

Station Extension Algorithm Considering Destinations to Solve Illegal Parking of E-Scooters

Jeongeun Song*, Yoon-Ah Song*, ZoonKy Lee*

*Student, Graduate School of Information, Yonsei University, Seoul, Korea

*Student, Graduate School of Information, Yonsei University, Seoul, Korea

*Professor, Graduate School of Information, Yonsei University, Seoul, Korea

[Abstract]

In this paper, we propose a new station selection algorithm to solve the illegal parking problem of shared electric scooters and improve the service quality. Recently, as a solution to the urban transportation problem, shared electric scooters are attracting attention as the first and last mile means between public transportation and final destinations. As a result, the shared electric scooter market grew rapidly, problems caused by electric scooters are becoming serious. Therefore, in this study, text data are collected to understand the nature of the problem, and the problems related to shared scooters are viewed from the perspective of pedestrians and users in 'LDA Topic Modeling', and a station extension algorithm is based on this. Some parking lots have already been installed, but the existing parking lot location is different from the actual area of tow. Therefore, in this study, we propose an algorithm that can install stations at high actual tow density using mixed clustering technology using K-means after primary clustering by DBSCAN, reflecting the 'current state of electric scooter tow in Seoul'.

▶ **Key words:** Personal Mobility, Shared electric scooter, LDA Topic Modeling, DBSCAN, K-MEANS

[요 약]

본 연구는 공유 전동 킥보드 불법 주차 문제를 해결하고 서비스 품질을 개선하기 위하여 새로운 스테이션 선정 알고리즘을 제안하는 것을 목표로 한다. 최근 도시 교통 문제의 해결 방안으로 대중 교통과 최종 목적지를 연결하는 퍼스트-라스트 마일 수단으로 공유 전동 킥보드가 주목받고 있다. 그에 따라 공유 전동 킥보드 시장도 급속도로 성장하면서 그로 인한 문제 또한 심화되고 있다. 이에 본 연구에서는 문제의 본질을 파악하기 위해 텍스트 데이터를 수집하여 'LDA 토픽 모델링'으로 공유 킥보드에 관한 이슈를 보행자와 이용자 관점에서 살펴보고 이를 바탕으로 스테이션 증설 알고리즘을 마련하려 한다. 기존에 이미 일부 주차장이 설치되어 있지만 기존의 주차장 위치는 실제 견인 다발 지역과 불일치한다. 따라서 본 연구에서는 '서울특별시 전동 킥보드 견인 현황' 데이터와 여러 요인을 반영하여 DBSCAN을 통한 1차 클러스터링 후, K-means를 적용하는 혼합형 클러스터링 기법으로 실제 견인 밀도가 높은 곳에 스테이션을 설치할 수 있는 알고리즘을 제안한다.

▶ **주제어:** 퍼스널 모빌리티, 공유 전동 킥보드, LDA 토픽 모델링, DBSCAN, K-MEANS

-
- First Author: Jeongeun Song, Yoon-Ah Song, Corresponding Author: ZoonKy Lee
 - *Jeongeun Song (stuplant@yonsei.ac.kr), Graduate School of Information, Yonsei University
 - *Yoon-Ah Song (dbsdk1683@yonsei.ac.kr), Graduate School of Information, Yonsei University
 - *ZoonKy Lee (zlee@yonsei.ac.kr), Graduate School of Information, Yonsei University
 - Received: 2023. 01. 26, Revised: 2023. 02. 08, Accepted: 2023. 02. 14.

I. Introduction

도시 교통 문제의 해결 방안으로 대중교통을 비롯한 친환경 교통수단이 각광을 받으면서 최근 ‘퍼스널 모빌리티(Personal Mobility)’ 시장이 급속도로 성장했다. ‘퍼스널 모빌리티(Personal Mobility)’란 전기를 동력으로 삼는 개인형 이동 장치를 말한다. 전동 휠, 전동 킥보드, 전기 자전거, 초소형 전기차 등을 예로 들 수 있다[1]. 특히 이 중 전동 킥보드는 2018년 서울을 시작으로 공유 킥보드 서비스가 시작되면서 그 수가 급격하게 증가하고 있다. 2020년 10월 기준 전년 동월 대비 전동 킥보드 이용자는 314% 증가했다[2]. 2018년 1개의 업체에서 시작하여 2021년 말 기준으로 국내에서 20여 개의 업체가 공유 전동 킥보드 서비스를 운영하고 있다[3]. 서울시는 2025년까지 전동 킥보드 40만대 이상 증가를 예상하고 있다[4]. 그에 따라 전동 킥보드로 인한 사회 문제 또한 발생하고 있다. 특히 공유 전동 킥보드 이용 종료 후 도로나 인도에 방치하는 문제가 집중적으로 관심을 받고 있다. 수도권 거주 20~50대 남녀 1천명을 대상으로 한 설문 조사에 따르면 거리에 무질서하게 주차된 킥보드에 대해 59.6%가 보행에 불편하고 미관상 좋지 않다고 응답했다[5].

이 문제의 원인은 공유 킥보드 운영 방식에 있는데 바로 ‘프리-플로팅(Free-Floating)’을 기반으로 한 ‘독리스(Dockless)’ 방식이다. ‘독리스(Dockless)’ 방식은 일정하게 머무는 곳 없이 공유 킥보드만 있으면 아무 곳에서도 대여 및 반납이 가능한 시스템이다[6]. 직전 이용자가 반납한 임의의 장소에서 대여하고 이용이 끝나면 그 자리에서 즉시 반납한다. 이러한 시스템은 이용자에게 편리할 수도 있겠지만 도로나 보도 한가운데 전동 킥보드를 방치할 가능성이 높다. 따라서 전동 킥보드 주차 문제가 발생할 가능성이 크며 보행자 입장에서는 통행 방해, 운전자 입장에서는 차량 파손 위험, 안전 사고 등 위험 요소를 동반한다. 관련 민원 또한 증가하고 있는 추세다. 서울시 공유 전동 킥보드 민원의 89%가 무단 방치 및 견인 관련으로 나타났다[7]. 또한 이용자에게도 편리함만을 보장하지는 않는다. GPS 오차나 사유지 방치로 인해 공유 전동 킥보드의 위치를 필요할 때 찾기 어려움 때문에 서비스 이용에 불편함을 겪을 수 있다[8].

이에 본 연구에서는 점차 사회 문제가 되고 있는 공유 전동 킥보드 불법 주차 문제를 해결하기 위해 먼저 보행자와 이용자 입장에서 LDA(Latent Dirichlet Allocation) 토픽 모델링을 통해 공유 전동 킥보드에 대한 인식을 분석한다. 이후 결과를 바탕으로 하여 보행자와 이용자 관점에

서 모두 편리한 도착지 중심 스테이션 증설 알고리즘을 제안한다.

기존 시스템에서 ‘독리스(Dockless)’ 방식을 고집하는 이유는 스테이션 마련 비용이 부담이 되기 때문이고 이용자의 편의를 위해서가 가장 크다. 따라서 본 연구의 목적은 스테이션 설치 비용을 최소화하고 이용자의 편의를 증진하면서 불법 주차로 방치되는 전동 킥보드를 줄여 견인으로 인한 공유 킥보드 산업의 위축을 방지하는 것이다.

기존에 설치된 전동 킥보드 주차장은 수가 적을 뿐만 아니라 전동 킥보드의 수요만을 고려하여 지하철역 주변으로 설치가 되어 있어 실제 주차가 필요한 지역을 반영하지 못하고 있다. 이에 ‘서울특별시 전동 킥보드 견인 현황’과 ‘서울 생활이동 인구’ 데이터를 분석하여 견인이 다량 발생하는 지역을 중심으로 살펴본다. 이를 기반으로 하여 로컬 상가와 제휴 또는 공유지 중심으로 스테이션 장소를 선정하는 알고리즘을 제안한다. 주차 스테이션은 견인 다발 지역을 대상으로 한 혼합형 클러스터링 기반 알고리즘을 통해 선정한다. 결과적으로 증설한 스테이션 간 이동으로 공유 킥보드의 고정적 공급을 가능하게 하여 대역자(이용자) 입장에서는 GPS에 의존하지 않아도 접근성이 높은 스테이션에서 서비스를 이용할 수 있으며 사업자 입장에서는 수거 및 충전 인력 비용 또한 절감할 수 있다. 그리고 지자체 입장에서도 견인으로 인한 사회적 비용 절약과 불법 주차로 인한 안전 문제를 해결할 수 있어 공동의 이익이 된다. 이를 통해 서비스 전반의 운영 방식과 부정적 인식 개선으로 공유 전동 킥보드 산업이 더욱 활성화 되기를 기대한다.

II. Research background

1. Related works

공유 전동 킥보드가 2018년 서울에서 첫 등장한 이후로 최근 연구 동향은 기술적 방면과 이용 특성, 수요에 관한 연구가 주를 이루었다.

조정훈 외(2020)에서는 공유 자전거와 공유 킥보드 이용 특성을 비교하고 수요 패턴을 알기 위해 회귀 방법을 통해 수요에 영향을 미치는 요인을 분석하였다. 그 결과 두 수단 모두 주말, 주중은 비슷하지만 시간대 별로 수요가 다르게 나타난다는 것을 알 수 있었다[9].

김수재 외(2021)에서는 무분별한 공유 전동 킥보드 서비스 운영으로 인한 문제의 해결점을 찾기 위해 전동 킥보드의 수요에 초점을 맞췄다. 이용량에 영향을 미치는 요소

에 대하여 분석하였는데 그 결과 연령층, 시간대, 지하철역 출구 수 등이 영향을 미친다는 사실을 밝혀냈다. 그리고 그 요인을 바탕으로 대중교통, 대학가 근처에 파킹 존을 마련할 것을 추천하였다[10].

장은진 외(2021)에서는 공유형 전동 킥보드의 별도 주차 공간 부재를 언급하며 적절한 위치의 주차 스테이션을 선정하여 스테이션 별 잔여 킥보드 정보를 제공하는 IoT 관리 시스템을 제안하였다[11].

최선일 외(2022)에서는 기술적인 측면으로 접근하여 자율 주행 장치를 전동 킥보드에 부착하는 방법으로 불법 주차 문제를 해결하려 하였다. 국내 업체에서 자율 주행 주차 킥보드를 개발하였지만 기존에 서비스 중인 전동 킥보드 전체를 교체하려면 초기 투자 비용이 부담된다는 점에서 부착식 장치를 해결 방안으로 제시하였다[12].

최해정, 장성만(2022)에서는 대중교통과 퍼스트-라스트 마일 수단으로서 전동 킥보드 연계 사용을 증진하기 위해 공유 전동 킥보드의 이용 패턴을 분석하여 수요를 기반으로 배치 방안을 제시하고자 하였다. 그 결과 특정 시간대에 이용이 집중된다는 것과 이용 목적이 출퇴근이라는 것을 추정할 수 있었다. 이를 바탕으로 수요 주도형으로 킥보드를 배치할 것을 제안하였다. 이 연구에서는 스테이션 마련보다는 수요가 많은 곳에 킥보드를 적절하게 배치하는 방안이 집중하였다[13].

한재원 외(2022)에서는 도시 교통 문제의 해결 수단으로서 퍼스널 모빌리티 공급을 긍정적 방안으로 주목했다. 따라서 지형이나 통행 실태 데이터를 바탕으로 수요를 분석하고 퍼스널 모빌리티 공급 적정 지역을 도출했다[14].

기존 연구는 주로 수요와 공급, 이용량에 중점을 두었지만 본 연구에서는 견인 데이터를 기반으로 견인 수량에 기반하여 그에 대한 인식과 영향을 미치는 요인에 대해 분석하고 이를 바탕으로 도착지 중심 스테이션 증설 알고리즘을 마련하려 한다.

Kurniawan 외(2017)에서는 설문, 인터뷰, 관련 연구 데이터를 활용하여 사회 문제를 K-means 알고리즘을 통해 군집화하고 이를 지리정보시스템으로 그룹화하여 지역별 사회 문제를 파악하는 연구를 진행하였다[15].

Annas 외(2022)에서는 도시 대기오염을 분석하기 위한 지역 클러스터링으로 K-means와 SOM을 통합하여 사용하였으며, 두 알고리즘의 결합이 GIS 시각화를 통해 적절하게 매핑될 수 있음을 증명하였다[16].

이병길(2003)에서는 분할기법, 계층적 기법, 밀도 기반 기법, 격자 기반 기법 등 다양한 공간 데이터 마이닝 기법들을 비교 분석하고, 공간 클러스터링을 통해 명확하지 않

은 상권의 범위와 의미를 파악하고자 했다[17].

본 연구에서 제안하는 방법은 특성이 가진 의미를 파악하고자 공간 데이터를 클러스터링한 여러 선행연구를 바탕으로 하여 스테이션 위치 선정을 위해 클러스터링 알고리즘을 활용하고자 한다. 밀도 기반 클러스터링 알고리즘인 DBSCAN과 중심점 기반 클러스터링 알고리즘인 K-means++를 결합한 혼합형 알고리즘으로 견인 데이터의 지리 정보를 클러스터링하여 견인 밀집 지역에 스테이션을 증설하고자 한다.

III. Awareness Analysis of Shared e-scooter

1. Analytical method

스테이션 알고리즘을 제안하기에 앞서 공유 전동 킥보드에 대한 대중들의 인식을 보행자와 이용자 입장에서 분석하여 스테이션 마련 알고리즘에 참고하고자 한다. 우선 보행자의 입장을 분석하기 위해 지역 주민 커뮤니티가 발달되어 있는 네이버 카페를 대상으로 게시글과 댓글 텍스트 데이터를 수집하였다. 수집 방법은 키워드 ‘공유킥보드’를 검색하여 견인 제도 시행 기간 1년 2021.07.15 ~ 2022.07.31 사이의 게시글과 댓글 텍스트를 크롤링하였다. 이용자의 입장을 분석하기 위해서는 구글 플레이 스토어 리뷰를 대상으로 텍스트 데이터를 수집하였다. 수집 방법은 키워드 ‘공유킥보드’를 검색 후 다운로드 100만 이상 상위 5개 앱 리뷰를 크롤링하여 문제점을 집중적으로 살펴보기 위해 평점 5점 만점 중 평점 3점 이하 리뷰만을 추출하였다.

Table 1. Condition of Text data collection

Keyword	Website	Type	Period	Number of data
공유킥보드	Naver Cafe	post comment	2021.07.15. ~ 2022.07.31	16,185
	Google Play Store	review	2018.09.18. ~ 2023.01.16	7,913

크롤링한 전체 텍스트 데이터는 Okt형태소 분석기를 사용하여 정제 과정을 거친다. 명사, 형용사, 동사를 추출하고 불용어, 복합명사, 동의어를 처리하였다. 정제 과정을 거쳐 분석에 사용한 ‘명사’, ‘형용사’, ‘동사’ 토큰 수는 Table 2.와 같다.

이후 ‘공유 킥보드’에 관한 토픽을 통해 대중들의 인식을 분석하고자 LDA(Latent Dirichlet Allocation) 토픽 모델링을 진행하였다. LDA 토픽 모델링은 텍스트에 내재한 토픽들을 확률 분포에 따라 찾아내는 방법이다[18].

Table 2. Tokens Used for Analysis

Website	Total	Type	Number of data
Naver Cafe	559,370	Noun Adjective Verb	16,185
Google Play Store	223,582		7,913

2. Results of LDA Topic Modeling

2-1. Pedestrian's point of view

보행자 관점에서 이슈를 분석하기 위해 네이버 카페에서 ‘공유 킥보드’를 키워드로 하는 게시글과 댓글을 수집하여 LDA 토픽 모델링으로 분석한 결과 3개의 토픽을 추출할 수 있었다. 추출된 토픽은 ‘사고’, ‘이용 행태’, ‘방치와 충전’으로 각 토픽에 속한 단어들을 텍스트 데이터에서 확인한 결과 주로 부정적인 의견이 대다수를 차지했다.

Table 3. Results of LDA Topic Modeling -Pedestrian

Topic	Word	Interpretation
Accident	운전, 사고, 경찰, 차량, 처벌, 사건, 신고, 도주, 보험, 갓길	- Vehicle damage accident caused by neglected e-scooter - Many accident references due to inexperienced e-scooter operation
Usage Behavior	사진, 공원, 주말, 아침, 저녁, 타고, 놀이터	- Confirmation of negative views on morning and evening e-scooter customer sightings - Sharing information on reports of neglected e-scooter photos
Abandonment and Charging Problem	주차, 주차장, 문제, 아파트, 인도, 충전, 도로, 전기차, 신고, 관리, 사용	- E-scooter illegal charging problem (Unauthorized use of electricity in apartment parking lots) - Many references to interference with the use of electric vehicle charging stations - Confirmation of negative views on e-scooters left unattended in apartments, roads and sidewalks

‘사고’ 토픽은 킥보드와 보행자, 차량 관련 사고로 부상이나 차량 파손에 관한 언급이 많았다. 특히 차량 파손 관

련해서는 방치된 킥보드가 넘어지거나 운전자가 불법 주차된 킥보드를 운전 중 발견하기 어려워 사고 위험이 높았다. ‘이용 행태’ 토픽에서는 공유 전동 킥보드 이용자에 대한 부정적인 시선을 확인할 수 있었다. ‘방치 및 충전 문제’ 토픽에서는 지상 차량 통행이 금지된 아파트 단지 내부 방치나 공용 전기를 이용한 불법 충전 문제를 살펴볼 수 있었다.

공유 킥보드에 대한 대중들의 부정적인 시각은 장기적으로 공유 전동 킥보드 서비스 사업의 위축을 가져올 우려가 있다. 실제 프랑스에서는 전동 킥보드 대여 금지가 검토 되기도 했다[19].

2-2. User's point of view

이용자 관점에서 이슈에 접근하기 위해 구글 플레이 스토어에서 ‘공유 킥보드’를 검색하여 다운로드 100만 이상 상위 5개 업체의 리뷰를 LDA 토픽 모델링으로 분석한 결과 3개의 토픽을 추출할 수 있었다. 추출된 3개의 토픽은 ‘애플리케이션’, ‘주차(반납)’, ‘위치’였다. 문제점을 파악하기 위해 평점 5점 만점 중 3점 이하의 리뷰를 추출하여 분석하였기 때문에 이용자의 불편 사항을 중점적으로 살펴볼 수 있었다.

Table 4. Results of LDA Topic Modeling -User

Topic	Word	Interpretation
Application	종료, 이용, 업데이트, 환불, 결제, 센터, 문의, 오류, 비싸요, 안되고	- Many error references to end of use - Opinion that the usage fee is high
Parking(Return)	주차, 반납, 이용, 구역, 서비스, 불가, 최악, 금지	- Parking and return areas are ambiguous and inconvenient - Insufficient guidance on no-parking areas
Location	인증, 위치, 가입, 지도, 회원, 안됨, 어떻게, 탈퇴	- Mismatch between actual e-scooter location and map location - Complain of difficulty locating the e-scooter

‘애플리케이션’ 토픽에서는 애플리케이션 기능적인 문제에 대한 언급이 대다수를 차지했다. ‘주차(반납)’ 토픽에서는 주차 가능 구역으로 지정되어 있더라도 GPS 오차로 인해 반납이 잘 되지 않거나 올바르게 반납을 했더라도 추가 요금이 발생하는 문제에 대한 불만이 많았다. ‘위치’ 토픽 또한 GPS 오차로 인해 킥보드를 찾을 때 지도 상에 표시된 곳과 실제 킥보드의 위치가 달라 불편을 호소하였다.

실제 고층 건물이 많은 도심 지역에서는 위치가 부정확하게 표시되는 경우가 많아 이는 사업자 입장에서도 키포드 회전을 떨어질 수 있어 비효율적이다[20].

2-3. Issue

보행자와 이용자 관점에서 텍스트를 기반으로 분석해 본 결과 전동 키포드 방치로 인한 안전 문제와 주차(반납), 위치 찾기 문제 등이 있었다. 실제 공유 전동 키포드 불법 주차 문제가 심화되자 서울시에서는 전국 최초로 2021년 7월 15일부터 불법 주정차 공유 키포드 견인을 시행했다. 6개 구(성동구, 송파구, 도봉구, 마포구, 영등포구, 동작구)를 시작으로 점차 제도를 확대해 나가고 있다. 견인된 공유 전동 키포드는 업체에 견인료 4만원과 보관료(30분당 700원)가 부과된다[21]. 제도를 시행한지 2달 반만인 10월에는 견인, 보관료만 4억 5,000만원에 달했다[22]. 이는 업계에 큰 부담이 되고 있어 서비스 지속이 불투명해지고 있는 실정이다. 이후에는 약관을 개정하여 견인 비용을 이용자에게 전가하고 있어 소비자(이용자)의 피해 또한 증가하고 있다[23].

보행자와 이용자 두 입장에서 공통적으로 이 문제의 원인은 현재의 공유 전동 키포드 운영 방식인 ‘독리스(Dockless)’ 방식에서 비롯한다. ‘독리스(Dockless)’ 방식은 키포드의 위치를 GPS에 의존할 수 밖에 없다. 하지만 앞선 리뷰 데이터 분석에 따르면 GPS 오차 때문에 주차 금지 구역 구별이 모호하여 같은 장소라도 때에 따라 주차 금지 구역 GPS 인식이 달라 이용자에게 혼란을 가중하고 있다. 따라서 본 연구에서는 GPS에만 의존하지 않는 일정한 스테이션 증설 알고리즘을 마련하여 공유 전동 키포드 서비스에 대한 신뢰성을 높이고자 한다.

IV. Station extension algorithm

1. Current status of e-scooter tows

‘서울특별시 전동 키포드 견인 현황(총 60,605건, 2021.07.15 ~ 2022.07.31)’ 데이터를 분석해 보면 제도 초기부터 동시 시행된 6개 자치구(성동구, 송파구, 도봉구, 마포구, 영등포구, 동작구) 중 전동 키포드 견인이 많은 상위권 3곳은 1위 마포구(7,408), 2위 송파구(6,492), 3위 영등포구(5,381)다.



Fig. 1. Map of e-scooter tows

2021년 기준 서울시 인구는 송파구가 663,965명으로 인구 밀도가 가장 높지만 견인 건수는 마포구(378,686명)가 7,408건으로 가장 다수이기 때문에 ‘서울 생활이동 인구’ 데이터를 활용하여 구별 유입인구를 비교해 보았다[24]. 총 유입인구 역시 1위 송파구(476,998,558), 2위 영등포구(334,691,734), 3위 마포구(311,507,042)로 송파구가 1위를 차지하였다[25]. 이에 연령별 유입인구를 고려하였는데 20~29세 유입인구가 전동 키포드 견인 순위와 가장 밀접한 연관이 있었다. Fig. 2.에서 확인할 수 있듯이 20~29세 자치구별 유입인구 순위가 공유 전동 키포드 견인 순위와 정확히 일치했다.

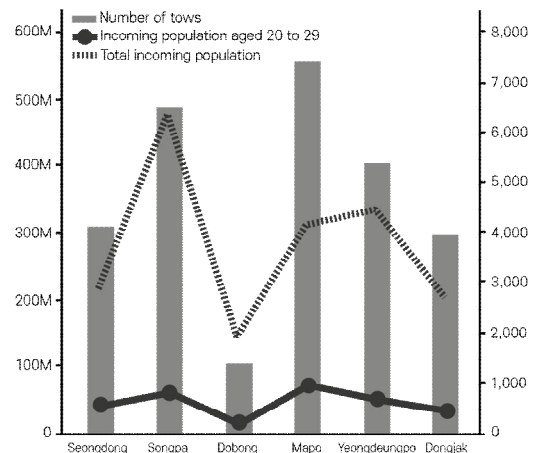


Fig. 2. Incoming population and number of tows

견인이 가장 많이 발생한 마포구 견인 현황을 동 별로 살펴보면 동 별 견인 개수와 기존 주차장의 개수가 불일치를 확인할 수 있다. 실제 견인이 가장 많이 발생한 곳은 서교동(1,581)인데 주차장은 상암동(11)에 가장 많이 분포하고 있었다. 따라서 현행 주차장은 실제 주차가 필요한 지역을 반영하지 못하고 설치되어 있음을 Fig. 3.에서 확인할 수 있다.

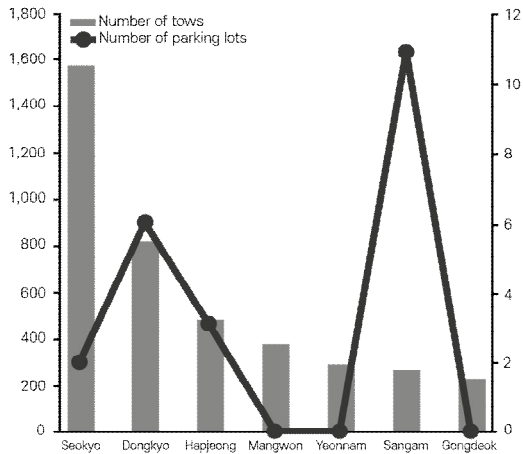


Fig. 3. Number of tows and number of parking lots

2. Theoretical Approach of Methodology

2-1. DBSCAN

DBSCAN 알고리즘은 밀도 기반 클러스터링 알고리즘으로, 반경과 최소점으로 이루어진 두 개의 하이퍼파라미터를 가진다[26]. 각 데이터는 하이퍼파라미터에 따라 코어 포인트, 경계 포인트, 이상치 중 하나로 분류된다. 만약 하나의 데이터인 포인트가 지정 반경 내에서 최소점 이상의 포인트를 갖게 될 경우 코어 포인트로 분류된다. 그리고 반경 내에 최소점 이상의 포인트는 갖지 않으나 다른 포인트 내에 속할 경우 경계 포인트로, 그 외는 이상치로 분류된다.

DBSCAN 알고리즘의 경우는 클러스터의 수를 지정하지 않고, 중심점 기반이 아닌 밀도를 기반으로 군집화하므로 고밀도 클러스터링 문제에 탁월하다[27][28]. 그리고 노이즈 포인트를 코어 또는 경계 포인트로 분류하는 대신 이상치로 처리하여 노이즈가 많은 클러스터링에 효과적이다 [29]. 그러나 밀도를 기반으로 Euclidean Distance가 적용되기 때문에 고차원 데이터가 주어지는 상황에서는 성능이 좋지 않은 단점이 있다[30]. 더하여 반경과 최소점을 설정해주어야 한다는 점에서 하이퍼파라미터에 대한 의존도가 존재한다.

2-2. K-means++

K-means++ 알고리즘은 중심점 기반의 클러스터링 알고리즘으로, K-means 알고리즘의 단점인 초기 중심점 선택을 개선한 알고리즘이다[31][32].

개선 버전인 K-means++는 초기 랜덤한 중심점을 전략적으로 선택하고 알고리즘의 속도를 개선하기 위해 제안되었으며, 정확도와 속도 측면에서 K-means보다 월등함이 입증되었다[31]. K-means와 동일하게 K-means++ 또한 군집의 개수, 즉 K를 하이퍼파라미터로 가지며 중심점

을 기준으로 Euclidean Distance로 거리를 측정하여 데이터를 군집화한다. 이때 군집은 초기 설정한 K에 의해 결정되는데, 첫 중심점은 랜덤으로 선택된다[33].

K-means 기반 알고리즘은 중심점에 대한 거리를 기반으로 거의 모든 데이터를 군집화하며 고차원 데이터 클러스터링 문제에 효과적이다. 그러나 K를 지정해주어야 하므로 파라미터 선정에 예민하고 초기 랜덤한 중심점 사용으로 인해 매 시험의 결과가 초기값에 민감히 반응한다 [34][35][36]. 또한 거의 모든 데이터를 군집화하기 때문에 중심점에서 멀리 떨어져 있는 노이즈를 판별하고 처리하는 데에 어려움이 있다[29].

2-3. Comparison

비지도 클러스터링 알고리즘 DBSCAN과 K-means++는 특성에 따라 각각 장단점을 가지나 어느 하나가 월등한 성능을 가지고 있다고 평가할 수 없다. 두 알고리즘을 비교한 많은 연구에서는 특정 실험에 대해서만 하나의 알고리즘이 뛰어난 성능을 보이거나[27][29][37][38] 두 모델의 장단점이 뚜렷하게 나타났[29][30][32][33][35][39][40]. 따라서 상황에 따른 적절한 모델 선정으로 장점을 극대화할 필요가 있는데, 다양한 연구에서 클러스터링 모델의 병합을 통해 각 모델이 가진 단점을 극복하였다. Fahim 외 (2021)에서는 DBSCAN을 통해 예상 군집 수를 구하여 K-means를 활용하였다[39]. 이성규 외(2006)에서는 평균 연결법과 K-means를 혼합하여 K-means의 단점을 보완하였다[41]. 본 연구에서는 많은 연구에서 입증한 바와 같이 비지도 클러스터링 알고리즘의 장점과 단점을 보완할 수 있는 혼합형 모델을 제안하고자 한다.

3. DBS-K++ Algorithm for Station Selection

본 실험에서는 ‘서울특별시 전동 키펠드 견인 현황’ 데이터를 활용하여 공유 키펠드 주차 스테이션 적정 지역을 선정하는 알고리즘을 제안한다. 데이터는 2021.07.15 ~ 2022.07.31 동안 발생한 60,605건으로, 클러스터링을 통해 불법 주차 견인 건 수가 많았던 지역을 선별하여 키펠드 주차 스테이션을 증설하고자 한다. 공유 키펠드 견인 및 관리는 각 자치구에서 진행하므로 견인 데이터 또한 구별로 구분하여 사용하였다.

실험에서 제안하는 방법은 DBSCAN과 K-means++를 혼합한 DBS-K++ 알고리즘이다. 실험의 프로세스는 Fig. 4과 같다.

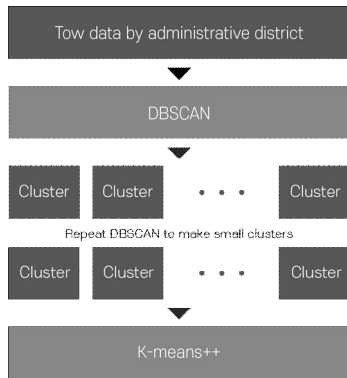


Fig. 4. Across-the-board process

우선 해당 데이터에서 견인 신고 건수가 적은 지역은 스테이션 선정 후보군에서 제외했다. 또한 총 60,605건의 데이터를 군집화할 때, 적정 수준의 하이퍼파라미터 K 를 선정하는 것에 어려움이 존재한다. 따라서 견인 밀도에 따라 클러스터링을 하기 위해서 첫 번째로 각 구에 대해 DBSCAN 알고리즘을 이용하였다. DBSCAN 알고리즘의 경우는 이상치를 잘 분별하며 군집 수를 자체적으로 설정하므로 K -means++로 진행할 때보다 명확한 군집을 마련할 수 있다. 이때 DBSCAN의 파라미터인 최소점은 5로 고정하였고 반경은 이상치와 클러스터링 수가 가장 적은 파라미터로 설정하였다. 다음으로 각 자치구에서 생성한 DBSCAN 군집들을 소 군집으로 구성하여, 각 소 군집을 대상으로 다시 클러스터링을 진행하였다. 그러나 소 군집을 대상으로 또 다시 DBSCAN을 활용하는 데에는 어려움이 있었다. DBSCAN의 경우는 반경과 최소점을 설정해 주어야 하는데, 크기가 모두 다른 다수의 소 군집을 대상으로 같은 알고리즘을 적용하기에는 많은 하이퍼파라미터의 설정이 필요하며 파라미터의 영향이 컸다. 따라서 소 군집을 대상으로는 K -means++ 알고리즘을 사용하였다. 특히 K 를 설정할 수 있을 정도로 군집이 작아졌기 때문에 평가 지표인 엘보우(elbow) 기법과 실루엣 계수를 통해 적절한 K 값을 고려할 수 있었다. 이때 소 군집은 1차적으로 DBSCAN을 끝낸 군집 전체에서 군집 내 데이터가 300개 이상인 군집을 대상으로만 K -means++을 진행했으며, K -means++ 중심값을 출력하여 스테이션 위치를 선정하였다.

4. Experiment result

DBS-K++는 이상치가 포함된 1년 간의 견인 데이터에서 이상치를 제거하기 위해 DBSCAN을 활용하고, 그 결과로 마련된 소 군집을 K -means++에 반복 적용하여 중심 스테이션을 찾아내는 알고리즘이다. 본 연구에서는 불법

주정차 공유 전동 킥보드 견인 제도를 동시 시행한 6개의 자치구를 대상으로 스테이션을 마련하였다.

서울시 전체 견인 데이터 60,605건 중 가장 다수인 7,408건을 차지한 마포구의 스테이션 적정 위치 선정 과정은 다음과 같다. 견인 데이터는 위도와 경도 좌표로 변환하여 사용하였으며 DBSCAN 클러스터링 시에는 반경의 영향을 줄이기 위해 표준화하지 않고 데이터를 사용하였다. 최적의 이상치 제거와 소 군집 마련을 위해 DBSCAN의 하이퍼파라미터인 반경은 0.001로 설정하였으며, 이때 이상치는 7,408건 중 165건으로 나타났다. 1차 클러스터링 결과 마포구 내에서 총 62건의 소 군집이 마련되었고, 반경 하이퍼파라미터에 따른 이상치와 군집 수는 표 Table 5와 같다.

Table 5. Outliers and Clusters by Hyperparameter

	Outliers	Number of clusters
0.0008	301	106
0.0009	216	88
0.001	165	62
0.0015	44	19
0.002	18	6
0.003	6	1
0.004	5	1

반경이 0.001을 미만일수록, 이상치와 군집 수가 점차 늘어나 제대로 군집을 형성하지 못하였다. 반면 0.001을 초과할수록, 이상치와 군집수는 줄었으나 군집 하나에 과도하게 분포하였다. 따라서 반경 파라미터는 0.001로, 최소점은 5로 설정하여 1차 클러스터링을 진행하였다. 결과적으로 62개의 소 군집이 형성되었고 이 중에서 300개 이상의 견인 지역을 포함하는 군집은 3개로, 각각의 군집은 4,189건, 1,137건, 349건의 견인 지역으로 구성되었다. 이 중 4,189건을 포함하는 군집은 K -means++를 적용하기에 밀도가 클 것으로 판단하였다. K -means++의 경우는 밀도가 높은 데이터 분석에 효율적이지 못하므로, DBSCAN을 반복 시행하였고 같은 과정을 통해 반경 0.001을 택하여 2,058건, 768건의 견인 지역을 포함하는 군집 2개로 나누어졌다. 이어 유의미한 4개의 소 군집을 대상으로 K -means++ 알고리즘을 적용하였고, 이때 반경을 설정할 필요성이 없으므로 알고리즘의 성능을 향상시키기 위해 좌표 데이터를 표준화하여 전처리하였다. 하이퍼파라미터 K 는 군집 수와 SSE의 관계를 보여주는 엘보우 기법과, 클러스터링 평가 척도인 실루엣 계수를 활용하여 설정하였다. 이에 따른 진행 결과는 Fig. 5와 Fig. 6와 같다.

클러스터링에서 군집 내 유사성은 높고 군집 간 유사성

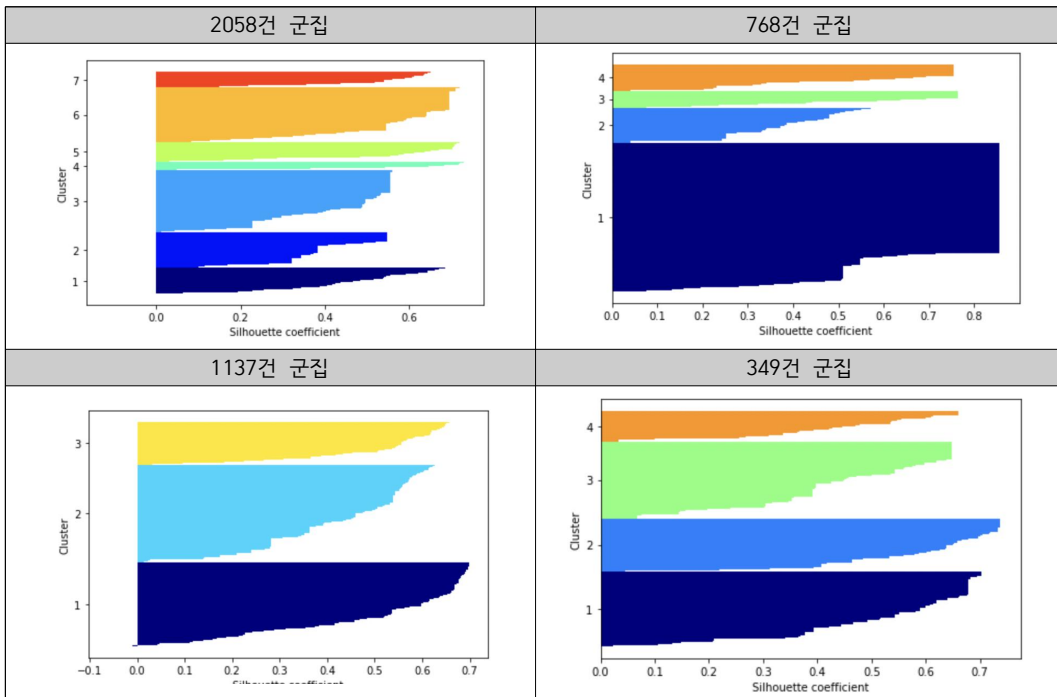


Fig. 5. Silhouette coefficient

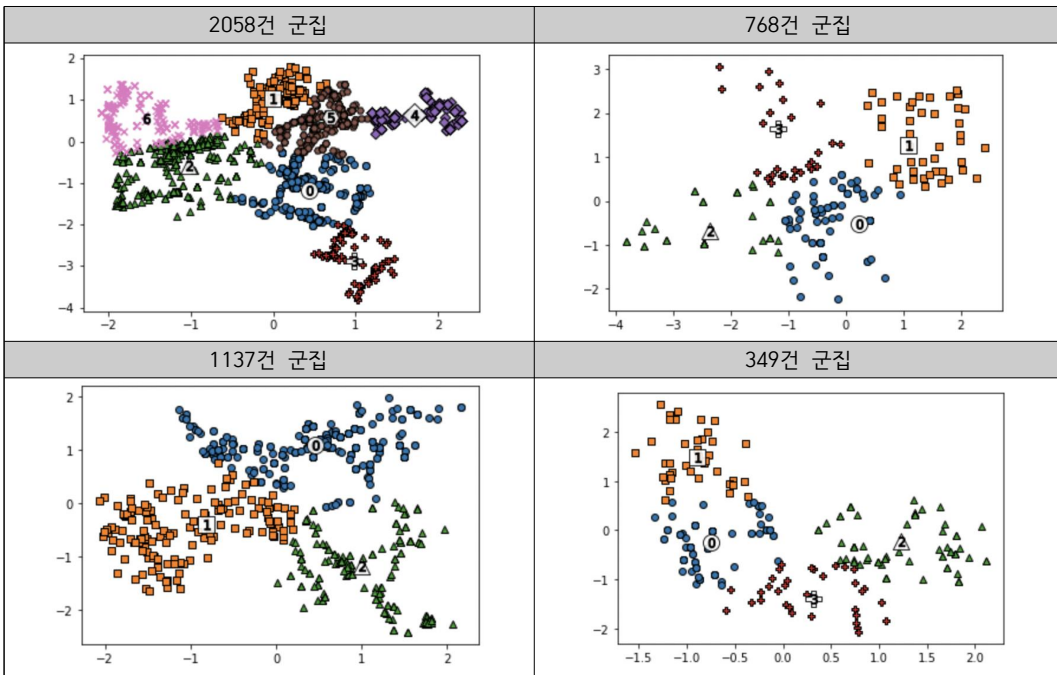


Fig. 6. Clusters and central points

은 낮을수록 좋다고 평가할 수 있는데[42], 이때 실루엣 계수가 1에 가까울수록 잘 군집화 되었다고 판단할 수 있다. 마포구 소재의 4개 소 군집의 경우는 엘보우 기법을 사용하여, 군집의 개수가 많은 순서대로 7개, 3개, 4개, 3개로 K를 설정했을 때 실루엣 계수가 가장 좋았다. Fig. 6.는 앞서 설명한 프로세스에 따라 K-means++를 적용한 결과로, 각 소 군집에서 마련한 클러스터의 중심점을 표시한 그림

이다.

결과적으로 4개의 소 군집에서 18개의 스테이션을 마련하였다. Fig. 7.은 기존 마포구의 주차 스테이션과 알고리즘으로 마련한 18개 스테이션을 비교한 결과이다. 초록색 마커는 기존 스테이션을, 빨간색 마커는 알고리즘을 적용한 스테이션을 의미한다. 마포구의 기존 스테이션은 주로 수요 측면만을 고려하여 지하철역 인근에 밀집하여 분포

해 있었다. 그러나 본 연구의 알고리즘을 적용하여 스테이션을 추가하면 기존과 달리 스테이션이 부족한 지역 및 건인 수량이 많은 지역을 중심으로 새로운 스테이션 위치가 선정되었음을 확인할 수 있다.

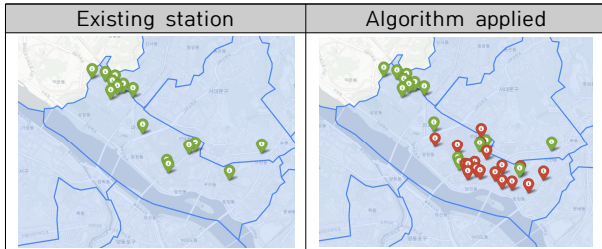


Fig. 7. Comparing Existing station with Station with algorithm applied

4-1. Detailed Experiment result

앞선 실험에서 2021.07.15 ~ 2022.07.31 기간동안의 마포구의 공유 전동 킥보드 건인 데이터와 혼합형 클러스터링을 사용하여 공유 전동 킥보드 불법 주차 해결을 위한 스테이션을 마련했다. 이와 함께 불법 주정차 공유 킥보드 건인이 마포구와 동시 시행되었던 나머지 5개의 자치구인 송파구, 영등포구, 성동구, 동작구, 도봉구를 대상으로도 스테이션 위치 선정 알고리즘을 적용하여 스테이션을 마련하였다. 진행 과정은 앞서 언급한 마포구의 경우와 동일하며, 선정된 스테이션 위치는 아래 표 Table 6.와 같다.

Table 6. Number of selected stations by district

Administrative district	Selected stations
Mapo-gu	18
Songpa-gu	11
Yeongdeungpo-gu	15
Seongdong-gu	10
Dongjak-gu	10
Dobong-gu	5

서울시 건인 데이터 중 6,492건을 차지하는 송파구의 경우는 11개, 5,381건을 차지하는 영등포구의 경우는 15개, 4,119건을 차지하는 성동구의 경우는 10개, 3,949건을 차지하는 동작구는 10개, 1,387건을 차지하는 도봉구의 경우는 5개의 스테이션 위치가 새롭게 선정되었다.

일부 지역에서는 공유 전동 킥보드 전용 주차장이 없거나 명확하지 않아 기존 스테이션과의 비교를 진행할 수 없었다. 새롭게 마련된 스테이션은 Fig. 8.에 제시된 바와 같다.

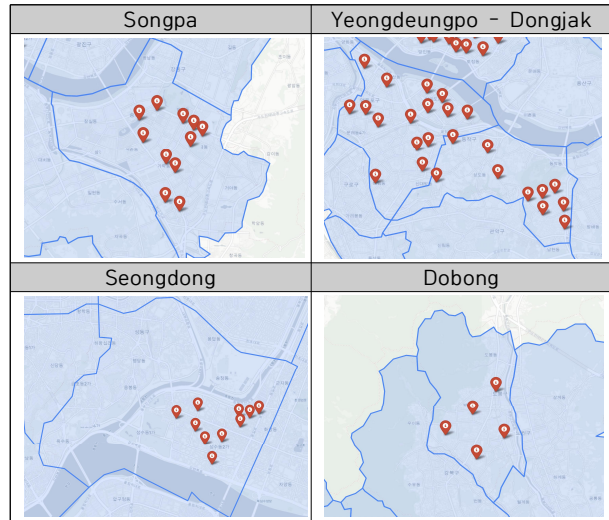


Fig. 8. Selected stations by administrative district

동작구의 경우는 마포구와 같이 공식적으로 기존 주차 구역 정보를 찾을 수 있어, 알고리즘 적용 전과 후를 비교할 수 있었다. 알고리즘 적용을 통해, 사당동 인근과 일부동에 추가 스테이션 위치가 선정된 것을 확인할 수 있다. 기존 스테이션과 알고리즘을 적용한 스테이션의 위치가 상이했던 마포구와 달리 동작구의 경우는 새 스테이션이 기존 스테이션 인근에 분포하고 있었다. 마포구의 건인 개수가 동작구보다 2배 이상 많다는 점에서 추정해 봤을 때 이는 마포구의 주차 구역이 수거 측면에서 비효율적이라는 것을 의미한다. 따라서 본 알고리즘을 통해 적절한 스테이션 위치가 선정되었다는 것을 확인할 수 있다.

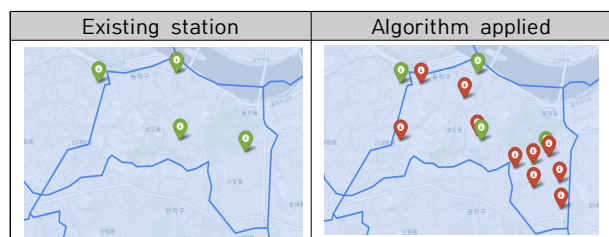


Fig. 9. Comparing Existing station with Station with algorithm applied in Dongjak

V. Conclusions

본 연구에서는 ‘독리스(Dockless)’ 방식으로 운영되는 공유 전동 킥보드에 관한 이슈를 LDA 토픽 모델링으로 파악하고, 혼합형 클러스터링으로 해결 방안을 제시했다. LDA 토픽 모델링 결과 보행자 입장에서는 ‘사고’, ‘이용행태’, ‘방치와 충전’이 주요 토픽으로 추출되었고, 이용자 입장에서는 ‘애플리케이션’, ‘주차(반납)’, ‘위치’가 주요 토픽

픽으로 추출되었다. 추출한 토픽을 분석한 결과 전동 키포드 방지와 키포드 주차 위치에 관한 문제를 파악할 수 있었다. 공유 전동 키포드 방지 문제가 심화되자 서울시 6개 자치구에서는 2021년 7월 15일부터 불법 주정차 공유 전동 키포드 견인을 시행했다.

그러나 견인 제도는 보행자의 입장만을 고려한 것으로 업계에는 큰 부담이 되고 있어 공유 전동 키포드 산업 전체에 큰 타격이 되고 있다. 또한 이용자 입장에서는 GPS 인식 오차 때문에 주차 구역 인식이 모호해 서비스에 대한 신뢰가 감소하고 있다. 따라서 본 연구에서는 적절한 지역에 스테이션을 설치할 것을 제안하며 스테이션 위치 선정 방법으로 DBSCAN과 K-means++ 알고리즘의 혼합형인 DBS-K++ 알고리즘을 제안하였다. 해당 알고리즘으로 '서울특별시 전동 키포드 견인 현황' 데이터를 활용하여 견인 밀집 지역을 클러스터링하고 주차 스테이션 위치를 선정하였다. 그 결과 공유 키포드 불법 주정차 견인 제도를 시행하는 6개 자치구를 대상으로 전체 69개의 스테이션 설치 적정 지역을 마련했다. 그 중 견인 건 수가 가장 많은 마포구에서는 18개의 스테이션이 새롭게 생성되었다. 기존 공유 전동 키포드의 수요 측면에서 마련된 주차 구역이 특정 지하철역 인근에 포진되어 있었다면 새롭게 추가된 18개의 스테이션은 기존 스테이션과 달리 견인 다발 지역을 고려하여 견인 밀도가 높은 곳으로 선정되었다. 이를 통해 전동 키포드 무단 방치로 인한 안전 문제를 해결하고 보행자와 이용자의 불편을 동시에 해소할 수 있을 것이라 예상된다. 더 나아가 이용 종료된 공유 전동 키포드의 효율적인 위치 관리가 가능해져 견인으로 인한 사회적 비용과 업계의 부담을 줄일 수 있을 것이다.

본 연구의 한계점은 견인을 시행한 지난 1년간의 공개된 데이터를 기반으로 하고 있어 더 장기적인 관점에서 접근이 부족했다는 점이다. 또한 공유 키포드의 운영 주체가 서울시가 아닌 각 사업체에 있어 이용자들의 서비스 사용행태와 관련한 폭넓은 데이터 확보에 어려움이 있었다. 하지만 '독리스(Dockless)' 방식의 운영으로 보행자 및 이용자 모두의 불편이 심화되어 사업의 지속에도 악영향을 미칠 우려가 있으므로 기존 운영 방식의 개선을 위한 다양한 후속 연구의 필요성이 있다. 견인 제도가 정착되고 데이터가 더 개방되어 연구의 한계점을 극복할 수 있다면, 향후 연구에서는 GIS 공간분석 방법론을 적용할 수 있을 것이다. 다양한 공간분석 방법론을 통해 지역의 특성 또한 고려하여 더 정교한 스테이션 위치 선정 알고리즘을 마련할 수 있으리라 기대한다.

REFERENCES

- [1] Naver, A Topical Dictionary, <https://terms.naver.com/>.
- [2] Nielsen KoreanClick, Monthly Topic, http://www.koreanclick.com/insights/newsletter_view.html?code=topic&id=599&page=1&utm_so.
- [3] Lee Jeongheun, Global companies packing in Korea in one year of strengthening regulations on electric scooters, <https://magazine.hankyung.com/business/article/202207061571b>.
- [4] Shin Huicheol, "Activation Plan for Future Personal Mobility," Monthly KOTI Magazine on Transport, Vol.268, pp. 31-37, Jun. 2020.
- [5] Baek Bongsam, An e-scooter accident... 9 out of 10 'have safety concerns', <https://zdnet.co.kr/view/?no=20201029140823>.
- [6] Korea Transportation Research Institute, "What is the solution to shared e-scooter parking?," Monthly KOTI Magazine on Transport, Vol.268, pp. 77-77, Jun. 2020.
- [7] Jo Jaehak, 'As civil complaints increase, processing costs rise' What should I do with a shared e-scooter?, <https://www.etnews.com/20220830000059>.
- [8] Bak Sunyeop, I came here after looking at the app but there is no e-scooter, <https://www.edaily.co.kr/news/read?newsId=01249686622554848&mediaCodeNo=257>.
- [9] Jo Jeonghun, Ham Seungu, Kim Donggyu, A Comparison study on Micro-mobility Usage Pattern : focusing on Bike-sharing service and E-Scooter share service in Seoul, Fall Conference of the Korean ITS, 237-242, Jeju, Korea, Nov. 2020.
- [10] Su jae Kim, Gyeong jae Lee, Sangho Choo, Sang hun Kim, "Study on Shared E-scooter Usage Characteristics and Influencing Factors," The Journal of The Korea Institute of Intelligent Transport Systems, Vol.20, No.1, pp.40-53, Feb. 2021.
- [11] Jang Eun-Jin, Shin Seung-Jung, "Design of a New IoT Management System for Efficient Recovery of Shared Electric Scooters," The Journal of the Institute of Internet Broadcasting and Communication, Vol.21, No.1, pp.189-194, Feb. 2021.
- [12] Choi Seonil, Kim Gyurin, Jeong Taehyeok, Shin Huijun, Hwang Byeonguk, Choi Hyeonjin, Modular Autonomous Parking System for Shared Scooters, ICROS Annual Conference, 693-694, Geoje, Korea, Jun. 2022.
- [13] Choi HyeJung, Jang Seongman, "A Study on the Demand-Driven Layout Method of Shared Personal Mobility to Increase Public Transportation Utilization," Journal of the Korean Cadastre Information Association, Vol.24, No.1, pp.126-137, Apr. 2022.
- [14] Han Jae-Won, Ahn Soo-Yeong, Kim Mi-Sung, Han Dong-Hyung, Lee Sugie, "Analysis of Priority Supply Areas for Personal Mobility in Seoul: Focusing on First-Last Mile Traffic Characteristics," Journal of Korea Planning Association, Vol.57, No.1, pp.42-56, Feb. 2022.

- [15] Kurniawan, Dwi Ely, A. G. U. S. Fatulloh, "Clustering of Social Conditions in Batam, Indonesia Using K-Means Algorithm and Geographic Information System," *International Journal of Earth Sciences and Engineering (IJEE)*, Vol.10, No.5, pp.1076-1080, Aug. 2017.
- [16] Annas, Suwardi, Uca, U., Irwan, I., Safei, R. H., & Raisl, "Using k-Means and Self Organizing Maps in Clustering Air Pollution Distribution in Makassar City, Indonesia," *Jambura Journal of Mathematics*, Vol.4, No.1, pp.167-176, Aug. 2022.
- [17] Lee Byeonggil, "Commercial district extraction using spatial data mining technique in business GIS," *Korean GIS Journal*, Vol.11, No.2, pp.171-184, Jul. 2003.
- [18] David M. Blei, Andrew Y. Ng, Michael I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, Vol.3, pp.993-1022, 2003.
- [19] Oh Jinsong, Paris considers e-scooter rental ban amid successive deaths, <https://www.yna.co.kr/view/AKR20221118061500009>.
- [20] Ju Yeongjae, Shared E-scooter 'War' in Gangnam, <https://www.khan.co.kr/economy/industry-trade/article/201908260927001>.
- [21] Seoul Official Blog, Illegal parking and sharing e-scooter tow is implemented, <https://blog.naver.com/haechiseoul/222440751201>.
- [22] Oh Areum, The e-scooter industry appeals for the burden of 'immediate tow' 450 million won for tow and storage only, http://autotimes.hankyung.com/apps/news.sub_view?nkey=202110251020431.
- [23] Lee Gibeom, Strengthening shared e-scooter regulations: 100 million monthly tow fee 'Cost is passed on to the consumer', <https://www.news1.kr/articles/?4426171>.
- [24] Seoul, Seoul Population Density Statistics(2021), <http://data.seoul.go.kr/dataList/10790/S/2/datasetView.do>.
- [25] Seoul, Seoul Living Migration, <https://data.seoul.go.kr/livPopu/html/dashboard.html>.
- [26] Ester, M., Kriegel, H. P., Sander, J., Xu, X., "A density-based algorithm for discovering clusters in large spatial databases with noise," In *kdd*, Vol.96, No.34, pp.226-231, Aug. 1996. DOI: 10.5555/3001460.3001507
- [27] Kremers, B. J., Ho, A., Citrin, J., & van de Plassche, K. L., "Two step clustering for data reduction combining DBSCAN and k-means clustering," *arXiv preprint arXiv:2111.12559*, Nov. 2021. DOI: 10.48550/arXiv.2111.12559
- [28] Murugesan, N., Cho, I., & Tortora, C., "Benchmarking in cluster analysis: a study on spectral clustering, DBSCAN, and K-Means," In *Data Analysis and Rationality in a Complex World*, Vol.16, pp.175-185, Feb. 2021. DOI: 10.1007/978-3-030-60104-1_20
- [29] Choi Eunseok, Kim Jeonghun, A Jijeu, Lee Sanghyeon, Gang Jeongtae, & Ryu Gwanhui, "DBSCAN-based manufacturing process data defect location detection," *Journal of Korean Contents Society*, Vol.17, No.7, pp.182-192. Jul. 2017.
- [30] Wang, K., Yang, R., Liu, C., Samarasinghalage, T., Zang, Y., "Extracting Electricity Patterns from High-dimensional Data: A comparison of K-Means and DBSCAN algorithms," In *IOP Conference Series: Earth and Environmental Science*, Vol.1101, No.2, pp.022007, Nov. 2022. DOI: 10.1088/1755-1315/1101/2/022007
- [31] Vassilvitskii, S., Arthur, D., k-means++: The advantages of careful seeding, In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pp.1027-1035, Jun. 2006.
- [32] MacQueen, J., Classification and analysis of multivariate observations, In *5th Berkeley Symp. Math. Statist. Probability*, pp.281-297, 1967.
- [33] Savvas, I. K., Stogiannos, A., Mazis, I. T., "A study of comparative clustering of EU countries using the DBSCAN and k-means techniques within the theoretical framework of systemic geopolitical analysis," *Int. J. Grid Util. Comput.*, Vol.8, No.2, pp.94-108. Aug. 2017. DOI: 10.1504/IJGUC.2017.10006818
- [34] Ran, X., Zhou, X., Lei, M., Tepsan, W., Deng, W., "A novel K-means clustering algorithm with a noise algorithm for capturing urban hotspots," *Applied Sciences*, Vol.11, No.23, pp.11202, Nov. 2021. DOI: 10.3390/app112311202
- [35] Kotyrba, M., Volná, E., Komínková Oplatková, Z., Comparison of modern clustering algorithms for twodimensional data, In *Proceedings-28th European Conference on Modelling and Simulation, ECMS*, pp.346-351, May. 2014. DOI:10.7148/2014-0346
- [36] Limwattanapibool, O., Arch-int, S., "Determination of the appropriate parameters for K-means clustering using selection of region clusters based on density DBSCAN (SRCD-DBSCAN)," *Expert Systems*, Vol.34, No.3 e12204, May. 2017. DOI: 10.1111/exsy.12204
- [37] Kim Taegyeong, Chung Jin-Hyuk, "A Study on the Identification of Accident Hot Spots Using DBSCAN - Focused on Gyeong-Bu Expressway," *Journal of Transport Research*, Vol.21, No.3, pp.55-63, Sep. 2014. 10.34143/jtr.2014.21.3.55
- [38] Hong Jun Cho, Yu Jeong An, Ju Hee Lee, Young Min Kim, "Analysis on a selection criterion of land transaction cases using K-means," *Journal of the Korean Data And Information Science Society*, Vol.30, No.3, pp.525-537, 2019. DOI: 10.7465/jkdi.2019.30.3.525
- [39] Fahim, A., "K and starting means for k-means algorithm," *Journal of Computational Science*, Vol.55, Oct. 2021. DOI: 10.1016/j.jocs.2021.101445
- [40] Fan, T., Guo, N., Ren, Y., "Consumer clusters detection with geo-tagged social network data using DBSCAN algorithm: a case study of the Pearl River Delta in China," *GeoJournal*, Vol.86, No.1, pp.317-337, Sep. 2019. DOI: 10.1007/s10708-019-10072-8
- [41] Seong-Kyu Yi, Sung-Eon Hong, Soo-Hong Park, "A Similar Price

Zone Determination of Public Land Price Using a Hybrid Clustering Technique,” Journal of the Korean Geographical Society, Vol.41, No.1, pp.121-135, Mar. 2006.

- [42] Rousseeuw, P. J., “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” Journal of computational and applied mathematics, Vol.20, pp.53-65, Nov. 1987. DOI: 10.1016/0377-0427(87)90125-7

Authors



Jeongeun Song received the B.S. degree in Korean Language and Literature from Myongji University, Seoul, Korea in 2010. She is currently pursuing the M.S. degree in Graduate School of information, Yonsei

University, Seoul, Korea. Jeongeun Song is interested in deep learning, NLP.



Yoon-Ah Song received the B.S. degree in Economics and Finance from Sangmyung University, Seoul, Korea in 2022. Yoon-Ah Song is currently a student of Business Analytics at Graduate School of Information,

Yonsei University, Korea, for M.S. degree. Yoon-Ah Song is interested in Reinforcement Learning, Deep Learning and Finance.



ZoonKy Lee received the B.S. degree in Computer Science from Seoul National University, Korea in 1985, M.S. in Social Psychology from Carnegie Mellon University, USA in 1991, and Ph.D in Management

Informatics from Southern California University, USA in 1999. He is currently a Professor in the Graduate School of Information at Yonsei University, Korea. ZoonKy Lee is interested in Big Data Analytics, Digital Transformation, and Open Collaboration.