

트위터에서 트윗 주기와 사용자 속도 사이 관계

전소영¹ · 이알찬² · 서고은² · 신원용^{3*}

Relationship Between Tweet Frequency and User Velocity on Twitter

So-Young Jeon¹ · Al-Chan Lee² · Go-Eun Seo² · Won-Yong Shin^{3*}

¹Department of Applied Computer Engineering, Dankook University, Yongin 448-701, Korea

²Department of Mobile Systems Engineering, Dankook University, Yongin 448-701, Korea

^{3*}Department of Computer Science and Engineering, Dankook University, Yongin 448-701, Korea

요 약

최근 위치 정보를 제공하는 온라인 소셜 네트워크 서비스들의 급증으로 인해 사용자들의 지리적 위치 데이터의 중요성이 강조되고 있다. 본 논문에서는 사용자들의 고 정밀 위치 정보를 알려주는 공간 태그된 트윗 (geo-tagged tweet) 정보를 활용하여 트위터 사용자들의 정확한 위치와 트윗 전송시각을 알아낸 후, 이를 통해 사용자의 평균 이동속도와 트윗 주기 (tweet frequency) 사이의 관계를 분석한다. 구체적으로, 트윗 빈도수 계산 알고리즘을 소개하며, 결과에 대한 분석은 국가별, 도시별로 나누어 진행한다. 주요 결과로써, 사용자 속도에 따른 트윗 주기가 멱 법칙 분포 (power-law distribution) (또는 Zipf의 법칙 분포, Pareto 분포)를 따름을 보인다. 또한, 미국과 일본에서의 결과를 비교할 때, 일본에서의 분포도 지수가 미국의 경우에 비해 작음을 확인한다.

ABSTRACT

Recently, the importance of users' geographic location information has been highlighted with a rapid increase of online social network services. In this paper, by utilizing geo-tagged tweets that provides high-precision location information of users, we first identify both Twitter users' exact location and the corresponding timestamp when the tweet was sent. Then, we analyze a relationship between the tweet frequency and the average user velocity. Specifically, we introduce a tweet-frequency computing algorithm, and show analysis results by country and by city. As a main result, it is shown that the tweet frequency according to user velocity follows a power-law distribution (i.e., Zipf⁷ distribution or a Pareto distribution). In addition, by performing a comparison between the United States and Japan, one can see that the exponent of the distribution in Japan is smaller than that in the United States.

키워드 : 공간 태그된 트윗, 사용자 속도, 온라인 소셜 네트워크, 트위터, 트윗 주기

Key word : Geo-tagged tweet, user velocity, online social network, Twitter, tweet frequency

Received 13 April 2015, Revised 22 April 2015, Accepted 07 May 2015

* Corresponding Author Won-Yong Shin(E-mail:wysin@dankook.ac.kr, Tel:+82-31-8005-3253)

Department of Computer Science and Engineering, Dankook University, Yongin 448-701, Korea

Open Access <http://dx.doi.org/10.6109/jkiice.2015.19.6.1380>

print ISSN: 2234-4772 online ISSN: 2288-4165

©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.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

온라인 소셜 네트워크 (OSN: Online Social Network)는 인터넷 상의 상호 의사 소통 및 정보 교류에 있어 유용한 사회기반 시스템이다. 이러한 OSN은 컴퓨터, 모바일 장치, 그리고 태블릿과 같은 인터넷이 가능한 장치 사용의 급증으로 인해 더욱 대중화되었다. OSN 사용자들은 그들의 프로필과 콘텐츠를 게시하고 사람들과 링크를 공유하는 등의 활동을 하며, 이러한 활동을 통하여 사용자들은 사회적 관계를 유지하게 된다[1]. OSN은 사용자들의 활동을 기반으로 수집된 빅 데이터를 가지고 있는데, 이는 여러 가지 분석에 유용하게 사용된다. OSN 중에서 “트위터 (Twitter)”는 가장 보편적으로 사용되는 마이크로 블로그 중 하나인데, 사용자들이 트윗 (tweet)이라는 140자 이내의 단문 메시지를 전송함으로써 개인의 견해나 정보를 공유하고 다른 사용자들을 팔로우 (follow)함으로써 그들의 트윗을 받을 수 있다[2]. 트위터에서는 다른 OSN와 달리 사용자 개인의 요구에 의해 타 사용자와의 일방적인 관계 형성이 가능하기 때문에 유동적인 정보의 흐름이 가능하다.

인간, 컴퓨터 바이러스, 교통 그리고 인구 예측 등으로의 다양한 응용에 대한 근본적인 중요성 때문에 인간 개개인의 이동성에 대한 많은 연구가 진행되어 왔다. 인간 이동성에 대한 이전 연구들에서는 모바일 폰[3-5], GPS (Global Positioning System) 수신기[6], Wifi 단말기[7], RFID 장치[8]와 같은 추적 기술을 사용하는 것에 초점을 맞추었다. 이전 연구들에서는 분석에 필요한 데이터가 부족하거나 사용자의 정확한 위치 정보를 얻지 못하는 등의 결점이 있었다.

반면, *공간 태그된 트윗* [9,10]은 사용자의 동의 하에 스마트폰의 GPS interface를 통해 10미터 이내의 정확한 공간 정보 (위도 및 경도)를 제공해주며, 거대한 샘플과 함께 포괄적인 메타데이터를 제공한다. 공간 태그된 트윗은 전체 트윗 양의 대략 1% 정도를 차지하지만, 스마트 기기 및 모바일 앱의 폭발적인 보급으로 인해 이는 분석에 유효한 데이터의 양을 충분히 제공할 수 있다. 따라서, 공간 태그된 트윗 (geo-tagged tweet)은 인간 이동성 분석에 대한 풍부한 특성을 제공하는 것을 가능하게 한다.

예를 들어, [9]에서는 전반적인 인간 이동성 유형이 폭넓게 발견되었고, 여러 나라에서의 이동성 특성에 대

한 비교 분석이 수행되었다. 게다가 [10]에서는 호주에서의 공간 태그된 트윗 데이터를 통해 인간 이동성이 작은 지역 내에서부터 대도시 및 도시 간 이동까지 다중 모드를 나타냄을 보였다.

II. 주요 제안사항

본 논문에서는 [9,10]에서와 마찬가지로 트위터 사용자들의 정확한 위치를 확인하기 위해 공간 태그된 트윗을 활용하고, 이를 통해 인간 이동성에 기반한 새로운 통계 분석인 *트윗 주기 (tweet frequency)와 사용자 이동 속도 사이의 상관관계*를 규명한다. 즉, 사용자의 평균적인 움직임에 따른 트위터 이용 빈도수에 대해 연구한다. 분석을 위해 미국과 일본 내 일부 대도시에서 트위터 사용자들로부터 생성된 거대한 양의 공간 태그된 트윗으로 구성된 데이터셋을 사용한다. 트윗 주기와 이동 속도 사이의 관계를 규명함으로써 사용자 이동 속도에 따른 일반적인 OSN의 사용 빈도를 유추할 수 있다. 또한, 이를 통해 향후 지역별 (나라별/도시별), 시간별 (오전/오후/출퇴근 및 통학시간 등), 교통상황별로 해당 OSN을 통해 사용자들에게 효율적으로 광고를 브로드캐스팅 (broadcasting)하는 등의 상업적 활용이 가능할 것이다.

먼저, 분석을 위해 특정 사용자가 생성한 연속된 공간 태그된 트윗의 위도와 경도 좌표 정보를 이용하여 두 지점 사이의 거리를 구한다. 거리를 계산할 때에는 지구는 완전한 구임을 가정하며 구면코사인법칙 (spherical law of cosines)을 적용한다. 또한, 연속된 트윗이 생성된 시간차이를 계산한다. 본 연구에서는 트윗 간 시간 구간이 1시간 이내인 이벤트를 채택하며 이는 분석하는데 부정확한 데이터를 제거하기 위한 것이다. 보다 구체적으로, 트윗 간 시간차이가 클 경우 계산한 평균 속도와 실제 사용자가 이동한 순간 속도 사이에 차이가 날 수 있기 때문에 이러한 이벤트는 제거하기 위함이고, 1시간 대신 다른 수치를 가정할 수 있겠으나 본 연구에서는 편의상 1시간 미만의 트윗 간 시간차이를 갖는 이벤트만을 고려한다. 계산된 거리와 시간을 이용하여 사용자의 평균 이동 속도를 구한다. 이를 위해 속도 계산 알고리즘을 제안한다. 다음 단계로써, 각 속도 구간에 따른 트윗 사용 빈도수 (즉, 트윗 주기)를

도출하는 계산 알고리즘을 제안한다. 사용 빈도수를 구하기 위해, 주요 도시에서의 표정 속도 (scheduled speed)를 고려하여 최대 50km/h 이하의 평균 속도를 가지는 데이터 내 이벤트만을 선별한다.¹⁾

주요 결과로써, 국가별 및 도시별로 나누어 사용자의 평균 이동속도와 트윗 주기와와의 관계를 보인다. 국가별 (국가 내 고려한 여러 도시 안에서의 분석 결과들의 평균치) 결과로는 실험적 데이터, 적합 함수 (fitting function)를 나타내었으며, 이 때 적합 함수를 구하기 위해 전통적인 최소 제곱 (least squares) 추정 방식을 사용한다. 도시별 결과로는 여러 도시에서의 실험적 데이터만을 보여준다. 그 결과, 사용자 속도에 따른 트윗 주기가 **멱 법칙 분포 (power-law)** (또는 Zipf의 법칙 분포, Pareto 분포)를 따름을 보인다. 이는 사용자들은 정지 상태일 때 트윗 사용 빈도가 아주 높은 반면, 속도가 증가할수록 사용 빈도가 다항적인 감소 (polynomial decrease)를 보이는 것을 뜻한다. 또한, 국가별 분석을 통해 일본인 사용자들이 미국인 사용자들에 비해 동적 상태에서 트위터를 더 활동적으로 이용한다는 것을 검증한다.

본 논문의 구성은 다음과 같다. III장에서는 데이터셋에 대해 소개한다. IV장에서는 속도 계산 과정 및 트윗 주기 계산 알고리즘을 포함한 연구 방법론을 제안한다. V장에서는 분석 결과를 보인다. VI장에서는 본 논문을 요약 및 마무리한다.

III. 데이터셋

트위터 분석을 위한 데이터 수집 프로세스로 어플리케이션의 접근 횟수에 대해 제한이 없고 실시간 이벤트 검출에 용이한 장점이 있는 트위터 Streaming API (Application Programming Interface)를 이용하여 트위터 네트워크 크롤링을 진행하였다. 그림 1에서는 데이터 수집의 전체적인 과정을 보여준다. 크롤러는 트위터 API로부터 R의 StreamR 패키지를 이용하여 일정시간 단위로 JSON (JavaScript Object Notation) 파일을 저장한다.

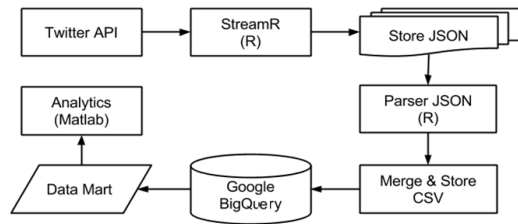


그림 1. 데이터 수집 과정
Fig. 1 Data collection process

표 1. 도시별 전체 트윗 수 및 사용자 수
Table. 1 The total numbers of tweets and users according to cities

	전체 트윗수	전체 사용자수
LA	115,050	23,187
New York	89,821	24,215
Chicago	82,184	16,709
San Francisco	28,209	8,316
Tokyo	90,180	21,049
Osaka	20,857	5,474

이후 수집된 JSON 파일에서 필요한 필드를 parsing 하여 하나의 분석용 표로 완성하고, 분석에 필요한 데이터마트를 구성한다. 이를 위하여 Google BigQuery를 사용하여 프로세스 시간을 획기적으로 줄일 수 있다. 분석 시에는 알고리즘 구현을 위해 Matlab 소프트웨어를 사용한다.

본 연구에서는 2014년 9월 5일부터 11일까지 (대략 일주일) 트위터 사용자로부터 기록된 거대한 양의 공간 태그된 트윗으로 구성된 데이터셋을 사용하였다. 일주일 가량의 단기간 데이터셋은 사용자 속도별 트윗 주기를 분석하기에 충분하다. 데이터 수집은 미국과 일본 안의 대표도시를 위주로 진행하였는데, 대표도시로는 미국의 Los Angeles (LA), New York, Chicago, San Francisco, 그리고 일본의 Tokyo와 Osaka이다. 도시별 수집된 전체 트윗 수 및 사용자 수는 표 1에서 정리된다. 데이터셋에서 각 트윗 기록은 트윗이 보내진 장소 및 시간을 가진다. 이 정보에 기반하여 각 사용자의 위치 정보를 $L = (x_{ij}, y_{ij}, t_{ij})$ 와 같이 나타낼 수 있다. 여기에서, x_{ij} 와 y_{ij} 는 각각 사용자 i 의 j 번째 트윗이 발생

1) 사람이 아닌 봇 (bot) [11]이 트윗을 생성하는 경우가 있는데, 이 경우 평균 이동 속도가 100km/h 이상으로 비현실적으로 클 수 있다. 따라서, 사람의 트윗 주기만을 분석하기 위해 평균 속도 50km/h 이하의 트윗 이벤트만을 고려한다.

되었을 때의 x -좌표 및 y -좌표이고, 이 위치 정보는 위도와 경도를 나타낸다. 또한, t_{ij} 는 사용자 i 의 j 번째 트윗이 발생된 시간을 나타낸다. 데이터셋은 사용자 트윗에 대한 많은 정보를 포함하고 있으며, 트윗의 메타데이터에서 본 연구에 필요한 필드 값을 아래와 같이 추출한다.

- *user_id str*: 송신자 ID
- *lat*: 송신자의 위도
- *lon*: 송신자의 경도
- *created_at*: 트윗을 전송 한 UTC/GMT 시각

IV. 연구 방법론

본 장에서는 속도 계산 과정과 이를 사용해 개발되는 트윗 주기 계산 알고리즘을 설명한다.

4.1. 속도 계산 과정

사용자 i 가 j 번째 트윗을 보냈을 때의 위치 정보가 (x_{ij}, y_{ij}, t_{ij}) 이고 $(j+1)$ 번째 트윗을 보냈을 때의 위치 정보가 $(x_{i(j+1)}, y_{i(j+1)}, t_{i(j+1)})$ 라 할 때, 두 지점 사이의 지리적 거리 d_{ij} 는 구면코사인법칙을 사용하여 다음과 같이 구할 수 있다.

$$d_{ij} = R \cos^{-1}(\sin x_{ij} \sin x_{i(j+1)} + \cos x_{ij} \cos x_{i(j+1)} \cos(y_{i(j+1)} - y_{ij})) \quad (1)$$

여기에서, 거리를 산출할 때는 지구가 완전한 구임을 가정하였으며, R [km]은 6,371로 주어지는 지구의 반지름을 나타낸다. 다음은 연속된 두 트윗 이벤트를 사용하여 속도 v_{ij} 를 구하는 방법이다.

$$v_{ij} = d_{ij} / (t_{i(j+1)} - t_{ij}) \quad (2)$$

여기에서, t_{ij} 와 $t_{i(j+1)}$ 은 각각 사용자 i 의 j 번째 트윗 발생 시간 및 $(j+1)$ 번째 트윗 발생 시간을 나타낸다. v_{ij} 는 j 번째 트윗과 $(j+1)$ 번째 트윗이 발생한 시간 사이의 평균 이동 속도이다.

본 연구에서는, 속도 계산 시 부정확한 데이터를 제거하기 위하여 몇 가지 기준을 설정하여 필터링을 수행한다. 먼저, 트윗 간 시간 구간이 1시간 이내인 이벤트만을

채택한다. 보다 구체적으로, 트윗 간 시간차이가 클 경우 계산한 평균 속도와 실제 사용자가 이동한 순간 속도 사이에 차이가 날 수 있기 때문에 이러한 이벤트는 제거하기 위함이다. 둘째, 주요 도시에서의 표정속도를 고려하여 최대 50km/h 이하의 평균 속도를 가지는 데이터 내 이벤트만을 선별하여 분석을 실시한다. 속도 계산 알고리즘은 표 2에서 보여진다. 이 때, I 는 공간 태그된 트윗을 전송한 총 사용자 수를 나타내며, \max_i 는 i 번째 사용자가 생성한 전체 트윗 수를 나타낸다.

표 2. 속도 계산 알고리즘

Table. 2 Velocity computing algorithm

```

입력:  $(x_{ij}, y_{ij}, t_{ij}), i \in \{1, \dots, I\}, j \in \{1, \dots, \max_i\}$ 
출력:  $v_{ij}$ 
초기화:  $d_{ij} \leftarrow 0, v_{ij} \leftarrow 0$ 
00: for  $i \leftarrow 1$  to  $I$  do
01:   for  $j \leftarrow 1$  to  $\max_i$  do
02:     if  $t_{i(j+1)} - t_{ij} < 1$  hour then
03:       식 (1) 계산
04:       식 (2) 계산
05:     end if
06:   end for
07: end for

```

표 3. 트윗 주기 계산 알고리즘

Table. 3 Tweet frequency computing algorithm

```

입력:  $v_{ij}, i \in \{1, \dots, I\}, j \in \{1, \dots, \max_i\}$ 
출력:  $F = \{f_1, \dots, f_{v_{\max}}\}$ 
초기화:  $f_n = 0, n \in \{0, \dots, f_{v_{\max}}\}, v_{\max} = 49$ 
00: for  $i \leftarrow 1$  to  $I$  do
01:   for  $j \leftarrow 1$  to  $\max_i$  do
02:     for  $n \leftarrow 0$  to  $v_{\max}$  do
03:       if  $v_{ij} \geq n$  and  $v_{ij} < n+1$  then
04:          $f_n \leftarrow f_n + 1$ 
05:       end if
06:     end for
07:   end for
08: end for

```

4.2. 트윗 주기 계산 알고리즘

본 절에서는 앞서 계산한 평균 속도에 기반하여 트윗 주기를 계산하는 알고리즘을 설명한다. 먼저, 속도를 0km/h부터 최대 속도 v_{\max} 까지 구간을 1km/h의 간격으로 나눈다. 변수 f_n ($n \in \{0, \dots, v_{\max}\}$)

을 $v_{ij} \in [n, n+1)$ 에 속하는 속도 구간에서 트윗이 발생한 총 횟수로 정의하자. 또한, 모든 $n \in \{0, \dots, v_{\max}\}$ 에 대해 집합 $F = \{f_1, \dots, f_{v_{\max}}\}$ 를 정의한다. 본 연구에서는 $v_{\max} = 49$ 를 사용하고, $F = \{f_1, \dots, f_{49}\}$ 가 된다. 트윗 주기 계산 알고리즘은 표 3에서 자세히 보여진다.

V. 분석결과

본 장에서는 국가별, 도시별로 사용자 속도에 따른 트윗 주기 분포도를 보이고, 결과적으로 멱 법칙 분포를 따름을 검증한다.

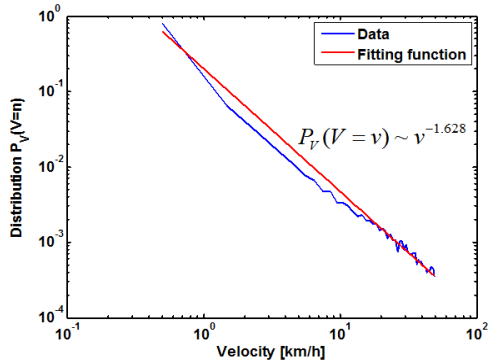


그림 2. 미국에서의 속도에 따른 트윗 주기 확률 분포 $P_V(V=v)$
Fig. 2 Probability distribution $P_V(V=v)$ of the tweet frequency according to the velocity in the United States

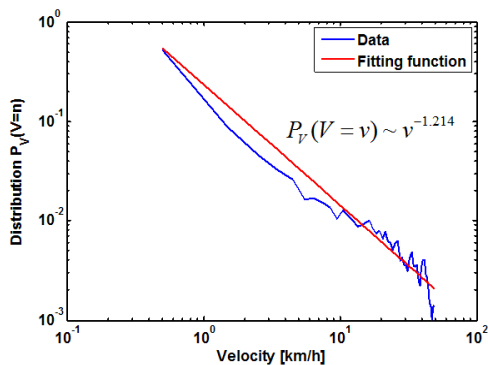


그림 3. 일본에서의 속도에 따른 트윗 주기 확률 분포 $P_V(V=v)$
Fig. 3 Probability distribution $P_V(V=v)$ of the tweet frequency according to the velocity in Japan

5.1. 국가별 분석

본 절에서는 표 1에서 보인 미국 내 4개 주요 도시 및 일본 내 2개 주요 도시의 데이터를 모두 사용하여 국가별 분석을 수행한다. 구체적으로, 국가별로 사용자 이동 속도 v [km/h]에 따른 트윗 주기의 확률 분포 $P_V(V=v)$ 를 찾는다. 그림 2와 3에서는 각각 미국 및 일본에서의 실험 데이터로부터 얻어지는 분포도 $P_V(V=v)$ 의 log-log plot을 보여준다. 또한, 각각의 실험 결과에 대하여 전통적인 최소 제곱 추정 방식을 사용하여 적합 함수를 찾는다. 주요 결과로써, 확률 분포는 단일 멱 법칙 함수 $P_V(V=v) \sim v^{-\alpha}$ ($\alpha > 0$)로 모델링될 수 있음을 확인할 수 있다. 이를 통해 정지 상태에서는 트윗 사용빈도가 아주 높은 반면, 이동속도가 증가할수록 사용 빈도가 다항적인 감소를 보이는 것을 확인할 수 있다.

분포도 $P_V(V=v)$ 의 지수 값 α 는 미국 및 일본 지역에서 각각 1.628과 1.214로 주어진다. 따라서, 일본에서의 트위터 사용자들이 미국 내 트위터 사용자에게 비해 동적 환경에서 트위터를 더 활동적으로 이용한다는 점을 추정할 수 있다. 뿐만 아니라, 그림 2와 3에서 확인할 수 있듯이 속도 v 가 큰 경우 분포도의 tail에서 통계적 잡음이 존재함을 알 수 있다.

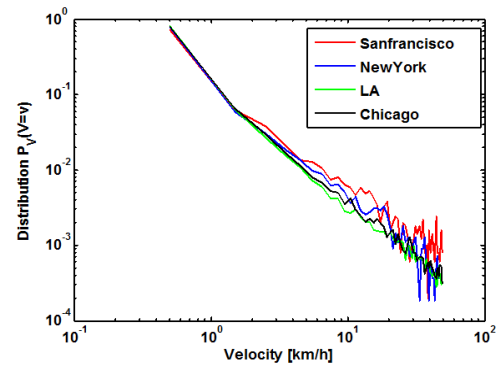


그림 4. 미국 내 도시별 속도에 따른 트윗 주기 확률 분포 $P_V(V=v)$
Fig. 4 Probability distribution $P_V(V=v)$ of the tweet frequency according to the velocity for each city in the United States

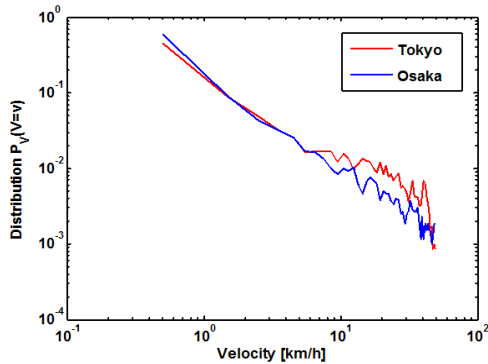


그림 5. 일본 내 도시별 속도에 따른 트윗 주기 확률 분포 $P_V(V=v)$
Fig. 5 Probability distribution $P_V(V=v)$ of the tweet frequency according to the velocity for each city in Japan

5.2. 나라 안 도시별 분석

그림 4와 5에서는 실험 데이터로부터 각각 미국 및 일본 내 도시별 사용자 속도에 따른 트윗 주기의 확률 분포 $P_V(V=v)$ 의 log-log plot을 보여준다. 그림 4로부터 미국 내 주요 도시별 확률 분포가 거의 유사하게 나타나지만, 속도 v 가 커짐에 따라 San Francisco에서의 결과 곡선이 다른 도시에서보다 다소 완만하게 감소함을 알 수 있다. 또한 그림 5로부터 일본 내 주요 도시별 확률 분포가 거의 유사하게 나타나지만, 속도 v 가 증가함에 따라 Tokyo에서의 결과 곡선이 Osaka에서보다 완만하게 감소함을 알 수 있다. 이는 Tokyo에서의 트위터 사용자들이 Osaka 내 트위터 사용자들에 비해 동적 환경에서 트위터를 더 활동적으로 이용한다는 점을 유추할 수 있다.

VI. 결 론

본 논문에서는 기존 연구[9,10]에서와 마찬가지로 공간 태그된 트윗을 활용하여 인간 이동성을 분석하였는데, 이를 활용해 기존 연구에서는 수행되지 않았던 사용자 이동 속도 계산 및 속도별 트윗 주기 계산을 수행하는 알고리즘을 제안하고 둘 사이의 상관 관계를 새롭게 규명하였다. 구체적으로, 사용자가 생성한 연속된 공간 태그된 트윗의 경위도 좌표 정보를 이용하여 지리학적 거리를 계산하고, 연속된 트윗이 생성된 시간 차

이를 계산함으로써 사용자의 평균 이동 속도를 구하는 알고리즘을 소개하였다. 그리고, 각 속도 구간에 따른 트윗 사용 빈도수를 도출하는 계산 알고리즘을 제안하였다. 주요 결과로써, 미국과 일본 내 일부 대도시 트위터 사용자들로부터 생성된 공간 태그된 트윗으로 구성된 데이터셋을 사용하여 속도에 따른 트윗 주기 확률 분포를 분석하였다. 그 결과, 기존 연구에서 보이지 않은 새로운 발견으로 사용자 속도에 따른 트윗 주기가 멱 법칙 분포를 따름을 보였다. 또한, 국가별 분석을 통해 일본에서의 분포도 지수 α 가 미국에서의 α 에 비해 작음을 관찰하였다.

ACKNOWLEDGMENTS

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (MSIP) (2012R1A1A1044151) and the Ministry of Education (2014R1A1A2054577).

REFERENCES

- [1] C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao, "User interactions in social networks and their implications," in *Proceedings of the 4th ACM European Conference on Computer Systems (EuroSys '09)*, Nuremberg, Germany, pp. 205-218, Mar./Apr. 2009.
- [2] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," in *Proceedings of the 19th International World Wide Web Conference (WWW2010)*, Raleigh, NC, pp. 591-600, Apr. 2010.
- [3] M. C. Gonzalez, C. A. Hidalgo, and A. L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779-782, June 2008.
- [4] S. Jiang, G. A. Fiore, Y. Yang, J. Ferreira, Jr., E. Frazzoli, and M. C. Gonzalez, "A review of urban computing for mobile phone traces: Current methods, challenges, and opportunities," in *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing (UrbComp 2013)*, Chicago, IL, Aug. 2013.

- [5] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi, "Human mobility, social ties, and link prediction," in *Proceedings of the 17th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD2011)*, San Diego, CA, Aug. 2011, pp. 1100-1108.
- [6] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, "In the Levy-walk nature of human mobility," *IEEE/ACM Transactions on Networking*, vol. 19, no. 3, pp. 630-643, June 2011.
- [7] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 606-620, June 2007.
- [8] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani, "Dynamics of person-to-person interactions from distributed RFID sensor networks," *PLOS ONE*, vol. 5, no. 7, pp. e11596, July 2010.
- [9] 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 Information Science*, vol. 41, no. 3, pp. 260-271, May 2014.
- [10] R. Jurdak, K. Zhao, J. Liu, M. AbouJaoude, M. Cameron, and D. Newth, "Understanding human mobility from Twitter," Preprint, [Online]. Available: <http://arxiv.org/abs/1412.2154>.
- [11] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of Twitter accounts: Are you a human, bot, or cyborg?," *IEEE Transactions on Dependable and Secure Computing*, vol. 9, no. 6, pp. 811-824, Dec. 2012.



전소영(So-Young Jeon)

2012년~현재 단국대학교 응용컴퓨터공학과 학사과정
※관심분야 : 빅데이터 분석



이알찬(Ai-Chan Lee)

2012년~현재 단국대학교 국제학부 모바일시스템공학전공 학사과정
※관심분야 : 통신이론, 빅데이터 분석



서고은(Go-Eun Seo)

2012년~현재 단국대학교 국제학부 모바일시스템공학전공 학사과정
※관심분야 : 통신이론, 빅데이터 분석



신원용(Won-Yong Shin)

2002년 연세대학교 기계전자공학부 학사
2004년 KAIST 전자전산학과 석사
2008년 KAIST 전자전산학부 박사
2008년 2월~4월 Harvard University 방문연구원
2008년 9월~2009년 2월 KAIST BK 정보전자연구소 박사후연구원
2009년 3월~4월 KAIST 고성능집적시스템연구센터 선임급 위촉연구원
2009년 5월~2011년 10월 Harvard University Postdoctoral Fellow
2011년 10월~2012년 2월 Harvard University Research Associate
2012년 3월~현재 단국대학교 국제학부 모바일시스템공학전공/대학원 컴퓨터학과 조교수
※관심분야 : 정보이론, 통신이론, 신호처리, 모바일 컴퓨팅, 빅데이터 분석