

서울시 지하철 2 호선 역세권의 토지피복도를 이용한 소셜 데이터의 통시적 분석

김세연*, 박기남, 김동민, 이동만

전산학부, 한국과학기술원

{ ha7198ha, parkinam321, underkoo, dlee }@kaist.ac.kr

Diachronic Analysis of Social Media Data using LULC Map of Seoul Metro No.2 Line neighborhood

Seyeon Kim*, Kinam Park, Dongmin Kim, Dongman Lee
School of Computing, KAIST

요 약

한 장소의 사회적 의미(Social meaning), 즉 장소성(Placeness)을 이해하는 것은 공간 기반 서비스를 제공하기 위한 핵심 요소로 대두되고 있다. 현재 정부로부터 제공받을 수 있는 토지피복도는 갱신되기 위해 오랜 시간이 걸리기 때문에 시간에 따라 나타나는 다양한 장소성을 포착하는 데에 한계가 있다. 이 논문에서는 다양한 정보를 지닌 축적된 소셜 미디어 데이터를 통시적으로 분석하여 특정 지역의 장소성과의 연관성을 관측하고자 한다. 최종적으로 소셜 미디어 데이터를 이용하여 특정 지역의 변화하는 장소성을 분석할 수 있는 토대를 마련하고자 한다.

1. 서론

정보통신기술(ICT)의 발전으로 오늘날 많은 형태의 공간 기반 서비스가 제안되고 있다. 양질의 공간 기반 서비스를 제공하기 위한 핵심 요소로서 장소가 갖는 사회적인 의미인 'Sense of place'[1] 즉 장소성(Placeness)의 중요성이 강조되고 있으며 이에 대한 연구가 활발히 진행되고 있다. 장소성을 식별하기 위해서는 해당 장소에서 어떤 사회적 행위(Social Activity)가 어떤 형태로 나타나는지에 대한 충분한 분석과 이해가 동반되어야 한다[2].

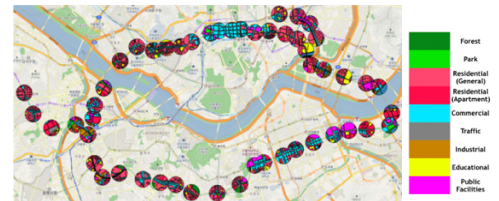
현재 정부에서는 토지피복도를 통해 특정 지역의 장소성에 대한 정보를 제공한다. 하지만 토지피복도는 갱신을 위해 상대적으로 긴 시간을 필요하기 때문에 시간에 따라 나타나는 다양한 장소성을 포착하는 데에 한계가 있다. 우리는 이러한 한계를 geo-tag 된 소셜 데이터를 이용하여 보완할 수 있는 가능성을 알아보고자 한다.

geo-tag 된 소셜 데이터는 게시물의 이미지, Timestamp 와 같은 다양한 정보를 제공한다. 본 논문에서는 이러한 축적된 소셜 데이터로부터 제공받은 여러 정보의 통시적 분석(Diachronic Analysis)을 통해 실제 해당 지역의 장소성과의 연관성을 확인해보고자 한다. 이를 통해 특정 장소에서의 장소성 변화 양상을 관측 및 예측할 수 있는 토대를 마련하여 기존의 한계를 보완할 수 있는 가능성을 확인하고자 한다.

2. 실험 장소 및 사용 데이터

본 논문에서는 서울 지하철 2 호선의 각 역세권을 실험 장소로 선정하였다. 통계청에 따르면 서울 지하철 2 호선은 하루 평균 약 224 만명이 이용하는 사람들이 가장 많이 이용하는 지하철 노선이다[3].

[그림 1]은 서울특별시에서 제공하는 2 호선 노선에 속한 역의 토지피복도이며 이는 해당 지역이 실제로 어떤 형태의 건물들로 사용되는지를 보여준다. 본 논문에서는 토지피복도를 이용하여 각 지하철역으로부터 200m 내의 지역에 한해 5 가지 지리적 특징의 분포를 추출하여 비율로



[그림 1] 서울시 지하철 2 호선 토지피복도

나타내었다. 측정된 5 가지 지리적 특징으로는 Business(업무지구), Apartment(아파트 단지), First(제 1 종 근린생활시설), Second(제 2 종 근린생활시설), Detached(주택 지역)가 있다.

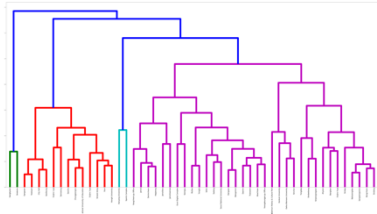
Social Media Data 의 경우 인스타그램에서 각 역의 위치에서 업로드된 포스트를 모두 크롤링하여 사용하였다[4]. 본 논문에서는 인스타그램 포스트의 이미지를 Microsoft Face++ API[5]를 이용하여 분석하였다.

3. 데이터 분석 방법

[그림 2]는 앞서 2 에서 언급한 토지피복도를 이용하여 지하철 2 호선 역을 계층적 군집 분석(Hierarchical Clustering)을 시행한 결과이다. 각 지하철역들은 5 개의 비슷한 군집(Cluster)으로 분류될 수 있다. 그 중에서도 관찰되는 큰 군집을 왼쪽부터 Cluster 1, Cluster 2, Cluster 3 라 차례대로 명명하였다. 다른 두 군집은 상대적으로 매우 작은 군집으로 예외로 인지하여 분석에서 제외하였다. Cluster 1 의 경우는 주로 업무지구의 특징을 지니고 있고, Cluster 2 는 주로 제 1/2 종 근린생활시설을 포함한다. Cluster 3 은 거주지역의 특징을 지닌다.

차례대로 명명하였다. 다른 두 군집은 상대적으로 매우 작은 군집으로 예외로 인지하여 분석에서 제외하였다. Cluster 1 의 경우는 주로 업무지구의 특징을 지니고 있고, Cluster 2 는 주로 제 1/2 종 근린생활시설을 포함한다. Cluster 3 은 거주지역의 특징을 지닌다.

인스타그램 포스트 내의 사회적 행위(Social Activity)를 알기 위해 우선 사회적 행위를 보편적인 6 가지의 카테고리로 분류



[그림 2] 지하철 2 호선 토지피복도의 계층적 군집 분석



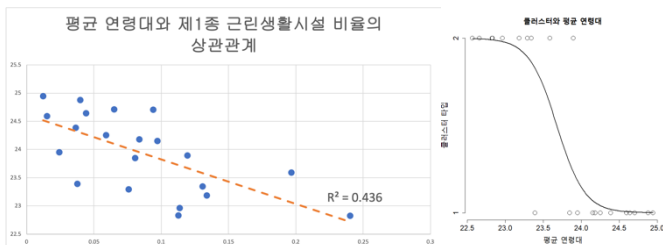
[그림 3] Transfer Learning 결과 중 일부

하였다. 6 가지의 사회적 행위에는 Dining(식사), Coffee(카페 방문), Drink(음주), Leisure(취미 생활), Shopping(쇼핑)과 Sightseeing(관광)이 있다. 각 카테고리 이미지 분류하기 위해 Transfer Learning [6]기법을 통해 Pre-Trained Resnet[7]모델을 학습시켜 분류 작업을 하였다. 본 논문에서는 이 중에서도 대부분의 지역에서 가장 많이 관찰되는 사회적 행위인 Dining(식사)과 Coffee(카페 방문)의 경우만 분석을 진행하였다. 본 논문에서는 이 두 카테고리의 데이터를 이용하여 장소의 특징과 인스타그램 데이터 사이의 상관관계를 확인하고자 하였다.

분류된 인스타그램 포스트의 이미지는 앞선 2 에서 설명한 것처럼 Microsoft Face++ API 를 이용하여 분석하였다. 이 API 는 인스타그램에 포스팅된 이미지 속의 사람의 얼굴을 분석하여 나이, 성별, 그리고 사진 속 사람들이 그룹으로 확인되는 비율(그룹성)에 대한 정보를 얻을 수 있다. 또한, 인스타그램 사진 속 Timestamp 를 이용하여 포스팅 시간 역시 알 수 있었다. 이러한 정보들을 바탕으로 장소의 특징과 인스타그램 데이터가 어떤 상관관계를 갖는지 알아보려고 하였다.

4. 분석된 데이터와 장소성의 연관성

4.1 연령대와 장소성의 연관성

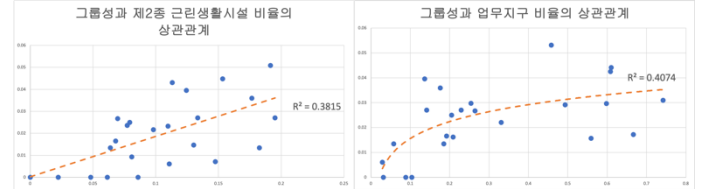


[그림 4] 평균 연령대와 제 1 종 근린생활시설 비율의 상관관계 그래프 및 Cluster 분포도

[그림 4]의 왼쪽 그래프는 분석된 평균 연령대와 Cluster 1, Cluster 2 의 제 1 종 근린생활시설 비율과의 상관관계를 나타낸 것이다. 제 1 종 근린생활시설의 비율이 높을수록 연령대가 낮아짐을 관찰할 수 있다. 또한, [그림 4]의 오른쪽

그래프와 같이 Cluster 2 와 Cluster 1 가 연령대를 통해 분류되는 것을 관찰할 수 있다. 이를 통해 제 1 종 근린생활시설인 필수 생활 시설이 밀집한 곳에는 비교적 낮은 사람들이 확인되고 제 1 종 근린생활시설 특징이 미약한 업무지구의 경우에는 비교적 연령대가 높은 사람들이 활동함을 알 수 있다.

4.2 그룹성과 장소성의 연관성

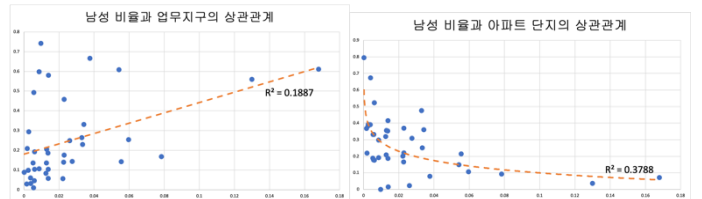


[그림 5] 그룹성과 장소적 특징의 상관관계 그래프

[그림 5]의 왼쪽 그래프는 그룹성과 Cluster 1, Cluster 3 의 업무지구 비율과의 상관관계를 나타낸 것이다. 이를 통해 업무지구에 가까울수록 사람들이 그룹으로 사회적 행위를 함을 알 수 있다.

[그림 5]의 오른쪽 그래프는 그룹성과 Cluster 2, Cluster 3 에 속한 역의 제 2 종 근린생활시설 분포 비율과의 상관관계를 나타낸 것이다. 이를 통해 제 2 종 근린생활시설 특징이 강하게 나타날수록 그룹성이 커짐을 알 수 있다.

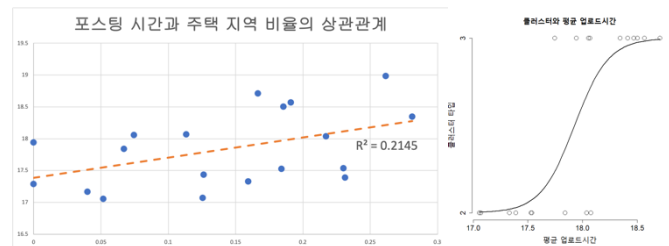
4.3 성별과 장소성의 연관성



[그림 6] 남성 비율과 업무지구 비율 및 아파트 단지의 상관관계 그래프

[그림 6]의 왼쪽 그래프는 업무지구 비율과 포스팅 된 남성의 비율과의 상관관계를 나타낸 것이다. 그래프에 따르면 업무지구의 비율이 높을수록 포스팅 된 남성의 비율이 다른 지역보다 상대적으로 많아지는 경향을 확인할 수 있다. 또한, [그림 6]의 오른쪽 그래프를 보면 업무지구와 반대로 아파트 단지 비율이 높을수록 포스팅 된 남성의 비율이 상대적으로 낮아짐을 알 수 있다.

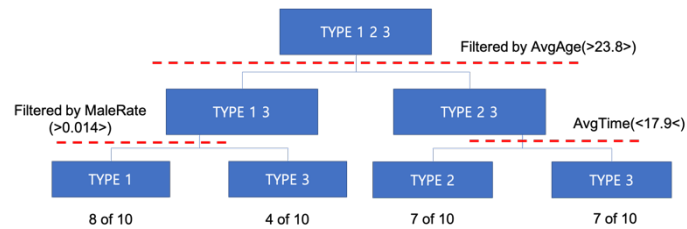
4.4 평균 포스팅 시간과 장소성의 연관성



[그림 7] 평균 포스팅 시간과 주택 지역 비율의 상관관계 그래프 및 Cluster 분포도

[그림 7]의 왼쪽은 Cluster 2 와 Cluster 3 의 주택 지역 비율과 인스타그램 평균 포스팅 시간과의 상관관계를 나타낸 그래프이다. 해당 장소에서 주택 지역 비율이 높아질 수록 평균 포스팅 시간이 늦어지는 경향을 보이며 [그림 7]의 오른쪽 그래프에 따르면 이를 통해 Cluster 2 와 Cluster 3 가 분류됨을 확인할 수 있다.

4.5 분석



[그림 8] Cluster 분류 모델

연령대의 경우 연령대가 높을수록 해당 역의 제 1 종 근린생활시설 분포의 비율이 낮은 경향을 확인할 수 있었다. 또한, 이를 이용하여 두 가지 Cluster를 신뢰적으로 분류할 수 있었다. 이는 제 1 종 근린생활시설인 필수 생활 시설이 밀집한 곳에 비교적 연령대가 낮은 사람들이 많이 활동한다는 점과 제 1 종 근린생활시설 특징이 미약한 업무지구의 경우는 연령대가 비교적 높은 직장인들이 활동함이 반영된 결과이다.

그룹성을 분석한 경우에는 그룹성이 높을수록 해당 역의 업무지구 분포 비율이 높은 경향을 확인할 수 있었다. 이는 회식과 같은 단체적 행동이 많은 업무지구의 특징이 반영된 결과라고 해석할 수 있다. 또한, 제 2 종 근린생활시설이 비율이 높을수록 그룹성 역시 높은 경향을 볼 수 있었다. 이는 제 2 종 근린생활시설인 생활 편의 시설이 밀집한 장소일수록 많은 사람이 함께하는 활동함이 반영된 결과임을 알 수 있다.

성별을 놓고 보았을 때에는 포스팅 된 남성의 비율과 업무지구 비율 및 아파트 단지 비율과의 유의미한 상관관계를 확인할 수 있었다. 이를 통해서 업무지구 비율이 높은 Cluster 1 과 아파트 단지 비율이 높은 Cluster 3 를 비교할 수 있게 한다. 이는 남성들이 다른 지역보다 업무 지구의 특징이 높은 지역에서 비교적 많은 활동함을 알 수 있는 결과이다.

평균 포스팅 시간과 주택 지역 비율과의 상관관계 역시 확인할 수 있었고, 이를 이용하여 Cluster 2 와 Cluster 3 을 분류할 수 있었다. 이는 근린생활시설이 밀집한 장소가 거주 지역이 밀집한 장소보다 더 이른 시간에 사람들이 주로 활동함을 알 수 있는 결과이다.

위의 결과들을 바탕으로 [그림 8]과 같이 인스타그램 사진의 통시적 분석을 통하여 지하철 2호선 역을 특정 값을 기준으로 분류할 수 있었다. 이 모델은 실제 토지피복도의 양상을 예측할 수 있다. 각 Cluster 별로 특징을 확인해보면 업무지구 비율이 높은 Cluster 1 의 경우에는 다른 Cluster 들과 비교했을 때 평균 연령대가 높은 사람들이 주로 활동하며 남성들의 활동이 다른 Cluster 에 비해 높다는 특징을 확인할 수 있다. 제 1 종 및 제 2 종 근린생활시설 비율이 높은 Cluster 2 의 경우에는 평균 연령대가 낮은 사람들이 주로 활동하며 비교적 이른 시간에 활동하는 장소의 특징을 갖는다. 주거지역 비율이 높은 Cluster 3 의 경우에는 여러 가지 복합적인 특징을 갖고 있는 것을 확인할 수 있다.

5. 결론 및 향후 연구

본 논문에서는 소셜 데이터의 통시적 분석과 토지 피복도를 통한 해당 장소의 실제 장소성과의 상관관계를 확인할 수 있었다. 이러한 결과는 같은 행위일지라도 목적에 따라 다른 사회적 행위가 나타나며 소셜 데이터의 통시적 분석을 통해 이를 확인할 수 있음을 시사한다. 결론적으로 소셜 데이터로부터 받는 여러 데이터의 통시적 분석 결과들을 조합하여 토지피복도로 분류되는 각 Cluster 의 특징을 확인할 수 있었다. 또한, 이를 바탕으로 해당 장소의 장소성을 확인하는 모델을 구상할 수 있다는 가능성을 확인할 수 있다.

소셜 데이터의 통시적 분석 결과를 토대로 해당 지역에서의 다양한 장소성의 변화를 확인하는 것 역시 매우 중요하다. 소셜 데이터의 통시적 분석과 더불어 물리 센서로부터 얻은 데이터를 함께 분석함으로써 보다 정확하고 민감하게 장소성을 확인하는 것이 중요한 연구가 될 것이다.

6. 사사 문구

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학 사업의 연구결과로 수행되었음"(2016-0-00018)

7. 참고문헌

- [1] Harrison, S., & Dourish, P. (1996, November). Re-place-ing space: the roles of place and space in collaborative systems. In Proceedings of the 1996 ACM conference on Computer supported cooperative work (pp. 67-76). ACM.
- [2] 김세연, 박기남, Gaurav Kalra, 이동만, (2018). 도시 공간에서의 사회적 행위와 물리 센서 데이터 간의 상관관계 분석 - 북촌지역 사례연구. 한국정보과학회 학술발표논문집, (), 1674-1676.
- [3] 조성근, et al. "서울지하철 2 호선 혼잡도 완화방안 기초연구." *한국철도학회 학술발표대회논문집* (2013): 193-203.
- [4] Minsang Yu, Gaurav Kalra, Dongman Lee, Meeyoung Cha, Daeyoung Kim, "Ballparking the Urban Placeness: A Case Study of Analyzing Starbucks Posts on Instagram", The 10th international Conference on Social Informatics (SocInfo2018), St.Petersburg, Russia, Sep 25-28, 2018
- [5] "Face API - 얼굴 인식 소프트웨어", Microsoft Face API, last modified April 2, 2019, accessed April 29, 2019, <https://azure.microsoft.com/ko-kr/services/cognitive-services/face/>
- [6] HE, Kaiming, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 770-778.
- [7] Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." *IEEE Transactions on knowledge and data engineering* 22.10 (2010): 1345-1359.