

# 트위터 사용자의 도시 내 활동반경과 거주지역의 탐색: 라스베이거스 사례

조 재 희\*, 서 일 정<sup>o</sup>

## Investigation of Twitter Users' Activity Radius and Home Region in the City: The Case of Las Vegas

Jaehee Cho\*, Il-Jung Seo<sup>o</sup>

### 요 약

본 연구는 전 세계에서 발생한 200,578,703건의 지오투잇을 수집하여 트윗 봇을 제거한 후, 인간의 도시 내 이동패턴을 분석하였다. 활동반경(Activity Radius)이라는 개념을 이용하여 트위터 사용자를 구분하였으며, 거주지역을 국내와 국외로 구분하고 국내는 다시 시내와 시외로 구분하였다. 그리고 활동반경과 거주지역에 따라 트위터 사용자의 활동성과 활동지역에 대한 통계적 특성을 기술하였고 지리적 분포를 시각적으로 표현하였다. 라스베이거스를 대상으로 하는 사례 분석을 통해, 거주지역에 따른 활동성과 활동지역의 차이를 확인하였다. 향후 본 연구의 방법에 따라 다양한 도시를 대상으로 분석을 수행하면, 인간의 이동성에 대한 다양한 이론을 도출할 수 있을 것이다.

**Key Words** : human mobility, big data, geotagged tweets, home region, activity radius, central location

### ABSTRACT

In this study, we collected 200,578,703 geo-tweets and removed the twitter bots. Using the concept of activity radius, Twitter users are classified. Users are also divided first into domestic and overseas, and again domestic ones are divided into locals and non-locals. Statistical characteristics of activity strength and active area of Twitter users are described according to activity radius and home region, and the geographical distribution is presented visually. Through a case study of Las Vegas, we have identified the difference in activity strength and active area by the user's home residence. We expect to derive theories about human mobility by analyzing various cities with the method proposed in this study.

### I. 서 론

소셜 네트워크 서비스(SNS)에서 발생하는 빅데이터는 활용가치가 높아져, 많은 연구자들이 유용한 정보를 산출하기 위한 다양한 분석 방법론과 모델을 적용하고 있다<sup>1,2)</sup>. SNS 중 트위터는 연구자에게 거의

무제한으로 데이터를 제공하고 있어서 트윗 데이터를 이용한 연구가 급증하고 있다. 특히 GPS 좌표가 포함되어 있는 지오투잇 데이터는 순도 100%의 시공간 데이터인데, 이 데이터를 이용한 인간의 이동 행태에 관한 연구가 새로운 전기를 맞이하고 있다. 이제까지는 휴대폰 데이터<sup>3)</sup>, 교통카드 데이터, 차량 GPS 데이

\* 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음 (IITP-2016-R2718-1 6-0014)

• First Author : Kwangwoon University College of Business, mis1@kw.ac.kr, 정희원

o Corresponding Author : Kwangwoon University College of Business, iljung.seo@gmail.com, 정희원

논문번호 : KICS2017-01-010, Received January 10, 2017; Revised January 28, 2017; Accepted January 31, 2017

터, 신용카드 데이터 등을 이용하여 이동성을 분석하였다. 하지만 이러한 비즈니스 데이터들은 고객정보 및 경영정보 유출 문제 때문에 일반에 제한적으로 공개되었다. 따라서 데이터를 얻게 되더라도 몇 가지 제약이 존재한다. 첫째, 공개되는 양이 적으며, 둘째, 국내 데이터이기 때문에 다른 국가와 비교연구가 어렵고, 셋째, 고객의 위치정보가 정밀하지 않다. 그에 반해 트위터 데이터는 무제한으로 다운로드 받을 수 있으며, 지역적 제한이 없어 글로벌 수준의 비교연구가 가능하며, GPS 위치 값이 포함된 지오투잇을 수집할 경우 개인의 상세한 시공간 이동성을 분석할 수 있다. GPS 좌표는 위도와 경도 값이 포함되며, 오차가 10미터 이내이기 때문에 도시 내, 국가 내, 그리고 국가 간 등 사람이 실제로 이동하는 공간에 대한 분석을 가능하게 한다. 이와 같은 다양한 수준의 이동성 분석은 사회·경제적으로 의미 있는 정보를 제공한다.

본 연구는 전 세계에서 발생한 200,578,703건의 지오투잇을 수집하여 트윗 봇을 제거한 후, 인간의 도시 내 이동패턴을 분석하였다. 활동반경(Activity Radius)이라는 개념을 이용하여 트위터 사용자를 구분하였으며, 거주지역을 국내와 국외로 구분하고, 국내는 다시 시내와 시외로 구분하였다. 그리고 활동반경과 거주지역에 따라 트위터 사용자의 활동성과 활동지역에 대한 통계적 특성을 기술하였고 지리적 분포를 시각적으로 표현하였다.

## II. 관련 연구

### 2.1 활동반경

트위터 사용자의 활동반경에 대한 연구는 거리를 측정하는 방법, 격자 셀을 방문한 개수로 측정하는 방법, 그리고 이 두 가지를 혼합한 연구가 있다. 거리를 측정하는 방법은 사용자의 시간 변화에 따른 위치 변화(이전 트윗 위치와 현재 트윗 위치)에 대한 정보를 이용하여 트윗 순서, 트윗 거리, 시간 간격, 이동 속도 등과 같은 이동성 지수를 계산하거나<sup>[4]</sup>, 사용자의 트윗들의 회전반경(중심점과의 거리)을 이용하여 활동반경을 측정한다<sup>[5]</sup>. [5]의 연구에 의하면, 뉴질랜드와 호주 같은 원격지에 떨어져 있는 국가의 사용자는 활동반경이 800km 가까이 월등하게 크고, 이스라엘은 400km로 3위인데, 주변 중동지역 내에는 갈 곳이 없고, 일단 이스라엘이라는 국가를 벗어나면 다른 대륙으로 이동해야만 하는 정치상황과 무관하지 않은 결과를 보인다. 격자 셀 방식에 의한 활동반경 연구는 분석 대상이 되는 지역을 일정 규격의 정사각형 셀로

분할한 후 사용자가 얼마나 많은 셀을 방문했는지 집계하는 방식으로 활동성을 평가한다<sup>[6,7]</sup>.

위 두 가지를 혼합한 방법을 사용한 [8]은 세계 주요 58개 도시를 대상으로, 트위터 사용자를 현지인(local)과 방문자(non-local)로 구분한 후, 두 집단별 트윗 평균 반경과 방문셀 수를 계산하였다. 그 결과 현지인은 방문자에 비해 활동반경과 방문셀 수 모두 현격하게 적다는 사실을 발견하였다.

### 2.2 거주지역

사용자의 거주지를 알 수 있다면, 사용자의 이동패턴과 라이프스타일에 관해 더 높은 수준의 통찰을 얻을 수 있다. 어떤 장소(예를 들어, 에펠탑이나 경복궁)를 방문하는 사람이 그 도시의 거주자인지, 아니면 시외 또는 해외 여행객인지 여부에 따라 활동반경과 이동패턴이 다르다. [8]은 어떤 장소(point)에서 발생한 트윗의 빈도를 면(area)이라는 공간 단위로 집계하기 위해서 방문셀(cell coverage) 개념을 도입했는데, 관광객은 현지인에 비해 약 3배 많은 장소를 방문한다고 보고하였다.

거주지를 판별·추정할 때 이용하는 트위터 데이터는 (1) 사용자의 트윗값, (2) 사용자의 프로필 정보, 또는 (3) 트윗 위치(트윗이 발생한 GPS 좌표값)이다. 사용자의 트윗글을 이용할 경우, 도시나 주와 같은 행정구역 정도를 판별하는 수준이거나<sup>[9]</sup>, 오차범위가 100마일 정도로 정교성이 떨어진다<sup>[10]</sup>. 사용자 프로필 정보를 이용한 [11,12]는 사용자가 임의로 입력한 주소 정보와 자동 생성된 시간대 정보를 비교하여 사용자 위치를 추정하였다.

트윗 위치를 얻기 위해서는 지오투잇 데이터를 수집해야 하며, 다양한 휴리스틱을 이용한다. [5]는 국가 간 이동(해외여행)에 관한 연구이기 때문에 트윗이 가장 많이 발생한 국가를 사용자의 거주지(출발지)로 판별하였다. [13]은 사용자가 가장 긴 시간 동안 가장 많은 트윗을 한 국가를 사용자의 거주지로 판별하였다. 또한 사용자가 하루 중 마지막 트윗을 한 곳(last destination of a day)을 거주지 측정 모델의 변수로 사용한 연구도 있다<sup>[14]</sup>. [7]은 트윗 수로 거주지를 판별할 때 생기는 문제점을 인식하여 트윗 일수(number of tweet days)가 가장 큰 곳을 거주 국가(한국 또는 타국)와 거주 도시(서울 또는 지방)로 판별하였다. [8]은 가로, 세로 각변이 100km인 그리드 5,000개를 지표면에 배치하고, 사용자별로 트윗 수가 가장 많이 발생한 그리드를 사용자의 거주지로 판별하였다.

### III. 분석 과정

#### 3.1 데이터 준비

트위터의 Public Streaming API를 이용하여 2015년 7월 14일부터 2015년 9월 1일까지 50일 동안 전세계에서 발생한 200,578,703건의 지오프윗을 수집하였다. 트위터 사용자의 도시 내 이동패턴을 분석하기 위하여 라스베이거스를 선택하였다. 라스베이거스는 미국 서남부 네바다주에 위치한 세계적인 관광 도시로 거주자와 방문자의 이동패턴을 비교하기에 적절한 도시이다.

전체 데이터에서 지오프윗의 경위도 좌표를 이용하여 라스베이거스에서 발생한 트윗을 선택하였다(위도: 35.944955~36.355835, 경도: -115.367373~-114.902359). 그리고 인간의 이동에만 집중하기 위하여 트윗 봇을 제거하였다. 트윗 봇에 관한 선행 연구를 검토하여 총 이동거리의 합이 100m 미만인 사용자, 이동속도가 시속 1,000km 이상인 사용자, 활동반경의 표준편차가 0.0m 이하인 사용자 등을 트윗 봇으로 판단하고 삭제하였다. 이상의 과정을 통해 12,191개 트위터 계정의 371,532개 트윗 위치를 활동반경의 계산에 사용하였다.

#### 3.2 활동반경 계산

사용자의 활동반경을 계산하기 위하여 우선 중심위치의 경위도 좌표를 산출하였다. 중심위치의 경위도 좌표는 수식 (1)과 같이 해당 사용자의 모든 트윗 위치의 경위도 좌표를 평균하여 계산하였다.

$$Lat_c = \frac{1}{n} \sum_{i=1}^n Lat_i, Lon_c = \frac{1}{n} \sum_{i=1}^n Lon_i \quad (1)$$

수식 (1)에서  $Lat_c$ 는 중심위치의 위도좌표이고  $Lon_c$ 는 경도좌표이다.  $n$ 은 사용자의 트윗 수이고  $Lat_i$ 와  $Lon_i$ 는 각각 트윗의 위도좌표와 경도좌표이다. 중심위치로부터 트윗위치까지의 거리인 활동반경  $R_i$ 는 수식 (2)와 같이 구하였다. 수식 (2)는 경위도 좌표를 이용하여 두 지점 사이의 거리를 구하는 대표적 공식인 하버사인 공식(Haversine formula)<sup>[15]</sup>을 이용하였다.

$$R_i = 2r \arcsin \left( \sqrt{\sin^2\left(\frac{A}{2}\right) + \cos(Lat_c)\cos(Lat_i)\sin^2\left(\frac{O}{2}\right)} \right) \quad (2)$$

수식 (2)에서  $A$ 는  $(Lat_i - Lat_c)$ 이고  $O$ 는  $(Lon_i - Lon_c)$

이다. 그리고  $r$ 은 지구 반경으로 평균값인 6371.01로 근사하여 거리를 계산하였다. 다음으로 수식 (2)를 통해 구한 활동반경을 이용하여 수식 (3)과 같이 활동반경의 평균(활동반경평균, Mean Activity Radius, MAR)을 구하였고 수식 (4)과 같이 활동반경의 표준편차(활동반경편차, Standard Deviation Activity Radius, SDAR)를 구하였다.

$$MAR = \frac{1}{n} \sum_{i=1}^n R_i \quad (3)$$

$$SDAR = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (R_i - \bar{R})^2} \quad (4)$$

활동반경평균(MAR)의 평균(4.16)과 활동반경편차(SDAR)의 평균(2.73)을 기준으로 그림 (1)에서 보는 바와 같이 트위터 사용자를 4개의 활동반경 그룹으로 구분하였다.

활동반경평균이 활동반경평균의 평균보다 작고 활동반경편차도 활동반경편차의 평균보다 작은 그룹을 도박사(gambler)로 명명하였으며, 활동반경평균은 크고 활동반경편차는 작은 그룹을 관광객(sightseer)으로 명명하였다. 그리고 활동반경평균과 활동반경편차가 모두 큰 그룹을 통근자(commuter)로 명명하였다. 마지막으로 활동반경평균은 작고 활동반경편차는 큰 그룹을 전업주부(homemaker)라고 명명하였다.

활동반경 그룹의 특성을 정리한 표 (1)을 보면, 사용자 비율은 도박사가 44%로 가장 높고 통근자, 전업주부, 관광객 순이다. 활동반경평균은 통근자가 평균 7.35km로 가장 크고 도박사가 평균 1.82km로 가장 작다. 그리고 활동반경편차는 평균적으로 통근자가 가장 크고 도박사가 가장 작다.

표 1. 활동반경 그룹의 특성  
Table 1. Characteristics of Activity Radius Groups

Activity Radius Groups	# of Users	% of Users	Avg. MAR (km)	Avg. SDAR (km)
Gambler	5,398	44	1.82	1.32
Commuter	3,690	30	7.35	4.55
Homemaker	1,755	15	2.67	3.87
Sightseer	1,348	11	6.75	1.86
Total	12,191	100	4.16	2.73

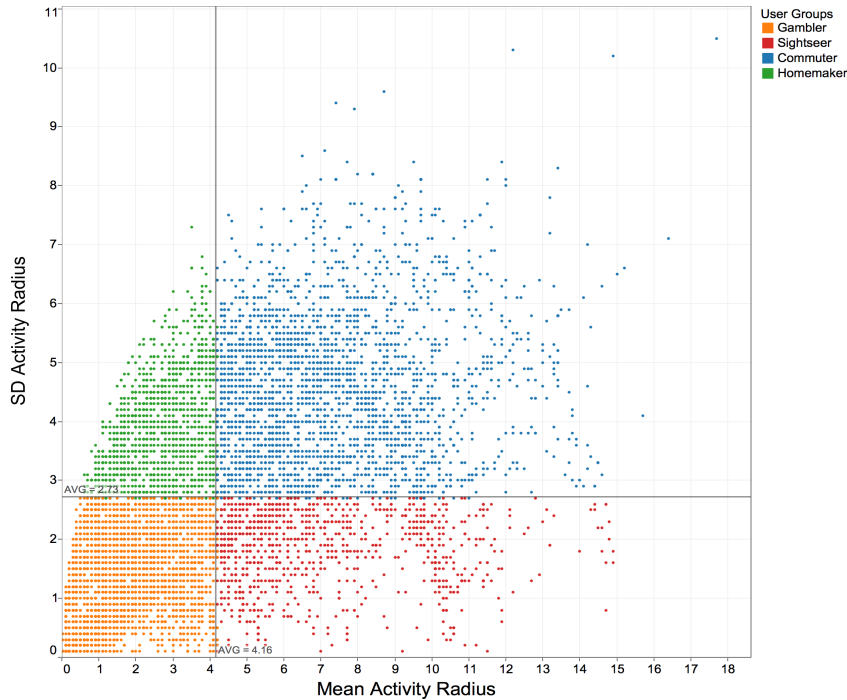


그림 1. 활동반경평균과 활동반경편차에 의한 사용자 그룹  
Fig. 1. User Groups by MAR and SDAR

### 3.3 거주지역 추정

일반적으로 거주지역(home region)은 일정 기간 동안 가장 오래 머무른 지역으로 정의할 수 있다<sup>7)</sup>. 본 연구에서는 데이터 수집 기간인 50일 중에서 가장 기간(일) 체류한 지역을 거주지역으로 정의하였다. 거주지역의 추정은 2 단계로 진행하였다. 먼저 거주국가를 추정하였으며, 거주국가가 미국인 경우 거주지역이 라스베이거스 시내인지 시외인지를 추정하였다. 다시 말해서 거주지역을 국외, 라스베이거스 시외, 라스베이거스 시내로 구분하였다. 거주지역의 국내외 추정은 지오투트 데이터의 국가코드(country\_code)를 이용하였고, 시내외 추정은 라스베이거스의 경위도 좌표 범위를 이용하였다(위도: 35.944955~36.355835, 경도: -115.367373~-114.902359).

거주일수는 해당 지역에서 트윅한 시작일과 최종일의 차이로 계산하였다. 그런데 사용자의 이동경로가 ‘미국-멕시코-미국’처럼 미국에 있다가 멕시코를 방문하고 다시 미국으로 돌아오는 경우에, 미국의 거주일수를 미국에서 트윅한 시작일과 최종일의 차이로 단순히 구하면 미국의 거주일수가 멕시코의 거주일수를 포함하는 오류가 발생한다. 이를 해결하기 위하여 지역(국가 또는 시내외)의 변화가 있을 때마다 거주일수

를 구한 후, 동일 지역의 거주일수를 합산하여 최종적으로 해당 지역의 거주일수를 계산하였다.

마지막으로, 지역별 거주일수를 비교하여 가장 기간(일) 체류한 지역을 거주지역으로 추정하였다. 가장 기간 체류한 지역을 단일 지역으로 특정하지 못할 경우에는 거주지역 불명(unknown)으로 판정하였다. 표 (2)는 거주지역 추정의 결과를 정리한 것이다. 표 (2)를 보면, 거주지역이 라스베이거스 시내인 사용자의 비율이 56%로 월등히 높고 국외 거주자의 비율이 5%로 가장 낮은 것을 알 수 있다. 활동반경평균은 라스베이거스 시내 거주자가 평균 4.32km로 가장 크고 활동반경편차도 라스베이거스 시내 거주자가 3.00km로

표 2. 거주지역별 사용자 그룹의 특성  
Table 2. Characteristics of User Groups by Home Region

Home Region	# of Users	% of Users	Avg. MAR (km)	Avg. SDAR (km)
Las Vegas	6,797	56	4.32	3.00
Other USA	4,544	37	4.11	2.46
Foreign	621	5	2.90	1.88
Unknown	229	2	3.86	2.18
Total	12,191	100	4.16	2.73

가장 크다.

### 3.4 중심구역 구성

활동반경과 거주지역의 지리적 분포를 정량적으로 집계하고 그 결과를 효과적으로 시각화하기 위하여, 사용자의 중심위치를 포함하는 지리적 상위 계층으로 ‘중심구역’을 구성하였다. 먼저 수식 (1)을 통해 계산한 중심위치의 경도 좌표와 위도 좌표를 각각 소수점 셋째 자리에서 반올림하여 중심구역의 경도 좌표와 위도 좌표를 계산하였다. 경위도 좌표에서 0.01도의 차이를 거리로 환산하면 1km 정도로 활동반경의 지리적 패턴을 파악하는 데 적절하다고 판단하였다. 다음으로, 중심구역의 경도 좌표와 위도 좌표를 결합하여 ‘중심구역\_ID’를 생성하였다.

예를 들어, 특정 사용자의 중심위치가 (36.1093, -115.1721)이면 소수점 셋째 자리에서 반올림하여 중심구역의 경위도 좌표는 (36.11, -115.17)이 된다. 중심구역의 경도 좌표와 위도 좌표를 결합한 중심구역\_ID는 ‘36.11\_-115.17’이 된다. 이러한 방법으로 총 693개의 중심구역을 생성하였다. 중심구역\_ID를 이용하여 사용자 수, 사용자 비율, 활동반경평균, 활동반경편차 등을 집계하였고 중심구역의 경도 좌표와 위도 좌표를 이용하여 사용자의 지리적 분포를 시각화하였다.

## IV. 분석 결과

### 4.1 활동반경과 거주지역

그림 (2)는 활동반경평균과 활동반경편차의 평균을

기준으로 구분한 활동반경 그룹별로 거주지역의 사용자 비율을 보여주고 있다. 전체적으로 보면 라스베이거스 시내 거주자의 비율이 57%로 가장 높다. 전체 방문자(시의 거주자와 국외 거주자) 중에서 국외 거주자의 비율은 12%이다. 라스베이거스 지역의 관광 마케팅 공식 기관인 LVCVA(Las Vegas Convention and Visitors Authority)는 2015년에 라스베이거스를 방문한 사람 중에서 외국인의 비율은 16%라고 보고 하였다<sup>16)</sup>. 이 사실을 본 연구의 결과와 비교하면 4%의 차이가 있지만 근사한 것으로 판단된다. 도박사 그룹은 다른 그룹과 비교하여 라스베이거스 거주자의 비율이 47%로 가장 낮고, 국외 거주자의 비율은 8%로 가장 높다. 관광객 그룹과 통근자 그룹은 거주지역의 사용자 비율이 유사하다. 전업주부 그룹은 라스베이거스 거주자의 비율이 73%로 가장 높고 시외 거주자의 비율이 24%로 가장 낮다.

그림 (3)은 그림 (2)와 차원을 달리하여 거주지역별로 활동반경 그룹의 사용자 비율을 보여주고 있다. 전체적으로 보면 도박사 그룹의 사용자 비율이 44%로 가장 높고 관광객 그룹의 비율이 7%로 가장 낮다. 거주지역별로 살펴보면, 라스베이거스 거주자는 다른 지역의 거주자보다 도박사 그룹의 비율이 36%로 가장 낮고 나머지 비율은 가장 높다. 시외 거주자는 라스베이거스 시내 거주자와 비교하여 전업주부 그룹의 비율은 9%로 크게 낮고 도박사 그룹의 비율은 53%로 크게 높다. 마지막으로 국외 거주자는 다른 지역의 거주자보다 도박사 그룹의 비율이 70%로 월등히 높고, 나머지 그룹의 비율은 다른 거주지역과 비교하여 가장 낮다.

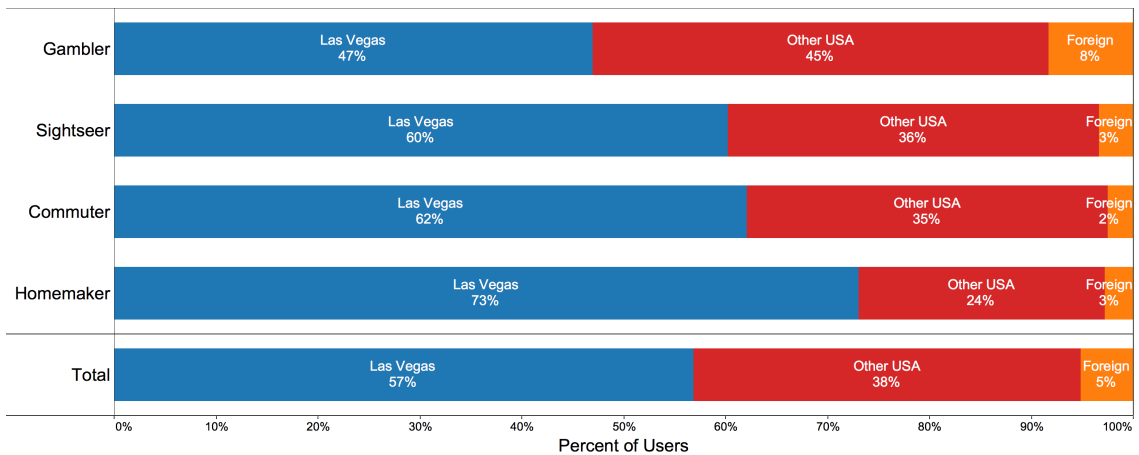


그림 2. 활동반경 그룹별 거주지역의 비율  
Fig. 2. The Percentage of Home Region by Activity Radius Group

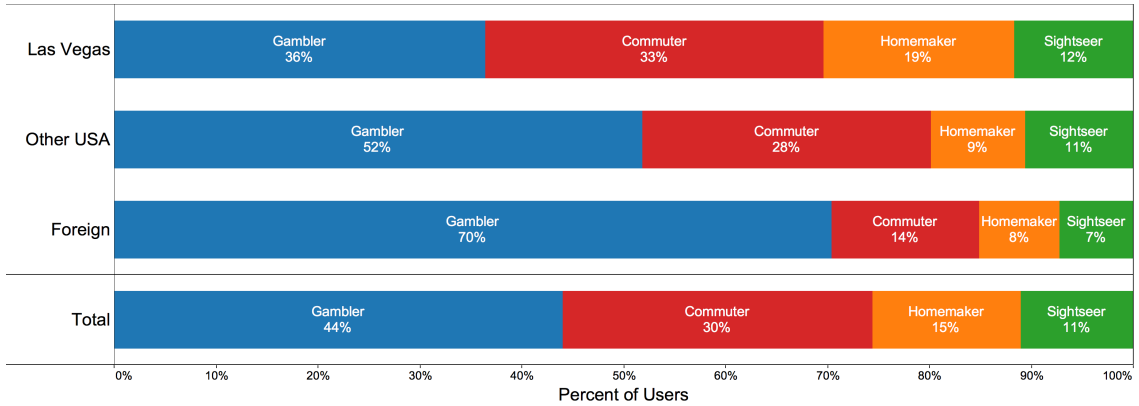


그림 3. 거주지역별 활동반경 그룹의 비율  
Fig. 3. The Percentage of Activity Radius Groups by Home Region

표 (3)은 활동반경 그룹과 거주지역을 교차하여 활동반경평균과 활동반경편차를 비교한 것이다. 활동반경평균이 가장 큰 집단은 라스베이거스에 거주하는 통근자 그룹(7.28)이고 활동반경편차가 가장 큰 집단은 라스베이거스에 거주하는 통근자 그룹(4.56)이다. 반대로 활동반경평균이 가장 작은 집단은 라스베이거스에 거주하는 도박사 그룹(1.71)이고 활동반경편차가 가장 작은 집단은 국외에 거주하는 도박사 그룹(1.12)이다.

#### 4.2 지리 분포

그림 (4)는 3.4에서 설명한 중심구역별로 활동반경평균과 활동반경편차를 지리적으로 시각화한 것이다. 원의 크기는 활동반경평균의 크기이고 색상은 활동반경편차의 크기를 표현하고 있다. 중심인 Paradise 지역에서 빨간색의 작은 원을 확인할 수 있는데, 이것은 Paradise 지역을 중심구역으로 하는 사용자는 활동반

경평균도 작고 활동반경편차도 작음을 뜻한다. 그리고 외곽의 North Las Vegas, Enterprise, Henderson 지역에서 푸른색의 큰 원을 확인할 수 있다. 이들 지역을 중심구역으로 하는 사용자는 활동반경평균도 크고 활동반경편차도 크다는 것을 알 수 있다.

그림 (5)는 사용자를 활동반경 그룹과 거주지역으로 구분하여 그들의 대표적 중심구역을 지리적으로 표현한 것이다. 대표적 중심구역을 표현하기 위하여 사용자 수가 50명 이상인 중심구역만 지도에 표현하였다. 원의 크기는 해당 구역의 상대적인 사용자 비율이다. 전체적으로 보면 활동반경 그룹별로 중심구역의 분포가 다름을 확인할 수 있다. 그리고 라스베이거스를 방문한 시외 거주자와 국외 거주자의 중심구역의 지리적 분포는 유사하고, 라스베이거스 시내 거주자의 지리적 분포는 방문자와 다른 패턴을 보인다. 세부적으로 살펴보면, 도박사 그룹에서 방문자의 중심구역은

표 3. 사용자 그룹의 활동반경 특성  
Table 3. Characteristics of Activity Radius by User Group

Activity Radius Group	Activity Radius	Las Vegas	Other USA	Foreign	Total
Gambler	MSR (km)	1.71	1.93	1.84	1.82
	SDAR (km)	1.47	1.21	1.12	1.32
Sightseer	MSR (km)	6.85	6.73	5.96	6.75
	SDAR (km)	1.87	1.85	1.93	1.86
Commuter	MSR (km)	7.28	7.52	6.61	7.35
	SDAR (km)	4.56	4.54	4.50	4.55
Homemaker	MSR (km)	2.58	2.94	2.78	2.67
	SDAR (km)	3.89	3.83	3.81	3.87
Total	MSR (km)	4.32	4.11	2.90	4.16
	SDAR (km)	3.00	2.46	1.88	2.73



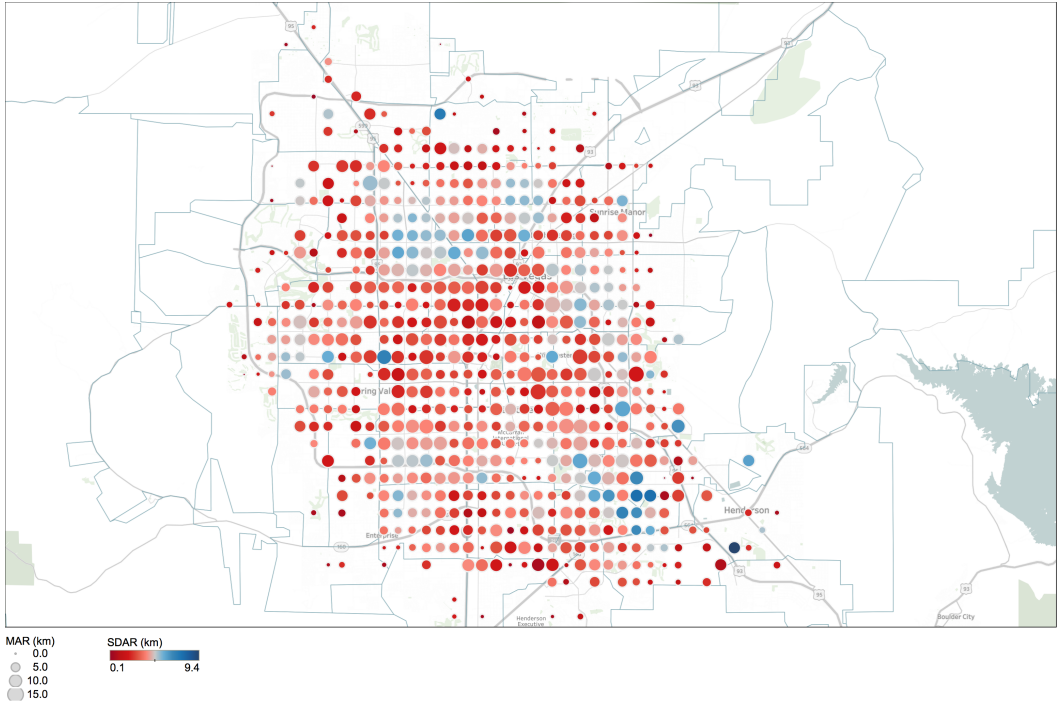


그림 4. 중심구역별 활동반경과 활동편차  
Fig. 4. MAR and SDAR by Central Area

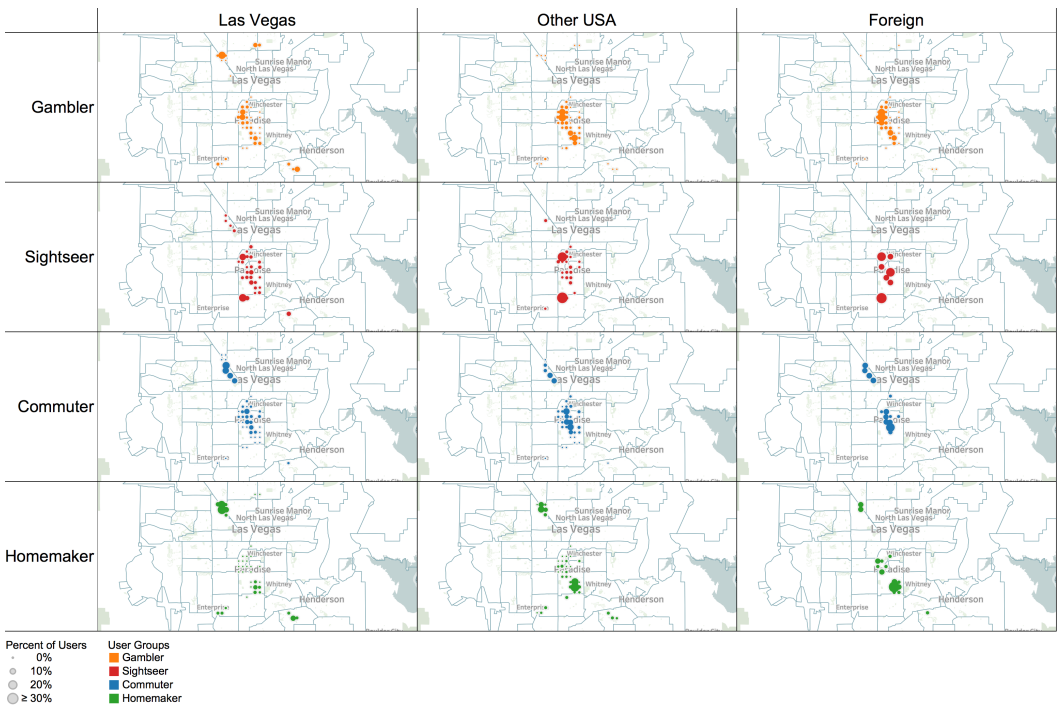


그림 5. 사용자 그룹별 중심구역의 지리적 분포  
Fig. 5. Geographical Distribution of Central Area by User Group

Paradise 지역에 집중되어 있고, 라스베이거스 거주자의 중심구역은 Paradise 지역뿐만 아니라 North Las Vegas, Enterprise, Handerson 지역까지 분포하고 있다. 관광객 그룹에서 국외를 거주지역으로 하는 사용자의 중심구역은 다른 지역에 거주하는 사람과 달리 7 개 구역에 집중되어 있다. 통근자 그룹에서 라스베이거스 거주자의 분포는 North Las Vegas 지역의 사용자 비율이 다소 높아 보이고 국내 방문자는 Paradise 지역의 비율이 다소 높아 보이고, 국외 방문자는 Paradise 지역과 North Las Vegas 지역의 분포가 비슷해 보인다. 전업주부 그룹에서 라스베이거스 거주자는 North Las Vegas 지역의 사용자 비율이 높고, 국외 방문자와 국내 방문자는 Paradise와 North Las Vegas 지역의 사용자 비율이 높다.

## V. 결 론

본 연구는 전 세계에서 발생한 지오투잇을 수집하여 인간의 도시 내 이동패턴을 분석하였다. 연구의 의의를 토의하면 다음과 같다. 첫째, 활동반경을 이용하여 인간의 활동성을 이해할 수 있는 지표를 개발하였다. 활동반경평균(활동반경의 평균)은 활동성의 크기를 나타내고, 활동반경편차(활동반경의 표준편차)는 활동성의 변화를 나타낸다. 둘째, 거주일수를 이용하여 거주지역(시내 거주, 시외 거주, 국외 거주)을 추정하는 방법을 고안하였다. 체류지역의 변화에 따라 트윗 시작일과 종료일을 확인하여 거주일수를 계산함으로써 거주지역 추정의 정확도를 높였다. 마지막으로, 인간의 거주지역과 활동성의 유의한 관계를 발견하였다. 라스베이거스를 대상으로 하는 사례 분석을 통해, 거주지역에 따른 활동성과 활동지역의 차이를 확인하였다. 본 연구의 방법에 따라 다양한 도시를 대상으로 분석을 수행한다면, 인간의 이동성에 대한 다양한 이론을 도출할 수 있을 것이다.

## References

[1] S. Park, Y. I. Kim, and S. R. Lee, "Hierarchical visualization of cloud-based social network service using fuzzy," *J. KICS*, vol. 38B, no. 7, pp. 501-511. Jul. 2013.

[2] J. Moon, I. Jang, Y. C. Choe, J. G. Kim, and G. Bock, "Case study of big data-based agri-food recommendation system according to types of customers," *J. KICS*, vol. 40, no. 5,

pp. 903-913, May 2015.

[3] J. T. Oh, "Personal environment service and technology based on smart phone," *J. KICS*, vol. 38C, no. 5, pp. 454-463, May 2013.

[4] J. H. Cho, "Tutorial: Geo-tweet analysis to understand mobility patterns of people," in *Proc. Korea Soc. IT Serv. Conf.*, pp. 419-428, Seoul, Korea, May 2016.

[5] B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti, "Geo-located twitter as proxy for global mobility patterns," *Cartography and Geographic Inf. Sci.*, vol. 41, no. 3, pp. 260-271, Feb. 2014.

[6] J. Yin, Y. Gao, Z. Du, and S. Wang, "Exploring multi-scale spatiotemporal twitter user mobility patterns with a visual-analytics approach," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 10, id. 187, pp. 1-19, Oct. 2016.

[7] J. H. Cho and I. Seo, "Comparing the spatial mobility of residents and tourists by using teotagged tweets," *J. Inf. Technol. Serv.*, vol. 15, no. 3, pp. 211-221, Sep. 2016.

[8] M. Lenormand, B. Gonçalves, A. Tugores, and J. J. Ramasco, "Human diffusion and city influence," *J. The Royal Soc. Interface*, vol. 12, no. 109, id. 20150473, pp. 1-9, Jul. 2015.

[9] J. Mahmud, J. Nichols, and C. Drews, "Home location identification of twitter users," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, id. 47, pp. 1-21, Jul. 2014.

[10] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: A content-based approach to geo-locating Twitter use," in *Proc. 19th ACM Int. Conf. Inf. and Knowledge Management*, pp. 759-768, Toronto, Canada, Oct. 2010.

[11] B. Hecht, L. Hong, B. Suh, and E. H. Chi, "Tweets from Justin Bieber's heart: The dynamics of the 'location' field in user profiles," in *Proc. SIGCHI Conf. Human Factors in Comput. Syst.*, pp. 237-246, Vancouver, Canada, May 2011.

[12] J. Kulshrestha, F. Kooti, A. Nikraves, and K. P. Gummadi, "Geographic dissection of the Twitter network," in *Proc. 6th Int. AAAI Conf.*



on *Weblogs and Social Media*, pp. 202-209, Dublin, Ireland, Jun. 2012.

- [13] A. Belyi, I. Bojic, S. Sobolevsky, I. Sitko, B. Hawelka, L. Rudikova, A. Kurbatski, and C. Ratti, *Global multi-layer network of human mobility* (2016), Retrived Jan. 6, 2017, from <http://arxiv.org/abs/1601.05532>.
- [14] N. Hossain, T. Hu, R. Feizi, A. M. White, J. Luo, and H. Kautz, *Inferring fine-grained details on user activities and home location from social media: detecting drinking-while-tweeting patterns in communities* (2016), Retrived Jan. 6, 2017, from <http://arxiv.org/abs/1603.03181>.
- [15] C. Robusto, "The cosine-haversine formula," *The Am. Math. Monthly*, vol. 64, no. 1, pp. 38-40, 1957.
- [16] LVCVA, *Las Vegas Visitor Profile Study* (2015), Retrieved Jan. 6, 2017, from <http://www.lvcva.com>.

**조 재 희 (Jaehye Cho)**



1983년 2월 : 연세대학교 경영학  
학사  
1985년 8월 : 마이애미대학교 경  
영학 석사  
1991년 12월 : 네브라스카대학  
교 경영학 박사

<관심분야> 트위터 데이터 분석, 데이터 시각화

**서 일 정 (Il-Jung Seo)**



1999년 2월 : 경기대학교 경영  
정보학 학사  
2006년 2월 : 광운대학교 경영  
정보학 석사  
2010년 2월 : 광운대학교 경영  
정보학 박사

<관심분야> 네트워크 분석, 데이터 시각화