

ARTICLE

공유형 자율차의 이동 경로 예측에 관한 기초연구: AI 분석 기반의 택시 경로와 도시 빅데이터를 이용하여

김건욱¹ · 김정화^{2*} · 김우진³ · 이승현⁴

¹(재)대구디지털산업진흥원 빅데이터활용센터 센터장, ²경기대학교 스마트시티공학부 도시·교통공학전공 조교수, ³경기대학교 스마트시티공학부 도시·교통공학전공 학부과정, ⁴대구 오픈스퀘어-D 매니저

Preliminary Study for Exploring SAV Route Using Taxi DTG and Urban Big-Data with Learning Based Approach in AI

KIM, Keunwook¹  · KIM, Junghwa^{2*}  · KIM, Woojin³  · LEE, Seunghyeon⁴ 

¹Director, Bigdata Center, Daegu Digital Industry Promotion Agency, Daegu 42250, Korea

²Assistant Professor, Department of Urban and Transportation Engineering, Kyonggi University, Gyeonggi 16227, Korea

³Student, Department of Urban and Transportation Engineering, Kyonggi University, Gyeonggi 16227, Korea

⁴Manager, Open Squire-D, Daegu Center, Daegu 42250, Korea

*Corresponding author: junghwa.kim@kyonggi.ac.kr

Abstract

This study analyzed the main travel routes of shared autonomous vehicles by using DTG (Digital Tachograph) data of Daegu Metropolitan City under the premise that the driving mode of the shared autonomous vehicle is similar to that of a taxi. Based on the taxi DTG data that recorded the driving route from the taxi user's entry point to the exit point, the ride rate analysis result was visualized as a heat map. The main travel routes of taxi users were found through the driving route. As a result of the loading rate analysis, it was confirmed that the loading rate was high in Dongseong-ro, the downtown area of Daegu Metropolitan City, and the nearby area of Dongdaegu Station. The following high passenger rate region was analyzed as the region with a high proportion of commercial facilities. In the case of movement routes, it was confirmed that they appear mainly along the same axis as subway lines and major arterial roads in the city. And through this, it was possible to obtain the implication that the infrastructure for shared autonomous vehicles could be built along arterial roads, auxiliary arterial roads, and subway lines in the future. In addition, this study established a regression model using machine learning techniques such as Random forest and LightGBM to confirm the variables that affect the user's choice to use a taxi. The distance to the subway station, the arterial road, the number of floating population, the card sales performance, and the distance to the bus stop were derived as the primary variables that affect the decision of the taxi route. As a result of model estimation, it was found that both regression models secured the fit. It is expected that the results of this study can be utilized to strategically arrange essential infrastructure related to the driving of shared autonomous vehicles in consideration of changes in these influencing variables.

Keywords: autonomous vehicle, big data, machine learning, shared mobility, spatial analysis

J. Korean Soc. Transp.
Vol.39, No.6, pp.766-779, December 2021
<https://doi.org/10.7470/jkst.2021.39.6.766>

pISSN : 1229-1366
eISSN : 2234-4217

ARTICLE HISTORY

Received: 18 October 2021

Revised: 25 October 2021

Accepted: 5 November 2021

Copyright ©
Korean Society of Transportation

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.

초록

본 연구에서는 대구광역시를 대상으로 택시 DTG(Digital Tachograph) 데이터를 활용하여 공유형 자율주행차의 주행 형태가 택시와 유사하다는 대전제를 두고 공유형 자율주행차의 주요 이동 경로를 예측하기 위한 기초연구를 수행하였다. 택시 이용자의 승차지점부터 하차지점까지의 주행 경로가 기록된 DTG 데이터를 바탕으로 승차율 분석결과를 히트맵으로 시각화하였으며, 주행 경로를 통해 택시의 주요 이동 경로를 확인하였다. 승차율 분석결과 대구광역시 도심인 동성로, 부도심인 동대구역의 승차율이 높게 나타났으며 그다음으로 승차율이 높은 지역은 상업시설의 비중이 높은 지역으로 확인되었다. 분석결과 지하철 노선과 도시 내 주간선도로와 유사한 축으로 나타나는 경향을 확인하였으며 이를 통해 향후 간선도로와 보조 간선도로, 지하철 노선을 따라 공유형 자율주행차를 위한 인프라가 구축 될 수 있다는 시사점을 얻을 수 있었다. 또한 머신러닝 기법인 랜덤포레스트와 LightGBM를 활용한 회귀모형을 구축하여 택시 이동 경로 결정에 영향을 주는 인자를 분석하였다. 모형 추정결과 2개의 회귀모형 모두 적합성을 확보한 것으로 나타났으며, 지하철역까지의 거리, 간선도로, 유동인구 수, 카드매출 실적, 버스정류장까지의 거리가 주요 변수로 도출되었다. 장래 이와 같은 영향변수들의 변화를 고려하여 공유형 자율주행차의 주행과 관련된 필수 인프라를 전략적으로 배치하는 데에 본 연구의 결과가 활용 될 수 있을 것으로 기대한다.

주요어: 자율주행차, 빅데이터, 머신러닝, 공유형 모빌리티, 공간분석

서론

자율주행 시대에 진입하면 공유형 자율주행자동차(Shared Autonomous Vehicles, SAV)의 시장이 확대될 것으로 예견되고 있다. 공유형 자율주행차 운행에 있어 도시 교통시스템, 인프라 등에 변화를 가져올 것으로 예상하며 이는 도시의 이동 행태를 변화시킬 것이다. 프란시스코 당국은 「City of San Francisco: Meeting the Smart City Challenge」 통해 통행수단이 전통적인 소유의 형태에서 공유형 자율주행차로 변화할 것을 예측하였으며, SECAV(Shared, Electric, Connected and Automated Vehicles) 모델을 제시함으로써 도로 공간의 인프라 대비를 준비하고 있다(SFMTA, 2016). 샌프란시스코 외 독일의 베를린, 영국의 런던, 미국의 뉴욕은 공유형 자율주행차 주행에 적합한 도로를 선정하였으며, 이에 대한 인프라 구축 계획을 설계하였다(Rothnie et al., 2016). 공유형 자율주행차에 대한 대비를 국외에서 진행하고 있는 상황을 고려하여 국내에서도 공유형 자율주행차에 필요한 대비로 주차공간, 전용차로 등의 인프라 설계 계획이 필요하다.

본 연구에서는 도시 전체를 대상으로 공유형 자율주행차의 인프라를 한 번에 구축하기 어려우므로 효율적 운영이 가능할 수 있도록 주요 이동 경로를 분석하고 인프라 구축의 대상이 될 수 있는 도로망을 검토하고자 한다. 또한, 주요 도로축 선정 외에도 공유형 자율주행차의 수요가 있는 토지이용 지점을 파악하여 토지의 활용도를 분석하고자 한다. 위의 결과를 도출하기 위하여 현재 택시와 SAV의 주행경로, 그에 따른 주요 이용 경로가 유사 할 것이라는 대전제를 두고 연구를 진행하며, 이에 따라 택시의 DTG(Digital TachoGraph) 데이터를 활용하여 주요 노선축과 택시 수요에 있어 영향을 주는 토지 이용 관련 변수를 머신러닝을 활용하여 분석한다.

선행연구 검토

공유형 자율주행차의 도입이 향후 인프라, 도로 용량 등에 어떠한 변화를 가져올지에 대해 예측하는 연구가 이뤄지고 있다. 그리하여 변화에 대응하기 위해 어떤 점이 변화할지 예측하는 연구와 예측된 연구결과를 바탕으로 시뮬레이션을 진행한 연구를 조사해볼 필요가 있다. 변화를 예측하는 연구로써 Milakis et al.(2017)은 자율주행차의 등장으로 단계별 변화가 있을 것으로 예측하였다. 1단계에서는 도로 용량과 여행비용 등이 변화되며, 2단계에서는 자율주행차를 소유할 것인지, 공유할 것인지에 대한 선택과 토지의 이용에 대한 변화를 예측하였다. 마지막 3단계 변

화에서는 공유형 자율주행차로 주차공간 감소 등 교통인프라에 변화가 발생할 것으로 예측하였다. 이러한 결과는 공유형 자율주행차의 도입은 교통인프라에 변화가 발생할 것을 내포하고 있으며, 공유형 자율주행차 주행에 적합한 교통인프라가 필요하다는 점을 알 수 있다. 예측된 결과를 바탕으로 Fagnant et al.(2015)은 텍사스 오스틴을 대상으로 하여 공유형 자율주행차 운영을 MATsim 프로그램을 통해 시뮬레이션을 진행하였다. 시뮬레이션 결과 공유형 자율주행차 이용자의 평균대기 시간은 5분 미만이었으며, 공유형 자율주행차 1대는 일반 차량 9대를 대체할 수 있다고 분석되어, 이는 도로 용량 감소에 효과가 있음을 의미한다. 또한, 공유형 자율주행차가 빈 상태일 때 다음 이용에 대해 예측을 하여 이용자가 곧바로 이용할 수 있도록 위치를 이동할 때 약 8% 더 많은 VMT를 생성하였다는 결과가 산출되었다.

앞선 선행연구검토 결과 단계별 변화 중 3단계에서 교통인프라에 대한 변화가 이뤄질 것으로 예측되고 있다는 점을 확인하였다. 따라서 이에 대해 공유형 자율주행차와 관련된 인프라 선행연구를 조사하였다. 공유형 자율주행차를 다양한 관점에서 분석한 Narayanan et al.(2020)은 공유형 자율주행차 관련 연구를 교통 및 안전, 여행특성, 경제, 운송, 토지 이용 등의 초점에 맞춰 검토를 진행하였다. 이 중 인프라와 관련되어 분석한 내용은 토지 이용 초점에 맞춰져 있었으며, 대부분의 연구는 주차공간과 관련된 연구인 것을 확인하였다. 주차공간과 관련된 대부분의 연구는 공유형 자율주행차 주행으로 인하여 주차공간이 감소할 것으로 예측하였다. 검토한 논문 중 Boesch et al.(2016)은 공유형 자율주행차의 도입 비율에 따라 시나리오를 작성하여 주차공간의 감소 비율을 예측하였으며 Spieser et al.(2014), Zhang et al.(2015)은 시뮬레이션을 진행하여 주차공간을 얼마나 감소시킬 수 있는지 분석하였다. 해당 선행연구 검토결과 완전자율주행차 시대의 공유형 자율주행차 100%일 경우 주차공간을 66%에서 93%까지 감소할 수 있다는 예측 결과를 확인하였으며, 이를 통해 공유형 자율주행차로 인해 교통인프라에는 주차공간이 가장 큰 영향을 받으며 주차공간을 다른 용도로의 활용 가능성이 있음을 알 수 있다. 주차 공간 뿐만 아니라 도시형태에도 공유형 자율주행차는 영향을 끼친다는 점을 연구한 Gelauff et al.(2017), Soteropoulos et al.(2018)은 공유형 자율주행차의 도입으로 도시와 비도시 간의 주택가격 격차가 좁혀지고, 도시화를 앞당긴다고 예측하였다. 그러나 Zhang and Guhathakurta(2018)은 공유형 자율주행차의 시뮬레이션 모델과 주택위치 선택 모델을 결합한 연구를 통하여, 앞의 연구들과는 조금 다른 예측 결과인 공유형 자율주행차의 도입은 도시화를 빠르게 앞당기는 스프롤 현상으로 이어지지 않는다고 하였다.

공유형 자율주행차를 운행하기 위해서 차량 충전은 필수사항이다. 인프라에 속하는 충전소는 공유형 자율주행차 시대에 필요하며 충전소의 유무, 위치, 성능에 따른 연구를 조사해볼 필요가 있다. Chen et al.(2016)은 공유형 전기자율주행차에 대한 연구를 진행하였으며, 공유형 전기자율주행차와 충전소의 유형, 위치에 따른 시나리오를 생성하고 시뮬레이션을 진행하여 비용적인 측면이 가장 적합한 시나리오를 도출하였다. 충전소는 충전속도에 따라 레벨을 구분하였으며 레벨2(240V의 충전속도)의 충전소를 중심으로 80마일 범위에서는 공유형 전기자율주행차가 일반 차량 3.7대를 대체하며 레벨3(480V의 충전속도)의 충전소를 중심으로 80마일 범위에서는 공유형 전기자율주행차가 일반 차량 5.4대를 대체하는 결과를 산출하였다. 또한, 공유형 전기자율주행차의 이용을 평균 7-10분 이내에 96-98% 이용할 수 있다는 결과를 확인하였다. 이 연구에서 공유형 자율주행차의 충전소의 레벨과 위치에 따라 대체할 수 있는 일반 차량의 대수가 변화한다는 점에 있어 충전소의 성능과 위치는 인프라 구축에 있어 중요한 요소인 점 또한 확인하였다. Zhang et al.(2020)은 앞서 텍사스 오스틴을 대상으로 한 시뮬레이션 연구를 대상으로 하여 주차공간, 충전소 등의 인프라를 포함한 다양한 모달 옵션을 비용과 시간을 최소화하는 알고리즘을 이용한 Behavior, Energy, Autonomy, and Mobility(BEAM) 시뮬레이션을 진행하여 공유형 전기자율주행차 충전소 위치를 충전속도, 충전용량별로 위치를 제시하였다.

본 연구에서 활용하고자 하는 데이터는 택시 DTG 데이터로 선행연구에서는 어떤 방식을 사용했으며, 어떤 결과를 도출하였는지 알아볼 필요가 있다. Cho et al.(2017)은 화물차량의 DTG 데이터를 활용하여 고속도로에서 졸음 운전으로 인한 위험 구간을 분석하였다. 전체 사고 건수, 연평균 일교통량, 평균 과속비율 등의 정보를 읍이항 회귀

모형을 이용하여 졸음운전사고 위험구간을 상위 5%, 10% 구간으로 나눠 결과를 도출하였다. Kwon and Kim(2018)은 택시 DTG 데이터를 통해 산출한 주요 도로와 현재 서울시의 주요 도로를 비교 분석하여 기존 주요 도로축을 확인하는 연구를 진행하였다. DTG 데이터 가공 후 경로별 통과차량수를 산출하고, 통과차량수가 많은 경로를 나열하여 서울시 주요 도로와 대부분 일치하는 결과가 도출되었다. DTG 데이터 활용 선행연구를 통하여 분석을 통해 이동 경로와 주요 이동 경로, DTG 데이터와 여러 변수를 활용한 회귀분석을 통해 변수의 영향력을 산출 가능한 것을 확인하였다.

택시의 시·공간적 통행특성 분석

1. 데이터의 수집

본 연구는 택시 승하차 통행 수요를 분석하기 위해 공간적 범위는 대구광역시로 설정하였으며, 시간적 범위는 2020년 1월 1일부터 2020년 4월 23일까지의 DTG 데이터 6,301,369건을 연구목적으로 수집하여 분석에 활용하였다. 또한, 택시가 이동하는 주요 경로와 토지이용 특성과의 관련성을 살펴보기 위해 공공데이터로는 용도별 건물정보, 대중교통 위치 정보를 이용하였으며, 민간 데이터로는 대구 빅데이터활용센터가 제공하고 있는 통신사 유동인구와 카드 매출 실적 자료를 적용하였다. 그리고 도로 네트워크의 경우 파이썬 기반의 공간분석을 위해 오픈소스 맵인 OpenStreetMap을 활용하였다.

Table 1에는 데이터의 세부적인 내용을 담고 있으며, DTG 데이터의 경우 통행 건별로 운행 시간대와 운행 경로, 택시 영업정보 등이 포함되어 있으며, 공차를 제외한 실차의 운행정보만을 포함하고 있다. 그리고 카드 매출과 통신사 유동인구 데이터의 경우 시간대별, 연령별, 구역별 유동인구 정보 등이 포함되어 있다.

Table 1. Data structure

Taxi boarding and alighting (2020.01-2020.04)	- Taxi departure and arrival times - Taxi departure and arrival coordinates - Taxi route
Public transport location information	- Subway station name · coordinates - Bus stop name · coordinates
Card sales performance (2020.01-2020.04)	- Card sales by time period - Card sales by time age
Building information by purpose	- Classification codes by building purpose (number of floors, height, location, form)
Road network	- Road type, Name, Length, Bridge, One-way traffic
Telecommunication company floating population (2020.01-2020.04)	- Floating population traffic by hour - Floating population traffic by age

2. 데이터 전처리

수집된 데이터는 본 연구의 분석 목적에 맞게 전처리 작업이 필요하며, 이에 대해 행정안전부가 빅데이터 분석을 수행할 수 있도록 제시한 공공빅데이터 표준분석모델 매뉴얼(Ministry of the Interior and Safety, 2019)을 기반으로 하여 데이터 전처리를 진행하였다. 세부적인 내용은 다음과 같다.

첫째, 출발, 도착 좌표가 공간적 범위인 대구광역시를 벗어나는 이상치(Outlier)가 일부 존재하여 229,274건의 데이터를 제거하였다. 둘째, DTG 데이터의 경우 실제 운행한 궤적 정보가 포함되어 있으나, 위경도 좌표 오차가 일부 존재하고, 포인트를 경로로 연결시 기존 도로와의 조인을 수행함에 있어 한계가 존재하였다. 따라서 출발지와 도착지 정보를 기반으로 파이썬 라이브러리인 OSMnx를 활용하여 최단경로(Shortest Path) 알고리즘으로 운행 경로를 추정하였다. 셋째, 특정 시간대, 요일별 통행 특성이 상이하게 분석되어, 그룹내는 유사하고 그룹간은 이질적인

데이터를 구분하였으며, 세부적으로는 평일 첨두시간대, 평일 비첨두시간대 그리고 주말 첨두시간대, 주말 비첨두시간대의 4가지 데이터로 구성하였다. 넷째, 다양한 데이터의 공간적 결합과 통신사 유동인구 데이터의 기지국 범위 등을 고려하여 가로 50M, 세로 50M의 격자를 생성하였으며, 용도별 건물, 도로 네트워크, 주요 이동 경로 자료와 위치에 의한 조인으로 속성 정보를 격자 내 결합하였다. 도로 네트워크의 경우 격자내 경로별 개수를 계산하였으며, 계산 방법은 Figure 1과 같다. 그리고 도로 유형 이외의 속성값은 누락 값의 비중이 높아 분석에 활용되기에는 부적합하다고 판단하여, 도로 유형의 값만 격자 내에 결합하였다. 다섯째, 대중교통 위치정보의 경우 격자 내 중심점에서 대중교통인 지하철, 버스정류장 위치까지의 유클리디안 거리(Euclidean distance)를 계산하여 대중교통과의 거리 자료를 생성하였으며, 산출 방법은 Figure 2와 같다. 마지막으로 카드 매출 실적과 통신사 유동인구 자료는 별도의 존 체계를 포함하고 있으며, 격자 내 한 개의 존이 포함될 경우 해당 값을 조인하였으며, 여러 개의 존이 포함될 경우는 존들의 평균값을 계산하여 데이터로 결합하였다.

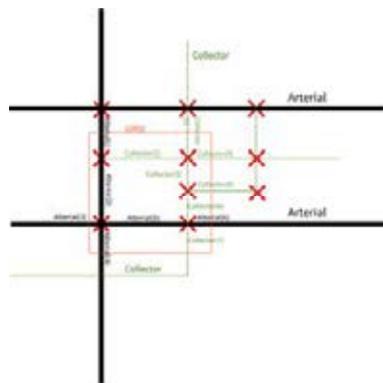


Figure 1. Calculation of the number of roads in the grid

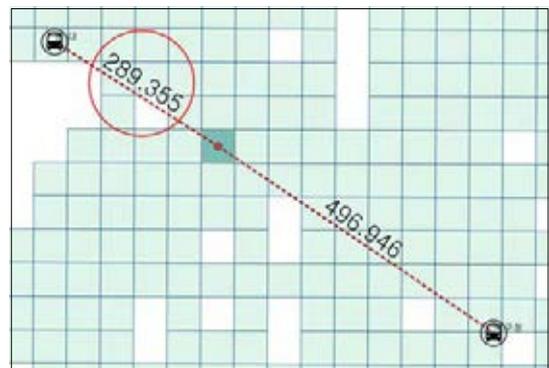


Figure 2. Calculation of distance from public transportation facilities

3. 택시의 통행특성 분석

1) 택시의 통행 패턴 분석

Figure 3은 전처리 된 택시 승하차 자료를 활용하여 시간대별 승차자 수를 나타냈다. 오후부터 심야시간까지 택시 이용자의 수요가 높은 첨두시간대로 나타났으며, 이는 일반 대중교통 이용 패턴과는 상이한 결과로 확인되었다.

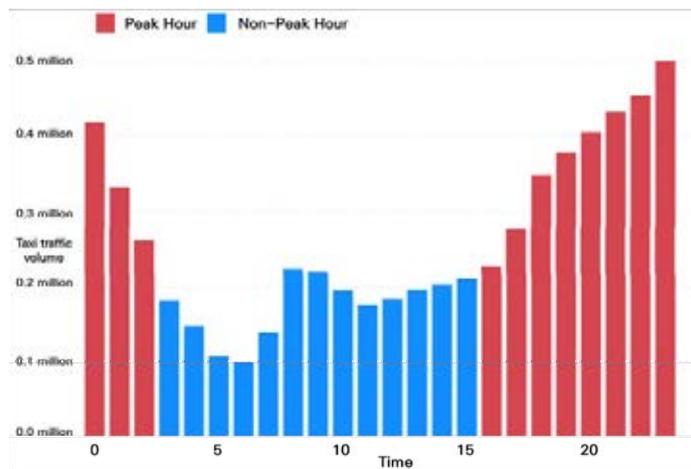


Figure 3. Taxi traffic volume by time of day

또한, 오전 08시에서 09시 사이에 일부 통행이 증가하나 일통행량에서 비중은 크게 높지 않은 것으로 나타났다. 첨두, 비첨두 시간대의 최소, 최대 기준을 살펴보면 첨두시간의 경우 시간당 최소 227,048건, 최대 498,790건으로 나타나며, 비첨두시간은 최소 98,221건, 최대 223,154건의 분포를 보인다. 이러한 분석결과를 통해 본 연구에서는 첨두시간을 16시부터 03시, 비첨두시간은 04시부터 15시까지로 정의하여 택시 이용자의 통행 특성을 구분하였다.

2) 택시의 요일별, 시간대별 통행 특성

요일별 시간대별 통행특성을 살펴보면 Figure 4와 같이 나타나며, 주말인 토요일과 일요일의 경우 평일과 대비하여 오전 08시에서 09시까지 택시 승차 통행량이 감소하며, 오후 첨두시간대의 경우 일요일이 다른 요일에 비해 상대적으로 낮게 나타났다. 이는 직장인 근로 환경에 따른 통행 차이인 점으로 보인다.

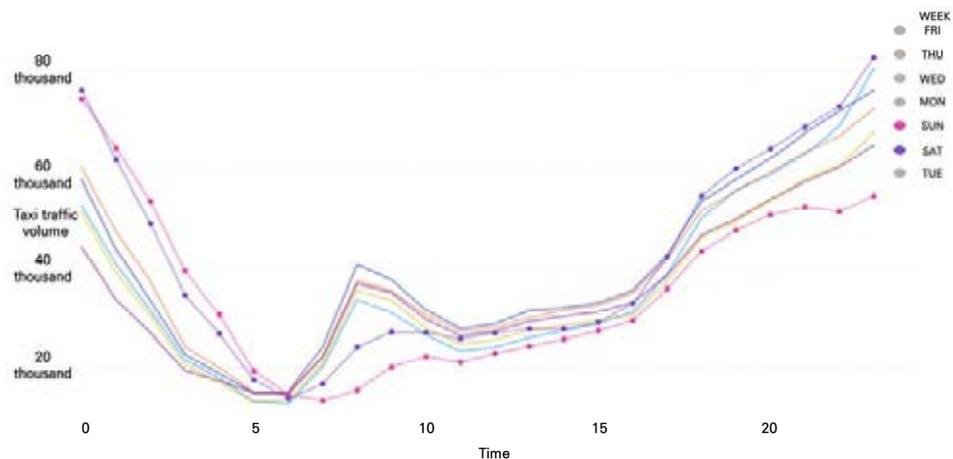


Figure 4. Taxi traffic volume by time of day for each day

3) 택시의 요일별 평균 운행시간

택시 승하차 통행 이력 자료의 경우 승차시간과 하차시간이 포함되어 있어, 이를 활용하여 운행시간을 도출하였다. 요일별 운행시간 분포를 살펴보면 Figure 5와 같이 나타나며, 오전 08시에서 09시에 평균운행시간이 11-12분으로 나타났으며, 오후 18시에서 19시에는 13-14분으로 운행시간이 분석되었다. 그리고 주말의 경우 평균운행시간이

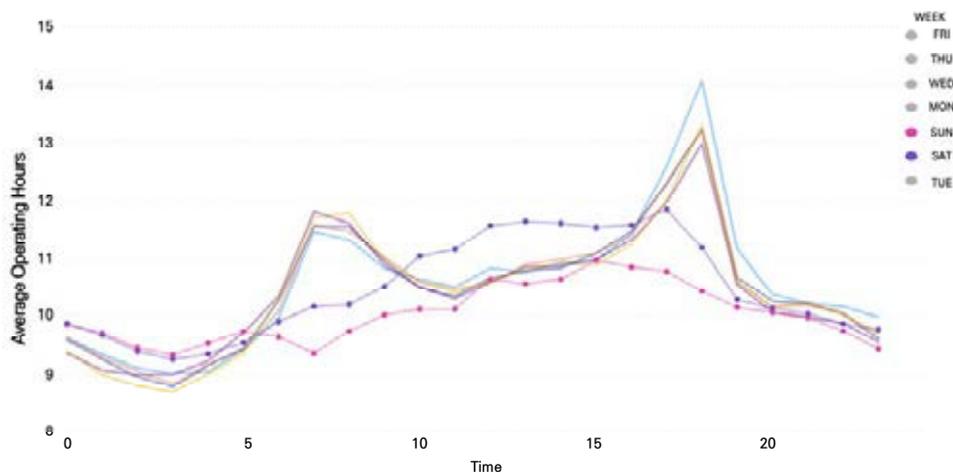


Figure 5. Average operating hours by time of day for each day

평일과 비교하면 첨도(Kurtosis)가 낮은 것으로 나타났으며, 이는 평일의 경우 오전, 오후 첨두시간대에 교통정체 및 도로환경에 기인한 것으로 판단되며, 평일과 주말 통행 목적이 상이하여 발생 된 결과라 보인다.

4) 택시의 시공간적 통행특성 분석

택시 이용자들의 전체 승차 지점을 공간분석 프로그램인 QGIS를 활용하여 히트맵으로 살펴보면 Figure 6과 같이 나타내었다. 대구광역시 도심인 동성로 일대와 부도심인 동대구역 지역의 승차가 높게 나타나며, 그 외에도 성서 산업단지, 대구광장, 서부정류장, 황금동, 상인동, 신천시장 등 상업시설 비중이 높은 지역을 중심으로 택시 승차 비중이 높게 나타났다.

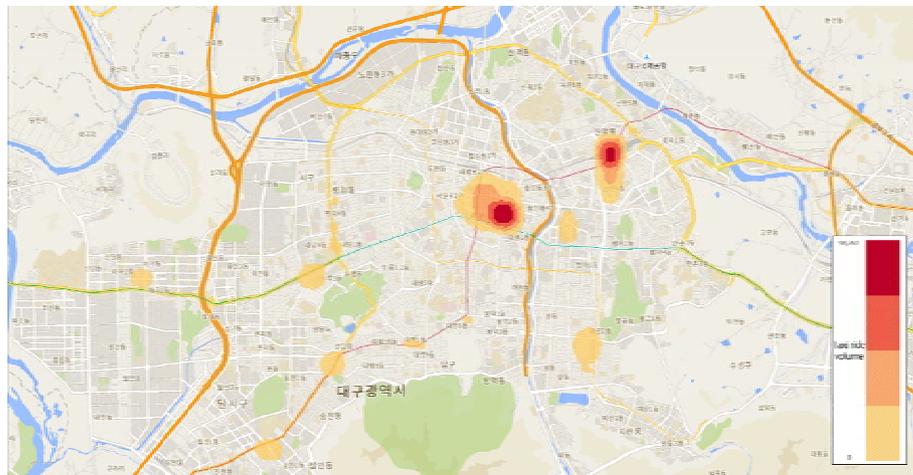


Figure 6. Time-space analysis of taxi ride volume (total)

요일별 시간대별 택시 이용자들의 공간적 분포를 살펴보면 평일 첨두시간(16-03시)의 경우 Figure 7과 같이 나타나며, 범어동, 평리동, 비산동 등 업무지구와 산업단지 주변에서 택시 승차의 통행량이 많게 나타남을 알 수 있다. 비첨두시간대(04-15시)를 나타낸 Figure 8에서 첨두시간대와 분포는 비슷하나 상업지역과 의료시설인 대구카톨릭대학교와 영남대병원을 제외한 업무, 산업단지 지역의 통행이 급감하는 것을 알 수 있으며, 이는 평일 근로자들의 통행 변화와 주변 토지이용 특성에 기인한 것으로 판단된다.

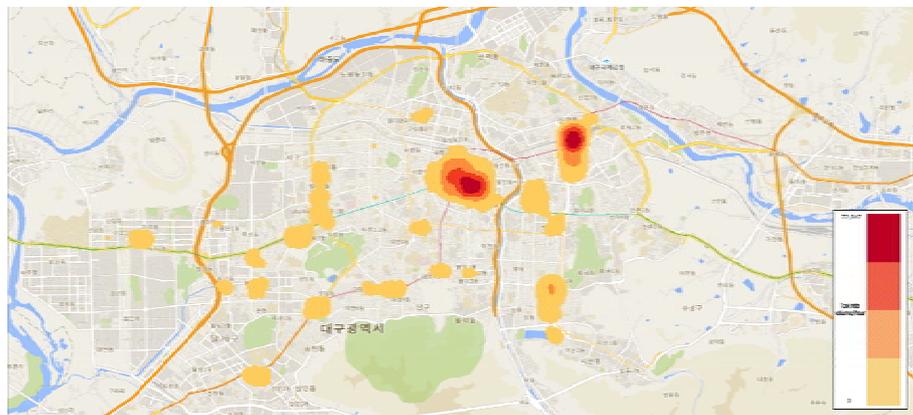


Figure 7. Time-space analysis of average number of taxi rides on weekday at peak hours (16:00-03:00)

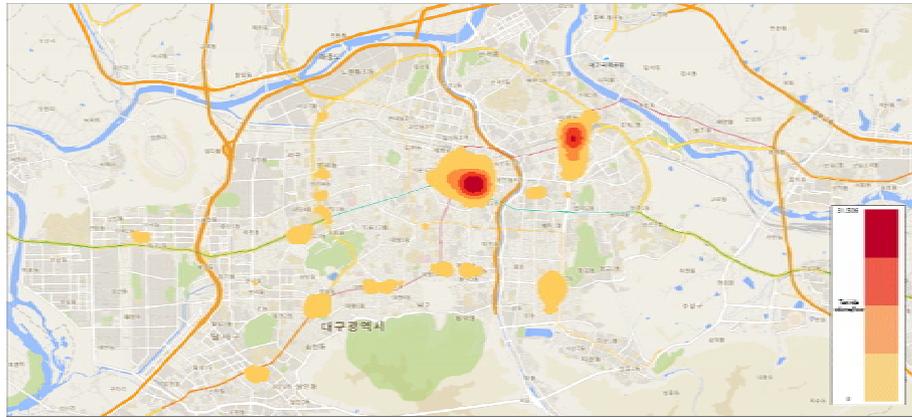


Figure 8. Time-space analysis of average number of taxi rides on weekday at non-peak hours (04:00-15:00)

주말 첨두시간대의 히트맵 분석결과는 Figure 9와 같이 나타나며, 택시 이용자의 공간적 분포가 상대적으로 특정 지역에 집중되어 있음을 알 수 있다. 대구시 상업시설 중심의 동성로와 동대구역을 제외하면 대구광장, 서부정류장, 황금동 등의 상업시설에서 일부 높게 나타남을 알 수 있다. 주말 비첨두시간대인 Figure 10에서는 동성로와 동대구역을 제외하면 통행 수요가 낮은 것을 확인하였다.

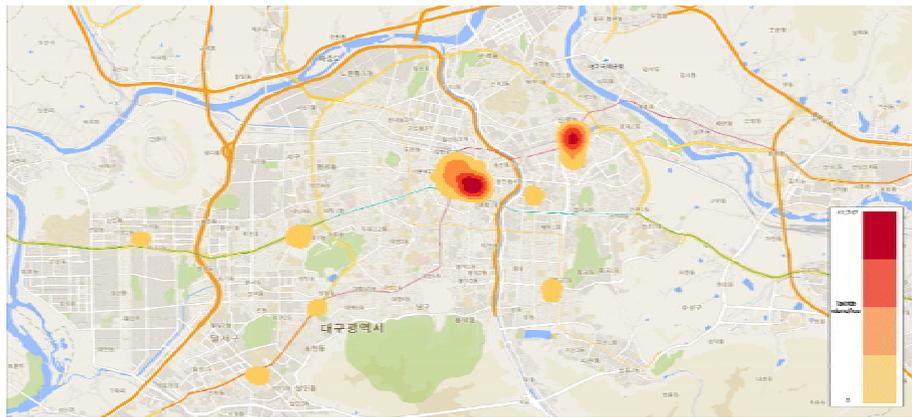


Figure 9. Time-space analysis of average number of taxi rides on weekend at peak hours (16:00-03:00)

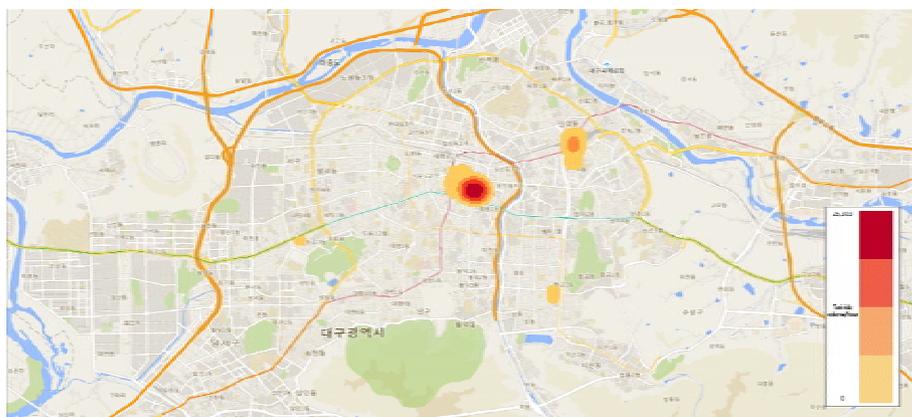


Figure 10. Time-space analysis of average number of taxi rides on weekend at non-peak hours (04:00-15:00)

5) 택시의 주요 이동 경로

택시 이용자들의 주요 이동 경로를 추출하기 위해 파이썬 경로분석 라이브러리인 OSMnx를 활용하였으며, 택시 승차 데이터의 출발지와 목적지를 기반으로 최단경로(Shortest Path) 알고리즘을 활용하여 운행 경로를 추정하여, 대표적인 단계 구분도인 네츄럴브레이크(Natural Breaks Classification)로 상위 40% 택시 이용자들의 주요 이동 경로를 시각화하여, Figure 11과 같이 도출하였다.

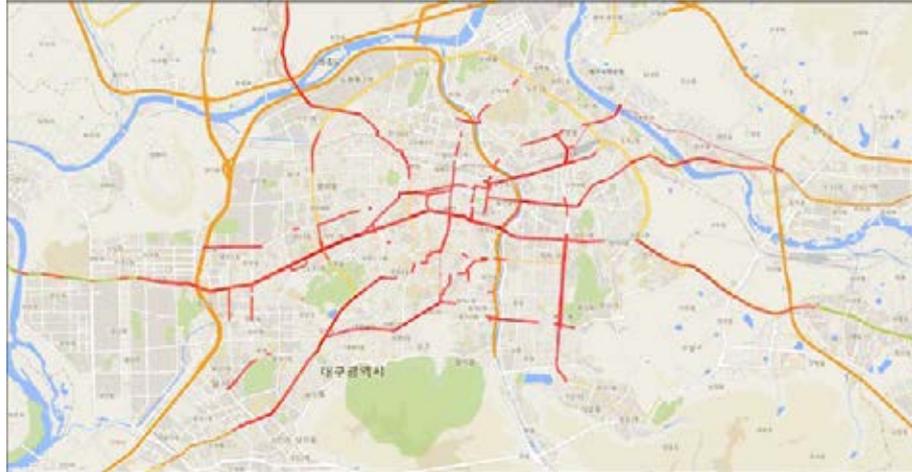


Figure 11. Main route for shared mobility

앞서 분류된 데이터인 평일 첨두·비첨두시간대, 주말 첨두·비첨두시간대를 기반으로 주요 이동 경로를 추출하였으나 분류된 데이터 모두에서 유사한 패턴을 보여 전체 데이터를 기반으로 주요 이동 경로를 추출하였다. 세부적으로 살펴보면 가로축은 지하철 1, 2호선과, 세로축은 지하철 3호선과 유사하게 주요 경로가 추출됨을 알 수 있다. 그 외에는 대구공항 접근도로 그리고 서부정류장에서 계명대학교 대명캠퍼스 구간이 택시들이 주로 이동하는 경로로 분석되었으며, 이는 택시 이용자의 주요 이동 경로가 도시 내 간선도로, 보조 간선도로 위주의 통행이 주를 이루는 것으로 판단된다.

주요 이동경로와 주변 토지이용 특성과의 관련성 분석

1. 머신러닝 기반의 모형 검토

격자 내 경로별 택시 이용자가 이동한 통행 수를 종속변수로 하고, 주변 토지이용 특성인 용도별 건물, 대중교통과의 거리, 도로 유형, 통신사 유동인구, 카드 매출 실적 등을 독립변수로 하였다. 독립변수 선정에 있어 용도별 건물 변수는 이용자의 통행 목적을 확인하기 위하여 선정하였으며, 대중교통과의 거리 변수는 택시 이용에 있어 대중교통과의 연관성을 확인할 수 있는 중요한 요소이므로 선정하였다. 도로 유형 변수의 경우 택시의 주요 동선이 어떤 도로에 영향을 받는지와 공유형 자율주행차의 인프라가 어느 도로축을 따라 구축되어야 하는지 파악하기 위하여 선정하였다. 통신사 유동인구와 카드 매출 실적 변수는 사회·경제변수로 선정하였으며, 성별과 연령별로 분류하여 택시의 주요 이용자를 파악하고자 선정하였다. 대표적인 머신러닝 기법인 랜덤포레스트(Random Forest)와 LightGBM 기반의 회귀모형을 구축하여 독립변수별 변수 중요도로 택시 주요 이동 경로와 주변 토지이용 특성과의 관련성을 살펴보고자 한다.

랜덤 포레스트의 개별 모형인 의사결정트리는 노드의 불순도를 가장 크게 감소시키는 변수로 마디를 형성하게 되며, 이를 측정하는 지니계수는 평균 감소량을 기반으로 특성 중요도를 판단하여 지니 평균 감소량이 클수록 중요도가

높은 변수가 된다. 또한 이는 의사결정나무(Decision Tree)를 여러 개 모아 데이터 분류 및 예측을 수행하는 알고리즘으로 어떤 데이터를 기반으로 분류 또는 예측을 시행한다고 할 때, 하나의 의사결정나무를 적용하는 것보다 여러 개의 의사결정나무를 결합하여 예측을 수행하면 단일 모형에 비해 높은 성능의 알고리즘 모형을 만들 수 있는 장점이 있으며, 이러한 기법을 앙상블(Ensemble) 기법이라 한다.

LightGBM은 일반 의사결정트리 계열의 트리 분할 방법과는 다르게 리프 중심의 트리분할(Leaf Wise) 방식의 알고리즘을 활용하고 있으며, 이러한 리프 중심의 트리분할 방식을 적용할 경우 트리 균형을 맞추지 않고 노드가 지속적으로 분할하면서 비대칭적인 규칙 트리가 생성된다. 이러한 알고리즘은 최대 손실 값을 가지는 리프 노드를 중심으로 지속적으로 분할함으로써 균형 트리 분할 방식에 비하여 상대적으로 예측력에 있어 우수하다는 평가를 받고 있다.

그 외에도 더 작은 메모리 사용량과 빠른 학습시간 등의 이유로 데이터 사이언스 분야에서 랜덤포레스트 모형과 더불어 분류와 회귀예측에 있어 대표적인 머신러닝 기법으로 활용되고 있다. 본 연구에서는 랜덤포레스트 기법뿐만 아니라 트리 분할 방식이 다른 LightGBM을 활용하여 다양한 측면에서 변수 중요도를 도출하여 관련성을 살펴보고자 한다.

2. 모형 추정

격자 내 경로별 택시 통행 수와 주변 토지이용과의 관련성을 살펴보기 위하여 공간정보상에 가로 50M, 세로 50M 격자62,062건을 생성하여 용도별 건물정보, 대중교통 위치정보, 도로 네트워크, 카드 매출실적, 통신사 유동 인구 등의 자료를 활용하여 독립변수들을 구성하였으며, 도로 네트워크의 경우 오픈소스 맵인(OpenStreetMap)에서 제공하고 있는 도로등급을 국내 용어로 변환하여 세부적인 내용은 Table 2와 같다.

Table 2. Model variable

	Division	Variable	Explanation	Type
Dependent variable	Taxi boarding and alighting data	Taxi traffic	Number of taxi trip	Continuous
Independent variable	Building information by purpose (33 categories)	Apartment house	Number of apartment houses in grid	Continuous
		Factory	Number of factories in grid	
		∴	∴	
	Public transport location information (2 categories)	Sales facility	Number of sales facilities in grid	Continuous
		Bus stop	Distance to bus stop (m)	
	Open street map road network (8 categories)	Subway station	Distance to subway station (m)	Continuous
		Motorway	Number of highways in grid	
		Trunk	Number of urban highways in grid	
		Primary	Number of arterial roads in grid	
		Secondary	Number of auxiliary arterial roads in grid	
Tertiary		Number of collection roads in grid		
Residential		Number of local maps in grid		
Living_street		Number of living roads in grid		
Credit card data	Card sales performance	Total card sales performance in grid	Continuous	
	Telecommunication company data	Telecommunication company floating population		Sum of floating population in grid

구축한 모형은 머신러닝에서 대표적으로 활용되고 있는 랜덤포레스트(Random Forest)을 적용하였고, 트리분할 기법에 따라 상이한 변수 중요도가 도출됨을 감안하여 LightGBM 모형의 추가 분석을 통해 다양한 측면에서의 변수 간 관련성을 분석하였다. 그리고 본 연구의 목적상 정확한 택시 이용자의 통행량을 예측하는 것이 아니므로 주변

토지이용 특성과의 관련성에 주된 결과 도출을 목적으로 하는 연구인 만큼 평가지표인 MAE(Mean Absolute Error), RMSE(Root Mean Square Error)를 통한 검증보다 커널밀도(Kernel Density) 함수를 통해 시각적으로 실제 값과 예측값의 분포도를 비교 분석하여 모형 예측의 적합도를 확인하였다.

3. 모형 추정의 결과

본 연구에서는 전체 데이터에서 랜덤하게 25% 데이터를 테스트 데이터로 분리하여 분리하는 Hold-Out 기법을 적용하였으며, 구축된 모형을 기반으로 예측을 하였을 경우 실제 값과 모형간 비교를 Kernel Density 함수를 통해 시각적으로 비교하였다. 실제 값인 택시 통행량이 빨간색 그래프로 나타나며, 추정된 랜덤포레스트 모형이 보라색 그래프, LightGBM은 노란색으로 시각화되어 보여진다. 이러한 분석결과를 통해 모형의 예측이 적합하다 판단되며, 실제 값과 모형 간 분포의 차이도 크지 않는 것으로 보인다(Figure 12 참조).

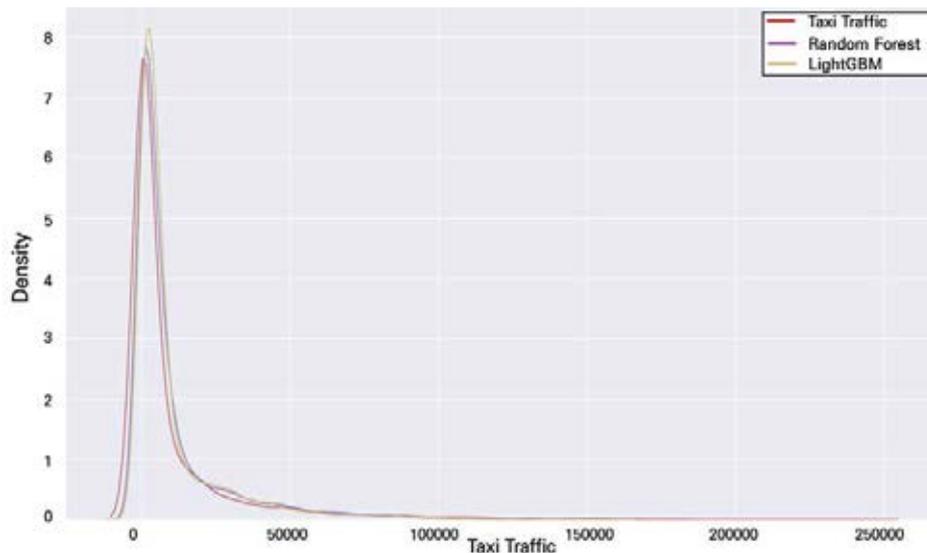


Figure 12. Comparison of fit between models (Kernel Density)

분석 결과 랜덤포레스트 모형의 변수 중요도(Feature Importances)를 살펴보면 지하철역과의 거리(Subway station), 간선도로의 개수(Arterial road), 유동인구(Floating population), 카드매출 실적(Card sales), 버스정류장과의 거리(Bus stop), 보조간선도로(Auxiliary arterial road), 집산도로(Collection road), 제1종 근린생활시설의 개수(Class1 neighborhood living facility) 순으로 나타났으며, 이는 택시가 이동하는 주요 경로가 대다수 간선도로와 보조 간선도로로 이루어져 있고 주변 토지이용 내 대중교통시설, 의료시설, 상업시설 등으로 구성되어 있다는 결과로 해석 가능하다(Figure 13(a) 참조).

LightGBM 모형의 추정 결과는 랜덤포레스트 모형과는 일부 상이한 변수 중요도를 보이고 있으며, 이는 트리 분할 방식이 리프 분할 알고리즘에 기반한 결과라 보인다. 랜덤포레스트 모형과의 변수 중요도 차이점은 도로 네트워크 변수보다 대중교통시설인 지하철역(Subway station), 버스정류장과의 거리(Bus stop), 그리고 민간데이터인 카드매출 실적(Card sales), 유동인구(Floating population) 변수의 중요도가 상대적으로 높게 나타났다(Figure 13(b) 참조).

데이터별 세부 중요도 속성은 두 모형 모두 유사하게 나타나 도로 네트워크 변수의 경우 간선도로(Arterial), 보조간선도로(Auxiliary arterial road), 집산도로(Collection road), 국지도로(Local road), 도시고속화도로(Urban express-

way) 등의 순으로 영향을 주는 것으로 분석되었다. 또한 대중교통시설은 지하철역(Subway station), 버스정류장과 거리(Bus stop)가 영향을 주는 주요 변인으로 확인되었다. 용도별 건물정보의 경우 제1종 근린생활시설(Class1 neighborhood living facility), 제2종 근린생활시설(Class2 neighborhood living facility), 단독주택(House), 공동주택(Apartment house) 순으로 변수 중요도가 높은 것으로 분석되었다.

용도별 건물정보의 경우 상대적으로 다른 데이터들의 변수들보다 변수 중요도가 낮은 것으로 나타났는데, 이는 토지이용의 면적인 특성보다는 도로 네트워크의 선적인 특성과 카드매출, 통신사 유동인구의 포인트 변수들이 택시 이동 경로 설명에 있어 주요한 변수로 작용하였음을 추정할 수 있다.

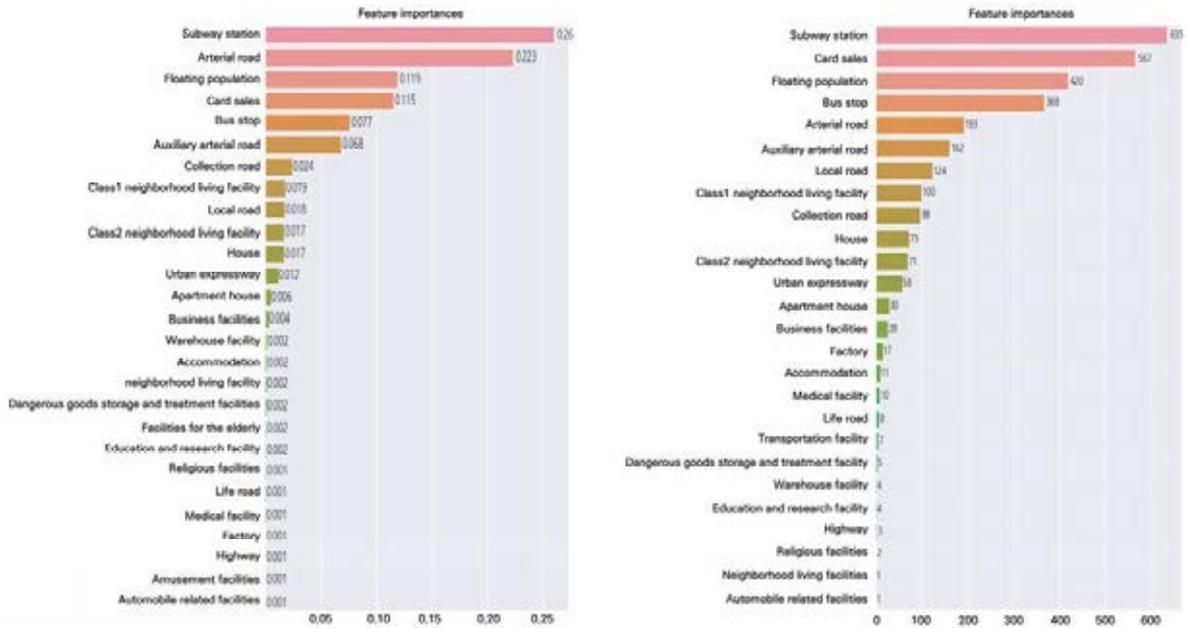


Figure 13(a). Variable importances from random forest model Figure 13(b). Variable importances from LightGBM analysis

결론

본 연구는 자율주행 시대에 진입하면서 공유형 자율주행차의 주행에 있어 필요한 인프라를 구축해야 하는 주요 노선을 거시적으로 분석했다는 점과 주요 이동 경로를 주변 토지이용 특성과 연계하여 장래 공유형 자율주행차의 통행에 관련된 필수 인프라를 전략적인 노선 배치와 더불어 효율적으로 구축할 수 있도록 방향성을 제시했다는 점에 의의를 두고 있다. 연구 내용을 정리하면 대구광역시를 대상으로 공유형 자율주행차의 주행 패턴이 현재 택시의 이용 패턴과 유사하다는 가정을 두고 분석을 진행하였으며, 택시 이용자의 승차지점부터 하차지점까지의 좌표와 시간 등의 데이터를 포함하는 택시 DTG 데이터를 활용하여 주요 경로와 이에 영향을 주는 요인을 토지이용과 관련된 빅데이터를 통해 분석하였다. 승차율 분석을 통해 대구광역시의 승차율 분포도를 맵 기반으로 파악하였으며, 택시의 주요 이동 경로를 도출하여 도시 내 간선도로와 보조 간선도로, 지하철 노선을 따라 공유형 자율주행차를 위한 인프라 구축이 선제적으로 진행 될 수 있다는 시사점을 얻을 수 있었다. 또한, 머신러닝 기법을 활용한 회귀모형을 통해 택시 이용 선택에 있어 영향력 있는 토지이용 요인을 파악하였으며, 이 점으로 향후 공유형 자율주행차를 위해 설계되는 인프라 위치를 주요 토지이용 변수를 고려하여 선정할 수 있다는 결론을 도출하였다.

다만 본 연구에서 활용한 택시의 DTG 데이터는 이용자가 승차하고 하차까지의 주행 기록으로 승객이 탑승하고 있지 않은 경우의 경로는 분석에 포함시키지 못했다는 점에서 일부 한계가 존재한다. 향후 이러한 데이터의 보완이

이루어진다면 공유형 모빌리티의 이용자 위치 예측과 자가 연료 충전을 위한 이동 경로 예측에 대한 연구가 가능할 것으로 판단된다.

Funding

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 21AMDP-C160881-01, Future Road Design and Testing for Connected and Autonomous Vehicles).

ORCID

KIM, Keunwook  <http://orcid.org/0000-0003-4181-7394>

KIM, Junghwa  <http://orcid.org/0000-0001-9771-5755>

KIM, Woojin  <http://orcid.org/0000-0002-1231-2533>

LEE, Seunghyeon  <http://orcid.org/0000-0003-4454-0407>

References

- Boesch P. M., Ciari F., Axhausen K. W. (2016), Autonomous Vehicle Fleet Sizes Required to Serve Different Levels of Demand, *Transportation Research Record: Journal of the Transportation Research Board*, 2542, 111-119.
- Chen T. D., Kockelman K. M., Hanna J. P. (2016), Operations of a Shared, Autonomous, Electric Vehicle Fleet: Implications of Vehicle & Charging Infrastructure Decisions, *Transportation Research Part A: Policy and Practice*, 94, 243-254.
- Cho J., Lee H., Lee J., Kim D. (2017), The Hazardous Expressway Sections for Drowsy Driving Using Digital Tachograph in Truck, *J. Korean Soc. Transp.*, 35(2), Korean Society of Transportation, 160-168.
- Fagnant D. J., Kockelman K. M., Bansal P. (2015), Operations of Shared Autonomous Vehicle Fleet for Austin, Texas, Market, *Transportation Research Record: Journal of the Transportation Research Board*, 2536, 98-106.
- Gelauff G., Ossokina I., Teulings C. (2017), Spatial Effects of Automated Driving: Dispersion, Concentration or Both, *The Hague: KIM-Netherlands Institute for Transport Policy Analysis*.
- Kwon M., Kim Y. (2018), Analysis of Traffic Flow Characteristics Using Taxi DTG Big Data, *J. Korean Soc. Transp.*, 36(6), Korean Society of Transportation, 415-428.
- Milakis D., van Arem B., van Wee B. (2017), Policy and Society Related Implications of Automated Driving: A Review of Literature and Directions for Future Research, *Journal of Intelligent Transportation Systems*, 21(4), 324-348.
- Ministry of the Interior and Safety (2019), *Public Big Data Standard Analysis Model Manual Selection of location for electric vehicle charging infrastructure*, 10.
- Narayanan S., Chaniotakis E., Antoniou C. (2020), Shared Autonomous Vehicle Services: A Comprehensive Review, *Transportation Research Part C: Emerging Technologies*, 111, 255-293.
- Rothnie A., Uffer S., Ghojeh M. (2016), *Urban Streets in the Age of Connected and Autonomous Vehicles*, 20-25.
- SFMTA (2016), *City of San Francisco Meeting the Smart City Challenge Volume 1*, 39.

- Soteropoulos A., Berger M., Ciari F. (2018), Impacts of Automated Vehicles on Travel Behaviour and Land Use: An International Review of Modelling Studies, *Transport Reviews*, 1-21.
- Spieser K., Treleaven K., Zhang R., Frazzoli E., Morton D., Pavone M. (2014), Toward a Systematic Approach to the Design and Evaluation of Automated Mobility-on-Demand Systems: A Case Study in Singapore, *Lecture Notes in Mobility*, 229-245.
- Zhang H., Sheppard C. J. R., Lipman T. E., Zeng T., Moura S. J. (2020), Charging Infrastructure Demands of Shared-use Autonomous Electric Vehicles in Urban Areas, *Transportation Research Part D: Transport and Environment*, 78, 102210.
- Zhang W., Guhathakurta S. (2018), Residential Location Choice in the Era of Shared Autonomous Vehicles, *Journal of Planning Education and Research*, 0739456X1877606.
- Zhang W., Guhathakurta S., Fang J., Zhang G. (2015), Exploring the Impact of Shared Autonomous Vehicles on Urban Parking Demand: An Agent-based Simulation Approach, *Sustainable Cities and Society*, 19, 34-45.