

누적합 관리도를 활용한 생활인구 이상치 탐색

김형준 · 손소영[†]

연세대학교 산업공학과

CUSUM Chart Applied to Monitoring Areal Population Mobility

Kim, Hyoung Jun · Sohn, So Young[†]

Department of Industrial Engineering, Yonsei University

ABSTRACT

Purpose: Certain places in Seoul such as Shinchon, Hongdae, and Gangnam, often suffer from sudden overflow of mobile population which can cause serious safety problems. This study suggests the application of spatial CUSUM control chart in monitoring areal population mobility data which is recently provided by Seoul metropolitan government.

Methods: Monitoring series of standardized local Moran's I enables one to detect spatio-temporal out-of-control status based on the accumulation of past patterns. Moreover, we visualize such pattern map for more intuitive comprehension of the phenomenon. As a case study, we have analyzed the female mobility population aged 25 to 29 appeared in 51 Jipgyegu near Hongik university on fridays from January, 2017 to June, 2018. They are validated by exploring related articles and through local due diligence.

Results: The results of the analysis provide insights in figuring out if the change of the mobility population is short-term by particular incident or long-term by spatial alteration, which allows strategic approach in constructing response system. Specific case near popular downtown near Hongik University has shown that newly opened hotels, shops of global sports brand and franchise bookstores have attracted young female population.

Conclusion: We expect that the results of our study contribute to planning effective distribution of administrative resources to prepare against drastic increase in floating population. Furthermore, it can be useful in commercial area analysis and age/gender specific marketing strategy for companies.

Key Words: Local Moran's I, CUSUM Control Chart, Outlier Detection, Areal Population

● Received 22 March 2020, 1st revised 20 April 2020, accepted 3 May 2020

† Corresponding Author(sohns@yonsei.ac.kr)

© 2020, Korean Society for Quality Management

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.

1. 서 론

서울시의 신촌, 강남, 홍대, 이태원 등의 특정 지역들에 특정 날과 시간대에 지역이 감당할 수 있는 수준 이상의 사람들이 몰리는 경우가 종종 발생한다. 이러한 비이상적인 인구 이동 패턴으로 여러 사람들이 몰렸을 때 안전사고가 발생할 수 있고, 심하게는 생명까지 잃을 수도 있어 초기 단계에서 이상치 관리가 중요하다(Zhang et al., 2018). 하지만 이전에는 사용할 수 있는 데이터가 비교적 큰 행정동 단위, 혹은 주민등록인구 등 밖에 존재하지 않아 연구의 범위가 매우 제한적이었다.

최근에는 휴대폰 이용량, 유동인구 등의 데이터가 사용 가능해지면서 연구가 활발하게 이뤄지기 시작하였다. 특히 서울시는, 서울의 공공데이터와 KT의 LTE 시그널을 이용하여 서울시 집계구¹⁾별로 해당 시각에 존재하는 “생활인구”²⁾를 추계하여 제공하기 시작하였다. 이는 다른 말로 해당 지역에 거주하거나, 직장, 관광, 의료, 쇼핑, 교육 등 목적으로 머무는 모든 인구를 포함한다. 서울의 인구는 서울에 거주하는 인구뿐만 아니라 위성도시에서 출퇴근하는 수많은 직장인과 관광객들이 혼재되어 있기 때문에 상주인구가 도시서비스의 수요를 잘 설명하지 못하기 때문에 생활인구의 필요성이 대두되었다. 본 연구는 서울시에서 제공하는 생활인구 추계 빅데이터를 모니터링하여 이상치 스파이크가 언제 어디에서 발생하는지를 규명하고자 한다.

우리는 공간연구에서 흔히 사용되는 국지적 Moran지수를 이용하여 주위와 대비하여 인구가 많이 몰렸는지 알아보고, 해당 지표를 표준화하여 기대치로부터 차이의 누적합(Cumulative Sum)을 시간대별로 관리하여 특정지역의 이상치 탐색을 시도하였다(Kracalik et al. 2011). 더불어 Yamada et al. (2009)의 연구에서 고안한 GeoSurveillance를 이용하여 집계구별 누적합을 시각화하였다. 해당 값들을 지도와 매칭하여 젊음의 거리이자 서울 문화의 집합체라 할 수 있는 마포구 서교동(홍익대학교 인근)에 예시로 적용한 결과 생활인구 원 데이터에서는 파악하기 어려웠던 시공간적 이상치를 탐지 할 수 있었다. 도출된 시간대와 지역은 인구 수요가 주위보다 혹은 과거보다 갑자기 많아진 곳으로써 제한된 경찰, 구급인원 등의 행정적 자원의 효율적인 분배를 정량적인 근거 아래 가능케 한다. 또 반복되는 패턴 파악을 통해 사전경고를 하여 시민들의 편의성을 증대시킬 수 있을 것으로 기대된다.

이어지는 본문에서는 인구이동 패턴 이상치 탐지와 누적합 관리도에 관련된 선행 문헌들을 고찰하였다. 다음으로 사용하고자 하는 프레임워크의 알고리즘과 생활인구 데이터에 대해 상세히 설명하고, 해당 방법론을 마포구 서교동에 적용하여 결과를 제시하고 해석하였다. 마지막으로 본 연구가 기여하는 점과 제한점에 대해 서술하였다.

2. 이론적 배경 및 선행연구

2.1 인구이동 패턴 이상치 탐지

도시에서 발생하는 인구이동 패턴의 이상치를 탐지하는 것은 도시 관리와 긴급 사태 대비 측면뿐만 아니라, 사회적인 서비스 차원에서 좋은 기회를 부여할 수 있다(Chae et al., 2014). Gundogdu et al. (2016)는 포아송 분포를 가장한 마코프 모델을 사용하여 휴대폰 사용량이 갑작스럽게 증가하는 시공간적 이상치를 탐지하였다. 코트디부아

1) 집계구란 통계청에서 통계조사 자료를 집계하기 위한 편의상의 구역으로 인구는 500명 정도이면서, 면적은 읍면동의 약 1/25이다. 서울시에는 19,153개의 집계구가 존재한다.

2) 생활인구 데이터는 서울열린데이터광장(<https://data.seoul.go.kr>)에 2017년 1월 1일 0시부터 2018년 6월 30일 23시까지 0세부터 70세 이상까지 성별 5세 구간, 시각별 1시간 구간으로 공개되어 있다.

르에서 실제 발생한 폭력적인 사태, 시위, 휴일 그리고 주요 스포츠 이벤트와 연결지어서 정성적으로 설명하였다. 마찬가지로, Chae et al. (2014)의 연구에서도 소셜 미디어 데이터의 토픽을 추출하여 해당 내용이 시공간적으로 집중되는 행태를 시각화하는 시각 분석방법론을 제안하였다. 최근에 Zhang et al. (2018)의 연구에서 도심에서의 예상치 못한 인구이동의 이상치를 찾는 것의 중요성을 강조하였다. 저자들은 단일 클래스 서포트 벡터 머신(one-class Support Vector Machine)을 이용하여 과거와는 다른 패턴을 보이는 지역과 시간을 탐색하였다. 다양한 연구들이 행해졌음에도 불구하고 해당 논문들은 공간자기상관성을 고려하지 않았고, 추가적으로 확률분포 가정이나 모형 학습을 위한 대용량 데이터를 필요로 하는 제한점이 존재했다.

국내에서는 해당 분야의 연구가 비교적 최근에서야 유동인구 데이터가 공개되면서, 연구논문과 보고서가 많이 나왔다. 유동인구는 시민들의 이동패턴을 잘 나타내기 때문에, 인구이동 패턴 탐지 이외에도 다양한 분야에서 적용되었다. Jeong & Moon (2014)은 시간대별 유동인구를 활용하여 서울에서 설정한 공간구조 체계가 인구 이동 특성을 얼마나 반영하고 있는지를 알아보았다. 실제로 유동인구 수가 많은 지역의 경우 생활밀접업종의 수와 유의한 관련이 있었다(The Seoul Institute, 2013, January 28; The Seoul Institute, 2014, December 1) 또 유동인구를 소비수요로 판단하고 상업 측면의 소비 객체로서 역할이 어떻게 되는지 알아보는 연구들도 활발히 진행되었다(Lee & Jung, 2014; Lee et al., 2014;). Kim et al. (2016)의 연구에서는 대피소 수용 능력과 같은 인구이동패턴에 따른 적절한 구조물 설치 여부를 파악하였다. 또 위험에 노출된 정도를 판단하기 위해 유동인구가 사용되기도 하였다(The Seoul Institute, 2013, February 12).

2.2 공간 통계량을 이용한 누적합(Cumulative Sum) 관리도

이상치를 감지하는 방법론에는 크게 두 가지가 존재한다(Rodeiro & Lawson, 2006; Rogerson, 1997; Yamada et al., 2009). 첫 번째로는 Retrospective change detection으로, 고정된 데이터 샘플을 이용하여 해당 기간 동안 변화가 일어났는지 확인하고, 그 변화의 원인이 무엇인지 판단하게 된다. 극단치 이론을 활용하여 서울시의 초미세 먼지 농도 수준을 지역별로 분석한 Oh & Lim (2019)의 연구가 좋은 예이다. 반면에 Prospective change detection의 경우 새로운 관측치가 이전 프로세스에 추가되어 해당 관측치가 이상치 인지 판단하거나 관측치를 계속 수집할 것인지 결정하게 된다. 후자의 시공간 이상탐지 방법론으로는 베이지안 모델(Rodeiro & Lawson, 2006), 공간통계량(Rogerson, 1997), 그리고 품질 관리도(Lund & Seymour, 1999; Yamada et al., 2009)가 있다. 이외에도 최근에는 머신러닝을 활용한 방법론들이 많이 연구되었지만(Zhang et al., 2018; Seo & Yun, 2019), 학습을 위해서는 다량의 데이터를 필요로 하고, 모형들의 1차적인 목표는 과거의 추세를 활용하여 미래를 예측하는 것이다(Han & Yu, 2019). 본 연구는 미래 예측이 아닌 짧은 기간의 데이터만으로도 과거에는 없었던 변동성에 민감하게 반응하는 방법론을 고려하고자 한다.

품질 관리도 중에서도 누적합 관리도가 공간 관리도에서 자주 사용된다. 본래 산업에서 품질 관리를 위해 사용되던 누적합 관리도는 Page (1954)에 의해 제안되었으며, 평균으로부터 미세한 변화를 탐지할 수 있다는 점에서 뛰어난 관리도라고 할 수 있다(Kim & Lee, 2005). 해당 방법론은 공정 분야 이외에 지속적인 모니터링이 필요한 공중보건 관리 분야에도 활용되었고(Rogerson, 1997), 조금 더 발전되어 공간상에서의 관리도로 발전하게 되었다(Rogerson, 1997; Rogerson, 2005; Sonesson, 2007). Han et al. (2008)은 누적합(Cumulative Sum, CUSUM) 관리도와 스캔 통계량의 성능을 비교하여, 누적합 관리도가 더 좋다는 점을 밝혀내기도 하였다. 하지만 국내에서는, Lee (2015)의 연구에서 시공간 탐지를 위한 관리도 활용에 대한 선행연구 고찰을 제외하고는, 대부분의 경우 여전히 관리도 개념이 공정 관리 분야에 한정적으로 사용되고 있다(Kwon et al., 2016).

공중 보건 분야에서 사용되는 공간 CUSUM 차트는 대부분 스코어 통계량을 기반으로 한다. 이 통계량은 포아송 분포를 가정하여 예측치보다 실제 관측치가 유의하게 많은 지역을 발견하는데 사용된다. 하지만 생활인구의 경우 Anselin(1995)에 의해 고안된 국지적 모란지수(Local Moran's I)를 사용한다면 확률분포에 대한 가정이 필요하지 않고, 동시에 공간 자기상관성을 고려할 수 있어, 공간 클러스터 탐지에서 우수한 성능을 얻을 수 있다(Kracalik et al., 2011; Manogaran & Lopez, 2018). 특정 지역의 생활인구가 인접한 지역에 크게 영향을 받을 수밖에 없는 것을 고려한다면 공간 통계량을 활용한 누적합 관리도가 본 연구에 가장 적합하다고 판단된다.

3. 제안 프레임워크

생활인구란 특정 시점과 지역에 존재하는 모든 인구를 의미한다. 이때 특정 시점이란 한 시각을 의미하고 하루에 0시, 1시, ..., 23시까지 24개의 데이터가 발생한다. 지역은 서울시에서의 19,153개의 집계구 단위로 측정된다. 집계구란 통계청에서 통계조사 자료를 집계하기 위한 편의상의 구역으로 인구는 500명 정도이면서, 면적은 읍면동의 약 1/25이다. 이는 각 지역의 서비스 인구를 정확히 측정하기 위하여 서울시가 보유하고 있는 공공빅데이터와 통신기업 KT에서 제공한 LTE 시그널 데이터를 이용하여 추계한 데이터이다. 상주인구 혹은 거주인구가 도시서비스의 수요를 잘 설명하지 못하기 때문에 생활인구의 필요성이 대두되었다. 서울의 인구는 서울에 거주하는 인구뿐만 아니라 위성도시에서 출퇴근하는 수많은 직장인과 관광객들이 혼재되어 있기 때문에, 생활인구 데이터의 활용범위는 무궁무진하다고 할 수 있다.

본 연구에서는 해당 생활인구 추계 데이터와 국지적 모란지수의 누적합 관리도를 이용하여 생활인구 이상치를 탐지하는 프레임워크를 제안하고자 한다. 우선, 서울시의 관심 행정동의 집계구 i ($= 1, 2, \dots, n$) 별로 17년-18년 상반기에 해당하는 t ($= 1, 2, \dots, T$) 주차 특정 요일, 시간대, 연령대, 성별 생활인구 데이터 $C_{i,t}$ 를 수집한다.

국지적 모란지수는 관찰하고자 하는 지역과 주변 지역의 공간자기상관성을 측정하는 지표이다. 전역적인 공간 자기상관성을 측정하는 모란지수와 대비하여, 지역적인 특성을 잘 파악할 수 있다. 공간적 자기상관이 주변에 대비하여 어떤 지역에서 높게 혹은 낮게 나오는지 알 수 있으며, 값의 부호에 따라 이웃 지역과의 유사정도를 알 수 있다(Kwon & Sohn, 2017). 지역의 국지적 모란지수는 다음 식 (1)과 같이 표현된다.

$$I_{i,t} = \frac{C_{i,t} - \overline{C_{i,t}}}{S_c^2} \cdot \sum_{j=1}^n [w_{i,j} \cdot (C_{j,t} - \overline{C_{j,t}})] \quad (1)$$

여기서 $\overline{C_{i,t}}$ 는 해당지역과 인접지역의 평균값, S_c^2 은 주변지역 값들의 분산을 나타내며, $w_{i,j}$ 는 공간가중행렬에서 공간 단위 i 와 j 의 인접 여부나 정도를 나타낸다. 이 때, 공간가중행렬을 인접성 여부 또는 거리 기준으로 구성할 수 있다(Anselin, 1988). 인접성 척도는 지리적 경계를 맞닿는 여부를 기준으로 하며, 경계선만 공유하고 있다면 Rook, 경계점까지 공유하고 있다면 Queen 공간가중행렬이라 한다(Lloyd, 2010). 거리 척도의 경우에는 지역 간 거리에 반비례하는 역거리행렬을 공간가중행렬로 사용할 수 있다(Kim & Lee, 2011). 식 (1)에 따르면 국지적 모란지수는 자신의 표준화된 값과 주변 지역의 표준화된 값의 합을 곱한 값이다. 이 값을 통해 관찰하고자 하는 지역과 인접한 지역의 유사성 또는 비유사성의 정도를 나타낼 수 있는 것이다. 표준화된 관리도를 사용 하는 것은 관리도 파라미터를 여러 개의 차트에 동일하게 적용될 수 있다는 점에서 장점이 있다(Montgomery, 2009). 따라서 t 시점 i 집계구의 표준화된 국지적 모란지수를 다음과 같이 구하여 사용한다.

$$z_{i,t} = \frac{I_{i,t} - \mu_i}{\sigma_i} \quad (2)$$

$$\text{where } \mu_i = \sum_{t=1}^T \frac{I_{i,t}}{T}, \sigma_i = \frac{1}{T-1} \sum_{t=1}^T (I_{i,t} - \mu_i)^2$$

CUSUM 관리도로 탐지하고자 하는 이상 프로세스의 평균을 μ_1 이라고 할 때, 목표 평균 μ_0 와 변동 수준을 나타내는 $\delta\sigma$ 로 μ_1 을 식 (3)과 같이 표현할 수 있다 (Montgomery, 2009). 식 (4)는 μ_0 와 μ_1 의 평균으로부터 변동의 절댓값이고, k 로 표시한다. 표준화된 누적합 관리도에서는 $\mu_0 = 0, \sigma = 1$ 이기 때문에, $\delta = 1$ 로 정했을 때 $k = 0.5$ 로 설정된다(Yamada et al., 2009).

$$\mu_1 = \mu_0 + \delta\sigma \quad (3)$$

$$k = \frac{|\mu_1 - \mu_0|}{2} = \frac{\delta\sigma}{2} \quad (4)$$

정상적인 $z_{i,t}$ 프로세스 (평균=0, 표준편차=1)에서, 프로세스 평균이 k 를 벗어나는 이상치 프로세스를 관리하기 위해 식 (5)와 같은 누적합 관리도를 사용한다. 보통 누적합 관리도는 부호에 따라서 쌍방향으로 관리도를 그려 양과 음으로 벗어나는 이상치를 각각 탐지한다. 하지만 우리의 관심사는 인구가 많이 몰리는 것에 있기 때문에 단방향 누적합 관리도를 사용하고자 한다.

$$S_{i,t} = \max(0, S_{i,t-1} + z_{i,t} - k), S_{i,0} = 0 \quad (5)$$

각 시점별 표준화된 국지적 모란지수의 변동치가 k 를 넘게 되면 해당 값들이 이전 시점의 누적합에 쌓이게 된다. 누적합이 특정 임계치 h 를 넘게 되면 이상치로 판단하고 경고하게 된다. 임계치 h 의 수준은 정상 프로세스임에도 불구하고 이상치라 판단할 오류(1종 오류, α)를 얼마나 허용하느냐에 관련된다고 할 수 있다. 관리도에서는 이를 오경보가 올린 시점 간 평균 기간인 in-control average run length (ARL_0)로 판단할 수 있는데, 이는 $\frac{1}{\alpha}$ 과 동일하다(Qiu, 2013). 만약 $ARL_0 = 100$ 이라면 100번의 구간마다 한번 오경보가 올리는 것이 기대된다는 뜻이고(Kim & Lee, 2005), 1종 오류가 0.01이라고 할 수 있다.

Rogerson (2005)은 주어진 수준의 k 와 ARL_0 이 있을 때, h 를 추정할 수 있는 근사식을 도출하였다. 하지만, 공간 누적합 관리도는 n 개의 관리도를 동시에 고려하기 때문에 Rogerson (2005)의 추정식에서 ARL_0 대신 $n * ARL_0$ 을 이용하여야 분석 지역에 대해 특정 관리 수준을 유지할 수 있다(Yamada et al., 2009). 따라서, 최종적으로 임계치 h 는 식 (6)과 같이 추정할 수 있다(Yamada et al., 2009).

$$h \approx \left(\frac{2k^2 n ARL_0 + 2}{2k^2 n ARL_0 + 1} \right) \cdot \frac{\ln(2k^2 n ARL_0 + 2)}{2k} - 1.166 \quad (6)$$

우리는 Kracalik et al. (2011)와 Manogaran & Lopez(2018), Yamada et al. (2009)의 연구에서 제안된 프레임 워크를 이용하여 국지적 모란지수의 CUSUM 관리도를 그리고, GeoSurveillance 프로그램(Yamada et al., 2009)을 이용하여 시각화 하고자 한다. 아래의 Table 1은 사용하고자하는 연구 방법론의 개괄적인 알고리즘을 나타낸다.

Table 1. Algorithm for CUSUM Control Chart using Local Moran's I

알고리즘: 국지적 모란지수의 CUSUM 관리도	
1. Input:	17년부터 18년 상반기 서울시 생활인구
2. Let:	$C_{i,t}$ 는 집계구 i , t 주차 특정 요일/시간대/연령대/성별 생활인구
3. Calculate:	국지적 모란지수, $I_{i,t} = \frac{C_{i,t} - \bar{C}_t}{S_c^2} \cdot \sum_{j=1}^n [w_{i,j} \cdot (C_{j,t} - \bar{C}_t)]$, $w_{i,j}$ 는 공간이웃행렬
4. Calculate:	표준화된 국지적 모란지수, $z_{i,t} = \frac{I_{i,t} - \bar{\mu}_i}{\sigma_i}$
5. Set:	파라미터 k (허용 범위), ARL_0 (1종오류) 설정
6. Calculate:	임계치 $h \approx \left(\frac{2k^2 n ARL_0 + 2}{2k^2 n ARL_0 + 1} \right) \cdot \frac{\ln(2k^2 n ARL_0 + 2)}{2k} - 1.166$
7. Initialize:	누적합 $S_{i,0} = 0$
8. Calculate:	누적합 $S_{i,t} = \max(0, S_{i,t-1} + z_{i,t} - k)$
9. Output:	누적합 $S_{i,t}$ 의 이상치(임계치 h 초과) 유무 판단
10. Plot:	집계구별 CUSUM 관리도 & 시점별 CUSUM 지도

4. 데이터

직관적인 결과 해석과 세분화된 인구가동 패턴을 확인하기 위하여 지역별, 연령대 별, 성별을 특정지어 연구를 진행하였다. 우선, 분석 대상지로는 서울 문화공간의 중심지라고 불리는 홍익대학교 인근을 선정하였다. 해당 지역은 문화활동이 가능할 뿐만 아니라 덴스클럽, 대형음식점, 소규모 음식점이 존재하는 등 다양한 볼거리와 먹거리가 공존하는 공간이다(Gim, 2013). 따라서, 홍대 인근은 주말(금요일부터 일요일) 늦게까지 20대 생활인구가 많이 존재하는 지역 중 하나다. 그 중에서도 트렌드에 민감한 여성의 이동 패턴을 고려하였다. 해당 지역은 그림 1에서와 같이 마포구 서교동으로 행정적으로 소속되어 있고, 그림 2에 나와 있듯이 50개의 집계구로 구획되어있다. 해당 연구에서는 2017년부터 2018년 상반기까지 금요일 저녁(17시부터 01시까지의 평균)에 해당하는 74일(인구의 변화가 큰 설날 및 추석 제외)을 25세에서 29세까지의 여성 생활인구를 추출하였다.

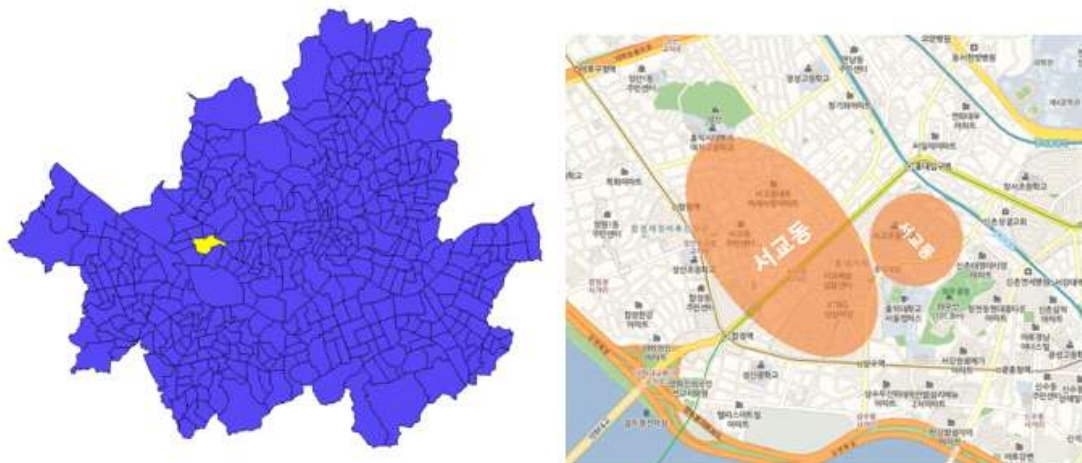


Figure 1. Map of 'dong's in Seoul and Seogyo-dong



Figure 2. Map of 'Jipgyegu' in Seogyo-dong

Figure 3은 수집한 서교동의 25세에서 29세까지 여성 생활인구를 금요일 17시에서 01시 동안의 평균값을 그래프로 나타낸 것이다. 특정 집계구에 월등히 높은 생활인구가 집중 되어 있는 것을 관찰할 수 있고, 몇몇 집계구에서는 생활인구가 거의 없는 것을 확인 할 수 있다. 생활인구의 높고 낮음이 비슷한 추이를 보이지만, 시간이 지날수록 생활인구가 늘어나기도, 줄어들기도 하는 것을 관찰할 수 있다. 전반적으로는 큰 변화가 관찰되지 않고 평균의 주변에서 크게 벗어나지 않은 것을 확인 할 수 있다.

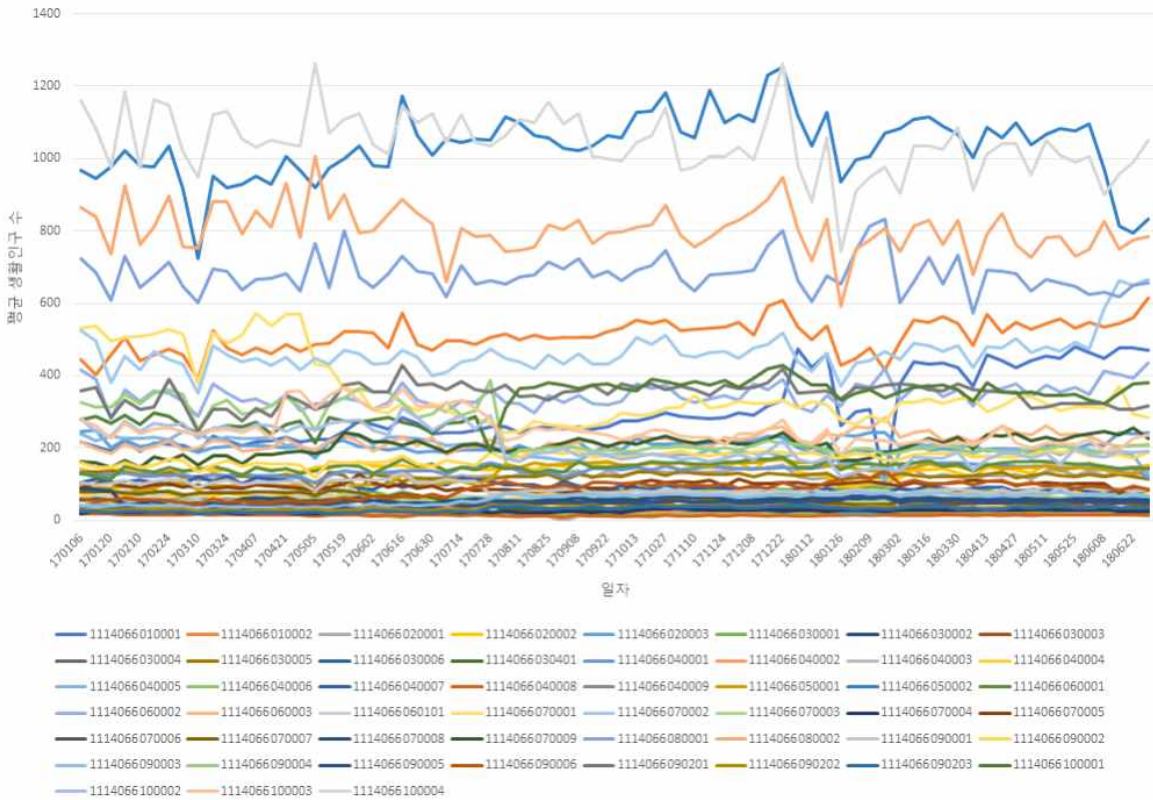


Figure 3. Average Areal Population of Female (aged 25-29) for each Jipgyegu in Friday Nights (17-01)

Table 3은 25세에서 29세까지 여성의 금요일 저녁의 평균 생활인구 기초통계량을 집계구별로 나타낸다. 이를 보았을 때, 특정 집계구들에 생활인구가 쏠려 있는 것을 확인 할 수 있다. 집계구 1114066100004는 상상마당 주위로 술집과 클럽이 다수 포진 되어 있는 곳이다. 마찬가지로 집계구1114066080002에는 다수의 식당과 클럽이 분포하고 있고, 집계구 1114066060002는 “걷고 싶은 거리”로 음악, 마술 등 다양한 버스킹 공연이 이루어지는 곳이다. 반면 공항철도 홍대입구역 쪽과 망원역 쪽 거주지역은 상대적으로 생활인구가 적은 것을 확인 할 수 있다. 또 최근에 생활인구가 많이 몰리고 있는 집계구들이 표준편차가 높게 도출되었다. 집계구 1114066050002, 1114066070001, 1114066010001은 각각 홍대입구역 1번 출구 부근, 경의선 기차길 창전동 부근, 경의선 기차길 연남동 부근으로 새로운 상점과 건물 혹은 기차길공원이 완공 되면서 상권이 활성화된 대표적인 지역이라고 할 수 있다.

Table 3. Descriptive Statics for Female (aged 25 to 29) Areal Population on Friday Nights from January 2017 to June 2018

집계구코드	평균	최대	최소	표준편차	집계구코드	평균	최대	최소	표준편차
1114066010001	304	479	103	95	1114066060003	220	285	181	21
1114066010002	509	613	390	44	1114066060101	72	106	56	9
1114066020001	125	152	88	12	1114066070001	355	572	242	94
1114066020002	142	173	91	15	1114066070002	464	667	351	53
1114066020003	208	250	174	17	1114066070003	61	74	49	6
1114066030001	80	96	59	9	1114066070004	26	35	18	4
1114066030002	54	71	36	8	1114066070005	97	126	75	10
1114066030003	46	55	32	4	1114066070006	48	74	28	8
1114066030004	38	48	16	6	1114066070007	61	84	39	12
1114066030005	123	136	87	9	1114066070008	58	89	41	8
1114066030006	21	25	16	2	1114066070009	205	255	147	26
1114066030401	29	37	21	3	1114066080001	680	834	571	49
1114066040001	141	194	105	17	1114066080002	804	1006	591	65
1114066040002	20	24	17	2	1114066090001	91	129	73	14
1114066040003	20	36	0	10	1114066090002	172	194	117	15
1114066040004	69	83	45	6	1114066090003	59	79	32	15
1114066040005	198	244	112	26	1114066090004	246	388	165	64
1114066040006	15	18	10	2	1114066090005	45	65	25	11
1114066040007	94	121	65	13	1114066090006	85	137	44	22
1114066040008	14	19	8	2	1114066090201	348	429	245	32
1114066040009	55	64	19	9	1114066090202	35	45	24	5
1114066050001	130	170	74	27	1114066090203	31	45	18	7
1114066050002	1032	1252	724	92	1114066100001	327	428	193	54
1114066060001	149	175	115	12	1114066100002	208	311	121	47
1114066060002	347	434	282	31	1114066100003	252	369	183	44
					1114066100004	1042	1264	742	85

5. 결과 및 고찰

주어진 데이터를 이용하여 $k = 0.5$, $ARL_0 = 100$ 혹은 1종오류율 $\alpha = 0.01$ 으로 설정하였고, 식 (3)으로 $h = 6.168$ 을 도출하였다. Table 4는 분석 구간(74개)동안 집계구별로 이상치가 발생한 횟수를 나타낸다. 공간가 중행렬로는 한 점이라도 맞닿으면 이웃으로 간주하는 Queen 방식을 택하였다. Table 4의 이상치 탐지 개수의 직관적인 이해를 돕기 위하여 지도화하여 Figure 4로 나타내었다. Table 4와 Figure 4를 통해서 2017년부터 2018년

상반기 동안 어느 지역에서 평소보다 그리고 주위보다 많은 변화가 있었는지 알 수 있었다.

가장 이상치가 많이 발생했던 구역으로는 Table 3에서 표준편차가 높은 지역으로 꼽혔던 홍대입구역 1번 출구 쪽이었다. 해당 지역에는 건물들이 리모델링을 하거나 새 건물들이 들어서게 되면서 대형프랜차이즈 음식점이나 유명 운동브랜드 신발 상점이 생겼다. 무엇보다도 3번 출구에 위치한 연남동의 경의선 숲길이 큰 영향력이 있었을 것으로 생각한다. 연남동의 상권이 사람들이 많이 몰리기 시작하여 급속도로 성장하면서 주변으로 상권이 빠르게 퍼져 나갔다. 상위권에 위치한 다른 지역으로는 평균치로 가장 많았던 상삼마당 근처를 들 수 있겠다. 앞서도 언급되었다시피 해당 지역을 필두로 하여 주변에 댄스클럽과 술집, 음식점이 다수 위치하고 있다. 최근 들어서 핫플레이스로 주목받고 있는 상수역과 합정역 주변도 이상치가 많이 탐지되었다. 상수역에는 개성있는 상점들이 많이 위치하고 있다. 특히, 독특한 카페와 악세사리/패션의류 상점들이 많이 들어서면서 상수역만의 새로운 상권이 형성되고 있다. 합정역 근처에는 복합 문화공간으로 사용되는 주상복합 건물을 필두로 망원역까지 이어지는 집계구에서 이상치가 많이 발생한 것을 확인 할 수 있다. 이 지역도 마찬가지로 아기자기한 카페와 분위가 좋은 레스토랑 등이 입소문을 타면서 새롭게 뜨는 상권이라 할 수 있다.

반면 하위권 지역을 보게 되면, 합정역과 망원역을 있는 큰 도로 뒤쪽의 주거지역에서는 거의 변화가 없었던 것을 확인 할 수 있다. 마찬가지로 동교동을 맞닿고 있는 경의선 숲길도 최근에서야 완공이 되면서 활기를 띄기 시작했기 때문에, 큰 변화를 감지할 수 없었다. 홍익대학교 바로 앞의 홍대 예술의 거리에서는 예상외로 이상치가 발생하지 않았는데, 이는 큰 변동없이 비슷한 수준의 생활인구 수를 유지했기 때문이라고 생각이 된다.

Table 4. The Number of outliers for each Jipgyegu during the study period

순위	집계구 코드	탐지 횟수	순위	집계구 코드	탐지 횟수	순위	집계구 코드	탐지 횟수
1	1114066050002	35	18	1114066090003	11	35	1114066050001	2
2	1114066100003	31	19	1114066010002	8	36	1114066060003	2
3	1114066060002	30	20	1114066040005	8	37	1114066070002	2
4	1114066090201	29	21	1114066040002	7	38	1114066030002	2
5	1114066100002	28	22	1114066040003	7	39	1114066020003	1
6	1114066010001	27	23	1114066090203	7	40	1114066070007	0
7	1114066100001	26	24	1114066070006	6	41	1114066070005	0
8	1114066090006	22	25	1114066020002	6	42	1114066070009	0
9	1114066030401	21	26	1114066080002	5	43	1114066070003	0
10	1114066060001	20	27	1114066070008	4	44	1114066080001	0
11	1114066030004	20	28	1114066090005	4	45	1114066090001	0
12	1114066040004	19	29	1114066020001	4	46	1114066090002	0
13	1114066070004	17	30	1114066040007	4	47	1114066090004	0
14	1114066060101	14	31	1114066070001	3	48	1114066030006	0
15	1114066100004	14	32	1114066030005	3	49	1114066030001	0
16	1114066040001	13	33	1114066030003	3	50	1114066040008	0
17	1114066090202	12	34	1114066040009	3	51	1114066040006	0

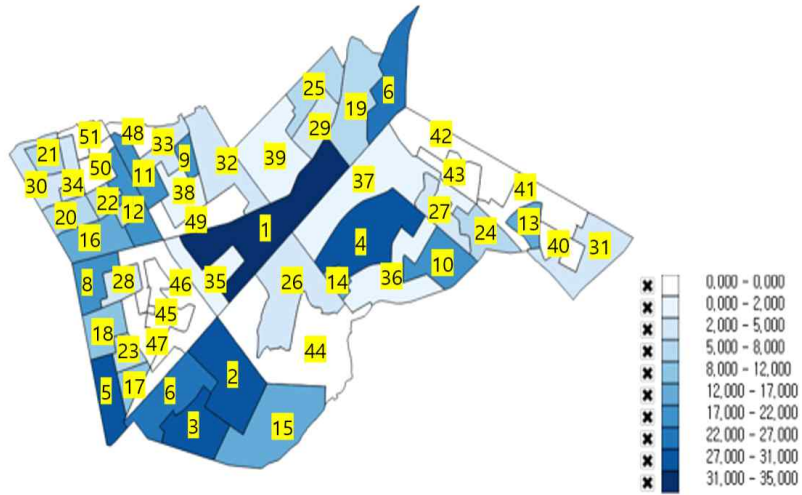


Figure 4. Map of Jipgyegu in Seogyo-dong based on the Number of Outliers

우리는 이상치가 가장 많이 나온 집계구 1114066050002에 대하여 상세한 분석을 위하여 원데이터와 국지적 Moran지표 데이터, 누적합 그래프를 그려보았다. 집계구 1114066050002은 앞서 언급되었다시피 홍대입구역 1번 출구 근처이다.

Figure 5는 해당 집계구에서 분석기간 동안의 금요일 저녁 평균 생활 인구수를 나타낸다. 크고 작은 변동이 존재하지만 1000명의 안팎의 생활인구가 존재하는 것을 확인할 수 있다. 3월 10일 최소치, 12월 22일은 최대치를 기록하였다. 12월 22일은 크리스마스인 25일이 월요일이 되면서 긴 크리스마스 연휴를 시작했기 때문에 많은 인원이 모인 것으로 생각할 수 있다.

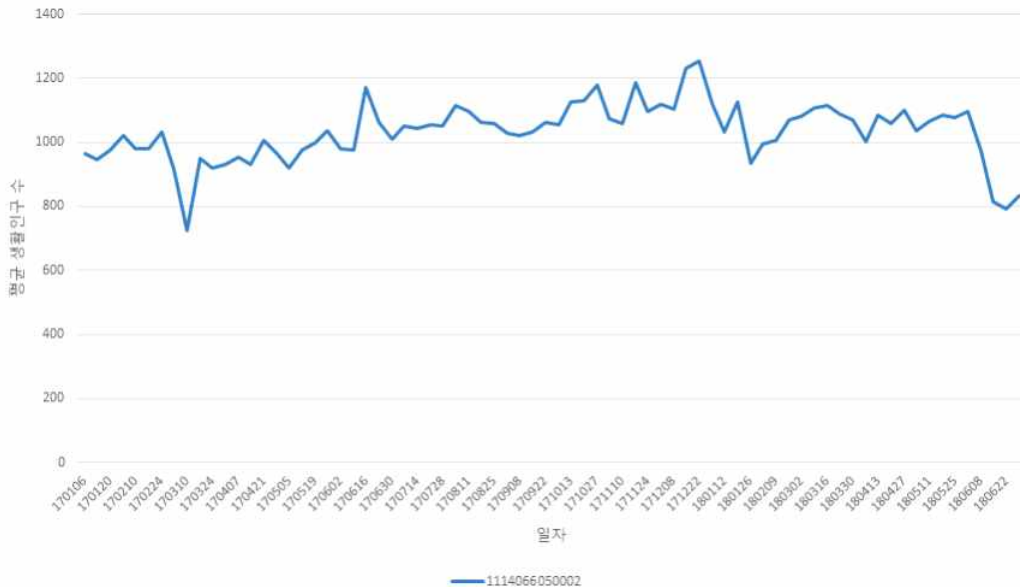


Figure 5. Female Areal Population (aged 25-29) of Jipgyegu 1114066050002 in Friday Nights (17-01)

해당 원데이터를 이용하여 Queen 공간가중행렬 기반으로 국지적 Moran지수를 구했다. Figure 6는 해당 집계구의 국지적 Moran지수의 추이를 보인다. 국지적 Moran지수는 원 생활인구수와 달리 증가하는 추세를 보인다. 이는 주변 집계구보다 사람이 모이는 수준이 더 증가하고 있다는 것을 뜻한다. 대표적으로 앞서 Figure 5에서 봤던 것과 달리 2017년 12월 22일의 경우 원데이터는 월등하게 높게 나타났지만, 국지적 Moran지수의 경우 주변의 집계구와 비교하여 그렇게 크게 몰린 지역이 아님을 알 수 있다. 생활인구가 최소지점이었던 17년 3월 10일 부근에도 국지적 Moran지수는 크게 변하지 않은 것으로 보아, 해당 지역과 인근 지역에는 전반적으로 사람이 적었던 것을 알 수 있다.



Figure 6. LISA for Female Areal Population (aged 25-29) of Jipgyegu 1114066050002 in Friday nights (17-01)

국지적 Moran지수를 평균과 표준편차를 이용하여 표준화하여 누적합 그래프를 그리면 다음과 같이 된다. 9월 이후로 이상치 탐색 구간인 h 를 넘겨 분석 기간 마지막까지 이상치로 기록된다. 앞서 설명한 것과 마찬가지로 홍대입구역 1번 출구 쪽에 이전보다 더 많이 인구들이 주변보다 몰리는 것을 탐지하였다. 어떤 변화가 일어났는지 확인하기 위하여 해당 시점 언론 기사탐색과 지역 실사를 실시하였다. 해당 지역에는 17년도 상반기에 건물이 두 개가 생기면서 사람들이 서서히 몰리기 시작했다 할 수 있다. 한 건물에는 젊은이의 문화 쉼터라는 주제로 상점들이 들어섰다. 특히, 17년도 8월 초에 글로벌 스포츠 브랜드의 농구 컨셉트 매장이 개장한 것이 의류, 농구화의 관심이 많은 젊은 층의 이목을 끈 것으로 예상된다. 이로 인해 해당 시점에 누적합에 변화가 있었던 것으로 추측된다. 또 9월부터는 동일 건물 3층에 홍대 인근에 최초로 주요 브랜드 서점이 새롭게 단장하였다. 날씨가 쌀쌀해지는 것에 힘입어 실내 문화공간에 사람들이 많이 몰렸다고 예상된다. 또 다른 건물에는 유명 호텔이 오픈(18년 1월)하였고, 해당 건물 1층에는 중국과 일본에서 인기 많은 모바일 메신저 앱의 오프라인 캐릭터 스토어가 오픈하면서 한 번 더 누적합이 급격하게 증가하는 것을 확인 할 수 있었다. 이렇게 젊은 층들을 끌 공간들이 생기면서 이러한 결과가 발생하였다고 생각된다. 17년도 하반기에 대비하여 해당 지역의 몰림 정도가 지속적으로 증가하여 다음 해 3월까지 증가 한 것을 확인할 수 있다. 18년도 3월 이후에는 날씨가 풀리면서 실내 문화공간으로써 반등했던 1번 출구 인근의 인구가 홍익대학교 인근과 연남동 근처 등으로 관심이 분산 된 것으로 예상된다.

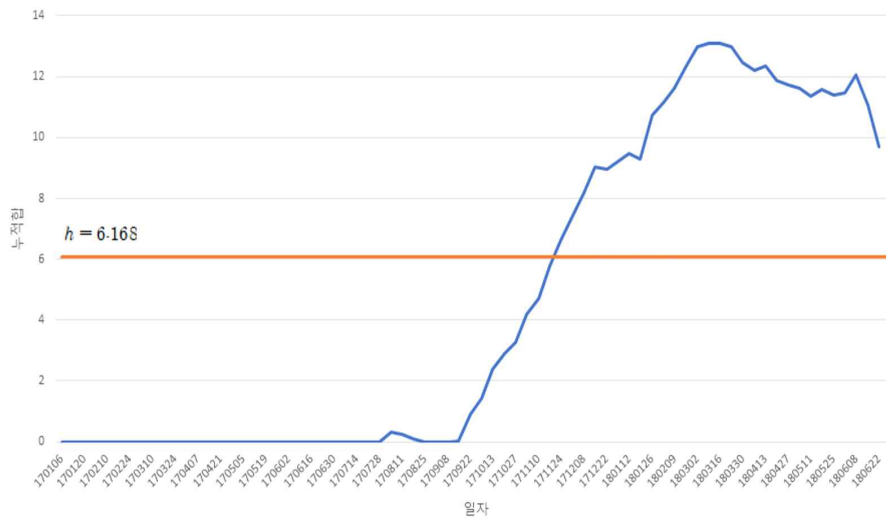


Figure 7. LISA CUSUM chart of Female Areal Population (aged 25-29) in Friday nights (17-01) for Jipgyegu 1114066050002

위와 같이 생활인구에서 큰 변화가 없더라도, 주변 지역에 대비해서 증가하는 것을 확인 할 수 있었다. 더불어 증가하는 국지적 Moran지수의 누적합 관리도를 통해서 미세한 변화도 잘 탐지하는 것을 확인 할 수 있었다. 더불어 증가하는 국지적 Moran지수의 누적합 관리도를 통해서 미세한 변화도 잘 탐지하는 것을 확인하였다. 도출된 이상치는 다양한 함의를 갖고 있었다. 우선 이상치가 분석기간 동안 지속적으로 유지되는지를 확인 할 수 있다. 이를 통해서 이상치가 특정사건으로 인해 단발적으로 발생한 것인지, 아니면 공간적인 특성의 변화로 성격이 변해서 생활인구가 유지되는지 알 수 있다. 이 정보를 이용하여 제한된 행정적 자원을 어떻게 효율적으로 배분 할 것인지에 대한 근거가 될 수 있을 것이다. 또, 이상치가 언제부터 또는 언제 발생했는지를 통해서 어떤 이유가 생활인구를 증가시켰는지 유추할 수 있을 것이다. 해당 정보를 이용하여 연령층 별로 생활인구를 끌어당기는 요소들을 알 수 있을 것이고, 상권을 구성하거나 활성화시키는데 도움이 될 수 있다. 기업의 입장에서는 이 방법론을 통해 오프라인 마케팅과 광고의 효과를 극대화 할 수 있을 것으로 기대된다.

Yamada et al. (2009)는 공간 누적합 관리도를 시각화하기 위하여 GeoSurveillance 프로그램을 디자인하였다. 공간 누적합 관리도는 공간 단위별로 관리도를 따로 그리기 때문에, 시각적으로 한 번에 비교하기가 어렵다. 이 프로그램을 이용하면 집계구별로 구해진 누적합 값을 시점별로 임계값을 넘었는지, 임계값 대비 몇 분위에 위치하는지 지도에 시각화가 가능하다. Figure 8에서 볼 수 있듯이 임계값 h 를 넘는 지역들을 “SIGNAL”로 빨간색으로 표시한다. 이상치 이외에도 각 누적합을 h 에 대비하여 몇 분위에 위치하는지에 따라 4분위로 나눠 25% 미만이면 흰색, 25% 이상 50% 미만이면 하늘색, 50% 이상 75% 미만이면 보라색, 75% 이상 100% 미만이면 파란색, 총 5단계로 지도화한다. 다시 말해 색깔이 진해질수록 누적합이 이상치에 가까워지는 것을 의미하고 눈여겨 볼 필요가 있는 지역임을 뜻한다. Figure 8은 집계구 1114066050002의 17년 5월 25일과 18년 3월 23일의 이상치 지도를 나타낸 것이다. 우선 왼쪽의 17년 5월 25일의 지도를 보면 지속적으로 언급되었던 상상마당 근처와 상수역, 망원역 등에서 이상치가 도출된 것을 볼 수 있다. 또 경의선 숲길 근처의 상권 등에서도 하늘색을 나타내며 누적합 대비 50% 정도에 위치하여 주시해야할 지역임을 나타내었다. 오른쪽 지도는 18년 3월 23일의 누적합 지도를 나타낸다. 왼쪽의 그림과는 달리 홍대입구역 주변과 홍대 예술의 거리에 이상치가 몰려 있다. 마찬가지로 경의선 숲길 주변도 누적합이

어느 정도 증가한 것을 관찰할 수 있다.

원활한 설명을 위하여 두 시점만 추출하였지만, 프로그램 내에서는 분석기간 동안의 결과 지도를 애니메이션으로 확인 할 수 있다. 이를 통하여 시간이 흐를수록 어느 지역에서 이상치 클러스터링이 나타나는지 직관적으로 관찰 할 수 있다. 더 나아가서 앞서서 Figure 7을 통해 확인했던 개별적인 누적합 관리도를 전반적으로 확인할 수 있다.

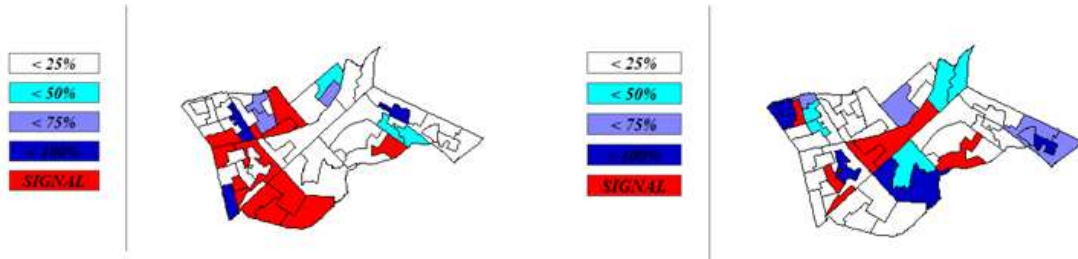


Figure 8. The result of Geosurveillance in May 26th, 2017 and March 23rd, 2018 for Jipgyegu 1114066050002

6. 결론

본 연구는 서울시 공공데이터에서 새롭게 제안하는 생활인구를 이용하여 일자별 이상치를 탐색하고자 하였다. 이전 공중보건 분야에서 자주 사용되었던 공간통계량의 누적합 관리도를 이용하였다. 특히 주변의 지역과 대비하여 측정치가 많은지 적은지 알 수 있는 국지적 모란지수를 활용하였다. 공간 통계량의 누적합 관리도를 통하여 시공간적인 이상치를 확인 할 수 있었다. 각각 집계구에 대해서 누적합 관리도를 구한 후, 발생한 이상치 횟수를 활용하여 어느 구역에 언제 이상치가 발생하였는지, 해당 이상치가 지속적으로 이어지는지 등에 대해 알 수 있었다. 해당 이상치들은 시간이 지나면서 특정 집계구들이 주변 집계구들에 대비하여 더 많이 몰리는지를 과거의 모란지수와 비교하여 증가하였는지를 확인 할 수 있었다. 집계구별 누적합 관리도를 시각화하여 직관적이고 전반적인 파악까지 할 수 있었다. 이는 지속적으로 변하는 공간적 성격을 빨리 탐지하여 행정적 자원을 효율적으로 배분하는데 정량적인 근거가 될 수 있을 것으로 기대된다. 예시로 금요일 저녁(17시-01시)의 마포구 서교동에 존재하는 25세에서 29세의 여성 생활인구를 분석 한 결과 누적합이 갑자기 증가하는 시점들에 해당 집계구에 어떤 일들이 있었는지 살펴 본 결과, 젊은 계층의 이목을 끌만한 상점들이 해당 시점에 생긴 것을 알 수 있었다. 연령대 별로 동일한 프레임워크를 적용한다면 새로 생기는 상점들이 상권에 어떻게 영향을 미치는지 알 수 있을 것으로 예상된다. 이를 활용한다면 해당 지역 상권의 창업 전략이나 기업들의 연령대별 마케팅 방향을 제안 할 수 있을 것이다. 세분화된 타겟 연령층의 인구 패턴을 활용하여 수익이나 광고효과의 극대화를 노릴 수 있다.

지역별, 연령대별 인구 변화 패턴을 빠르게 파악하는 것 이외에도 해당 프레임워크는 다른 분야에서도 다양하게 활용 될 수 있을 것으로 기대된다. 제안된 방법론을 대중교통 사용량 데이터에 직접 적용한다면, 대중교통 이용 집중 시간대에서의 수요와 공급 관점에서 더 많은 정보를 제공 할 수 있을 것이다. 이는 대중교통 노선 및 관련 정책 수립에 도움이 될 수 있다. 대중교통 사용인구의 이동에 알맞게 버스, 지하철 운행시간을 조절 하거나 운행노선을 적절히 조정하여 대중교통 이용에 편의를 제공할 수 있을 것이다. 혹은 유해공기물질(미세먼지, SOx, NOx 등) 집중 시간대에 적용하여, 관련 재난에 대한 경고 및 대처를 미리 할 수 있을 것으로 기대된다.

해당 연구는 위와 같이 여러 기여점이 있지만, 여전히 몇 가지 제한점이 존재한다. 첫 번째로는 공간연구는 작위적인 공간단위의 설정에서 자유로울 수 없다는 점이다. 해당 연구는 보유하고 있는 가장 최소 단위인 집계구로 분석을

진행하였지만, 한 집계구에서 생활인구는 시시각각 변할 수 있다. 따라서 조금 더 정확하고 해석가능한 분석을 하려면, 지점별 생활인구를 이용하여야 할 것이다. 뿐만 아니라 직관적인 이해를 위하여 예시 분석을 서교동에 국한지어 분석했기 때문에 지도의 끝에 위치한 집계구들은 공간단위에 매우 의존적이고 편향된 결과를 나타낼 수 있다. 실질적인 활용을 위해서는 전 집계구를 고려하여 해당 문제를 최소화할 수 있는 것으로 보인다. 두 번째로는, 왜 이상치가 발생했는지 정성적으로 평가해야하는 것이 해당 방법론에 대한 단점이라고 할 수 있다. 정량적으로 특정 집계구들이 이상치로 판별된 것을 갖고 행정적 자원을 배치하기엔 근거가 되지만, 그 주장에 힘이 실리려면 정성적인 근거가 뒷받침 되어야한다. 앞서 선행되었던 연구들 중 소셜미디어 데이터와 토픽모델링을 결합한 것과 같이, 키워드와 토픽 등으로 시공간적 이상치를 설명하는 연구가 차후에 진행될 수 있을 것으로 생각된다. 더 나아가, 본 연구는 데이터가 시간대별 생활인구를 제공함에도 불구하고, 세밀한 변화를 탐지하는 CUSUM 관리도를 택하고 있기 때문에 일별 이상치를 탐색하였다. 주어진 데이터를 잘 활용하려면, 시간대별 이상치를 탐지하기 위한 공간관리도에 관해서 조금 더 많은 연구가 되어야 한다. 현재의 방법론으로는 시간대별로 표준편차가 커서 공간관리도로는 탐지하기가 어렵다. 해당 부분들은 이후의 연구에서 채워질 것으로 기대해본다.

REFERENCES

- Anselin, L. 1995. Local indicators of spatial association—LISA. *Geographical Analysis* 27(2):93-115.
- Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., Ebert, D. S. 2014. Public Behavior Response Analysis in Disaster Events Utilizing Visual Analytics of Microblog Data. *Computers & Graphics* 38:51-60.
- Gim, S. 2013. Discursive Construction of the Seoul Metropolitan Cultural Space: Focused on the Hongdae Area. *The Seoul Institute Policy Research*. pp. 1-89.
- Gundogdu, D., Incel, O. D., Salah, A. A., Lepri, B. 2016. Countrywide Arrhythmia: Emergency Event Detection Using Mobile Phone Data. *EPJ Data Science* 5(1):25.
- Han, M., Yu, S. J. 2019. Prediction of Baltic Dry Index by Applications of Long Short-Term Memory. *The Korean Society for Quality Management* 47(3):497-508.
- Han, S. W., Mei, Y., Tsui, K. L. 2008. A Comparison between SCAN and CUSUM Methods for Detecting Increases in Poisson Rates. School of ISyE, Georgia Institute of Technology, Atlanta, GA. Technical report.
- Jeong, Y. Y., Moon, T. H. 2014. Analysis of Seoul Urban Spatial Structure Using Pedestrian Flow Data - Comparative Study with '2030 Seoul Plan' -. *Journal of the Korean Regional Development Association* 26(3):139-158.
- Kim, H. J., Lee, S. W. 2011. Determinants of 5 Major Crimes in Seoul Metropolitan Area: Application of Mixed GWR Model. *Seoul Studies* 12(4):137-155.
- Kim, M., Kang, S., Kim, S. P., Sohn, H. G. 2016. A Spatial Analysis of Shelter Capacity Using Floating Population. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography* 34(1):1-10.
- Kim, S., Lee, Y. D. 2005. Cusum Control Chart for Monitoring Process Variance. *The Korean Society for Quality Management* 33(3):149-155.
- Kracalik, I., Lukhnova, L., Aikimbayev, A., Pazilov, Y., Temiralyeva, G., Blackburn, J. K. 2011. Incorporating Retrospective Clustering into a Prospective Cusum Methodology for Anthrax: Evaluating the Effects of Disease Expectation. *Spatial and Spatio-temporal Epidemiology* 2(1):11-21.
- Kwon, D. S., Sohn S.Y. 2017. Spatial Analysis of Small Churches in Seoul Using Geographically Weighted Regression. *Journal of the Korean Geographical Society* 52(5):625-643.

- Kwon, H., Hong, S. H., Lee, M. K., Lim, S. 2016. Literature Review on the Statistical Quality Control in Journal of the KSQM for 50 Years. *The Korean Society for Quality Management* 44(1):1–16.
- Lee, K. M., Jung, C. M. 2014. The Effect of Time Period Pedestrian Volume on Store Location – Focused on the Suwon Retail Stores and Restaurants. *Journal of the Architectural Institute of Korea Planning & Design* 30(8):47–55.
- Lee, M. L. 2015. Comparison of Multivariate CUSUM Charts Based on Identification Accuracy for Spatio-temporal Surveillance. *The Korean Society for Quality Management* 43(4):521–532.
- Lee, Y. S., Park, H. S., Lew S. H., Kang J. M. 2014. An Analysis of the Location Factors that Affects the Sales of Campus Commercial District. *Seoul Studies* 15(1):17–34.
- Lloyd, C. 2010. *Spatial Data Analysis: An Introduction for GIS Users*. Oxford University press.
- Lund, R., Seymour, L. 1999. Assessing Temperature Anomalies for a Geographical Region: A Control Chart Approach. *Environmetrics: The Official Journal of the International Environmetrics Society* 10(2):163–177.
- Manogaran, G., Lopez, D. 2018. Spatial Cumulative Sum Algorithm with Big Data Analytics for Climate Change Detection. *Computers & Electrical Engineering* 65:207–221.
- Montgomery, D. C. 2009. *Introduction to Statistical Quality Control*. John Wiley & Sons : New York.
- Oh, J. W., Lim, T. J. 2019. Regional Analysis of Extreme Values by Particulate Matter(PM2.5) Concentration in Seoul, Korea. *The Korean Society for Quality Management* 47(1):47–57.
- Page, E. S. 1954. Continuous Inspection Schemes. *Biometrika* 41(1/2):100–115.
- Qiu, P. 2013. *Introduction to Statistical Process Control*. Chapman and Hall/CRC.
- Rogerson, P. A. 1997. Surveillance Systems for Monitoring the Development of Spatial patterns. *Statistics in Medicine* 16(18):2081–2093.
- Rogerson, P. A. 2005. Spatial Surveillance and Cumulative Sum Methods. *Spatial and Syndromic Surveillance for Public Health*. pp. 95–114.
- Seo, M. K., Yun, W. 2019. Condition Monitoring and Diagnosis of a Hot Strip Roughing Mill Using an Autoencoder. *The Korean Society for Quality Management* 47(1):75–86.
- Sonesson, C. 2007. A CUSUM Framework for Detection of Space-time Disease Clusters Using Scan Statistics. *Statistics in Medicine* 26(26):4770–4789.
- Vidal Rodeiro, C. L., Lawson, A. B. 2006. Monitoring Changes in Spatio Temporal Maps of Disease. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 48(3):463–480.
- Yamada, I., Rogerson, P. A., Lee, G. 2009. GeoSurveillance: A GIS-based System for the Detection and Monitoring of Spatial Clusters. *Journal of Geographical Systems* 11(2):155–173.
- Zhang, H., Zheng, Y., Yu, Y. 2018. Detecting Urban Anomalies Using Multiple Spatio-Temporal Data Sources. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2(1):54.

저자소개

김형준 연세대학교 산업공학과에서 학/석사 학위를 취득하였으며 연구 관심분야는 기술경영 및 공간 빅데이터 마이닝이다. 현재 NICE평가정보에서 고도화된 신용평가모형 개발 중이다.

손소영 연세대학교 산업공학과 교수로 재직하고 있으며 최근 관심분야는 공간 빅데이터 마이닝, 기술경영, 품질공학으로 상세 정보는 <http://isl.yonsei.ac.kr>에서 확인할 수 있다.