



사물인터넷 기반 도시데이터 센서를 활용한 서울시 유동인구의 측정 및 입지특성에 관한 기초연구*

A Study on the Measurement of Floating Population and Locational Characteristics Using IoT-based Urban Data Sensors in Seoul

박진홍** · 강민규***

Park, Jin-Hong · Kang, Min-Gyu

Abstract

This study aims to analyze the relationship between S-DoT (Smart Seoul Data of Things) floating population and locational characteristics of urban sensors in Seoul. First, we defined 'floating population' as "the number of people walking in a certain space and time" by comparing it with various population-related terminologies. We showed that only the 2nd-generation S-DoT data correspond to meaningful floating population data. The data from these urban sensors were analyzed through time-series clustering and ANOVA to determine the features of the floating population and sensor location. The time-series clustering analysis results suggested that S-DoT sensors could be categorized into three groups. The ANOVA results indicated that the level of population concentration is higher at places in which large commercial areas exist. These locations showed a rapid outflow of people after business hours. In contrast, places where the ratio of residential areas are higher showed a lower concentration of floating population with consistent night time volume of people. We suggested that new S-DoT sensor infrastructure should be installed at more suitable locations to estimate the dynamics of the floating population. Additionally, 1st-generation S-DoT sensors should be re-examined in terms of a measurement method. Future studies should investigate the long-term relationship between the floating population and urban spatial characteristics.

주제어 스마트도시, 스마트서울 도시데이터 센서(S-DoT), 유동인구, 시계열 군집분석

Keywords Smart City, Smart Seoul Data of Things(S-DoT), Floating Population, Time Series Clustering

1. 서론

최근 한국을 비롯한 전 세계 주요 국가에서 정보통신기술과 도시 인프라의 융합을 통해 도시 문제를 해결하려는 시도인 스마트 도시 전략이 크게 주목받고 있다(박배균, 2020). Townsend (2019)에 따르면, 스마트도시란 “정보기술이 도시의 인프라나 건축물, 일상용품, 심지어 우리의 몸과 결합하여 사회적, 경제적,

환경적 문제들을 해결해 나가는 장소”를 일컫는다. 철근과 콘크리트, 유리로 건설되어 온 기존의 근대 도시는 보이지 않는 네트워크와 소프트웨어로 촘촘히 재배열된 스마트도시로 급격히 변모 중이다.

이러한 스마트도시 전략을 구현하기 위해서는 사물인터넷, 인공지능, 드론, 3D 프린팅 등 4차 산업혁명의 주요 기술에 대한 이해가 필요하며, 특히 이러한 핵심요소 중 하나는 디지털 시대의

* 이 논문은 2022년도 한국도시계획학회 상반기 학술대회에서 발표한 내용을 수정·보완하여 작성하였음.

** Master's Student, Department of Urban Administration, University of Seoul (First Author: parkjh1020@uos.ac.kr)

*** Assistant Professor, Department of Urban Administration, University of Seoul (Corresponding Author: mgkang23@uos.ac.kr)

원유라 불리는 빅데이터다(손정렬, 2020). 도시가 다양한 빅데이터 - 소셜미디어·교통·통신·GPS 등에서 추출된 로그파일, 위치 정보, 웹 데이터 등 - 가 무수히 생산되고 소멸하는 “데이터 공장”과 같다면 이러한 도시 빅데이터를 측정 및 관리하는 시스템을 갖추는 것 역시 중요한 도시관리 전략이라 할 수 있다(Rabari and Storper, 2015).

이러한 새로운 도시관리 전략의 일환으로 서울시는 스마트도시 생태계를 선도하기 위한 인프라로서 ‘스마트서울 플랫폼 6S¹⁾’를 제안하였다. 이는 서울시 전역에 설치된 정보통신 인프라를 기반으로 3D 가상공간 및 AI 기반의 지능형 행정서비스를 제공하기 위한 스마트도시 전략사업을 의미한다.

이 중 정책의 체감도와 활용 가능성이 높은 사업 중 하나는 사물인터넷(IoT) 기반의 도시센서 설치 및 데이터 구축사업인 S-DoT이다. 이는 스마트서울 도시데이터 센서(Smart Seoul Data of Things, S-DoT)의 줄임말로, 도시 빅데이터를 실시간으로 수집하기 위한 센서 인프라를 서울시 전역에 구축하고, 이를 통해 도시 문제를 해결하는 것을 주요 사업 목표로 삼고 있다. 현재 S-DoT 센서는 약 1,100대가 설치되어 있으며 기온, 습도, 미세먼지 및 유동인구 등 17종의 다양한 도시환경 데이터를 실시간으로 수집하고 있다. S-DoT 센서는 데이터를 실시간으로, 그리고 자동으로 측정한다. 이러한 센서 기반의 데이터 측정 및 수집 방식은 사람의 수고와 주관에 개입되지 않아 경제적, 시간적 비용을 절약할 수 있고 직접 계수(計數)하는 방법보다 정확성이 높다는 장점이 있다.

이 가운데 유동인구는 도시의 다양한 경제, 사회, 문화적 현상을 분석할 때 중요한 지표로 기능한다. 가령, 유동인구는 도시의 활력을 증진시키고, 이것이 다시 상권 활성화로 연결된다는 논리에 따라 상권 분석 및 부동산 가치 평가의 핵심적 요소로 널리 사용되어 왔다(최막중·신선미, 2001; 이명호 외, 2019). 도시계획 및 설계 분야에서도 상업용 건물을 비롯한 도시시설의 수요예측 및 입지분석을 수행하거나 보행 중심의 도시공간구조 및 가로환경을 설계하기 위한 중요한 근거 자료로 활용되었다(최막중·이재우, 1999; 이연수 외, 2013). 또한 공공기관도 같은 맥락에서 행정 서비스 수요를 합리적으로 예측하고 이를 효율적으로 공급·배분하는 의사결정에 유동인구를 활용하면서 자료의 중요성이 증대되었다(변미리·서우석, 2011; 임화진·박성현, 2020). 이처럼 유동인구 데이터는 도시 및 지역계획 분야를 포함한 다양한 영역에서 학술적·정책적 연구가치가 있는 중요한 데이터라 할 수 있다.

현재까지 S-DoT 유동인구 데이터를 활용한 선행연구가 거의 이루어지지 않았고 S-DoT 인프라 구축 및 활용방안에 대한 논의 역시 초기 단계에 있다. 그러므로 본 연구는 데이터의 신뢰도를 높이기 위한 전처리 작업과 그 과정에서 확인된 이슈 및 처리 기법을 정리하는 것을 선행 과제로 다루었다. 또한 유동인구의 시계열적 패턴 분석을 기반으로 센서가 위치한 지역의 입지 특성을

살펴보았으며, 이를 통해 향후 수행될 응용 연구가 참고할 수 있는 활용사례를 마련하고자 하였다.

논문의 구성은 다음과 같다. 먼저 본격적 분석에 앞서, 유동인구의 정의와 의미에 대해 고찰하였다. 이는 기존의 다양한 인구 관련 용어 및 서울시 인구조사자료 간의 비교를 통해 명료하게 정리하였다. 이후 정밀한 전처리 과정을 거쳐 구득한 무결점 데이터를 여러 기준에 따른 시각화 작업, 시계열 군집분석, 분산분석을 통해 검토하였다. 마지막으로 분석된 결과를 토대로 S-DoT 유동인구 센서의 입지특성 및 데이터에 대한 장단점을 고찰하고 정책적 함의를 제시하였다.

II. 유동인구의 정의 및 특성

1. 유동인구의 의미와 정의

분석에 앞서 본 연구에서는 유동인구의 의미를 명확히 정의하기 위해 다양한 인구관련 용어를 비교하는 작업을 수행하였다. 인구와 관련된 용어에는 대표적으로 상주인구(야간인구), 주간인구, 현재인구 그리고 유동인구가 있다(표 1).

먼저 ‘상주인구(resident population)’는 한 지역에 거주하는 인구를 뜻하므로 ‘야간인구(nighttime population)’라고도 불린다. 2015년 이후에는 주민등록인구로 대체되었다. 1990년도부터 상주인구에 주간유입인구와 주간유출인구를 적용하여 ‘주간인구(daytime population)’ 또는 ‘주간활동인구’라는 용어의 통계 자료가 제작되기 시작하였다. 이는 단순히 거주지 등록현황을 반영하였던 상주인구에 통근 및 통학과 같은 인구 유출입 요인을 추가하여 인구자료의 동태적 성격을 제고하기 위한 것으로 볼 수 있다(은기수, 2001).

이 인구자료는 직접 모든 가구를 방문하지 않고 주민등록부나 건축물대장 등의 행정자료를 이용하여 생산하는 통계자료인 등록센서스(register-based census)라고 한다. 상주인구는 100% 행정자료만을 활용하여 전수조사가 가능한 완전(fully) 등록센서

Table 1. Terminologies for population

Terminology	Definition
Resident(nighttime) population	The number of people who reside in a given area and are typically present during the evening and nighttime hours
Daytime (commuter-adjusted) population	The number of people who are present in an area during normal business hours, including workers
De facto population	The number of people who are recorded to the geographical area where they were present at a specified time
Floating population ²⁾	The number of people* who are traveling in a particular location and time

*In a narrow sense, it would mean ‘pedestrian volume(count)’

스이고 주간인구는 행정자료와 함께 20% 표본가구를 대상으로 현장조사를 병행하여 작성하는 부분(partly) 등록센서스이다.

이러한 센서스 데이터는 그간 많은 정책 및 계획의 근거 자료로 활용되었다. 그러나 이 데이터의 공간 단위는 자치구이므로 세밀한 분석이 제한되어 있고, 특히 주간인구 자료는 표본가구에 대한 설문조사를 통해 구득하게 되므로, 결과가 주관적(subjective)이고 정확하지 못하다는 비판에서 벗어날 수 없다. 또한 주간인구의 이동은 단순히 통근·통학과 같은 주요 통행목적 이외에도 여가, 쇼핑, 출장 등 자발적이고 우연한 통행의 비중이 작지 않다(변미리·서우석, 2011). 유연하고 신속한 정책수립 및 평가가 필요한 상황에서 다양한 이동패턴을 고려한 인구통계자료의 생산이 요구되었다.

현재인구는 이러한 인구총조사(census) 자료의 정태적 속성 및 기술적 한계를 보완한다. 현재인구란 용어 그대로 특정 시점에 실제로 어느 곳에 위치하였는지를 기록하기 위한 개념이다. Swanson and Tayman(2011)은 사람들이 어디에 거주하는지보다 어디에 있는지(people where they are found rather than where they usually reside)에 초점을 맞추어야 한다는 것을 강조하였다. UN(2008) 또한 특정 지역에 거주하지 않으나 체류, 방문 및 업무 등의 다양한 도시활동을 경험하는 인구를 정책 대상으로 중요하게 고려하고자 서비스 인구(service population) 개념을 제안하기도 하였다.

이러한 유형의 인구자료는 획기적으로 발전한 정보통신기술로 인하여 집계 및 생산이 과거보다 용이해졌다. 가령, 서울시에서 공개하는 '생활인구'는 대표적인 현재인구 데이터다. 이 자료는 모바일 신호를 통해 서울 인구의 위치 현황을 시간 단위에 맞추어 실시간으로 기록한다. 다시 말해, 인구가 이동함에 따라 달라지는 체류 현황을 시점별로 수치를 기록한 자료이므로 실시간으로 달라지는 인구 현황의 공간적 '분포(distribution)'를 보여준다고 할 수 있다.

그렇다면 유동인구란 무엇인가? 사전적 정의에 따르면 유동인구란 '일정한 기간 동안 한 지역을 오가는 사람의 수'이다. 이는 특정한 시공간적 범위 내에서의 '통행량(volume counts)'을 뜻한다. 유동인구의 경우, 특정한 시간 및 공간적 범위 내에서 사람들이 얼마나 통행하였는지에 관한 횟수를 측정하는 것이 중요하다. 김경태 외(2015)는 유동인구를 단어 그대로 이동하는 인구로 보아 통계청이 제시하는 주간인구와 가깝다고 보았으나, 이는 유동인구가 통행량을 나타내는 지표라는 점을 간과했다는 한계가 있다.

정리하자면, 인구와 관련된 용어는 크게 상주인구, 주간인구, 현재인구와 유동인구가 있다. 과거에는 상주 및 주간인구 등의 센서스 자료가 주로 활용되었으나 측정방식의 관점에서 시공간적 범위에 대한 제약이 많았다. 통신자료 기반의 현재인구 자료는 다양한 주간활동 유형을 반영할 수 있을 뿐만 아니라 실시간

자료를 제공한다는 장점이 있다. 한편, 유동인구는 제한된 시공간적 범위 내의 통행량을 계수하는 것이므로 인구의 지리적 위치와 분포를 파악하기 위한 다른 인구용어와 비교할 때 그 성질 및 특성이 다르다.

유동인구는 '특정한 공간적 범위에서 일정한 시간 동안 발생한 통행량'을 의미하므로 넓은 의미에서는 보행, 자전거, 대중교통, 자동차 등 다양한 교통수단에 따른 이동을 포함할 수 있다. 반면 좁은 의미에서 유동인구는 '보행자 통행량(pedestrian volume)'으로 제한하여 정의된다. 가령, 후술될 서울시 유동인구조사 보고서, 빅데이터 캠퍼스 KT 유동인구 데이터 및 S-DoT 유동인구 데이터는 측정대상을 보행자에 한정하므로, 연구 목적과 자료 특성에 따라 유동인구 데이터의 이동수단 범위를 명확히 하는 것이 필요하다.

2. 서울시 인구자료의 측정방법 및 특징

서울시는 다양한 측정 방식을 통해 생산한 다수의 인구조사자료를 보유하고 있다. 이를 측정하는 방법에는 크게 간접적인 방법과 직접적인 방법이 있다. 전자는 신용카드 매출액이나 교통카드 이용내역을 이용하여 인구이동패턴을 간접적으로 측정하는 것이다(성명준 외, 2010; 김경근·염명배, 2017; 이승민, 2020). 그러나 이 방법은 자료의 시공간적 제약이 많아 유동인구의 동태적 양상을 제대로 파악할 수 없다는 한계가 있다.

반면에 인구수를 직접 계수하거나 통신 신호 및 도시 센서를 활용하여 다양한 유형의 인구 데이터를 실시간으로 측정한 자료도 있다. 그러므로 2장에서는 서울시를 대상으로 한 인구자료의 측정방식과 이에 따른 특징을 분석하였다. 이는 앞서 정의한 유동인구의 의미에 가까운 데이터를 선별하고 이를 바탕으로 심층적인 기술 분석을 준비하기 위함이다.

〈표 2〉는 2장에서 다루고자 하는 서로 다른 4개의 인구 데이터를 정리한 것이다. 이 자료들은 측정 방법에 따라 관찰계수 자료, 통신 자료 및 센서 자료로 구분된다. 요약하면, 이 중 연구 대상인 유동인구에 속하는 것은 '서울시 유동인구조사'와 'S-DoT 유동인구 데이터'인 반면, '서울시 빅데이터 캠퍼스 유동인구 자료'와 '서울 생활인구'는 엄밀한 의미에서 현재인구 개념을 바탕으로 한 다른 성질의 인구자료라고 볼 수 있다.

1) 관찰 계수 자료

관찰 계수는 현장에서 사람이 직접 인원수를 세서 측정하는 (manually counted) 조사 방법이다. 서울을 대상으로 한 2009년과 2012-2015년 동안 발행된 '서울시 유동인구조사'는 대표적인 관찰 계수 자료 중 하나이다.

서울시는 2009년에 10,000개 지점을 대상으로 보행량에 관한 첫 인구조사를 시작하였고 이후 양호한 입지를 중심으로 1,000개

Table 2. Comparing various population datasets published by Seoul

Title	Seoul pedestrian flow survey	Floating population data in Seoul Bigdata Campus	De facto population of Seoul	S-DoT floating population
Measurement method	Manual counting	Automatic & Mobile signal counting		Automatic & Video counting
Terminology	Floating population (Pedestrian volume)	De facto population		Floating population (Pedestrian volume)
Measurement range	Max. 10,000 spots	50×50 m grid 10m interval of road	50×50 m grid Output area	within 150 m buffer
Attributes	- Manually counting in fields - three times per week - Summarizing for 10 min. after researching for 5 min.	- Utilizing 3G signal data(call, SMS) of KT-SKT - Measuring with the location and speed of mobile signals	- Utilizing LTE mobile signal data(KT) and public bigdata of Seoul - Adjusting through multiple regression	- Measuring visitors within a 150 m buffer every 10 min. - Classifying people and vehicles with motion sensors
Feature	- Reinvestigating 1000 spots sampled at first survey locations - Studying attributes of pedestrians through individual interviews with questionnaires	- Only did KT classify pedestrians and non-pedestrians with speed of signals - Including info. such as age, gender and coordinates	- Short measurement intervals making lots of signals and no difference between individuals compared to 3G data - Including info. such as age, gender, admin. boundary and output area	- Analyzing CCTV videos(Edge Computing) - Adding 50 new sensors after 2021.01. - Collecting PM, temperature, humidity on environ. sensors

지점을 표본 추출하여 재조사하였다. 달라진 부동산 및 상권과 같은 사회경제적 여건이나 교통수단의 변화 등의 현황을 보정하기 위해 주기적으로 자료를 재조사함으로써 사업의 연속성을 유지하였다(이향숙 외, 2014).

이 자료는 가장 먼저 배포된 서울시 유동인구 조사자료였기에 이를 활용한 다양한 연구가 진행되었다. 김수현 외(2015)는 소매 상업시설의 매출액과 여러 결정요인 간의 관계를 분석할 때 해당 자료를 인구특성의 보행량 지표로 활용하였다. 같은 맥락에서 보행량과 가로 시설 요소 또는 토지이용혼합 유형 간의 관계에 대한 도시계획 및 설계 관점의 분석도 진행되었다(윤나영·최창규, 2013; 임하나 외, 2016). 특히 변미리·서우석(2011)은 본 연구와 유사하게 이 자료에 대한 심층적인 기술 분석을 수행하였다.

이처럼 관찰 계수를 통한 조사방식은 최근에도 빈번히 활용되고 자동화를 통한 자료 계수 및 구축이 불가능한 지역에서 유용하게 사용되는 조사방식 중 하나이다(최막중·신선미, 2001; 김용훈·양승우, 2017; 유승수 외, 2018). 그러나 사람이 직접 유동인구 수를 집계하는 관찰 계수 방법은 자료 조사와 구축에 방대한 시간과 비용이 소요된다. 서울시의 유동인구 조사사업은 공공이 막대한 예산을 투입하였기에 가능하였으나 그만큼 지속적이고 정기적인 조사를 실시하기 어렵다는 한계가 있다. 그로 인해 데이터의 시공간적 연속성이 부족하여 세밀한 자료 분석이 어렵다는 단점이 따른다(김홍선·김성문, 2016; 임수명 외, 2019).

2) 통신 자료

관찰 계수 자료는 지리적 제약이 없고 특별한 측정 설비가 필요 없다는 장점이 있다. 그러나 사람이 직접 계수한 주관적인(sub-

jective) 데이터이므로 자료의 신뢰도와 조사의 지속성이 낮다는 한계가 있다. 이와 달리 통신자료를 통해 인구수를 추정한다면 이용자 대다수의 자료를 다양한 공간적 범위에서 실시간으로 조사할 수 있다. 또한 관측 계수 방법에 비해 예산을 절감하고 시공간적인 연속성을 확보할 수 있다는 장점이 있다(김경태 외, 2015).

통신 자료는 관찰 계수 자료에 비하여 객관적으로(objectively) 인구를 추정한다는 장점이 있으나 엄밀한 의미에서 유동인구를 측정한 자료라고는 보기 어렵다. 통신 신호에 기반하여 인구 위치 현황을 파악하고 할당된 공간 단위 내의 인구수를 집계하는 것은 현재인구의 개념에 가깝기 때문이다.

이러한 통신 신호 기반의 서울시 인구조사자료에는 '서울시 빅데이터 캠퍼스 유동인구 데이터'와 '서울 생활인구'가 있다. 이 자료들은 통화, 문자, LTE 신호 등을 활용하므로 통신신호에 수반되는 성별과 연령대 등과 같은 인구속성정보를 구득할 수 있다. 이외에 시간 정보는 시간, 일, 월 등으로 제공되고, 공간 정보는 10m 도로, 50m 그리드, 거주지 또는 행정동 단위 등으로 측정된다.

먼저, 서울시 빅데이터 캠퍼스에서 제공하는 306개의 자료 목록 중 유동인구와 관련된 데이터는 총 4개로 그 내용은 <표 3>과 같다. (1)-(3)은 SKT와 KT 통신사에서 50m 간격의 격자를 기준으로 집계한 인구자료다. 이중 시공간적 범위가 가장 세밀한 것은 (3) SKT 일별 유동인구 자료다. 서울시가 유동인구 데이터를 도로구간으로 재집계한 것이므로 값이 평면의 격자가 아니라 선형의 도로링크 ID에 할당된다. 필요에 따라 성별, 연령, 시간대 또는 거주지·행정동 단위로 분류할 수 있다.

Table 3. Datasets of Seoul Bigdata Campus

Dataset	Provider	Period
(1) Seoul 50×50 m monthly SKT floating population	SKT	2016.01 ~ 2016.12
(2) Seoul 50×50 m monthly KT floating population	KT	2016.01 ~ 2018.01
(3) Seoul 50×50 m daily SKT floating population	SKT	2015.08 ~ 2016.07
(4) Seoul 10 m interval of roads floating population	Seoul	2015.01 ~ 2016.12

이때의 유동인구란 통신사별 측정기법에 따라 다소 차이가 있으나 기본적으로 상주인구를 제외한 것이다. 특이사항은 KT는 5km/h의 속도로 특정 기지국에 일정한 분(min.)단위 이상 체류한 신호를 보행인구로 판단하였다는 점이다. 즉, 해당 자료에서는 유동인구의 범위를 전체 인구에서 상주인구와 비보행인구를 제외한 보행인구로 제한한 것이다[유동인구(=보행인구)=전체인구-상주인구-비보행인구]. 이와 달리 SKT는 유동인구를 ‘해당 지역 거주자 및 근로자를 제외한 방문자’로 정의하였다. 최근 3개월간 신호 위치가 주중 주간(9~18시)과 야간의 전체 시간 중 60% 이상이 동일한 위치에 있으면 각각 근무자와 거주자로 판단하여 제외하였다.

KT 데이터의 경우 보행자에 중점을 두었으나, 빅데이터 캠퍼스의 유동인구 데이터는 전술한 바와 같이 ‘현재인구’의 개념에 가깝다. 이러한 한계를 보완하기 위해 서울시는 SKT 유동인구 자료를 비롯한 교통카드, 도보 및 택시 통행량, 2014 서울시 유동인구조사 보고서 및 다양한 집객시설의 정보를 바탕으로 가중치를 산정한 후 (4) 길 단위 유동인구를 추정하여 제공하였으나, 이 역시 특정한 시공간 내의 통행량을 지칭하는 본래의 유동인구 개념이라고 보기는 어렵다.

빅데이터 캠퍼스의 유동인구 데이터는 2018년을 끝으로 더 이상 제공되지 않았다. 이후 생활인구 데이터가 지속적으로 공개되며 현재인구 개념을 적실히 반영한 인구조사자료로 널리 활용되고 있다. 서울시는 자체 보유하고 있는 공공 빅데이터와 협력 통신사로부터 제공받은 민간 통신데이터를 종합가공한 ‘서울 생활인구(de facto population in Seoul)’ 자료를 2017년도부터 제공하고 있다.

서울 생활인구란 “조사시점 현재 서울에 머무르고 있는 ‘현재인구(de facto population)’로 서울에 주소지를 둔 사람은 물론 업무, 관광, 의료, 교육 등 일시적으로 서울을 찾아 행정수요를 유발하는 인구”를 총칭한다(서울시·KT, 2018). 이때, 현재 인구란 ‘조사시점에 실제로 있었던 장소에 귀속시켜 계산한 인구’를 뜻하므로 어떠한 지역에 거주하지 않더라도 다양한 통행 목적으로 특

정 지역에 방문한 사람은 상주인구가 아닌 현재인구로 집계될 수 있다.

다시 말해, 생활인구란 상주인구와 다양한 통행 목적을 지닌 주간활동인구 모두를 포괄하는 인구 개념인 것이다. 이는 상주·주간인구의 한계를 보완하는 개념이며, Swanson and Tayman (2011)가 제시한 현재인구(de facto population)나 UN의 서비스인구(service population)에 부합하는 개념이라 볼 수 있다.

이 데이터는 서울시 열린데이터 광장에서 구득할 수 있고 행정동, 집계구, 성별, 연령, 시간대 기준으로 분류된다. 통화, 문자 기반의 3G 자료와 달리 LTE신호 기반의 통신 자료를 활용하여 측정 주기가 짧고 신호 발생량이 많아 개인 간 차이가 균일하다는 특징이 있다(서울시·KT, 2018).

이러한 생활인구 데이터 역시 보행량을 뜻하는 유동인구라고는 볼 수 없다. 다만, 빅데이터 캠퍼스 유동인구 데이터에 비하여 최신 자료가 주기적으로 제공되고 품질과 정확성도 제고되었다는 장점이 있다. 게다가 상주인구가 포함되어 있으므로 서울 전역의 인구 패턴을 파악하기에 적합한 자료라고 볼 수 있다. 이러한 자료를 활용한다면 광범위한 공간적 범위에서 발생하는 인구의 사회경제적 현상을 분석하거나 통신 신호에 포함된 개인적 특성을 활용한 응용연구를 다양하게 수행할 수 있다.

예를 들어, 김미경 외(2016)는 격자 기반의 통신사 유동인구 자료로 강남 3구 내 대피소의 수용 능력을 평가하였다. 이를 통해 재난 상황이 발생할 경우, 수많은 인파가 신속히 대피할 수 있는지와 관련된 시간거리 접근성 및 추정된 유동인구로 최대 수용인원이 적정인지 판단하는 기준 등을 제시할 수 있다. 박예림·강영욱(2019)은 도보관광코스에 할당된 개별 코스의 장소, 시간, 성별과 연령 등을 활용하여 관광행태와 이용자 특성 간의 관계를 분석하였다.

그리고 서울 전역 또는 일부 장소에 대한 인구패턴과 도시환경 및 지역특성 간의 관계도 분석할 수 있다. 오다원(2019)은 상권 내 토지용도가 인구 패턴에 어떤 영향을 미치는지 분석하였다. 정재훈·남진(2019)과 윤메술(2020)은 생활인구 자료로 군집분석을 실시하여 서울 전역을 4-5개의 권역으로 분류하고 각 지역의 인구 동태적 특성을 밝혔다. 한편, 생활인구 자료는 통신 데이터에 기반하므로 내국인과 외국인의 신원을 구별할 수 있다. 이를 활용하여 외국인만을 대상으로 이들의 이동에 대한 지역적 또는 시공간적 패턴을 파악하기 위한 연구도 진행되었다(허정원·장주영, 2020; 김형준 외, 2021). 박성경(2018)은 10m 도로구간별 추정 유동인구 데이터를 통해 기존의 인구학적 조사가 미흡하였던 토지용도별 특성 및 이동패턴 연구를 보완하였다.

3) 센서 자료

서울시는 2019년도 말부터 S-DoT 센서를 설치하여 과학행정을 추진하기 위한 근거자료로서의 도시데이터를 수집하는 센서

인프라 구축사업을 시작하였다. 이를 통해 미세먼지, 온도, 습도, 조도, 소음, 진동, 자외선, 풍향, 풍속 등의 환경정보와 유동인구 데이터를 측정하고 있다(표 4). 환경정보는 2분, 유동인구는 10분 주기로 측정되며 서울 열린데이터 광장에서 각각 60분, 10분 간격으로 적재되고 있다.

S-DoT 유동인구 데이터는 2020년 5월부터 서울 열린데이터 광장에서 공개되기 시작하였다. 유동인구의 경우, 2020년 5월 4일 이후부터 매주 10분 주기의 자료가 지속적으로 업로드되고 있다. 해당 데이터의 예시는 <표 5>와 같다.

<그림 1>에서 알 수 있듯이 약 1,100여개의 모든 S-DoT 센서 중 유동인구 센서는 총 100개다. 이 중 2개는 과천시 소재의 서울대공원에 입지한 것으로 확인되었는데 해당 공원이 서울시에서

관할 및 운영하기 때문인 것으로 보인다.

유동인구는 2992~4050의 시리얼 번호를 가진 총 100개의 특수 센서에서만 측정된다. 이 중 4001~4050번대 센서의 데이터는 2021년 1월 이후에 추가되기 시작하였다. 이 시점을 전후로 100개의 유동인구 센서를 1차 센서(2992~3041) 50개와 2차 센서(4001~4050) 50개로 구분할 수 있다. 이 센서들은 설치 지점에서 반경 150m 이내의 인구수를 측정한다.

서울연구원(2020)의 보고서와 센서 위치정보를 참고하면, 1차 센서는 상업공간, 특히 전통시장을 중심으로 설치된 것으로 파악된다. 2차 센서에서는 일관적인 입지 특성을 찾기 어려웠으나 1차 센서가 위치하지 못한 나머지 지역에 고루 분포하되, 다양한 상권, 행정구역과 공공기관의 수요를 반영하여 신설된 것으로 추정된다.

주목할 점은 1차와 2차 센서의 데이터 측정 방식이 다르다는 것이다(서울시, 2022). 1차는 공공 WiFi 기반으로 수신된 통신 신호의 연결 횟수를 측정하고 2차 센서는 CCTV 영상분석기법을 활용한다. 2차 센서의 경우, 각 센서에 모션 감지 센서(GMS-100)가 설치되어 차량과 사람의 보행량과 이동을 구분하여 감지할 수 있고 데이터는 수집 지점에서 컴퓨팅을 수행하는 분산 시스템인 Edge computing으로 처리된다(서울시, 2021). 이 중 카메라로 보행자의 통행 여부를 확인하여 계수하는 2차 센서의 데이터가 보행량을 뜻하는 협의의 유동인구 의미에 부합한다고 볼 수 있다.

현재까지의 S-DoT 데이터를 활용한 선행연구는 주로 기온, 소음 및 미세먼지와 같은 환경변수를 대상으로 한다. 이 연구들은 데이터에 대한 기술 분석과 함께 인근 도시환경 특성 간의 관계를 분석하였다(박진철 외, 2020; 서울연구원, 2020; 박병훈 외, 2021). 박해경(2021)은 S-DoT과 AWS의 기온 데이터와 공간 해상도에 대한 비교 분석을 진행하였다. 이와 비교한다면, 본 연구는 S-DoT 유동인구 데이터에 대한 최초 분석이자 기존의 다양한 인구 자료 간 비교를 통해 유동인구의 의미 및 해당 데이터의 특성을 명료화하였다는 의의가 있다.

Table 4. The types and features of S-DoT sensors

Data	Type	Measurement object	Interval
Natural environ.	Base	PM _{2.5/10} , Temperature, Relative humidity, Illuminance, UV, Noise, Vibration	2 min.
	Special	Globe temperature Wind direction/speed	
Floating pop.	Special	The number of visitors (within a 150 m buffer)	10 min.

Table 5. The example of S-DoT population data

Institution	Model	Serial	Visitors	Registration date
Seoul	SDOT001	4027	8	2021-07-08 09:01:00 AM
Seoul	SDOT001	4027	14	2021-07-08 09:11:00 AM
Seoul	SDOT001	4027	27	2021-07-08 09:21:00 AM
...
Seoul	SDOT001	4041	5	2021-07-08 11:04:00 PM

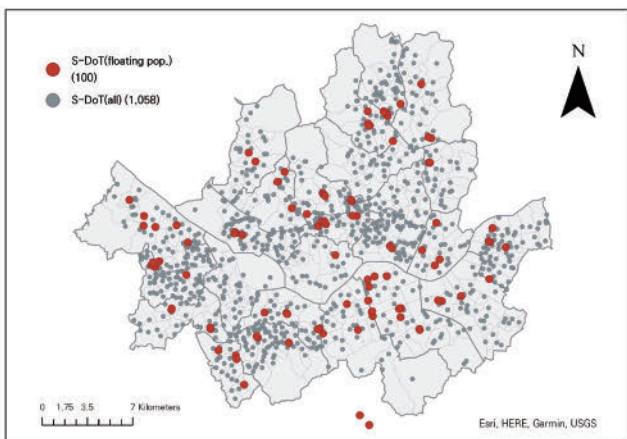


Figure 1. The location of S-DoT sensors

III. 데이터 및 분석방법

1. 연구의 범위

본 연구의 시간적 범위는 주중과 주말을 대표하는 데이터로서 각각 2021년 7월 8일 목요일과 2021년 7월 10일 토요일을 선정하였다. 두 일자만을 분석한 이유는 하루 내 24시간 데이터가 모두 시간당 6회로 온전히 측정된 날짜를 찾는 동시에 센서 외부의 다양한 변수를 고려하였기 때문이다. 예를 들어, 2차 센서 자료 업로드되기 시작한 2021년도 자료 중 코로나19 유행기간을 제외하였고 기온, 강수 및 미세먼지 등 다양한 날씨 요소를 고려하여 야외 활동에 제약이 비교적 적은 적합한 분석 일자를 선정하였다.

Table 6. The scope of study and sensor information

Scope of study	Weekday	2021.07.08 Thur.		
	Weekend	2021.07.10 Sat.		
		Seoul		
Subject	97 S-DoT floating population sensors			
	Type	Serial	Period	Note
	1st	2992-3041	2020.05~	Located mainly in traditional market
	2nd	4001-4050	2021.01~	Excepting 4020, 4024, 4025 sensors

연구의 공간적 범위가 서울특별시이므로 2차 센서 중 과천시 소재의 서울대공원 센서 2개(4024, 4025)를 제외하였다. 또한 선정된 분석 일자 내 특정 센서(4020)의 데이터가 공통으로 부재하여 시각화 결과물에서 자동으로 제외되었다. 따라서 연구 대상은 시공간적 범위에 따라 3개의 센서가 제외된 나머지 97개의 센서이며 연구 범위는 <표 6>과 같다.

2. 데이터 전처리(pre-processing)

S-DoT 도시센서 데이터와 같은 실시간 빅데이터는 이용자가 정해진 측정주기로 끊임없이 생산되는 방대한 양의 원자료를 신속히 구득할 수 있다는 장점이 있다. 그러나 이는 기계적 결함, 기상요건 등과 같은 센서의 다양한 내외부적 이유로 인해 결측치 및 이상치를 수반한다는 한계를 수반한다. 이러한 빅데이터를 활용하기 위해 연구자는 누락된 데이터를 제거 및 대체(imputation) 하는 전처리 과정을 거쳐야 한다(Prakash et al., 2018; Dalad and Dahiya, 2020). S-DoT 유동인구 데이터 또한 무결점 기반 자료를 통한 분석을 위해 전처리 및 정제작업을 수행하였으며 그 과정을 정리 및 공유함으로써 데이터 이해도를 증진하는 데에 기여하고자 하였다.

첫 번째 단계는 분석에 적합한 데이터를 취합하여 기초 정제작업을 진행하는 것이다. 1차와 2차 센서의 데이터가 업로드된 시점이 다르므로 늦게 설치된 2차 센서에 기준을 맞추어 2021년 이후의 개별 원시자료를 하나의 데이터셋으로 병합하는 과정을 진행하였다. 다음으로 중요한 변수열인 시리얼 번호, 방문자수와 등록일자를 추출하고 각각을 숫자(numeric) 또는 날짜(date-time) 등의 데이터형으로 변환하였다. 이때 날짜열에서 R의 lubridate 함수를 통해 월, 일, 시간, 분 등의 시간 정보를 추출하여 별도의 열로 생성하였다. 또한 약 130,000개의 중복되는 행 및 결측치를 추가로 제거하였다.

두 번째 단계는 데이터 가공이다. 유동인구 데이터는 10분 주기 자료이므로 시간당 6회가 측정되어야 온전하다고 볼 수 있다. 앞서 별도로 추출한 월, 일, 시간 등의 날짜열을 통해 특정 월(M)-일(D)-시(H)로 정렬하여 집단별 관측치의 개수를 집계하였

다. 이를 통해 각각의 집단별로 관측 횟수를 요약할 수 있다. 그 결과, 관측 횟수는 1번에서 61번까지 다양했으며 이 중 시간당 6회가 온전히 측정된 데이터는 전체 중 약 74%로 가장 많은 비중을 차지하였다. 많게는 시간당 61번이 측정된 경우도 있었는데, 이는 0분에서 59분까지 1분 간격으로 측정되고 13분이 두 번 측정되어 나타난 결과로 밝혀졌다. 본 연구에서는 S-DoT 유동인구 데이터에 대한 기술 분석에 집중하여 시간당 6회가 온전히 측정된 시간대의 자료만을 활용하였다.

세 번째 단계는 데이터셋의 구조를 변환하는 작업이다. 데이터는 크게 세로형(long-form) 또는 가로형(wide-form) 구조로 저장된다. 세로형 구조는 이름 그대로 데이터 형태가 세로로 길게 뻗어 있는 형태를 취하는 것으로, 손쉽게 데이터를 가공 및 저장할 수 있다는 장점이 있다. 주로 센서별 또는 시간대별 데이터를 전체적으로 개관하거나 이 값들을 집계한 결과를 시각화할 때 사용한다. 센서별로 데이터를 개관하는 작업에서는 이상 센서를 제외하지 않고 모든 센서를 분석 대상으로 선정하였다. 그러나 시간별 데이터가 필요한 나머지 시각화 작업에서는 각 관측치를 집계하는 과정이 필요하여 이상 센서로 판단한 센서(3009, 4019)를 제외하였다.

가로형 구조는 시계열 분석과 같이 여러 변수들을 특정 기준에 따라 비교할 때 사용되는 데이터 형태다. 본 연구에서는 기존의 세로형 데이터 중 주중 자료를 가로형 구조로 변환한 후 이를 시계열 군집분석을 진행할 때 활용하였다. 해당 데이터는 데이터의 분포와 주중 및 주말 비교 그래프를 시각화하는 작업에 활용된 세로형 구조의 시간별 데이터를 피벗(pivot) 기능으로 전환해주는 작업을 통해 준비되었다. 이 과정에서 군집화 분석에 포함될 필요가 없는 센서는 제외하였다. 이와 같이 전처리된 데이터셋의 구조는 분석기법과 밀접한 연관이 있으며 종합적인 데이터 처리 과정을 <그림 2>와 같이 정리하였다.

3. 분석 기법

1) 시계열 군집분석(Time-Series Clustering)

시계열 군집분석은 시계열 자료를 대상으로 군집화를 수행하는 분석기법이다. 강영민 외(2021)에 따르면, 군집분석은 비지도 학습(unsupervised learning)의 대표적 사례로, 지도 학습과 달리 정답지 역할을 수행하는 레이블(label) 없이 컴퓨터 스스로 학습 알고리즘을 입력하여 특정한 데이터 구조를 발견하는 분석방법에 속한다.

시계열 군집분석 역시 이러한 비지도학습에 해당하며, 시간의 변화에 따라 관측된 시계열 자료의 순간적인 변화를 포착하는 작업에 특화된 방법론이다. 이는 여러 시점의 방대한 시계열 자료에 대하여 패턴을 분석하거나, 개별 군집의 표준형(prototype)에 대한 핵심 정보를 요약할 때 주로 사용된다(D'Urso et al.,

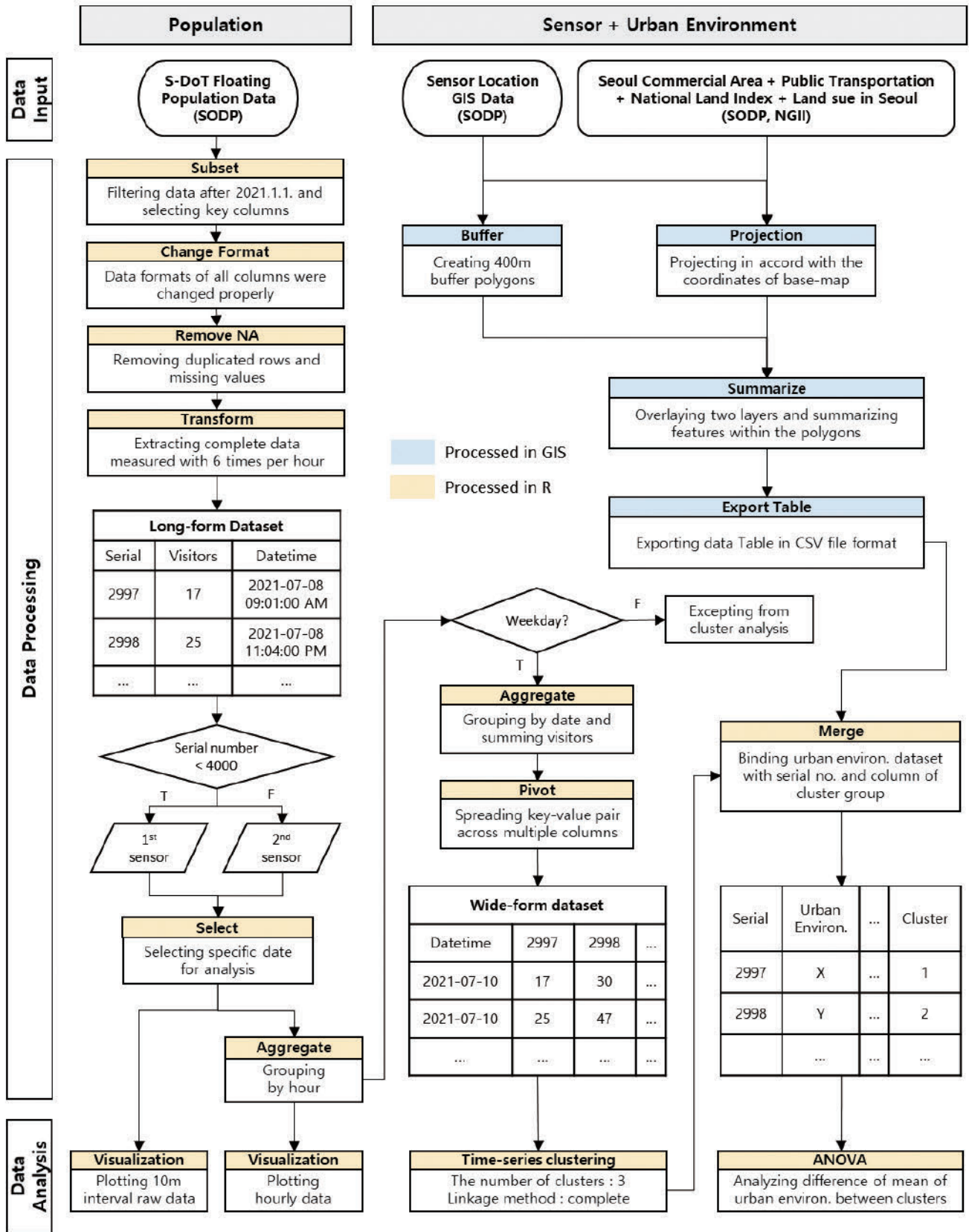


Figure 2. The study design containing procedures of data processing and analysis techniques

2018). 이는 특히 여러 시계열 개체 간의 기하학적 형상(geometric profiles)의 유사성을 발견할 때 유용하게 활용될 수 있다. 다만, 분석과정에서 데이터의 양과 시간적 범위가 지나치게 커지

않도록 주의가 필요하다(Maharaj et al., 2019).

이때, 동적시간왜곡(Dynamic Time Warping, DTW) 알고리즘이 시계열 개체 간의 유사성을 판단하는 기준으로 빈번하게

활용된다. 이 알고리즘으로 변형된 파형을 감지하여 누적 거리를 최소화하는 비틀림(warping) 경로를 찾아낼 수 있다. 이때 warping 경로 $W(= w_1, w_2, \dots, w_k, \dots, w_k)$ 는 다음과 같은 조건을 만족해야 한다. 즉, (1) 서로 다른 시계열 개체 X, Y를 매핑하는 경로 W가 시간의 흐름에 따라 단절없이 이어져 있어야 하고, (2) 정해진 step size만큼만 이동해야 하며, (3) 시계열이 음의 방향으로 이동하지 않아야 한다는 것이다(나중화, 2020).

시계열 군집분석은 유사한 데이터끼리 병합해 나가는 응집형 계층적 군집(agglomerative hierarchical clustering) 알고리즘을 사용하므로 분할적 군집화(partitional clustering) 알고리즘과 같이 사전에 군집 수를 지정하지 않아도 된다. 그러므로 연구자는 덴드로그램(dendrogram)을 통해 적절한 군집의 수(k)와 군집 연결법을 선택하게 된다(박노진·고한석, 2020).

2) 분산 분석

분산 분석(Analysis of Variance, ANOVA)은 그룹 간(between-group) 분산과 그룹 내(within-group) 분산을 비교하여 집단 간에 유의미한 평균 차이가 있는지 확인할 수 있는 기초적 통계분석 기법이다. 하나 또는 둘 이상의 독립변수와 그에 상응하는 종속변수 간의 관계를 분석하고, 요인(factor)이자 명목변수인 독립변수는 상호 배타적으로 나뉘며 종속변수는 등간 척도 또는 명목척도 등으로 측정된다(이희연·노승철, 2020).

분산분석 수행을 위한 본 연구의 독립변수는 크게 두 종류다(표 7). 첫 번째는 시리얼 번호에 따른 센서구분(1, 2차)이다. 이를 통해 사업 연도에 따른 센서의 입지특성을 조사하고자 하였으며, 모든 센서를 대상으로 분산 분석을 수행하였다. 두 번째는 시계열 군집분석을 통해 산출된 군집 범주다. 이는 센서 근방 도시환경 변수의 평균 차이에 따라 유동인구 수에 차이가 나타나는지를 분석하기 위함이다. 이때 분석 대상은 엄밀한 의미의 유동인구 데이터라 볼 수 있는 2차 센서만으로 제한하였다.

분산분석에 사용된 종속변수는 4개 범주에 해당하는 18개의 도시환경변수이다. 서울열린데이터광장과 국토정보플랫폼에서 제공하는 교통 인프라, 서울 상권, 국토지표 및 용도지역의 공간정보 데이터를 구축하여, 각 지리정보 객체의 속성에 따라 개수, 길이 또는 면적의 합계 또는 평균으로 집계하였다. 이때 각 변수는 보행권으로 설정한 센서 400m 반경 이내의 데이터만으로 구성되었다(표 8).

Table 7. Criteria of classification for ANOVA

Independent variables	Cluster's members	Values
Types of sensors	All sensors	1, 2
Time series clusters	2 nd sensors	1-3

Table 8. Urban environmental variables for ANOVA

Category	Dependent variable	Content(unit)*	Source
Traffic infra.	Sidewalk	Sum of length(km)	NGII
	Bus station	Sum of count	
Seoul trade area ³⁾	Traditional market	Sum of area(acre)	Seoul Open Data Plaza (SODP)
	Golmok trade area		
	Baldal trade area		
National index ⁴⁾	Floor area ratio(FAR)	Mean of ratio(%)	National Geographic Information Institute (NGII)
	Gross floor area(GFA)	Mean of area(m ²)	
	Building compactness	Mean of ratio(%)	
	Building complexity	Mean of count	
Land use	Class I exclusive	Sum of area(acre)	Seoul Open Data Plaza (SODP)
	Class I general		
	Class II general (over 7th floor)		
	Class II general (7th floor or less)		
	Class III general		
Cml. area	Quasi		
	Central		
	General		

*All contents were measured in 400 m buffers for each sensor

IV. 분석 결과

1. 데이터 전처리 결과 및 특징

〈그림 3〉의 A-D 패널은 센서별 10분 주기 원시자료를 토대로 S-DoT 유동인구 데이터의 품질과 특징을 분석한 그래프다. 4개의 그래프는 2개의 센서 종류와 2개의 분석 일자로 분류하여 각각 시각화하였다. A는 1차 센서의 주중 그래프, B는 1차 센서의 주말 그래프이고, C는 2차 센서의 주중 그래프, D는 2차 센서의 주말 그래프다.

1차 센서 그래프인 A, B의 인구수를 의미하는 y축 범위가 2차 센서 그래프인 C, D보다 전반적으로 큰 것을 알 수 있다. 또한 A, B가 C, D에 비하여 시계열 개체가 밀집한 구간의 수치도 높다는 것을 확인할 수 있다.

2. 데이터 분포 파악

〈그림 3〉의 E-H 패널은 센서 구분 없이 모든 관측치의 분포를 나타낸 그래프다. 통행량을 뜻하는 유동인구의 정의에 부합하기

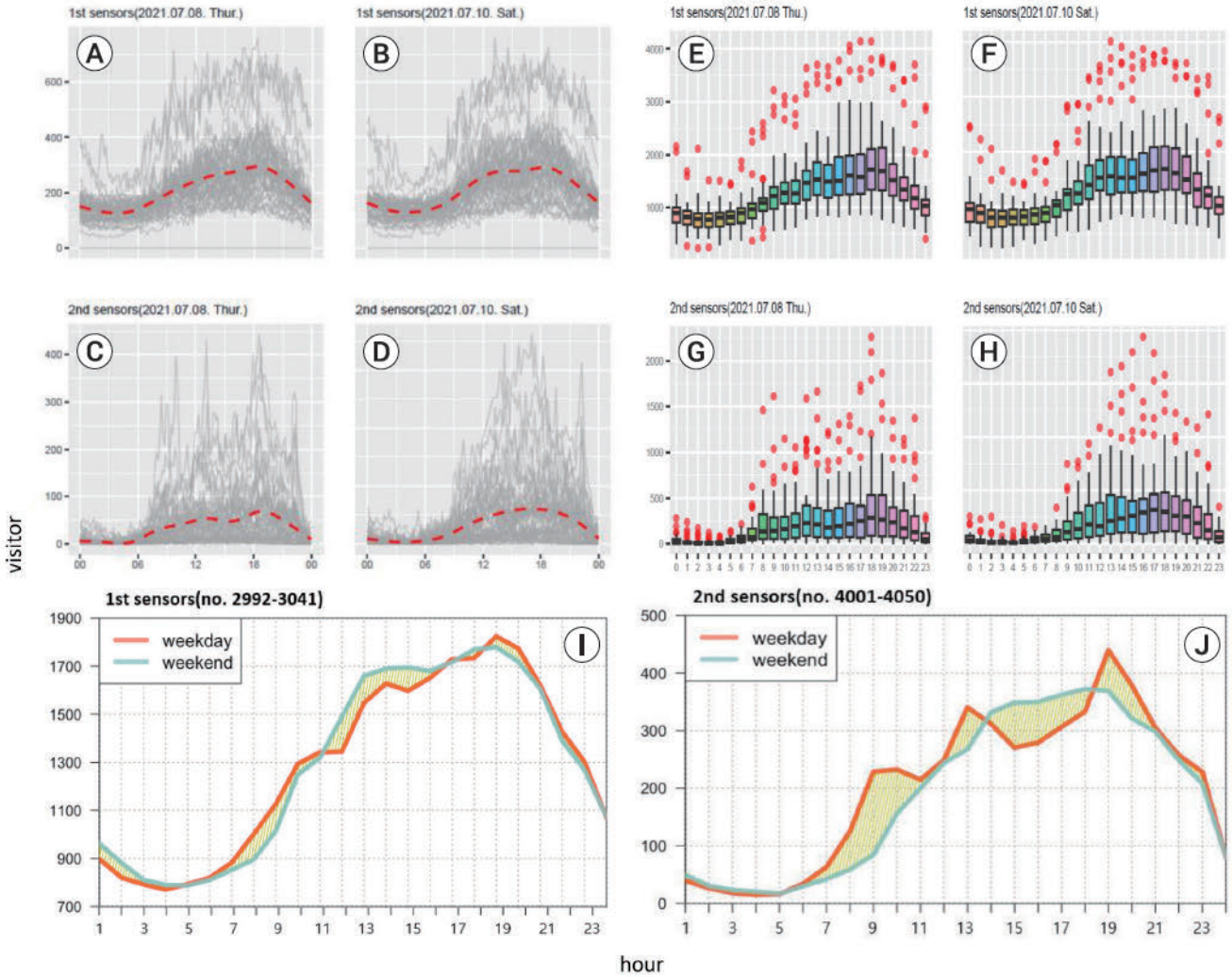


Figure 3. Basic visualization of S-DoT floating population sensors: (1) [A-D] Plotting 10 min-interval raw data for each sensor(grouped by dates and sensor types) (2) [E-H] Distributions of hourly data based on dates and sensor types (3) [I-J] Graphs comparing between weekday and weekend datasets

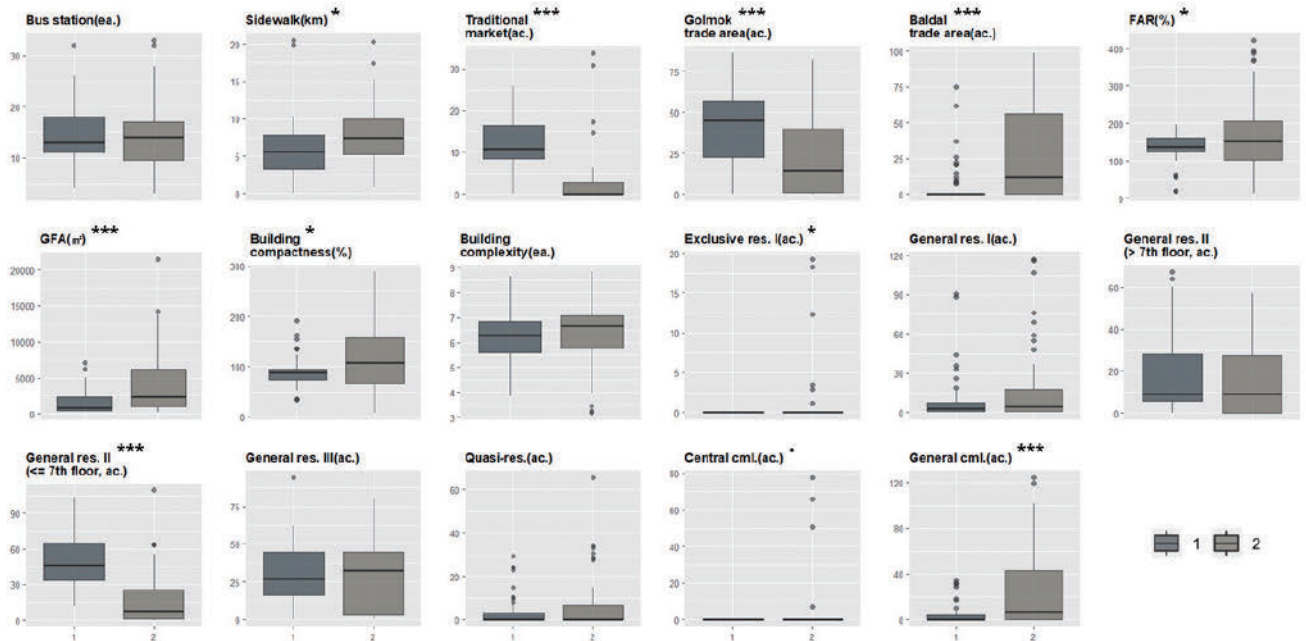


Figure 4. Boxplots of urban environ. between categories of sensors (Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1)

위하여 10분 주기의 원시자료를 시간별로 합산한 데이터를 사용하였다. 시간당 6회 측정된 여러 관측치에 대한 집계이므로 A-D 패널의 그래프보다 y축 범위의 최댓값이 높음을 알 수 있다. 이때 모든 관측치는 한 시간 동안 센서의 반경 150m 이내를 통행한 인구수를 의미한다.

E-H 그래프 또한 동일하게 센서 종류와 분석 일자를 기준으로 데이터를 분류하였다. 이는 상자 그림(Box plot) 그래프이므로 A-D에 비하여 개별 관측치가 밀집한 구간인 사분위수 범위와 이상치를 손쉽게 확인할 수 있다. 이를 통해 1차 센서 그래프인 E-F가 2차 센서 그래프인 G-H보다 관측치가 월등히 높은 것으로 분석되었다.

3. 주중 및 주말 데이터 비교

〈그림 3〉의 I와 J 패널은 각각 1차와 2차 센서 데이터로 크게 분류하되, 주중 및 주말 그래프를 중첩시켜 두 일자 간 유동인구수를 비교한 결과다. 이를 위해 앞서 E-H 패널에서 사용된 시간별 합산 데이터를 센서 구분 없이 평균한 데이터를 사용하였다. 그러므로 개별적인 센서를 구분하지 않고 1차와 2차 센서 군의 주중 및 주말 그래프의 평균적인 차이를 파악하고자 하였다.

I와 J는 E-H의 모든 개별 관측치의 시간별 평균치들의 꺾은 선 그래프라고 볼 수 있다. 다시 말해, I는 E와 F의 시간별 평균 그래프를 비교한 것이고 J는 G와 H의 시간별 평균 그래프다.

I에서 주중과 주말 그래프가 교차하는 지점은 크게 5시, 11시, 17-18시다. 1차 센서의 경우, 5시 이후부터 주중 그래프가 주말 그래프를 상회하기 시작하여 11시에 역전된 후, 17-18시 구간을 지나 야간 시간대 동안 주중 데이터가 지속적으로 높은 것을 확인할 수 있다.

2차 센서 데이터인 J도 마찬가지로 5시 이전에는 주중 및 주말 데이터가 근소한 차이를 유지하다가 5시 이후부터 14시까지 주중 데이터가 상회한다. 이후 14-18시 동안 주중 데이터가 상회하다가 그 이후부터 야간까지 주중 그래프의 값이 더 높은 것을 알 수 있다.

4. 센서 종류에 따른 입지특성 분석

S-DoT 유동인구 센서의 입지특성을 사업 연도에 따른 구분을 기준으로 하여 분산 분석을 수행하였다. 그 결과로 보도, 전통시장, 골목상권, 발달상권, 용적률, 연면적, 토지이용 압축도, 제1종 전용주거지역, 제2종 일반주거지역(7층 이하), 중심상업지역과 일반상업지역 변수가 최대 10%의 유의수준 이내에서 센서 종류 간 평균의 차이가 통계적 유의성을 확보한 것으로 분석되었다(그림 4).

〈그림 4〉의 상자 그림(Box plot)을 통해 1, 2차 센서의 전반적

인 변수별 분포를 확인할 수 있다. 1차 센서는 전통시장, 골목상권과 제2차 일반주거지역(7층 이하)의 면적 비중이 높다. 그리고 2차 센서는 보도의 길이, 발달상권·연면적·제1종 전용주거지역·중심상업지역·일반상업지역의 면적, 그리고 용적률·토지이용 압축도의 비율이 높은 것으로 나타났다.

5. 시계열 군집분석 및 군집별 입지특성 분석

시계열 군집분석에는 2장에서 논의한 바와 같이 유동인구 자료에 적합하다고 판단된 2차 센서 데이터만을 활용하였다. 그 결과, 총 연구 대상인 센서는 서울대공원 센서 2개와 분석 기간 내 결측 센서(4020번)를 제외한 나머지 총 47개로 확정하였다.

최종 군집 결과(그림 5)는 최장연결법과 3개의 군집 수로 결정되었다. 먼저, 덴드로그램과 시계열 그래프를 통해 각 군집별 중심선(점선)⁹⁾의 대표성과 군집 내 다른 개체 간의 응집성을 육안으로 확인하였다. 그 결과, 최장연결법(complete linkage method)이 다른 연결법보다 개체 간 군집성이 높은 것으로 나타나 이를 최종 군집연결법으로 선정하였다.

또한 동일한 군집연결법을 사용한 분석결과 중 객관적 기준을 통해 적절한 군집 수를 선정하고자 하였다. Sard'a-Espinosa (2022)에 따르면, 시계열 개체 간의 모든 거리를 연산하는 경우에는 Silhouette, Dunn, 그리고 COP 지표가 적합한 것으로 알려져 있다. 이때 Silhouette와 Dunn 지표는 최대화되고, COP 지표는 최소화될 때 분석모형의 성능이 우수하다고 볼 수 있다.

이러한 지표들을 활용하여 군집 평가를 수행한 결과는 〈표 9〉와 같다. 전반적으로 군집 수가 적어질수록 모든 지표의 값이 증가하는 것을 확인하였다. 군집 수가 3개일 때 Silhouette와 Dunn 지표의 값이 가장 높은 반면에 COP 지표는 군집 수가 6일 때 가장 낮았다. 따라서 군집 수가 3일 때 3개 중 2개의 지표가 우수한 것으로 분석되었다. 따라서 이때를 시계열 군집분석의 성능이 가장 높은 것으로 판단하여 최종 군집 수(k)를 3으로 결정하였다.

시계열 군집의 그래프 패턴 및 공간적 입지에 대한 분석결과는 각각 〈그림 5〉, 〈그림 7〉과 같다. 각 군집별로 유동인구가 집중되는 시간대와 그 통행량이 다른 것으로 나타났다. 모든 군집이 공통적으로 출근 시간대인 8시경부터 인구이동이 시작되었을 뿐만

Table 9. Clustering evaluation with validity indices

K	Linkage	Silhouette	Dunn	COP
3		0.5015249	0.2539501	0.1908627
4		0.4601104	0.1331500	0.1182132
5	Complete	0.4515431	0.1566426	0.1115950
6		0.4468735	0.1813974	0.0988266
7		0.4399819	0.1977890	0.0928807

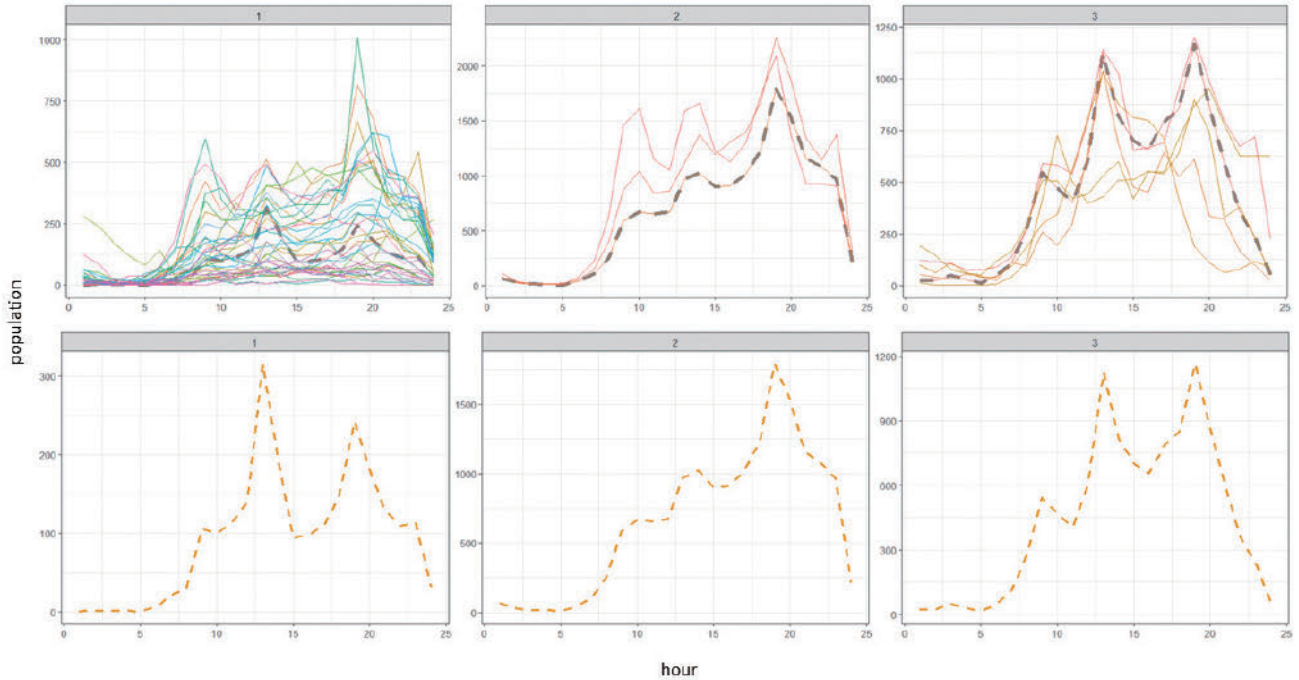


Figure 5. Results of time series clustering with 2nd S-DoT population sensors (k=3, linkage method="complete")

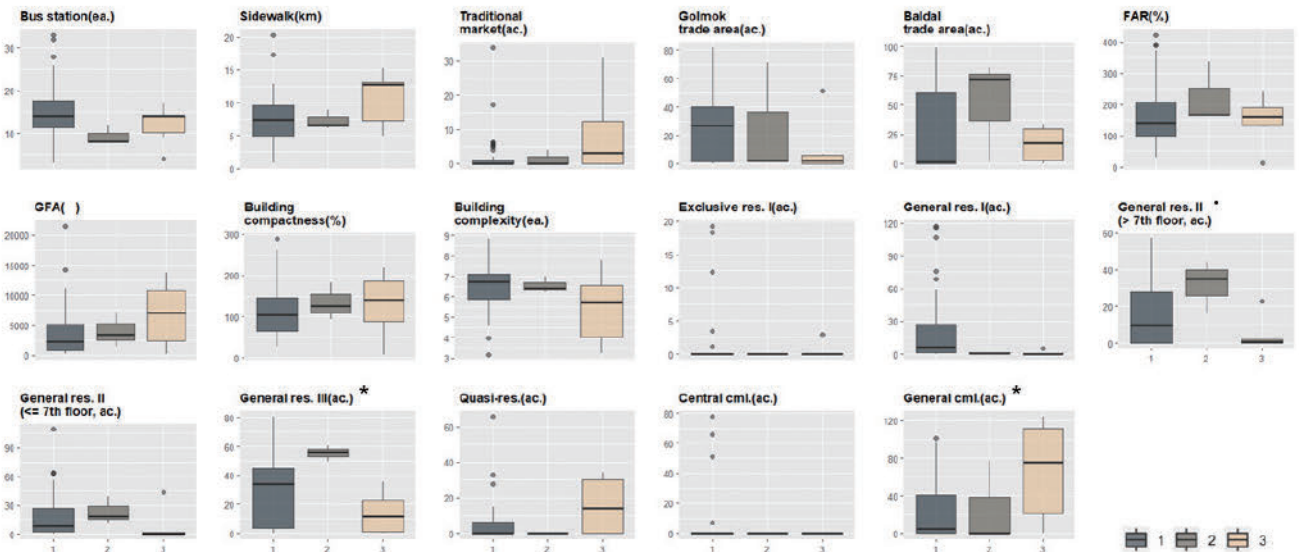


Figure 6. Boxplots of urban environ. between time series clusters (Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1)

아니라 점심 시간대인 13-14시와 퇴근 시간대인 18-19시에 통행량이 집중되었다.

군집 1의 개체 수는 38개로, 가장 많은 개체를 포함한 군집으로 확인되었다. 다양한 개체가 포함되어 있으므로 모든 값의 범위는 최대 약 1000명에 이르나 중심선의 최대치는 약 300명인 것으로 분석되어 모든 군집 중 인구수의 범위가 가장 낮은 것으로 나타났다. 8시부터 인구가동량이 증가하며 오전 시간대에 잠시 정체하였다가 점심시간에 급격히 증가하는 것을 알 수 있다. 이후 오후에는 그 값이 급감 내지 정체한 이후 퇴근 시간대인 18-19시에 증가하였다. 이후 약 22시까지 점진적으로 감소하는 것으로 나타났다.

군집 2의 개체 수는 3개로, 가장 적은 개체를 포함한 동시에 가장 큰 범위의 통행량을 보이는 군집으로 확인되었다. 모든 침투 시간대에 이동량이 급증하는 동시에 19-20시를 정점으로 값이 점진적으로 커진 것을 알 수 있다. 이후 시간이 지나며 인구수가 감소하나 22시 이전까지 평균적으로 약 1000명 이상의 야간인구가 활동하였던 것으로 분석되었다.

군집 3의 개체 수는 6개로, 모든 침투 시간대에서 평균적으로 인구 집중도가 높은 것으로 나타났다. 그중 출근 및 점심 시간대에 가장 높은 통행량으로 그래프가 정점에 달했고 퇴근 시간대 이후 인구이동량이 급감하였던 것을 확인할 수 있다.

시계열 군집분석으로 분류한 결과에 따라 군집별 도시환경특

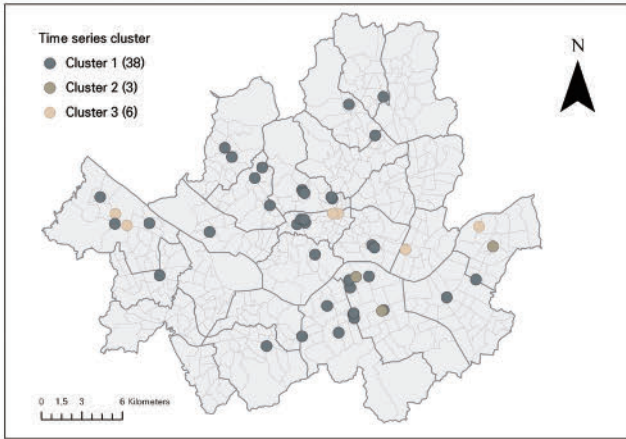


Figure 7. The location of S-DoT floating population sensors classified by time-series clustering

성에 대한 분산 분석을 수행하였다. 앞서 분석하였던 1, 2차 센서 구분에 따른 분산 분석의 결과와 달리 3개의 토지용도 변수만 유의성을 확보하였다. 제2종 일반주거지역(7층 초과)은 10%의 유의 수준에서, 제3종 일반주거지역과 일반상업지역은 5% 이내에서 통계적으로 군집 간 평균 차이를 보이는 도시환경변수로 확인되었다.

〈그림 6〉의 상자그림을 통해 각 변수별 군집 간 평균 차이를 시각적으로 확인할 수 있다. 제2종 일반주거지역(7층 초과)과 제3종 일반주거지역에서는 군집 2, 1, 3의 순서로 변수의 평균치가 높고, 일반상업지역에서는 군집 3의 평균이 다른 두 군집에 비해 월등히 높은 것을 알 수 있다.

정리하면 군집 1과 2는 상업지역의 면적에서 차이를 보이지 않으나 군집 2가 1에 비하여 두 주거지역의 면적이 높다. 한편, 군집 3은 주거용 면적이 가장 낮으나 상업용 면적이 가장 높은 것으로 나타났다.

V. 논의 및 결론

4장 1-3절을 통해 개별 센서 단위, 분석 일자 또는 사업 연도에 따른 센서 종류와 같은 다양한 세분화 기준을 통해 데이터의 구조, 내용 및 품질 등을 확인하였다. 그 결과, 센서의 인구 측정방식에 따라 데이터의 구조 및 크기가 큰 차이가 있음을 발견하였다. WiFi 기반의 1차 센서는 영상분석기법을 활용한 2차 센서에 비하여 인구수의 최대치가 높은 것으로 확인되었다.

이는 1차 센서가 통신신호 횡수를 기반으로 측정하기에 WiFi 이용자의 연결 상태가 체류 시간에 따라 누적되어 집계되었기 때문인 것으로 생각된다. 그러므로 〈그림 3〉의 I와 J를 통해서도 알 수 있듯이 1차 센서군의 시간별 평균 그래프는 주중과 주말의 차이가 미미할 뿐만 아니라 각각의 그래프에서 침투 시간대의 인구 집중도를 분별하기 어려운 것으로 보인다.

한편, 4장 4절에서는 센서 종류에 따른 입지 특성에 대하여 분

산 분석을 수행하였다. 이는 현재 센서의 위치를 분석하여 설치 사업의 목적을 역으로 추적하기 위함이다. 그 결과, 1차 센서는 서울연구원(2020)의 설명과 유사하게 전통시장, 골목상권 등과 같이 좁은 길에 면하고 소형 점포의 밀집도가 높은 상권의 비중이 높은 것으로 나타났다. 이 두 변수가 주로 저층주거지역에 포함되어 있기에 제2종 일반주거지역(7층 이하)에서도 통계적 유의성이 확인되었다.

반면에 2차 센서는 주로 규모가 비교적 큰 상업시설이 있는 발달상권뿐만 아니라 업무시설도 입지할 수 있는 중심·상업지역 등의 면적 비중이 높은 것으로 분석되었다. 이에 따라 용적률, 연면적 및 토지이용 압축도 등 건물군의 높이가 반영되는 변수에서 통계적 유의성이 확보되어 두 센서군의 입지 특성 간 인구활동 양상과 수직적 도시경관을 유추할 수 있다. 즉, 1차 센서군은 주로 좁은 길에 면한 소형 점포가 밀집한 저층공동주택지역에 있고 2차 센서군은 비교적 규모가 크고 중고층 이상의 집객시설이 입지 가능한 상업지역에 있음을 알 수 있다.

4장 5절은 센서의 인구이동패턴에 따라 유사한 개체들로 군집화한 이후, 시계열 군집 간 입지특성을 비교한 내용이다. 군집 1은 가장 많은 개체를 포함하고 있기에 2차 센서의 전형적인 인구이동패턴을 반영하고 있다고 볼 수 있다. 반면에 군집 2와 3은 개체 수는 적으나 각각 주거지역과 상업지역 면적의 비중이 특징적임을 알 수 있다.

먼저, 군집 3은 모든 군집 중 유동인구수의 크기와 집중도가 가장 두드러진 것으로 나타나 가장 높은 상업지역의 비중이 유동인구에 영향을 미칠 것으로 예상된다. 다시 말해, 군집 3에서는 업무시설이 입지할 수 있는 일반상업지역의 비중이 높아 근무 시간 이외의 침투 시간대에 집중도가 높은 동시에, 퇴근 이후에는 인구유출이 급격히 이루어지기 때문인 것으로 추측한다.

물론 나머지 두 군집인 1과 2도 기본적으로 침투 시간대에 유동인구가 집중되나, 군집 3과 비교하면 그 정도가 약한 것을 알 수 있다. 이 군집의 특징은 야간시간대에 군집 3과 달리 인구가 급격히 감소하지 않고, 퇴근 시간대의 절반 수준을 평균적으로 유지하는 것이다. 상자그림을 통해 두 군집의 주거지역 비중이 높은 것을 확인할 수 있는데, 해당 지역에 주거용 토지면적이 높음에 따라 늦은 시간에도 귀가로 인한 인구유입 내지 여가를 위한 야간 활동이 이루어졌기 때문인 것으로 생각된다.

본 연구는 IoT 기반 S-DoT 센서의 입지특성과 이로부터 생산되는 유동인구 데이터에 대한 심층적인 기술 분석을 진행하였다. 분석 기준을 다양하게 설정함으로써 서울시 유동인구의 동태적 양상을 관찰하고 군집분석을 통해 핵심적인 시계열 특성을 도출하였다. 또한 센서 인근의 보행과 관련된 도시환경특성을 추출함으로써 지역 특성과 유동인구 데이터 간의 연관성을 분석하고자 하였다. 이러한 분석은 정밀한 전처리 과정으로 구득한 무결점 데이터를 산출하였기에 가능하였다.

이를 바탕으로 한 정책적 제언은 다음과 같다. 첫째, 현재 병행되는 S-DoT 유동인구 센서의 측정 방식을 일원화하는 것이 필요하다. 2장에서 검토하였듯이, 이 데이터는 유량(flow)을 의미하는 보행 통행량이므로 CCTV 영상 기반의 분석기법이 적합한 것으로 판단된다. 따라서 WiFi 기반의 1차 센서 데이터는 저장(stock) 개념의 현재인구에 가까우므로 이에 대한 이용자 공지와 적절한 활용사례 발굴이 필요하다. 둘째, 센서 인프라 설치 시 유동인구 데이터의 의미를 효과적으로 활용할 수 있는 장소에 입지시키는 것이 중요하다. 예를 들어, 신호등, 교차로, 가로변 및 지하철역 등 작은 면적 대비 높은 인구가 집중되는 곳을 설치 장소로 탐색하는 과정이 필요하다.

한편, 본 연구에서는 시간당 6회로 측정된 무결점 데이터만을 분석용 자료로 활용하였기에 몇 가지 한계가 수반된다. 먼저 2일의 24시간 데이터만을 활용하였기에 분석결과를 다른 시간적 범주의 데이터에 적용하기 어렵다는 한계가 있다. 특정 일자에 한하여 분석된 결과는 대상 일자가 바뀔 경우 일정 수준 차이를 보일 가능성이 높기 때문이다. 이처럼 충분한 기간의 시계열 데이터를 확보할 수 없는 상황에서, 인과관계를 추정하는 통계적 방법을 적용하는 것은 적절치 못하므로, 도시환경특성과 유동인구 간의 관계를 명확히 설명하는 데에도 한계가 있다.

또한 시계열 군집분석 결과, 개체 수가 일부 군집에 편중되었기 때문에 이에 따른 분석 결과의 한계가 존재할 수 있다. 이는 도시 내 인구이동이 주로 출퇴근, 점심시간 등 시간대에 따라 유형화된 생활패턴이 반영된 현상에 기인한 것으로 생각된다. 물론 분석 방법의 원리에 따르면 매개 변수인 k값을 증가시킬 수 있으나, 미세한 값의 차이로 군집 수가 늘어나게 되면 오히려 직관적인 해석과 합의의 도출이 어려워지기 때문에 그 수를 제한하였다.

향후 센서 인프라 확충 및 고도화에 따라 좀 더 높은 수준의 품질을 갖춘 데이터가 확보된다면, 가용 데이터의 범위를 확대하여 일별 또는 월별 등 다양한 분석 기간 설정을 통한 후속 연구를 진행할 수 있을 것이다. 가령, 센서 유동인구 데이터를 기반으로 코로나19와 같이 이동패턴 및 사회경제적 활동에 직간접적으로 영향을 미치는 외생적 요소에 따라 시민들이 어떻게 반응하였는지 그 동태적 양상 등을 파악할 수 있을 것이다. 나아가 장기간에 걸친 시계열 데이터가 확보된다면, 시계열 간 관계를 분석하는 후속 연구도 기대할 수 있을 것이다.

- 밀집업종을 점포 및 지역 특성에 따라 분류한 4개의 상권 종류를 의미한다. 분석 시 관광특구는 데이터가 적어 제외하였다(서울시, <https://galmok.seoul.go.kr/introduce.do>).
- 전통시장: 오랜 기간에 걸쳐 일정한 지역에서 자연발생적으로 형성된 상설시장이나 정기시장
- 골목상권: 대로변이 아닌 거주지 인근의 좁은 도로를 따라 형성되고 점포 밀집도가 높은 상권
- 발달상권: 2000m² 이내 50개 이상의 상점이 분포하고, 배후지 고려 없이 도보이동이 가능한 상가업소밀집지역
- 주4. 국토교통부 국토지리정보원은 「국토기본법」 제25조 국토조사에 따라 2017년부터 500m 이하의 격자 단위로 인구, 건물, 생활인프라 등에 대한 국토지표를 구축하여 공개하고 있다. 기존의 통계가 단순 수치로 공간데이터를 제공하는 것과 달리, 국토지표는 변화하는 법정 경계와 달리 격자 단위의 세밀한 공간데이터를 항구적으로 제공 및 분석할 수 있다는 장점이 있다(박진호, 2021). 용적률과 연면적은 용도 구분 없이 최소 100m 격자 단위로 제공하고 토지이용(건물) 압축도 및 복합도는 500m 격자 단위만으로 공개된다. 이때 압축도는 '{격자 내 건물 연면적 합 ÷ 격자면적(0.25km²)} × 100%' 복합도는 '격자 내 건물용도 개수'다.
- 주5. 그래프 내 점선은 각 시계열 군집의 표준형(prototype) 그래프를 뜻한다. 이는 같은 군집 내 다른 시계열 군집 간의 모든 거리의 합이 최소가 되는 중심선(medoid)을 의미한다(Haulamaki et al., 2008).

인용문헌 References

- 강영민·박동규·김성수, 2021. 「옴므 머신러닝」, 파주: 생능출판사.
Kang, Y.M., Park, D.K., and Kim, S.S., 2021. *Prime Machine Learning*, Paju: Life and Power Press.
- 김경근·염명배, 2017. “신용카드 빅데이터를 활용한 지역별 소비 유출입 특성 연구”, 「경제연구」, 35(4): 129-153.
Kim, K.K. and Yeom, M.B., 2017. “A Study on the Characteristics of Regional Consumption by using Big Data of Credit Card Big Data”, *Journal of Economics Studies*, 35(4): 129-153.
- 김경태·이인묵·곽호찬·민재홍, 2015. “유동인구 추정 시 통신 자료의 활용에 관한 연구”, 「서울도시연구」, 16(3): 177-187.
Kim, K.T., Lee, I.M., Kwak, H.C., and Min, J.H., 2015. “Application Study of Telecommunication Record Data in Floating Population Estimation”, *Seoul Studies*, 16(3): 177-187.
- 김미경·강신혜·김상필·손홍규, 2016. “유동인구를 활용한 대피소 수용 능력 분석”, 「한국측량학회지」, 34(1): 1-10.
Kim, M.K., Kang, S.H., Kim, S.P., and Sohn, H.K., 2016. “A Spatial Analysis of Shelter Capacity Using Floating Population”, *Journal of the Korean Society of Survey, Geodesy, Photogrammetry and Cartography*, 34(1): 1-10.
- 김수현·김태현·임하나·최창규, 2015. “소매업의 매출액을 결정하는 보행량 및 건조 환경 요인에 관한 연구: 서울시 편의점, 화장품소매점, 커피전문점을 중심으로”, 「국토계획」, 50(3): 299-318.
Kim, S.H., Kim, T.H., Im, H.N., and Choi, C.K., 2015. “Pedestrian Volume and Built Environmental Factors on Sales of

주1. 6S는 스마트도시 인프라의 여섯 가지 구성 요소인 S-Net, S-Data, S-Brain, S-Map, S-Security과 S-DoT를 총칭한다.

주2. 국내에서 주로 'Floating population'으로 번역하고 있으나 Cambridge dictionary에 따르면 이주 또는 장거리 여행의 의미도 포함한다. 한국의 맥락에서는 '유동인구'의 번역어로 정착하여 본 연구에서도 이를 주된 번역어로 채택하였다.

주3. 서울상권이란 서울시 또는 외부 기관의 빅데이터를 통해 100개의 생활

- Convenience Stores, Cosmetic Shops and Coffee Shops in Seoul”, *Journal of Korea Planning Association*, 50(3): 299-318.
6. 김용훈·양승우, 2017. “서울 삼청동길의 보행인구 데이터를 활용한 구간별 상업가로 활성화 요인에 대한 실증연구 - 서울시 종로구 삼청동길을 중심으로”, 『한국도시설계학회지 도시설계』, 18(1): 63-77.
Kim, Y.H. and Yang, S.W., 2017. “An Empirical Research on the Vitalization Factors of the Commercial Street with the Walking Population Data - Focused on Seoul city Jongno-gu Samcheong-dong Road”, *Journal of the Urban Design Institute of Korea Urban Design*, 18(1): 63-77.
 7. 김형준·원미리·이정민·채호동·박인권, 2021. “외국인 거주자의 시공간적 분포패턴 및 결정요인: 서울시 생활인구 데이터 분석”, 『LHI 저널』, 12(1): 101-117.
Kim, H.J., Won, M.L., Yi, J.M., Chae, H.D., and Park, I.K., 2021. “Spatio-Temporal Distribution Patterns of Foreign Population and Their Determinants: Analyzing Seoul’s De Facto Population Data”, *LHI Journal*, 12(1): 101-117.
 8. 김홍선·김성문, 2016. “지하철 빅데이터를 이용한 수요 창출형 건축물의 유동인구 증가 효과 분석”, 한국경영학회 통합학술발표 논문집, 124-137.
Kim, H.S. and Kim, S.M., 2016. “Analyzing the Effect improving floating populations of Demand-generating Building Using Subway Bigdata”, paper presented at Conference of Korean Academic Society of Business Administration, 124-137.
 9. 나중화, 2020. 『R 응용 시계열분석』, 파주: 자유아카데미.
Na, J.H., 2020. *Applied Time Series Analysis*, Paju: Free Academy.
 10. 박건철·김유신·고설태·강희주·박지혜·박은주·지태훈·조창훈·임관목·남궁동원·주연진, 2020. 『스마트서울 도시데이터복합센서(S-DoT) 수집데이터 특성 및 활용 방향 분석연구』, 서울 디지털재단.
Park, G.C., Kim, Y.S., Ko, S.T., Kang, H.J., Park, J.H., Park, E.J., Ji, T.H., Cho, C.H., Im, K.M., Namgung, D.W., and Ju, Y.J., 2020. *Analyzing for Characteristics of S-DoT Data and Use Cases*, Seoul Digital Foundation.
 11. 박노진·고한석, 2020. “합성곱 오토인코더 기반의 응집형 계층적 군집 분석”, 『멀티미디어학회 논문지』, 23(1): 1-7.
Park, N.J. and Ko, H.S., 2020. “Agglomerative Hierarchical Clustering Analysis with Deep Convolutional Autoencoders”, *Journal of Korea Multimedia Society*, 23(1): 1-7.
 12. 박배균, 2020. “스마트 도시론의 급진적 재구성: 르페브르의 ‘도시혁명’론을 바탕으로”, 『공간과 사회』, 30(2): 141-171.
Park, B.K., 2020. “Re-reading Smart City Discourses through the Lefebvrian Concept of ‘Urban Revolution’”, *Space and Environment*, 30(2): 141-171.
 13. 박병훈·유다영·박동주·홍정열, 2021. “IoT 도시빅데이터를 활용한 도로교통특성과 유해환경요인 간 영향관계 분석”, 『한국 ITS학회논문지』, 20(5): 130-145.
Park, B.H., Yoo, D.Y., Park, D.J., and Hong, J.Y., 2021. “Impact of Road Traffic Characteristics on Environmental Factors Using IoT Urban Big Data”, *The Journal of The Korea Institute of Intelligent Transport Systems*, 20(5): 130-145.
 14. 박성경, 2018. “주거·공업 혼재지역의 성격 변화에 따른 성수동 유동인구 특성 분석”, 서울대학교 대학원 석사학위논문.
Park, S.K., 2018. “Analysis of the Characteristics of Seongsu-dong Floating Population according to Character Alteration of the Mixed Residential and Industry Area”, Master’s Dissertation, Seoul National University.
 15. 박예림·강영옥, 2019. “통신 데이터를 활용한 도보관광코스 유동인구 추정 및 분석”, 『지적과 국토정보』, 49(1): 181-195.
Park, Y.R. and Kang, Y.O., 2019. “Estimation of Flow Population of Seoul Walking Tour Courses Using Telecommunications Data”, *Journal of Cadastre & Land Information*, 49(1): 181-195.
 16. 박진호, 2021. “국토조사”, 『월간 국토』, 2021(6): 46-47.
Park, J.H., 2021. “National Research”, *Planning and Policy*, 2021(6): 46-47.
 17. 박혜경, 2021. “도시 열섬 강도 분석을 위한 도시데이터 센서와 국가기상관측 간 기온 및 공간해상도 비교 연구”, 『한국방재학회 논문집』, 21(3): 39-48.
Park, H.K., 2021. “Comparison of Temperatures and Spatial Resolutions between Urban Sensors and National Weather Observations (ASOS, AWS) for Urban Heat Island Intensity Analysis”, *Journal of Korean Society of Hazard Mitigation*, 21(3): 39-48.
 18. 변미리·서우석, 2011. “도시 거리의 주간활동인구 측정과 해석: 서울시 유동인구 조사 사례”, 『조사연구』, 12(2): 27-50.
Byun, M.R. and Seo, U.S., 2011. “How to Measure Daytime Population in Urban Streets?: Case of Seoul Pedestrian Flow Survey”, *Survey Research*, 12(2): 27-50.
 19. 서울연구원, 2020. 『도시데이터센서를 이용한 서울시 기온 분포 분석』, 서울.
The Seoul Institute, 2020. *Analyzing for the Distribution of Temperature in Seoul Using Urban Data Sensors*, Seoul.
 20. 서울특별시·KT BigData 사업지원단, 2018. 『서울 생활인구추계 매뉴얼』, 서울.
Seoul Metropolitan Government, KT BigData Business Support Group, 2018. *Manual for De Facto Population in Seoul*, Seoul.
 21. 서울특별시, 2021. 『CCTV 설치 공사지침서(CCTV지주, CCTV 스마트폴) Version 1.0.』, 서울.
Seoul Metropolitan Government, 2021. *Construction Guidelines for CCTV Installation Version 1.0.*, Seoul.
 22. 서울특별시, 2022. 『'22~'23년 도시데이터 센서(S-DoT) 유지관리 착수보고』, 서울.
Seoul Metropolitan Government, 2022. *'22-'23 S-DoT Maintenance Report*, Seoul.
 23. 성명준·엄진기·최명훈·이준, 2010. “스마트 카드 데이터를 이용한 통행패턴 분석”, 한국철도학회 2010년도 정기총회 및 추계학술대회.
Sung, M.J., Eom, J.K., Choi, M.H., and Lee, J., 2010. “Analysis of Travel Patterns using Transit Smart Card Data”, paper presented at Conference of The Korean Society for Railway, Jeju: Jeju Phoenixisland.
 24. 손정열, 2020. “빅데이터 연구에서의 빅마켓, 빅네임 그리고 빅네트워크: 국제학술지를 통해 본 도시 빅데이터 연구동향”, 『대한

- 지리학회지], 55(2): 161-179.
- Sohn, J.Y., 2020. "Big Markets, Big Names and Big Networks in Big Data Research: Urban Big Data Research Trends in International Academic Journals", *Journal of the Korean Geographical Society*, 55(2): 161-179.
25. 오다원, 2019. "주·야간 상권의 토지이용과 이용인구: 서울시 생활인구 자료 분석을 중심으로", 서울대학교 대학원 석사학위논문.
- Oh, D.W., 2019. "Land Use and Using Population of Commercial Areas by Day and Night: Focusing on Analysis of De Facto Population in Seoul", Master's Dissertation, Seoul National University.
26. 유승수·문준경·김민혜·채병선, 2018. "도시재생사업 지역의 보행량 변화 실증분석: 전주시 도심의 쇠퇴지구를 중심으로", 「국토연구」, 96: 103-123.
- You, S.S., Mun, J.K., Kim, M.H., and Chai, B.S., 2018. "An Empirical Analysis of Pedestrian Volume Change in Urban Regeneration Project: Focusing on the CBD of Jeonju-si", *The Korea Spatial Planning Review*, 96: 103-123.
27. 윤나영·최창규, 2013. "서울시 상업가로 보행량과 보행 환경 요인의 관련성 실증 분석", 「국토계획」, 48(4): 135-150.
- Yun, N.Y. and Choi, C.G., 2013. "Relationship between Pedestrian Volume and Pedestrian Environmental Factors on the Commercial Streets in Seoul", *Journal of Korea Planning Association*, 48(4): 135-150.
28. 윤매술, 2020. "서울시 인구이동패턴을 통한 지역 분석: 통신 및 교통 데이터를 활용한 멀티레벨 군집분석을 기반으로", 서울대학교 대학원 석사학위논문.
- Yoon, M.S., 2020. "Extracting Human Dynamics from Smart Phone and Transportation Data", Master's Dissertation, Seoul National University.
29. 은기수, 2001. "상주인구와 주간인구의 계층구조 비교분석: 서울을 중심으로", 「한국인구학」 24(1): 41-65.
- Eun, K.S., 2001. "Social Stratification of the Great Seoul Area: A Comparative Study Using Two Types of Population", *Korea Journal of Population Studies*, 24(1): 41-65.
30. 이명호·홍상기·신동빈, 2019. "서울특별시 소지역 상권 매출에 영향을 미치는 요인분석을 위한 공간가중회귀모형 적용에 관한 연구 - 공간 빅데이터 활용을 중심으로 -", 「국토지리학회지」, 53(2): 213-224.
- Lee, M.H., Hong, S.K., and Shin, D.B., 2019. "A Study on Application of Spatial Weighted Regression for Analyzing the Influential Factors of Sales in Sub-region Commercial Area of Seoul - Focusing on Geo-Spatial Big Data Utilization -", *The Geographical Journal of Korea*, 53(2): 213-224.
31. 이승민, 2020. "교통카드 빅데이터 분석을 통한 서울시 통행 유동 클러스터 탐지 및 유형화", 서울대학교 대학원 석사학위논문.
- Lee, S.M., 2020. "Travel Flow Cluster Detection and Characterization Based on the Analysis of Transport Smartcard Big Data in Seoul", Master's Dissertation, Seoul National University.
32. 이연수·추상호·강준모, 2013. "서울시 생활권별 보행량 변화에 미치는 요인 분석", 「국토계획」, 48(5): 197-208.
- Lee, Y.S., Choo, S.H., and Kang, J.M., 2013. "Analysing Key Factors to Affect Change of Pedestrian Volumes by Neighborhood Units in Seoul", *Journal of Korea Planning Association*, 48(5): 197-208.
33. 이향숙·김지윤·추상호, 2014. "서울시 유동인구조사자료를 활용한 보행특성 분석: 서울시 5개 생활권역을 중심으로", 「대한교통학회지」, 32(4): 315-326.
- Lee, H.S., Kim, J.Y., and Choo, S.H., 2014. "Analyzing Pedestrian Characteristics Using the Seoul Floating Population Survey: Focusing on 5 Urban Communities in Seoul", *Journal of Korean Society of Transportation*, 32(4): 315-326.
34. 이희연·노승철, 2020. 「고급통계분석론」, 고양: 문우사
- Lee, H.Y. and Noh, S.C., 2020. *Advanced Statistical Methodology*, Goyang: Moonwoosa.
35. 임수명·김현철·김동준·안영수, 2019. "다중회귀모형을 이용한 보행가로별 유동인구의 추정에 관한 연구 - 서울 강남·서초구를 중심으로 -", 「서울도시연구」, 20(2): 121-139.
- Yim, S.M., Kim, H.C., Kim, D.J., and An, Y.S., 2019. "A Study on the Estimating of the Floating Population of Walking Street by the Multiple Regression Model - Focusing on Gangnam and Seocho-gu in Seoul -", *Seoul Studies*, 20(2): 121-139.
36. 임하나·이수기·최창규, 2016. "서울시 토지이용 혼합유형과 보행량의 연관성 실증분석", 「국토계획」, 51(7): 21-38.
- Im, H.N., Lee, S., and Choi, C.K., 2016. "Empirical Analysis of the Relationship between Land Use Mix and Pedestrian Volume in Seoul, Korea", *Journal of Korea Planning Association*, 51(7): 21-38.
37. 임화진·박성현, 2020. "유동인구분석을 위한 이동통신 공간빅데이터 활용성 고찰 -일본 동경도 타마뉴타운을 사례로-", 「한국지적정보학회지」, 22(1): 95-107.
- Lim, H.J. and Park, S.H., 2020. "Analysis of mobile spatial big data at Tama New Town", *Journal of the Korean Cadastre Information Association*, 22(1): 95-107.
38. 정재훈·남진, 2019. "위치기반 빅데이터를 활용한 서울시 활동인구 유형 및 유형별 지역 특성 분석", 「국토계획」, 54(3): 75-90.
- Jung, J.H. and Nam, J., 2019. "Types and Characteristics Analysis of Human Dynamics in Seoul Using Location-based Big Data", *Journal of Korea Planning Association*, 54(3): 75-90.
39. 최막중·신선미, 2001. "보행량이 소매업 매출에 미치는 영향에 관한 실증분석", 「국토계획」, 36(2): 75-83.
- Choi, M.J. and Shin, S.M., 2001. "An Empirical Analysis of the Effect of Pedestrian Volume on Retail Sales", *Journal of Korea Planning Association*, 36(2): 75-83.
40. 최막중·이재우, 1999. "비주거인구 및 역세권인구에 의한 상권점유 규모와 수요특성", 「국토계획」, 34(1): 51-64.
- Choi, M.J. and Lee, J.W., 1999. "Non-resident and Transit Population-based Trade Areas: A Shopping Travel and Activity Analysis", *Journal of Korea Planning Association*, 34(1): 51-64.
41. 허정원·장주영, 2020. "코로나19 확산시기 서울시 외국인 밀집지역의 지역특성과 생활인구 변화", 「공간과 사회」, 30(3): 99-137.
- Huh, J.W. and Jang J.Y., 2020. "Mapping Community-level Mobility Changes of Koreans and Immigrants Using Bigdata of 'de Facto Population of Seoul': Six Ethnic Enclaves in Seoul Under COVID-19 Epidemic", *Space and Environment*,

- 30(3): 99-137.
42. D'Urso, P., De Giovanni, L., and Massari, R., 2018. "Robust Fuzzy Clustering of Multivariate Time Trajectories", *International Journal of Approximate Reasoning*, 99: 12-38.
 43. Dalal, S. and Dahiya, V., 2020. "Big Data Preprocessing: Needs and Methods", *International Journal of Engineering Trends and Technology*, 68(10): 100-104.
 44. Hautamaki, V., Nykanen, P., and Franti, P., 2008. "Time-series Clustering by Approximate Prototypes", Paper presented at 2008 19th International Conference on Pattern Recognition, Florida: Tampa.
 45. Maharaj, E.A., D'Urso, P., and Caiado, J., 2019. *Time Series Clustering and Classification*, Boca Raton: CRC Press.
 46. Prakash, A., Navya, N., and Natarajan, J., 2019. "Big Data Preprocessing for Modern World: Opportunities and Challenges", Paper presented at International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018, India: Coimbatore.
 47. Rabari, C. and Storper, M., 2015. "The Digital Skin of Cities: Urban Theory and Research in The Age of The Sensored and Metered City, Ubiquitous Computing and Big Data", *Cambridge Journal of Regions, Economy and Society*, 8(1): 27-42.
 48. Sard'a-Espinosa, A., 2022. "dtwclust: Time Series Clustering Along with Optimizations for the Dynamic Time Warping Distance", R package version 5.5.11. <https://CRAN.R-project.org/package=dtwclust>
 49. Swanson, D.A. and Tayman, J., 2011. "On Estimating a De Facto Population and Its Components", *Review of Economics & Finance*, 1: 17-31.
 50. Townsend, A.M., 2019. 「스마트시티, 더 나은 도시를 만든다」, 도시이론연구모임 역, 서울: MID출판사.
Townsend, A.M., 2019. *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia* (Urban Theory and Research Club, Trans.), Seoul: MID Press.
 51. United Nations, 2008. Principles and Recommendations for Population and Housing Censuses(Revision 2), New York: United Nations.

Date Received	2022-05-08
Reviewed(1 st)	2022-08-11
Date Revised	2022-08-29
Reviewed(2 nd)	2022-09-13
Date Accepted	2022-09-13
Final Received	2022-09-26