗 데이터가 집중적으로 분포하고 있는 지역은 전통적인 부산의 도심인 남포동 지역, 신도심이 형성된 서면 지역, 경성대와 부경대가 위치한 대연동 지역, 그리고 최근 급격히 성장하고 있는 광안동과 해운대 지역 등이다. 이들 지역은 부산시에서 가장 많은 인구가 활동하고 있는 지역이기도 하다.

공간적으로 위치한 트윗 데이터의 공간 분포 특성을 파악하기 위하여 커널을 이용한 공간밀도 분석을 수행하였다. 대상 지역을 일정 간격으로 분할한 후 점의 밀도 표현을 부드럽게 평활화하는 커널 함수를 이용하여 공간 밀도를 분석하였다. 커널 밀도 분석의 수식은 다음과 같다.

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

여기서 n은 점의 수이며, h는 평활화를 위한 밴드의 폭, K는 커널 함수, x는 추정할 지점의 위치, x_i 는 점 사상의 위치이다.

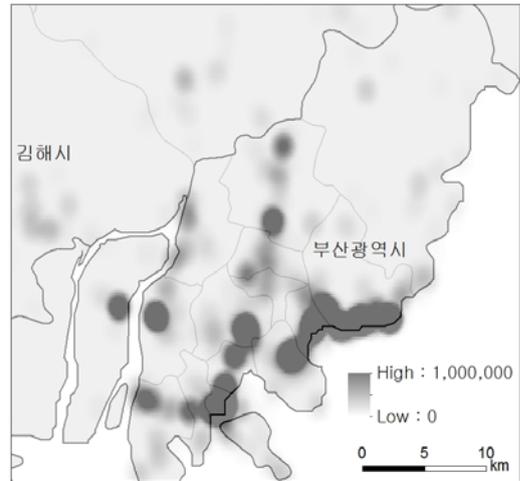


Figure 5. Density map of Tweet data in Busan area

공간 밀도의 표현 결과는 Figure 5와 같다. 그림과 같이 부산광역시에서 남포동, 대연동, 서면, 광안동, 해운대 등의 지역이 높은 밀도를 보이고 있으며, 이들 지역은 부산광역시 지역에서 대중의 일상활동이 가장 활발한 지역이라 할 수 있다.

4. 트윗 데이터의 시간대별 공간패턴 분석

4.1. 트윗 데이터의 시간대별 분석

본 연구에서 수집한 부산시 지역의 트윗 데이터를 시간대별로 분석하기 위하여 4가지 유형으로 구분하였

다. 우선 날짜를 기준으로 평일과 휴일(주말과 공휴일)로 구분한 후, 오전 6시부터 오후 6시까지를 주간, 그리고 나머지 시간을 야간으로 구분하였다. 이러한 구분에 의하여 평일 주간, 평일 야간, 휴일 주간, 휴일 야간의 4가지 시간대를 설정하였다. 이는 사람의 활동 패턴이 주간과 야간이 서로 다르며, 업무와 관계된 활동의 경우 평일과 휴일의 활동 패턴이 다르기 때문이다.

본 연구에서 수집한 트윗 데이터를 요일별로 활동 시간대에 따라 구분하고 트윗 데이터의 수를 집계한 결과는 Table 1과 같다. 표와 같이 전체 191,934건의 트윗 수 중에서 평일 주간이 89,658건으로 전체의 약 46.7%를 차지하고 있으며, 평일 야간은 36,660건으로 19.1%를 차지하고 있다. 따라서 사람의 일상활동이 가장 활발한 시기인 평일 주간에 가장 많은 트윗 데이터가 게시된다고 할 수 있다. 요일별로는 월요일과 목요일에 트윗 데이터의 게시가 많았으며, 화요일과 금요일은 상대적으로 게시 수가 적었다. 주말 또는 휴일의 경우, 주간은 45,863건으로 전체의 23.9%를 차지하고 있으며, 야간은 19,753건으로 나타났다. 토요일과 일요일에는 게시 건수가 평일 수준으로 나타났으나, 추석 연휴(14일~16일)의 게시 건수는 매우 적게 나타났다.

Table 1. Number of Tweet data by day and time.

Week days	Mon	Tue	Wed	Thu	Fri	Total
day	25,885 (13.5%)	12,681 (6.6%)	17,474 (9.1%)	19,039 (9.9%)	14,579 (7.6%)	89,658 (46.7%)
night	7,582 (4.0%)	6,467 (3.4%)	7,346 (3.8%)	8,183 (4.3%)	7,082 (3.7%)	36,660 (19.1%)
total	33,467 (17.4%)	19,148 (10.0%)	24,820 (12.9%)	27,222 (14.2%)	21,661 (11.3%)	126,318 (65.8%)

Weekend or Holiday	Sat	Sun	Holiday	Total
day	17,085 (8.9%)	17,674 (9.2%)	11,104 (5.8%)	45,863 (23.9%)
night	6,679 (3.5%)	7,333 (3.8%)	5,741 (3.0%)	19,753 (10.3%)
total	23,764 (12.4%)	25,007 (13.0%)	16,845 (8.8%)	65,616 (34.2%)

부산시 지역의 트윗 데이터를 시간대별로 파악한 결과 사람의 일상활동이 가장 활발한 평일 주간에 가장 많이 게시되었으며, 휴식 등으로 일상활동이 상대적으로 덜 활발한 휴일 야간의 경우 트윗 데이터의 수가 적었다.

시간대별로 구분된 트윗 데이터로부터 주요 단어를 추출하고 그 빈도수를 측정하여 버블 차트로 표현한 결과는 Figure 6과 같다. 그림에서 평일 주간에 지진, 속보 등 지진 관련 문구가 가장 많았으며, 평일 야간에 교수, 사이트, 민주주의 등 이념과 관계된 단어도 많이 나타나고 있다. 또한 휴일 주간에 민주주의, 시민동맹 등의 이념과 관계된 단어가, 휴일 야간에 사이트, 라이브, 게임 등 휴식이나 여가 활동과 관련된 내용이 트윗 데이터에 많이 나타나고 있다.

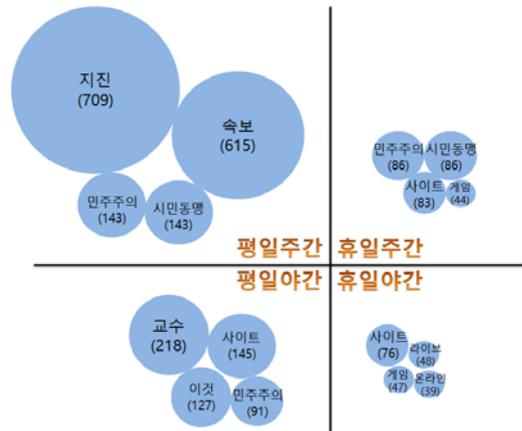


Figure 6. Frequent words on Tweet by day and time

이처럼 일상생활이 가장 활발한 평일 주간에는 안전과 같은 생활과 직접적으로 관련 있는 내용이, 평일 야간과 휴일 주간에 경우 학생의 토론과 관계된 활동이, 휴일 야간에 경우 온라인 게임과 같은 여가 활동과 관계된 내용이 많이 나타나고 있어, 시간대별로 일상생활의 특성을 나타내고 있다고 할 수 있다.

4.2. 시간대별 트윗 데이터의 공간 분포

부산시 지역의 트윗 데이터를 대상으로 시간대별로 위치 정보를 지도화하고, 공간 분포 분석을 수행하였다. 본 연구에서 수집한 트윗 데이터 중에서 위치 정보를 포함한 3,286건의 자료를 대상으로 평일 주간, 평일 야간, 휴일 주간, 휴일 야간 등 4개의 시간대별로 구분하고 공간분포를 파악하였다. 각 시간대별 트윗 데이터의 공간 분포 결과는 Figure 7에서 Figure 10과 같다. Figure 7은 평일 주간, Figure 8은 평일 야간, Figure 9는 휴일 주간, Figure 10은 휴일 야간의 트윗 데이터의 위치와 함께 밀도가 높은 지역을 지도로 표현한 것이다. 각 지도에서 검은 트윗 데이터의 위치를 표현한 것이며, 다각형 지역은 커널 밀도 분석 결과 상위 20%에 해당하는 고밀도 지역을 표현한 것이다. 커널 밀도는 3절에서 제시한 수식을 이용하였으며, 고밀도 지역을 일정 면적을 가진 다각형으로 표현하기 위하여 커널 밀도 값이 상위 20%에 해당되는 지역을 선정하였다.

평일 주간에서 트윗 데이터의 밀도가 높은 지역은 남포동과 대연동, 그리고 광안동으로 나타났다. 부산

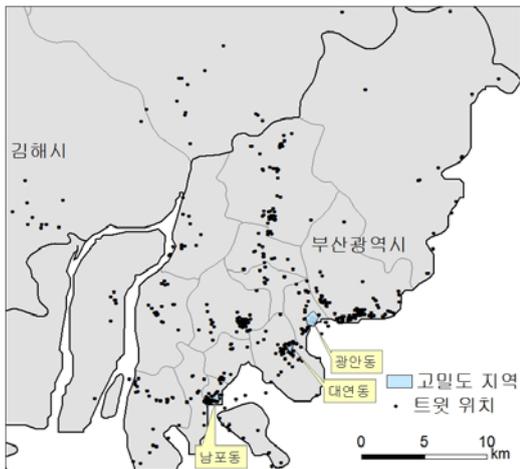


Figure 7. Spatial distribution of Tweet data at weekday day-time

의 전통적인 도심인 남포동은 많은 인구가 활동하고 있으며, 대연동은 경성대와 부경대 캠퍼스가 있어 대학생들이 많이 동하고 있는 공간이다. 또한 광안동은 최근 급격히 성장하고 있는 지역이다. 따라서 평일 주간에는 일상활동이 활발한 지역에서 트윗 데이터가 많이 발생하는 것으로 나타났다.

평일 야간의 경우 대연동과 광안동, 감전동 지역에서 트윗 데이터의 밀도가 높은 것으로 나타났다. 감전동의 경우 주거단지이고, 대연동과 광안동은 요식업 등이 분포하여 야간 활동이 활발한 지역이라 할 수 있다.

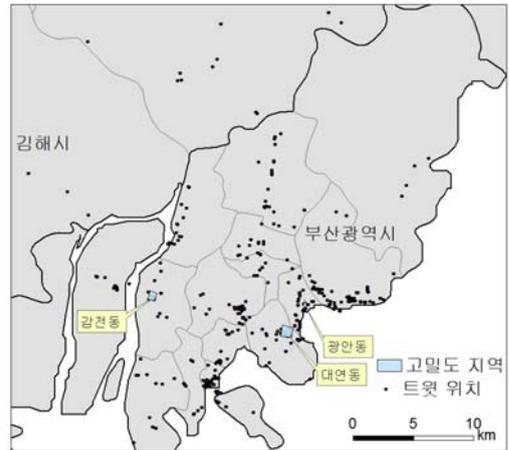


Figure 8. Spatial distribution of Tweet data at weekday night-time

휴일의 경우 주간과 야간 모두에서 해운대 지역의 트윗 데이터 밀도가 높게 나타났다. 해운대 지역은 부산의 대표적인 관광지로 휴일에 여가 활동이 활발히 이루어지고 있음을 나타내고 있다. 휴일 야간의 경우 남포동과 감전동 등도 트윗 데이터의 밀도가 높아 활동이 활발한 지역임을 알 수 있다.

부산시 지역의 트윗 데이터를 시간대별로 파악한 결과, 평일 주간에는 업무와 교육 기능이 높은 도심 지역에서 트윗 데이터의 밀도가 높았으며, 야간이나 휴일에는 휴식이나 여가와 관련된 주거지나 관광지에서

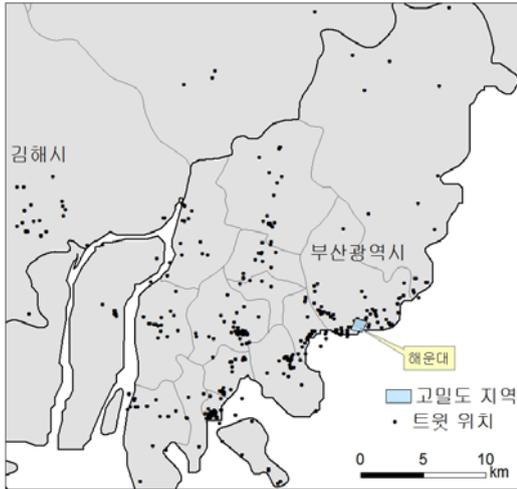


Figure 9. Spatial distribution of Tweet data at holiday day-time

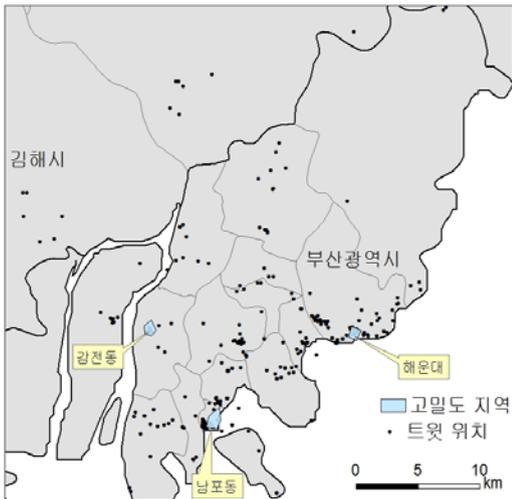


Figure 10. Spatial distribution of Tweet data at holiday night-time

트윗 데이터의 밀도가 높은 것으로 나타났다.

따라서 트윗 데이터의 밀도와 도시내부의 토지이용 간에는 어느 정도의 관계가 있는 것으로 볼 수 있다.

4.3. 시간별 트윗 데이터의 공간분포 특성

부산시 지역에서 수집한 트윗 데이터를 대상으로

시간대별로 공간 분포를 파악하고 그 결과를 부산시 지역의 토지이용 특성과 비교하였다. 평일의 경우 대연동, 광안동 등 평상시 유동인구가 많은 지역에서 많은 트윗 데이터가 게시되었으며, 휴일의 경우 여가 활동이 활발한 해운대에서 트윗 데이터가 집중되어 나타났다. 이는 트윗 데이터의 시간과 위치 데이터를 이용하여 그 지역의 토지이용 특성을 파악할 수 있음을 보여주고 있다.

트윗 데이터가 시간대에 따라 공간적으로 밀집하여 분포한 지역이 토지이용 특성과 어느 정도 일치하는가를 확인하기 위하여 이광국·조승규(2007)가 연구한 부산시 지역의 상업기능 분포 결과와 비교하였다. 그 결과 음식점이 특화된 지역인 남포동, 광안동, 서면 등의 상업 지역이 평일 주간의 트윗 데이터 밀집 지역과 일치하고 있으며, 일반 소매업이 특화된 감전동 등 주거 지역은 평일 야간의 트윗 데이터 밀집 지역과 일치하는 것으로 나타났다. 또한 숙박업이 특화된 해운대 지역은 휴일의 트윗 데이터 밀집지역과 일치하고 있다. 따라서 평일 주간에는 음식점 등 유동인구가 활발한 상업 지역에서, 평일 야간에는 주거 지역에서, 휴일에는 여가와 관계된 지역에서 트윗 데이터가 집중적으로 게시되고 있음을 알 수 있다.

트윗 데이터의 시간별 공간분포 패턴을 통계적으로 파악하기 위하여 트윗 데이터의 게시 시간을 숫자로 변환하고 핫 스팟 분석을 실시하였다. 핫 스팟 분석은 비슷한 속성을 가진 개체가 공간적으로 군집되어 있는가를 통계적으로 분석하는 기법으로, 본 연구에서는 Getis-Ord의 GI* 통계를 이용하여 분석을 수행하였다. GI* 통계의 수식은 다음과 같다.

$$GI_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{[n \sum_{j=1}^n w_{i,j}^2 - (\sum_{j=1}^n w_{i,j})^2]}{n-1}}}$$

여기서 x_j 는 공간사상 j 의 속성 값이고, $w_{i,j}$ 는 i 와 j 의 공간 가중치이며, n 은 전체 사상의 수이다. 또한 \bar{X} 와 S 는 다음과 같다.

$$\bar{X} = \sum_{j=1}^n x_j, S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2}$$

핫 스팟 분석의 결과는 Figure 11과 같다. 그림에서 속성 값의 표준 점수가 99% 유의 수준보다 높은 값이 군집된 지역을 핫 스팟(hot spot), 속성 값의 표준 점수가 99% 유의 수준보다 낮은 값이 군집된 지역을 콜드 스팟(cold spot)으로 표현하였다.

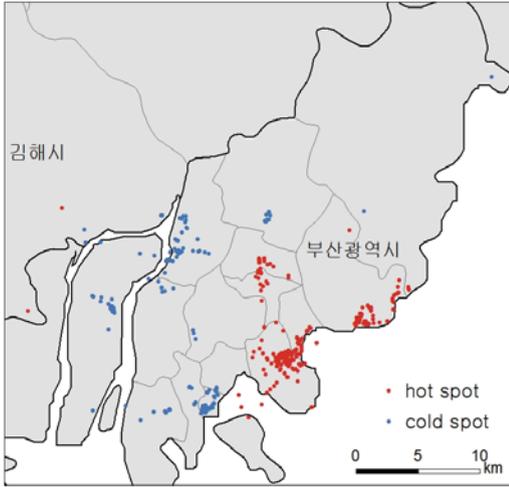


Figure 11. Hot spot analysis of Tweet data by time

부산시 지역의 시간별 트윗 데이터를 분석한 결과, 높은 값이 군집된 핫 스팟은 대연동, 서면, 해운대, 광안동 등 주로 부산의 동쪽 지역에 분포하고 있으며, 낮은 값이 군집된 콜드 스팟은 남포동, 감전동 등 주로 부산의 서쪽 지역에 분포하는 특징을 보이고 있다. 속성 값이 높은 개체는 트윗 게시 시간이 늦은 시간의

미하며, 낮은 개체는 트윗 게시 시간이 이른 시간의 미한다. 따라서 핫 스팟 지역은 오후의 활동이 활발하고 콜드 스팟 지역은 오전 활동이 활발하다고 할 수 있다. 즉 남포동과 감전동 지역은 오전 활동이 활발하며, 대연동, 해운대 지역은 오후 활동이 활발하다고 할 수 있다.

부산시 지역에서 수집한 트윗 데이터의 시간대별 공간분포를 파악한 결과 트윗 데이터를 게시한 시간대에 따라서 공간적인 분포 특성이 다르게 나타나고 있으며, 트윗 데이터가 게시된 지역의 토지이용 특성을 어느 정도 반영하고 있다고 할 수 있다. 특히 도심 지역과 상업 지역의 경우 트윗 데이터의 밀도가 평일 주간에는 높았으며, 거주 지역의 경우 평일 야간, 여가 지역의 경우 휴일에 많은 트윗 데이터가 게시되고 있다. 이는 트윗 데이터를 이용하여 그 지역의 토지이용 특성을 파악할 수 있다는 가능성을 보여주고 있다. 향후 보다 많은 소셜 데이터를 분석한다면 도시 지역의 일상활동 공간을 분석할 수 있으며, 나아가 토지 계획이나 도시 계획 등의 분야에도 소셜 미디어 자료를 활용할 수 있을 것으로 전망된다.

5. 결론

정보의 홍수 속에 살고 있는 현대 사회에서 소셜 미디어를 통해 생성되는 수많은 정보는 공간정보 분야에서 또 다른 정보의 원천으로 활용할 수 있다. 공간정보 빅 데이터 분석을 통하여 대중들의 공간적인 활동이나 관심 등을 분석할 수 있기 때문이다.

본 연구에서는 공간정보 빅 데이터로서 소셜 미디어의 활용 가능성을 확인하고 토지 계획이나 도시 계획 등 다양한 분야에서의 활용 사례를 제시하고자 하였다. 구체적으로 부산시 지역의 트윗 데이터를 이용하여 시간대별 공간분포 패턴을 분석하여 그 특성을 파악하였다. 이를 위하여 부산시 지역의 트윗 데이터를 게시 시기에 따라 평일 주간, 평일 야간, 휴일 주간,

휴일 야간으로 구분하고 각 시간대별로 트윗 데이터의 내용과 공간적 분포 특성을 비교하였다. 또한 트윗 데이터의 분포 특성과 도시내부의 토지이용 특성을 비교함으로써 공간정보 분야에서 트윗 데이터의 활용 가능성을 파악하였다.

본 연구의 결과는 다음과 같다. 첫째, 부산시 지역 트윗 데이터의 내용을 분석한 결과 시간대에 따라 일상생활의 특성과 관계된 내용이 나타났다. 평일 주간에는 안전이나 이념에 대한 내용이, 휴일이나 야간에는 여가 활동과 관계된 내용이 주로 나타나고 있어, 시간대별로 트윗 데이터의 특성을 파악할 수 있다. 둘째, 트윗 데이터의 시간대별 공간분포를 파악한 결과 시간대에 따라서 공간적으로 집중하는 지역이 특징적으로 나타나고 있다. 평일 주간에는 업무와 교육 기능이 높은 지역에서, 야간이나 휴일에는 휴식이나 여가와 관련된 지역에서 트윗 데이터가 밀집하고 있어, 트윗 데이터의 밀도와 도시내부 토지이용 특성과의 관련성을 보여주고 있다. 셋째, 트윗 데이터의 시간대별 공간 분포 특성과 토지이용 특성을 비교한 결과 트윗 데이터의 시간대별 공간분포는 해당 지역의 토지이용 특성을 다소 반영하고 있다. 도심 지역과 상업 지역은 평일 주간의 트윗 데이터와, 거주 지역은 평일 야간의 트윗 데이터와, 관광 지역은 휴일의 트윗 데이터와 관련을 보여주고 있어, 트윗 데이터의 시간대별 공간분포 패턴과 토지이용 특성은 어느 정도 관계하고 있다고 할 수 있다. 따라서 트윗 데이터의 시간대별 공간 분포 분석을 통하여 해당 지역에서의 일상생활 패턴과 특성을 파악할 수 있으며, 나아가 그 지역의 토지이용 특성도 어느 정도 파악할 수 있다.

본 연구는 소셜 미디어 자료 중에서 트윗 데이터만을 한정하여 분석하였으며, 수집된 자료 또한 30일간의 트윗 데이터로 한정되어 있어, 본 연구의 결과를 도시 지역의 특성으로 일반화하기에는 한계가 있다. 또한 본 연구에서 수집한 트윗 데이터 중에서 위치 정보를 가지고 있는 데이터는 전체 자료의 1.7%에 불과하

여 소셜 미디어 자료가 가지는 공간 희박성의 한계를 가지고 있다. 이러한 공간 희박성의 문제를 해결하기 위하여 소셜 미디어 자료의 위치를 추정하는 연구가 활발히 진행되고 있다(강애띠·강영옥 2015; Kim et al. 2016). 향후 이러한 연구의 성과로 공간 희박성의 문제가 해결되어 보다 풍부하고 다양한 소셜 네트워크 데이터를 이용하여 공간 분석을 수행할 수 있을 것이다.

이러한 한계에도 불구하고 본 연구에서는 트윗 데이터의 활용 가능성을 제시함으로써 향후 다양한 분야에서 소셜 미디어 자료를 활용할 수 있기를 기대한다. 보다 풍부한 소셜 미디어 데이터를 분석한다면 일상생활 공간을 보다 정확하게 분석할 수 있으며, 나아가 도시 지역의 다양한 특성을 보다 정확하게 파악할 수 있을 것이다. 이러한 소셜 미디어 데이터의 시간대별 공간분포 분석은 향후 토지 계획이나 도시 계획 분야에서도 활용될 수 있을 것으로 사료된다.

참고문헌

References

- 강애띠, 강영옥. 2015. 타임라인 데이터를 이용한 트위터 사용자의 거주 지역 유추방법. 한국공간정보학회지. 23(2): 69-81.
- Kang AT, Kang YO. 2015. Location inference of Twitter users using timeline data. *Journal of Korea Spatial Information Society*. 23(2): 69-81.
- 구자용. 2015. 공간정보 빅 데이터의 지도화와 공간적 분포 특성에 관한 연구 : 서울시 지역의 트윗 데이터를 사례로. 국토지리학회지. 49(3): 349-360.
- Ku CY. 2015. The study on the mapping and spatial distribution analysis for spatial big data: the case study on the twit data of Seoul area. *The Geographical Journal of Korea*. 49(3): 349-360.

- 김대중. 2014. 정부 3.0을 위한 공간 빅 데이터 구축 및 활용방안. 국토. p. 42-51.
- Kim DJ. 2014. *Building and Application of spatial big data for government 3.0*. Planning and Policy. p. 42-51.
- 박우진, 유기윤. 2015. 위치기반 소셜 미디어 데이터의 텍스트 마이닝 기반 공간적 클러스터링 분석 연구. 한국지형공간정보 학회지. 23(2): 89-96.
- Park WJ, Yu KW. 2015. Spatial clustering analysis based on text mining of location-based social media data. *Journal of Korean Society of Geospatial Information Science*. 23(2): 89-96.
- 박재희, 강영옥. 2014. 트윗을 이용한 서울시 주거환경 만족의 공간적 특성 분석 - 도시정책지표 보안을 위한 활용방안 모색. 한국도시지리학회지. 17(1): 43-56.
- Park JH, Kang YO. 2014. An analysis of spatial characteristics of residential satisfaction in Seoul using Tweet data: an applicability of Tweet data for complementing urban policy indicators. *Journal of the Korean Urban Geographical Society* 17(1): 43-56.
- 신정엽. 2014. 정보 격차의 맥락에서 트윗 데이터의 이론적 고찰과 실증적 공간 탐색: 미국 킹 카운티를 사례로. 한국지도학회지. 14(2): 89-106.
- Shin JY. 2014. Theoretical review and quantitative spatial exploration of Tweet data in the context of digital device: case of King county, US. *Journal of Korean Cartographic Association*. 14(2): 89-106.
- 안종욱, 이미숙, 신동빈. 2013. 공간 빅 데이터 개념 및 체계 구축방안 연구. 한국공간정보학회지. 21(5): 43-51.
- Ahn JW, Yi MS, Shin DB. 2013. Study for spatial big data concept and system building. *Journal of Korea Spatial Information Society*. 21(5): 43-51.
- 이광국, 조승구. 2007. 부산시 상업기능의 도시공간적 분포변화 특성. Journal of the Korean Data Analysis Society. 9(3): 1443-1456.
- Lee KK, Jo SK. 2007. A Study on characteristics of the periodical change by the urban spatial location of commercial function in Busan. *Journal of the Korean Data Analysis Society*. 9(3): 1443-1456.
- 조영임. 2013. 빅 데이터의 이해와 주요 이슈들. 한국 지역정보학회지. 16(3): 43-65.
- Cho YI. 2013. Understanding and main issues of big data. *Journal of Korean Association for Regional Information Society*. 16(3): 43-65.
- 홍일영. 2015. 국내 지오투잇의 공간분포. 한국지도학회지. 15(2): 93-101.
- Hong IY. 2015. Spatial distribution of Korean Geotweets. *Journal of Korean Cartographic Association*. 15(2): 93-101.
- 홍일영. 2016. 국내 위치기반 소셜 네트워크 데이터의 공간분포. 한국지도학회지. 16(2): 95-104.
- Hong IY. 2016. Spatial distribution of LBSN data in Korea. *Journal of Korean Cartographic Association*. 16(2): 95-104.
- Kim MG, Kang YO, Lee JY, Koh JH. 2016. Inferring tweet location inference for twitter mining. *Spatial Information Research*. 24(4): 421-435.
- Lee R, Wakamiya S, Sumiya K. 2013. Urban area characterization based on crowd behavioral lifelogs over Twitter. *Personal Ubiquitous Computing*. 17:605-620.
- Li L, Goodchild M, Xu B. 2013. Spatial, temporal and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic*

- Information Science*, 40(2): 61-77.
- Malleson N, Andersen MA. 2015. The impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science*, 42(2): 112-121.
- Miller HJ. 2004. Activities in space and time. in Hensher DA, Button KJ, Haynes KE, Stopher P.(eds) *Handbook of Transport Geography and Spatial Systems*. Elsevier. Lodon.
- Steiger E, Westerholt R, Resch B, Zipf A. 2015. Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban System*, 54: 255-265.
- Sui D, Goodchild M. 2011. The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographic Information Science*, 25(11):1737-1748.
- Xu C, Wong DW, Yang C. 2013. Evaluating the “geographical awareness” of individuals: an exploratory analysis of twitter data. *Cartography and Geographic Information Science*, 40(2):103-115.

2016년 10월 07일 원고접수(Received)

2016년 10월 31일 1차심사(1st Reviewed)

2016년 11월 22일 2차심사(2st Reviewed)

2016년 12월 07일 게재확정(Accepted)

초 록

최근 공간 정보 분야에서 소셜 미디어와 같은 공간 빅 데이터의 분석과 처리에 많은 관심이 집중되고 있다. 본 연구에서는 공간 빅 데이터 분석의 한 사례로서 트윗 데이터가 가지고 있는 위치 정보와 시간 정보를 바탕으로 시간대별로 공간분포를 분석하고 그 특성을 파악하였다. 부산시 지역의 트윗 데이터를 수집하고, 시간대별 공간분석을 통하여 그 특성을 파악하여, 그 지역의 토지이용 특성과 비교하였다. 부산시 지역의 트윗 데이터를 시간대에 따라 평일 주간, 평일 야간, 휴일 주간, 휴일 야간으로 구분하고, 각 시간대별로 공간적 분포 특성을 파악하여, 공간적으로 집중된 지역의 토지이용 특성과 비교하였다. 본 연구의 결과 트윗 데이터는 시간대에 따라 공간분포가 다르게 나타나고 있으며, 이는 그 지역의 일상생활 패턴과 토지이용 특성을 어느 정도 반영하고 있었다. 본 연구에서는 공간정보 분야에서 트윗 데이터와 같은 소셜 미디어 자료의 분석을 통한 활용 가능성을 제시하였다. 향후 토지 계획이나 도시 계획 등의 분야에서 다양한 소셜 미디어 자료를 활용할 수 있을 것으로 전망된다.

주요어 : 공간 빅 데이터, 소셜 미디어, 시간대별 공간분포, 트윗 데이터