

모바일 데이터를 활용한 지역간 수단통행 분류 방법론 개발

A Regional Trip Modes Classification Methodology Using Mobile Phone Data

김 규 혁* · 한 효 림** · 김 동 호*** · 송 태 진****

* 주저자 : 충북대학교 도시공학과 박사과정
** 공저자 : 충북대학교 도시공학과 석사과정
*** 공저자 : 한국교통연구원 교통빅데이터연구본부 책임전문원
**** 교신저자 : 충북대학교 도시공학과 부교수

Kyuhyuk Kim* · Hyorim Han** · Dongho Kim*** · Tai jin Song****

* Dept. of Urban Eng., Univ. of Chungbuk
** Dept. of Urban Eng., Univ. of Chungbuk
*** Transport Big Data Division, The Korea Transport Institute
**** Dept. of Urban Eng., Univ. of Chungbuk

† Corresponding author : Tai jin Song, tj@chungbuk.ac.kr

Vol. 23 No.4(2024)
August, 2024
pp.77~93

pISSN 1738-0774
eISSN 2384-1729
<https://doi.org/10.12815/kits.2024.23.4.77>

Received 26 March 2024
Revised 18 April 2024
Accepted 17 July 2024

© 2024. The Korean Society of
Intelligent Transport Systems. All
rights reserved.

요 약

최근 데이터 수집 기술의 발달은 모바일 데이터 등의 실제 통행이 반영된 데이터가 폭발적으로 증가시키며, 이러한 데이터들을 활용한 교통분야 연구 또한 촉진 시킨다. 본 연구에서는 모바일 데이터 궤적정보와 수단별 주요 교통유발시설 데이터를 매칭하여 모델링 없이 지역간 통행수단을 추정하는 방법론을 개발하였다. 개별 수단통행은 통행계획 내 교통유발시설 유무, 통행시간, 접근통행 유무 등의 규칙에 따라 분류되었다. 서울, 청주에 거주하는 908명의 개별 통행계획 데이터에 제안된 방법론을 적용하였고, 통행 다이어리 데이터의 기록된 수단과의 비교 결과를 Type I and II error 결과표로 나타내어 방법론의 정확도 검증을 수행하였다. 검증 결과, 본 연구에서 개발한 방법론은 항공수단 1.00, 철도수단 0.955, 버스수단 0.73 수준의 F1-Score를 보이는 것으로 나타났다.

핵심어 : 수단선택, 모바일 데이터, 교통인프라, 지역간 통행, 교통인프라

ABSTRACT

The recent development of data collection technology, which conveys various travel data in real-world such as mobile data and probe vehicle data, facilitates transportation planners identifying specified spatio-temporal travel patterns. In this study, an easily implementable travel mode classification methodology was proposed to classify inter-regional trip-modes without modeling by superimposing trajectories generated from mobile phone signaling and transportation infrastructure points into a polygon scale of a shapefile in a GIS system. Each regional mode trip was classified according to the rules such as the presence of transportation infrastructure in the trip trajectory, travel time, and the presence of access trips. An accuracy test generates Type I and Type II error results table to verify the proposed methodology. As a result, it was found that the methodology developed showed the F1-Score of the air mode 1.00, rail mode 0.95, bus mode 0.73.

Key words : Mode split, Mobile data, Transportation infrastructure, Regional trip

I. 서론

장래 통행량 예측을 통해 SOC 도입의 타당성을 평가하는 데 주로 사용되는 4단계 수요추정법은 통행발생, 통행분포, 수단선택, 통행배정 순의 절차로 수행된다(Currans and Clifton, 2015 ; Etu and Oyedepo, 2018; Eromietse and Joseph, 2019). 이 중 3단계에 해당하는 수단선택은 통행발생, 통행분포 단계에서 예측한 통행량을 기반으로 개별 통행자의 통행특성 형태를 반영하여 교통수단을 분류하는 단계로, 4단계 수요 추정 절차 중 매우 중요한 단계이며(Qu et al., 2015), 수단선택 단계에서의 결과 값은 교통서비스 제공 및 관리, 도시 계획에 주로 사용된다(Wang et al., 2010). 국내에서는 개별 통행 이동수단을 조사 후 전수화한 뒤, 대중교통, 전세버스, 택시 등 수단의 수송실적 자료를 이용하여 수단별 통행량을 보정하는 것이 수단선택을 위한 일반적인 방법론이다. 수단선택 단계 뿐만 아니라 4단계 수요추정의 전반적 과정은 가구통행실태조사와 같은 조사 기반 데이터를 활용하여 수행되어져 왔다(Qu et al., 2015; Kalatian and Shafahi, 2016; Andersson et al., 2022).

그러나 조사 기반 데이터는 자동으로 기록되는 것이 아닌 사람이 수동적으로 기입하는 자료를 활용하기 때문에 주요 통행을 제외한 통행에서 자료 누락이 발생된다. 특히, 이 현상은 장거리 통행에서 빈번하다. 또한 교통인프라가 빠르게 변화하는 국가 또는 도시는 조사 기반 데이터를 수요추정에 활용하면 지속적으로 변화하는 수단분담율을 반영할 수 없으며(Qu et al., 2015), 조사에 소요되는 시간과 비용이 크게 소비되어 데이터의 표본이 낮아지는 한계점이 존재한다(Doyle et al., 2011). 추가로, 통행분포 모형 개발 시 주로 중력모형 또는 엔트로피 극대화 모형을 사용한다. 이때 두 교통존(Traffic Analysis Zone, TAZ) 간의 통행거리는 주요 변수로 사용되기 때문에 장거리 통행이 과소추정되는 경향이 있다. 이로 인해 전통적인 수단선택 기법은 그 자체로 한계점을 지니게 된다.

최근에는 모바일 기술의 발달로 휴대폰과 같은 개인 장치에서 엄청난 양의 사회적 데이터가 생성되고 있다(Qu et al., 2015). 모바일 데이터는 시공간적 해상도가 기존 조사 기반 데이터에 비해 월등히 높으며, 세밀한 정보를 제공할 수 있다는 장점이 존재한다. 이러한 장점을 반영하여, 최근 모바일 데이터를 활용해 4단계 수요예측 프로세스 중 3단계에 해당하는 통행자 이동수단을 추정하려는 연구가 활발히 진행되고 있다.

본 연구는 이와 같이 시공간적 해상도가 조사 기반 데이터에 비해 높은 모바일 데이터를 활용해 인프라 변화에 구애받지 않는 이동수단 추정 방법론을 개발개발된 방법론의 검증하는 것을 목적으로 한다. 전통적인 교통수요예측 방법론과 조사 기반의 통행데이터 수집법은 상당수의 장거리 통행을 누락시킬 가능성이 존재한다. 따라서, 본 연구에서는 전체 수단 중 장거리 이동수단에 해당하는 항공, 철도, 버스 수단을 우선적으로 분류하는 방법론을 개발하고자 한다.

II. 선행연구 고찰

1. 국내 이동수단 추정 연구

국내 이동수단 추정 관련 연구는 대부분 조사 기반의 데이터를 중심으로 기존 4단계 수요추정 방법론 내 수단분담모형을 개선하는 방향으로 진행되었다. Cho et al.(2017)은 철도의 시간 및 비용 변화에 따른 지역간 통행수단의 수단분담비율 및 통행시간가치의 변화를 확인하였다. 수단별 효율을 도출하기 위해 SP (Stated Preference) 조사를 실시하였으며, 다항로짓모형과 혼합로짓모형을 활용해 수단분담비율과 통행시간가치를

도출하였다. 해당 연구의 결과에 따르면, 철도수단의 지역간 수단분담율은 시간과 비용 수준이 개선됨에 따라 21% 수준이 도출될 것으로 전망했다. Ryu et al.(2012)은 사회경제적 변수, 토지이용 변수, 교통인프라 관련 변수 등 다양한 변수를 활용한 수단선택모형을 개발하였다. 장래 변수값 추정이 가능한 행정구역 특성, 도시화 수준, 택시사업구역 동질성, 가구당 승용차 보유대수 및 인구밀도, 시가화 면적비율, 수용학생 수 비율, 종사자 수 비율, 연령대별 인구비율, 지하철역 거리 등이 변수로 선정되었다. 분석 결과, 해당 연구에서 개발한 수단분담모형은 비용변수를 기반으로 한 기존의 수단분담모형에 비해 높은 설명력을 가지는 것으로 나타났다. Lee and Choe(2010)는 가구통행실태조사 데이터를 기반으로 자전거 수단분담을 추정하는 모형을 개발하였다. 가구통행실태조사에서 자전거를 이용한 통행은 '1'로, 그렇지 않은 통행은 '0'으로 하여 자전거 수단분담 추정이 가능한 로지스틱 회귀모형을 개발했다. 가공된 자료에 따르면, 자전거의 존내통행 수단분담률은 약 2.41%이며, 존간통행 수단분담률은 약 1.5%에 달했다.

Shin(2014)은 존 간 통행량 미반영, 혼잡상태 미반영, 교통존 특성변수 미반영 등 기존 수단분담모형의 한계점을 보완한 형태의 수단분담모형을 개발하였다. 해당 연구의 모형은 존 간 통행량이 종속변수에 반영되었으며, 택시수단이 추가되었다. 또한, 교통혼잡이 반영된 수단별 통행시간이 적용되었다. 분석 결과, 모형의 적합도 등이 실제 적용 가능한 수준으로 도출되었다. Kim et al.(2021a)은 접근통행, 생활권 통행에서 공유 전동킴보드의 수단선택모형을 개발하였다. 549명의 응답자를 대상으로 SP (Stated Preference) 조사를 수행하였으며, 모형을 개발하기 위해 로지스틱 회귀분석을 적용하였다. 접근통행에서 공유 전동킴보드의 선호도는 버스, 도보에 비해 낮은 선호도를 보였으며, 생활권 통행에서 공유 전동킴보드의 선호도는 승용차, 자전거에 비해서는 높으나, 도보, 대중교통에 비해서는 낮은 것으로 분석되었다.

2. 국외 이동수단 추정 연구

표본이 전수에 가까운 모바일 데이터를 이동성 분석에 활용할 수 있게 되며, 통행자의 이동수단을 추정하는 분야에서도 관련 연구가 국외를 중심으로 활발히 진행되었다. 모바일 데이터를 활용한 통행자 이동수단 추정 관련 연구의 경우, 크게 ① 다양한 통계 기법을 적용하여 이동수단을 추정하거나, ② 도로, 철도망 등 네트워크와 데이터와의 경로 일치성을 기반으로 이동수단을 추정하는 연구로 구분되었다.

먼저, 통계 기법을 활용한 이동수단 추정 연구는 다음과 같다. Wang et al.(2010)은 출발지에서 목적지로 이동하는 통행자의 이동시간을 기반으로 수단별 통행비율을 추정하였다. 같은 거리를 이동할 때, 특정 이동시간대의 통행자 수는 다른 범위보다 상당히 많다는 사실을 기반으로, 밀도에 따라 통행자 그룹을 군집화하였다. 군집된 그룹은 실제로 각기 다른 교통수단에 대응되었다. 통행자의 이동시간은 K-means clustering 기법을 통해 군집화 되었으며, 통행시간이 13.4분 미만인 통행은 차량을, 13.4분 이상 42.4분 미만인 데이터는 대중교통을, 42.4분 이상인 데이터는 도보를 도보를 이용하는 것으로 분류할 것을 제안하였다. 해당 값은 구글 맵에 기록된 평균 차량 이용시간(9분), 대중교통 이용시간(47분)과 허용할 만한 차이를 보이는 것으로 나타났다. Xu et al.(2011)은 모바일 데이터를 기반으로 개인교통수단인 자동차, 자전거, 보행 등 세 가지 수단을 분류하는 방법론을 제시하였다. GPS 좌표를 기반으로 데이터의 평균속도를 구하여 전처리한 뒤, Markov model 을 기반으로 혼잡 및 비혼잡 등 두 가지 교통 조건에서 수단을 분류하는 방법을 다르게 하였으며, 정상적인 교통 조건에서는 Speed distribution law (SDL), 혼잡한 교통 조건에서는 Cumulative prospect theory (CPT)를 활용해 수단을 분류하여 수단 분류의 정확도를 높였다. 수단 분류의 정확도는 F1-Score를 기준으로 차량 0.93~0.95, 자전거 0.75~0.79, 보행 0.89~0.91 수준으로 나타났다. 해당 연구의 방법론은 개인교통수단을 분류하는 데 있어 높은 성능을 보이는 것으로 나타났지만, 차량을 자동차와 버스 등으로 분류할 수 없다는 한계

가 존재한다. Kalatian and Shafahi(2016)은 모바일 데이터 셀 커버리지 영역에 보로노이 기법을 적용 및 보정하여 통행량을 집계하였으며, 특정 시간대 도보권에 해당하는 특정 기종점에서 발생하는 통행을 그룹으로 병합하고, 해당 그룹을 각각 특정 수단을 나타내는 군집으로 분할하였다. K-means clustering을 사용하여 이동 속도를 기준으로 데이터를 군집한 결과, 군집의 중심은 각각 보행, 대중교통, 자가용 수단을 대표하는 5, 17, 29km로 나타났으며, 각 수단의 분담율은 4.5, 22.4, 72.1%로 나타났다. Breyer et al.(2021)은 지리정보 기반 수단분류 방법과 기계학습 기반 수단분류 방법을 비교하여 도로와 철도 수단을 분류하고 각 수단분류 방법의 성능을 평가하였다. 분석 결과, 도로와 철도가 인접한 경우 기계학습 기반의 방법이 지리정보 기반의 방법보다 도로와 철도 수단을 분류하는 데 더 우수한 성능을 가지는 것으로 나타났다. 이에 따라 최근에 수행된 상당 수의 연구에서도 기계학습 방법론을 활용하고 있다(Alam et al., 2023; Fourez et al., 2023; Bandyopadhyay et al. 2024).

기계학습 방법론의 경우 모델을 개발하는 데 활용되는 컴퓨팅 시간이 데이터 크기에 비례하여 커진다는 한계가 존재한다(Jun, 2011). 통행수단 분류에 활용되는 모바일 데이터의 궤적 정보는 매우 큰 크기를 가지기 때문에, 기계학습 방법론을 적용하는 컴퓨팅 시간 또한 그에 비례하여 커질 것이다. 통행자의 경로 유사성 기반 방법론은 통행자 경로 인근의 인프라 정보, 거리 등을 규칙화하여 통행수단을 파악하는 방법론으로, 단순한 규칙에 기반하여 통행수단을 추론하기 때문에 컴퓨팅 시간이 짧다는 장점이 존재한다.

통행자의 경로 유사성을 기반으로 이동수단을 추정 한 연구들은 다음과 같다. Doyle et al.(2011)은 통행수단 추정에 Virtual Cell Path (VCP)를 활용하였다. VCP는 휴대폰 사용자가 이동하는 동안 지시간 TAZ의 목록을 의미한다. 개별 VCP가 실제 도로 및 철도와 일치하는지의 여부에 따라, 철도 및 도로 수단을 추정하였으며, 나머지는 Unknown 데이터로 처리하였다. Phithakkinukoon et al.(2017)은 거주지와 직장 사이의 통근목적 통행에 초점을 맞추어 자동차 및 대중교통수단 추정 방법론을 제시하였다. Google maps direction api를 사용하여 거주지와 직장 위치를 출발지 및 도착지 매개 변수로 선정하고 선택 가능한 경로를 확인하였다. 선택 가능한 경로 중, 인접 경유지 중 휴대폰 사용 빈도가 높은 경유지가 가장 많은 경로를 통행자의 실제 이동 경로로 가정하였으며, 통행자가 경로를 따라 오랜 시간 일괄적으로 이동하였다면, 해당 경로는 통행자가 선택한 경로일 가능성이 높다고 판단하여 자동차 수단을 이용한 것으로 추정하였다. Qu et al.(2015)은 미국 보스턴 지역 내에 생성된 Call Detailed Record (CDR) 데이터와 도로망, 지하철 및 버스망을 기반으로 통행수단 추정 방법론을 제시하였다. 먼저, CDR 데이터에서 3주간 20~07시 사이에 사용자가 발견되는 장소는 주거지, 09~17시 사이에 발견되는 장소는 업무지로 정의하고, 통근통행을 분석하였다. 통근시간을 추정하기 위해, 3주간의 거주지-목적지 간 이동시간을 분석하였으며, 시차가 가장 짧은 시간을 통근유효시간으로 사용하였다. 해당 연구에서는 자동차, 대중교통, 도보 등 세 가지 교통수단만을 고려하였으며, 모든 사용자가 이동 중에 단일 수단을 사용한다고 가정하였다. 15km는 도보와 대중교통 및 자동차를 구분하는 속도의 임계치로 사용된다. 출발지 또는 목적지에서 500m 이내에 지하철역이나 버스정류장이 없는 경우자동차를 이용하였다고 판단하였고, 그렇지 않은 경우에는 로짓 모델을 사용하여 대중교통 및 자동차 이용 여부를 판단하였다. Brändle(2021)은 통행수단 추정 개발에 스마트폰 이용 경로 및 이동수단 조사에 관한 어플리케이션을 소개하고, 모바일 데이터를 활용한 수단추정 기법의 한계를 1) “셀 타워 단위 데이터 수집으로 인한 낮은 공간적 해상도로 철도와 도로 등이 인접한 경우 수단 분류가 어려움”, 2) “휴대폰을 사용할 때만 데이터가 수집되어 낮은 시간적 해상도를 보여 통행자의 경로를 완성하기 어려움” 등으로 정의하고 통신사를 대상으로 검증 데이터 수집 촉구, API 제공 회사에 대해 휴대폰 위치 데이터에 대한 액세스의 지속적 허용을 촉구하는 등의 시사점을 도출하였다. Andersson et al.(2022)는 모바일 데이터를 활용해 장거리 통행수단을 추정하는 두 가지 방법을 제안하였다. 두 방법 모두 우도 함수를 정의하고 최대우도 추정을 사용하는 로짓 공식을 기반으로 하

였다. 첫 번째 방법은 각 통행의 수단을 다른 순서로 추론하여 추정을 간단하게 할 수 있으나, 통행수단을 확정할 수 있는 것은 아니므로 결과가 편향될 수 있다. 두 번째 방법은 휴대전화의 이동이 감지된 안테나의 순서에 따라 우도 함수를 설정한다. 해당 방법의 핵심은 고해상도 데이터가 필요하다는 것이다.

3. 연구의 차별성

본 연구의 차별성은 다음과 같다. 첫째, 고해상도의 모바일 데이터를 활용했다. 본 연구에서 사용한 모바일 데이터는 Sightings 데이터로, 기존 모바일 데이터 기반 통행수단 추정 연구에서 활용된 CDR 데이터가 통화, 문자 등 휴대폰을 사용할 때만 데이터가 수집되므로 데이터의 활용도가 낮은 데 반해(Toole et al., 2012; Chen et al., 2014 ; Sadeghvaziri et al., 2016 ; Anda et al., 2017; Jiang et al., 2017; Bwambale et al., 2021), Sightings 데이터는 휴대폰을 사용하지 않더라도 일정 시간 간격으로 데이터가 기록되어 공간적으로 높은 해상도를 가지므로 활용도가 높다. 국내에서는 모바일 데이터와 같은 빅데이터와 인프라 근접성에 기반한 방법론을 수단추정 분야에 활용한 사례가 없기 때문에, 이러한 시도는 매우 의미있을 것으로 판단된다.

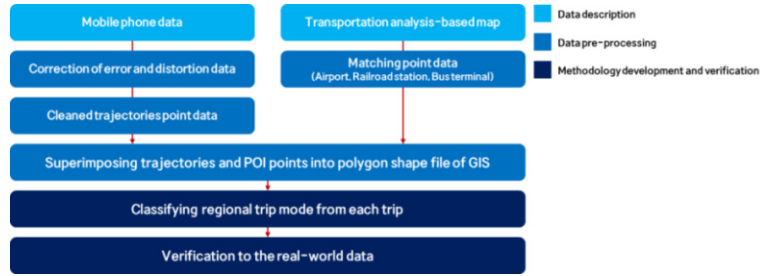
둘째, 1km 수준의 면적을 가지는 고해상도 TAZ에 각 수단별 기반시설 정보를 매칭하여 TAZ 내 기반시설의 존재 유무만으로 지역간 통행수단을 추정하는 방법론을 개발하였다. 본 연구에서는 수단별 기반시설과 모바일 통행궤적 정보를 집계하기 위한 공간 단위로 기존 읍면동 단위 TAZ에 비해 공간적 해상도가 높은 기지국 단위 TAZ (Cell-tower based TAZ)를 활용하여 경로 유사성에 기반한 통행수단 추정 방법론의 신뢰성을 높였다.

셋째, 수단통행 추정 방법론의 검증은 수행하였다. 대부분의 통행수단 추정 관련 연구의 경우, 통계 기법 또는 경로 일치성 기반으로 통행자의 수단통행량만을 추정하고 검증은 수행하지 않았다. 유일하게 Xu et al.(2011)의 연구에서 실제 데이터와의 검증을 수행했으나, 해당 연구에서는 차량, 자전거, 보행 등 세 가지 수단에 한해 연구를 수행했다. 본 연구에서는 항공, 철도, 버스 등 장거리 대중교통수단을 추정하는 것을 목표로 하며, 3주간 약 900명의 개개인이 실제 이용 수단을 기록한 통행 다이어리 데이터를 모바일 데이터와 매칭하여 기존 통행수단 추정 관련 연구에서 시도하지 못했던 통행수단 추정 방법론의 검증을 수행하였다.

III. 연구 방법론

1. 분석 과정

본 연구는 <Fig. 1>과 같은 흐름으로 진행된다. 첫 번째, 모바일 데이터의 특성을 설명하고 왜곡 데이터 보정 및 개별 통행궤적 생성 방법을 설명하였다. 다음으로, 본 연구에서 활용한 기지국 단위 TAZ의 제작 과정에 대해 설명하였다. 추가로, TAZ에 공항, 철도역, 버스터미널 등 주요 교통유발시설의 지리정보를 기지국 단위 TAZ 상에 DB화하였다. 두 번째, 주요 교통유발시설 정보가 DB화된 TAZ에 개별 통행궤적을 매칭하였다. 세 번째, 개별 통행궤적과 주요 교통유발시설의 정보를 기반으로 항공, 철도, 버스 등 지역간 통행수단별 OD를 분류하는 방법론을 개발하였다. 마지막으로, 개발 방법론을 샘플 데이터에 적용하여 수단별 집계 OD를 산출하고, 통행 다이어리에 기록된 실제 통행수단의 비교를 통해 방법론의 정확도를 검증하였다.



<Fig. 1> Overall research flow

2. 모바일 데이터

모바일 데이터란, 개인 휴대전화와 기지국이 주고받는 신호자료를 의미하며, 크게 통화, 문자, 인터넷 접속 등 휴대전화를 사용할 때만 로그 기록이 저장되는 CDR 데이터와 휴대전화를 사용하지 않아도 일정 시간 간격으로 로그 기록이 저장되는 Sightings 데이터로 구분된다(Wang et al., 2010; Chen et al., 2014; Chen et al., 2016; Won et al., 2021). 본 연구에서 사용하는 모바일 데이터의 유형은 Sightings 데이터로, 로그 기록이 저장되는 빈도가 높기 때문에 CDR 데이터에 비해 시공간적 해상도가 높다는 장점이 있다(Chen et al. 2014; Chen et al. 2016; Won et al., 2021). 모바일 데이터의 원천데이터에는 NULL값과 같은 오류 데이터와 핸드오버와 같은 왜곡 데이터가 존재한다. 핸드오버는 휴대전화 사용 중 기지국 수신 범위를 이탈하는 경우 신호 세기가 약해지는 현상을 방지하기 위해 적절한 기지국으로 신호를 변경하는 것을 의미한다. 핸드오버는 통행자 이동이 아닌 이상 신호에 의해 기록된 데이터이므로, 실제 통행정보를 왜곡시키기 때문에 반드시 보정이 필요하다. 본 연구에서는 이러한 오류 데이터를 제거하고 왜곡 데이터를 보정한 뒤 모바일 데이터의 개별 통행케적을 생성하였다.

본 연구에서는 6월 18일부터 7월 8일까지의 서울특별시, 청주시 거주 908명의 통행케적이 기록된 SK 모바일 데이터와 해당 인원들이 작성한 통행 다이어리 데이터를 분석 및 검증에 사용하였다. 통행 다이어리 작성은 참여자의 휴대전화 어플로 진행되었으며, 외출하여 통행하는 경우 통행 다이어리 어플을 활성화하고 이

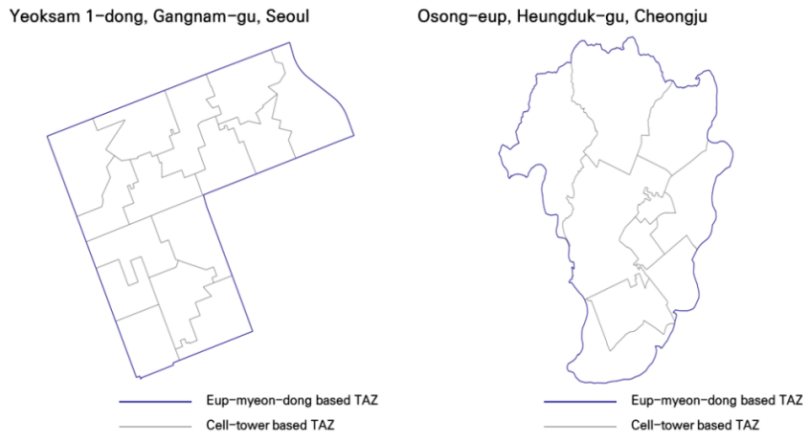
<Table 1> Mobile cell-based dataset

Columns	Description	Example
ID	The respondent's number	14
Date	The date of GPS recorded	20-06-19 10:28:00 AM
Lat.	GPS location information_latitude	37.4971
Long.	GPS location information_longitude	127.0417
Departure / Arrival	The point of trip departure or arrival	Departure
Departure / Arrival Place	The place of trip departure or arrival	Hyundai department store Cheongju
Departure / Arrival time (hh)	The time of trip departure or arrival (hh)	12
Departure / Arrival time (mm)	The time of trip departure or arrival (mm)	38
Trip purpose	Trip purpose on departure (Each purpose is transformed by number)	4
Trip mode	Trip mode between departure and arrival (Each mode is transformed by number)	9

동장소와 출발지 출발시간, 목적지 도착시간, 통행목적, 통행수단 등을 기입하였다. 제출된 통행 다이어리 자료는 참여자의 통행패적과 매칭되었다. 매칭된 데이터셋의 주요 내용은 <Table 1>과 같다.

3. 기지국 단위 TAZ

본 연구에서는 모바일 데이터에 내재된 사람들의 통행 정보와 경로 일치성 기반 수단추정의 단서가 되는 교통인프라 정보를 고해상도의 공간 정보에 집계하기 위해 기지국 단위 TAZ (Cell-tower based TAZ)를 제작하였다. 기지국 단위 TAZ는 다음과 같은 단계를 거쳐 제작되었다. 먼저, 보로노이 기법을 활용하여 주 기지국의 가상 셀 반경을 설정하고, 기지국 셀 반경과 집계구를 1:1 혹은 N:1 비율로 매칭하여 행정구역과 매칭을 용이하게 한다. 다음으로, 면적이 지나치게 작거나, 동일한 TAZ ID를 부여받았으나 이격되어있는 경우 해당 TAZ를 보정하였으며, 마지막으로 강 위에 위치한 TAZ는 동일한 행정구역에 속한 인접 TAZ와 병합하여 전국 단위의 TAZ를 구축하였다. 구축된 TAZ는 <Table 2>와 같이 읍면동 단위보다 작고 집계구 단위보다 큰 면적을 가진다(Kim et al., 2021b).

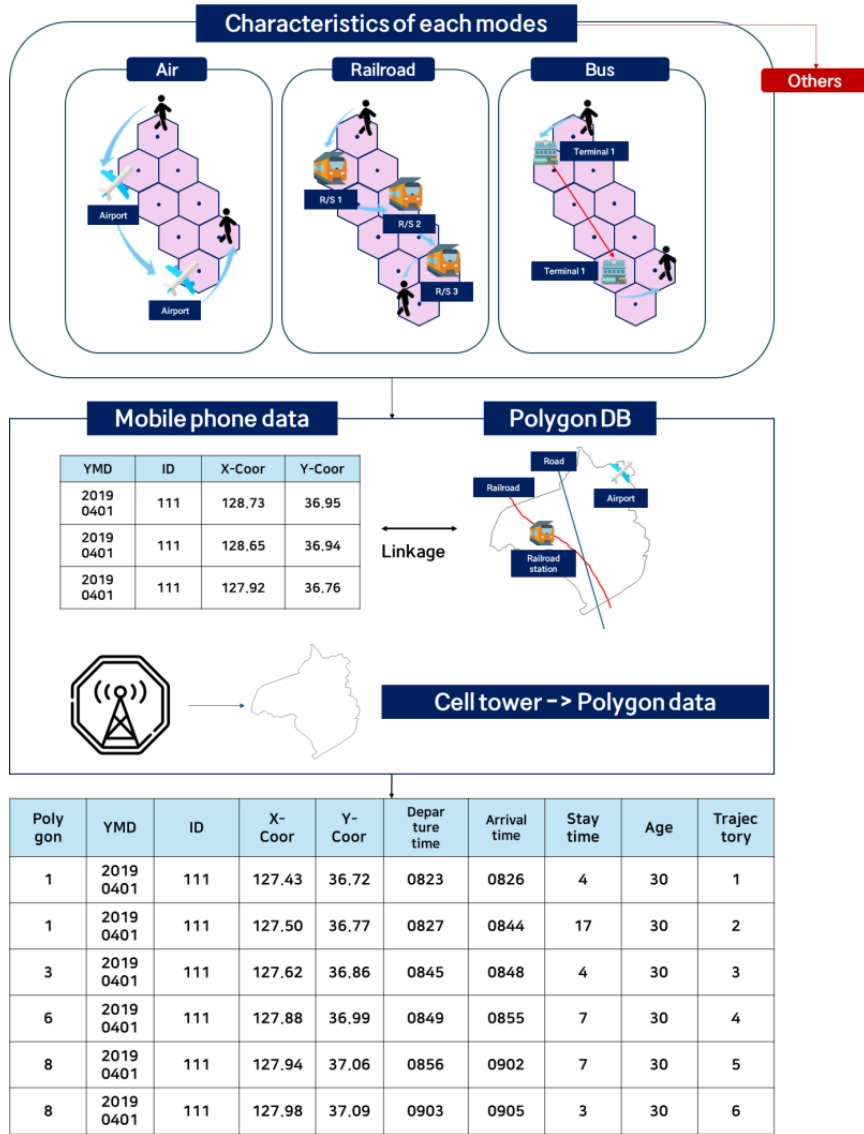


<Fig. 2> The case of high spatial resolution of cell-tower based TAZ
(Kim et al., 2021b)

4. 통행패적과 교통인프라 데이터 결합

본 연구에서 개발하는 지역간 수단통행 분류 방법론은 수단별 운행특성을 기반으로 한다. 예를 들어, 항공 수단통행의 경우, 출발지와 도착지가 모두 공항이고, 운행시간은 보통 90분을 초과하지 않는다. 철도수단통행과 버스수단통행의 경우, 출발지와 도착지가 모두 철도역 또는 터미널이며, 정해진 노선(철도망, 도로망)을 따라 이동한다는 특성이 있다.

이와 같은 지역간 교통수단의 운행특성에 기반한 통행수단 분류 방법론 개발을 위해, 본 연구에서는 <Fig. 3>과 같이 기지국 기반 TAZ 자료에 주요 공항, 철도역, 터미널 등 주요 교통인프라 정보를 DB화하는 작업을 수행하였다.



<Fig. 3> The schematic conceptual proposed map matching process

먼저, 수단별 교통인프라를 조사하고, 공간정보화 되어있는 교통인프라 자료를 수집 또는 제작하였다. 그 후, 수집 또는 제작된 교통인프라 공간데이터의 속성을 GIS 상에서 기지국 기반 TAZ와 매칭하였다. 다음으로, 매칭된 교통인프라가 TAZ 내 존재할 경우, 0 또는 1 이상의 정수형으로 DB화하였으며, 철도수단의 경우, 노선에 따라 순차적으로 역 또는 터미널을 이동한다는 점에서 착안, 각 역사 또는 터미널별로 순서를 부여하였다. 마지막으로, TAZ 번호, 교통인프라 존재유무, 교통인프라명 등을 제외한 불필요한 컬럼을 모두 삭제하였다.

첫 번째로, 항공수단의 경우, 주요 교통인프라는 공항이며, GIS에 도식화하기 위한 별도의 향로 자료를 구득하기 어려우므로 공항 이외 별도로 수단을 추정하기 위한 단서가 존재하지 않는다. 따라서, 구글 맵의 좌

표검색 기능을 활용하여 공항명, 공항 좌표 등으로 구성된 공항 위치데이터를 별도로 구축하였다. 2020년 4월 대한항공 운항노선도를 조회한 결과, 국내에 위치한 공항은 총 14개이나, 인천국제공항이 국제노선 여객수송만을 담당하는 것을 고려하여 제외하고 총 13개의 공항 위치정보를 DB화 하였다. 그 후, 제작된 점 형태의 공항데이터를 기지국 단위 TAZ와 매칭하고, 공항명을 문자형에서 정수형으로 변경하였다. 예를 들어, 광주공항은 1, 군산공항은 2와 같이 변경하였다.

두 번째로, 철도의 경우, 고속철도역 위치정보를 한국교통연구원에서 제공하는 대한민국 철도망 DB에서 추출하였다. 추출한 고속철도역 정보를 DB화하여 기지국 단위 TAZ 자료와 매칭하였다. 철도역은 TAZ 내 존재 유무에 따라 0 또는 1 이상의 정수형으로 표시하며, 노선별로 철도역 경우 순서가 정해져 있는 철도수단의 특성에 따라 정수형으로 순서를 부여하였다. 예를 들어, 서울역은 1번, 광명역은 2번으로 표시하였다.

마지막으로, 버스의 경우, 버스터미널 데이터를 구글 맵의 좌표검색 기능을 활용하였다. 추출한 버스터미널, 톨게이트 및 도로망 정보를 DB화하여 기지국 단위 TAZ 자료와 매칭하였다. TAZ 내 터미널 존재 유무에 따라 데이터는 0 또는 1 이상의 정수형으로 표시하였다.

IV. 분석 결과

1. 지역간 통행수단 추정 방법론 개발

본 연구에서는 주요 교통인프라 정보와 기지국 단위 TAZ, 개별 통행궤적을 결합하여 전체 통행 중 지역간 통행수단인 항공, 철도, 버스 수단통행을 분류하는 방법론을 개발하였다. 방법론을 개발하기에 앞서, 교통인프라 속성이 결합된 기지국 단위 TAZ와 모바일 데이터의 개별 통행궤적을 매칭하였으며, 각 수단별로 비집계 방식의 지역간 통행수단 분류 방법론을 개발하였다.

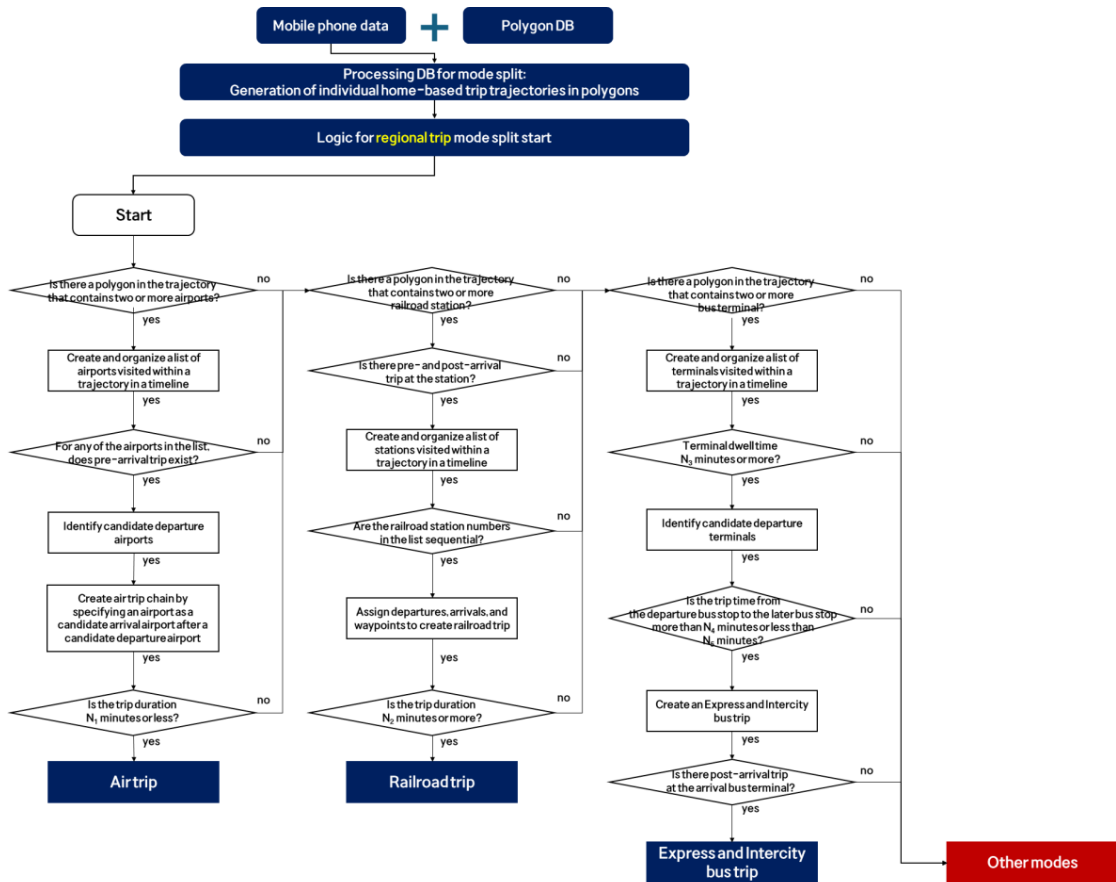
먼저, 항공수단 통행을 분류하기 위한 방법론의 구성은 다음과 같다. 첫 번째, 통행궤적 내 공항이 2개 이상 존재하는지 판단하여 존재하는 경우, 두 번째 단계로 넘어간다. 두 번째, 통행궤적 내 방문한 공항을 시간순으로 정렬하여, 공항 도착 이전 통행 여부를 통해 출발지 공항 후보를 도출한다. 세 번째, 출발지 후보 공항 이후 방문한 공항을 도착지 공항 후보로 지정하고 공항 도착 이후 통행 여부를 확인하여 통행을 생성한다. 마지막으로, 생성된 통행의 통행시간이 N_1 분 이하인지 여부를 판단한다. 내륙 간을 통행하는 경우, 통행시간이 일반적인 국내선 항공수단의 통행시간을 상회하는 경우 타 수단을 이용하였을 확률이 높으므로 타 수단을 이용한 것으로 판단한다. 네 번째, 이동 간에 통과한 TAZ 개수를 판단한다.

철도 통행 분류 방법론의 구성은 다음과 같다. 첫 번째, 통행궤적 내 철도역이 2개 이상 존재하는지 판단하여 존재하는 경우, 두 번째 단계로 넘어간다. 두 번째, 각 철도역 도착 이전과 이후 통행의 존재 여부를 판단하여 존재하는 경우, 세 번째 단계로 넘어간다. 세 번째, 통행궤적 내 방문한 철도역을 시간순으로 정렬하여, 철도역의 역 번호가 순차적인지 확인하는 단계를 통해 출발역과 도착역, 경유역 후보지를 포함한 통행을 생성한다. 마지막으로, 출발 후보역에서 도착 후보역까지의 통행시간이 N_2 분 이상인지 여부를 판단한다.

고속·시외버스 통행 분류 방법론의 구성은 다음과 같다. 첫 번째, 통행궤적 내 버스터미널이 2개 이상 존재하는지 판단하여 존재하는 경우, 두 번째 단계로 넘어간다. 두 번째, 버스터미널에서의 체류시간이 N_3 분 이상인지 판단하여 출발 버스터미널 후보지를 도출한다. 버스터미널의 경우, 그 수가 많아 통행을 과대추정할 수 있으므로, 이동 중에 통행궤적 내에 버스터미널이 단순히 존재하는 것인지 혹은 실제 버스 탑승 대기를 위해 버스터미널에 체류한 것인지를 구분하기 위해 출발터미널에서의 체류시간을 확인하였다. 세 번째,

통행계획 내 방문한 버스터미널을 시간순으로 정렬하여, 출발 버스터미널 후보 이후 방문한 버스터미널 사이 통행시간이 N_4 분 이상 N_5 분 미만인지 여부를 판단하여 통행을 생성한다. 네 번째, 도착 버스터미널 후보 이후 통행이 있는지 확인한다.

항공, 철도, 버스 등 지역간 통행수단을 분류한 뒤 나머지 수단통행은 그 외 수단통행으로 분류하였다. 그 후, 수단별 비집계 지역간 통행수단 분류 정보를 추출 및 DB를 구축하고, 마지막으로 수단별로 지역간 통행수단을 분류하여 집계하였다. 전체 방법론의 흐름도는 <Fig. 4>와 같다.



<Fig. 4> The proposed a regional trip mode classification methodology

2. 방법론 적용 및 검증

통행다이어리를 작성하는 908명의 참여자들(서울 620명, 청주 288명)은 각각에 통행에 이용한 수단을 정수형으로 기입하며(예 : 항공 → 10), 이는 방법론을 통해 도출한 통행수단과 매칭이 가능하다. 그러나 작성 기간 중 누락된 건이 일부 존재하여 전체 지역간 수단통행에 대한 검증은 불가능하며, 통행다이어리가 기록된 통행에 한정하여 검증을 진행하였다. 본 연구에서는 지역간 수단통행 분류 방법론 검증에 가설 검정 이론 중 하나인 Type I and II error 표를 수단별로 작성하여 총 72건의 통행을 검증하였다.

Type I error는 1종 오류라 하고, Type II error는 2종 오류라 하며, 각각 귀무가설을 잘못 기각하는 오류

(False positive)와 귀무가설을 잘못 채택하는 오류(False negative)를 뜻한다. 1종 오류는 귀무가설이 실제로 참이지만, 귀무가설을 기각하는 오류로, 실제 거짓(Negative)인 것을 참(Positive)으로 판정한다. 반대로, 2종 오류는 귀무가설이 실제로 거짓이지만, 귀무가설을 채택하는 오류로, 실제 참(Positive)인 것을 거짓(Negative)으로 판정한다. 본 연구에서는 예를 들어, 항공수단통행을 분류하는 경우 지역간 통행수단 분류 방법론으로도 출된 통행수단(측정값)이 항공이나, 실제 통행 다이어리에 기입된 수단(실제값)은 항공이 아닌 경우를 1종 오류로 판단하고, 방법론으로 도출된 수단(측정값)은 항공이 아니나, 실제 통행 다이어리에 기입된 수단(실제값)이 항공인 경우를 2종 오류로 판단하였다. 철도수단의 추정 결과를 검증할 때는 철도안전법 시행규칙 1조의 4(운행장애의 범위)에 의거하여 고속철도의 경우 20분, 일반철도의 경우 30분의 철도운행지연시간을 고려하여 검증을 수행했다.

검증 결과, 항공수단의 경우 <Table 2>와 같이 측정값과 실제값이 같은 통행이 5건, 1종 오류 0건, 2종 오류 0건으로 나타났으며 F1-Score는 1.000으로 나타났다. 철도수단의 경우 측정값과 실제값이 같은 통행이 26건, 1종 오류 3건, 2종 오류 0건으로 나타났으며 F1-Score는 0.945로 나타났다. 버스수단의 경우 측정값과 실제값이 같은 통행이 23건, 1종 오류 16건, 2종 오류 1건으로 나타났으며, F1-Score는 0.730으로 나타났다.

<Table 2> Verification results of each mode

Air mode		Reality		Precision	1.000
		True	False	Recall	1.000
Measured	True	5	0	Accuracy	1.000
	False	0	31,980	F1-Score	1.000

Rail mode		Reality		Precision	0.897
		True	False	Recall	1
Measured	True	26	0	Accuracy	0.999
	False	3	31,951	F1-Score	0.945

Bus mode		Reality		Precision	0.590
		True	False	Recall	0.958
Measured	True	23	1	Accuracy	0.999
	False	16	31,911	F1-Score	0.730

본 연구에서 개발한 항공, 고속철도, 고속·시외버스 수단통행 분류 방법론은 항공 수단통행을 성공적으로 분류하였으나, 철도 및 버스 수단통행을 분류하는 데 있어 일부 오류를 보였다. 방법론 검증에 활용된 모바일 데이터와 통행 다이어리를 검토한 결과, 오류의 주된 요인은 다음과 같았다.

첫 번째, 모바일 데이터 상에서 특정 시간대 통행궤적 정보가 누락된 케이스이다. 이러한 케이스는 통행 다이어리 상에는 버스 혹은 철도를 이용했다고 기록되었으나, 실제 모바일 데이터 궤적상에서는 버스 혹은 철도 이용시간대 통행궤적 정보가 누락되어 버스 혹은 철도 이용을 검증할 수 없는 경우에 해당한다. 철도 수단 오류 중 3건, 버스 수단 오류 중 2건의 해당 케이스 오류가 발생한 것으로 나타났다.

두 번째, 고속버스 혹은 시외버스를 이용하지 않았음에도 통행 다이어리 상에는 고속버스 혹은 시외버스를 이용했다고 기입한 케이스이다. 이러한 케이스에 해당하는 이용자들은 주로 시내버스나 광역버스를 이용

한 것으로 추정된다. 대부분의 케이스에서 출도착 지점 200m 이내에 버스정류장이 존재하였으며, 궤적 내 고속버스 혹은 시외버스터미널을 이용한 증거가 없는 것으로 나타났다. 버스 수단 오류 중 총 8건의 해당 케이스 오류가 발생한 것으로 나타났다.

세 번째, 통근버스를 이용했으나 통행 다이어리 상에는 고속버스 혹은 시외버스를 이용했다고 기입한 케이스이다. 이러한 케이스에 해당하는 이용자들은 주로 출도착지가 주거단지-산업단지로 일정한 패턴을 보이며, 일과시간에 해당하는 7~16시 어간에 통행궤적 정보가 발생하지 않는 것으로 나타났다. 버스 수단 오류 중 총 4건의 해당 케이스 오류가 발생한 것으로 나타났다.

V. 결 론

본 연구에서는 모바일 데이터 통행궤적과 주요 교통인프라, TAZ 자료를 매칭하여 기계학습에 의존하지 않고 지역간 통행수단을 추정할 수 있는 방법론을 개발하였다. 개발 방법론의 핵심 논리는 출발지·목적지 교통인프라의 존재 여부, 실제 운행예정시간과 통행궤적 내 통행시간의 일치 여부, 철도, 버스(시외) 수단의 경우 경유지의 노선 일치 여부, 항공, 철도 수단의 경우 출·도착 이전·이후의 통행 존재 여부, 버스 수단의 경우 출발 이전 체류시간 존재 여부 등이 포함된다. 방법론의 적용 및 검증을 위해 공항, 철도역, 버스터미널, 톨게이트, 도로망, 철도망 등의 주요 교통인프라 데이터를 모바일 데이터 기반맵 및 모바일 데이터 통행궤적과 매칭하였다. 매칭한 자료를 서울, 청주시역 908명의 통행 다이어리 자료와 비교하여 방법론을 검증한 결과, F1-Score 기반의 방법론 성능은 항공수단 1.00, 철도수단 0.95, 버스수단 0.73 수준으로 나타났다. 이와 같이 본 연구는 실제 통행을 기록한 통행 다이어리 데이터와의 검증을 통해 지역간 통행수단 분류 방법론이 항공 및 철도 수단을 분류하는 데 있어 우수한 성능을 보임을 확인하였다. 본 연구에서 개발한 방법론은 모바일 데이터 개별 통행궤적과 고해상도의 TAZ를 사용함으로써, 지역적, 인적 특성을 모두 고려할 수 있으며, 향후 주차장을 기·중점으로 하는 공유자동차 수단 또는 도심항공모빌리티 수단의 OD도 추정할 수 있을 것으로 사료된다.

그러나 본 연구에서 제안된 방법론은 수단별 교통인프라와의 일치성에 초점을 맞추고 있으므로, 출발지·도착지 정류장이 존재하지 않는 개인 교통수단과 비교적 통행거리가 짧은 교통수단의 추정을 어려울 수 있으며, 지역내 통행량을 추산할 수 없기에 지역간 통행량이 지역내 통행 대비 얼마나 되는지 등의 통계량을 예측할 수 없다. 또한, 고속철도와 일반철도, 고속버스와 시외버스, 광역버스 등 철도 수단과 버스 수단을 세분화해서 추정하지 못한다는 한계점이 존재한다. 따라서 향후 연구에서는 모바일 데이터 기반맵 이상의 공간적 해상도를 가지는 집계구 단위의 기반맵을 사용하거나, 대중교통카드, 공유자동차, 공유킵보드 이력데이터, 내비게이션 데이터 등 이종데이터와의 결합을 통해 지역내 통행수단을 추정하는 연구가 수행된다면, 교통수요추정의 수단선택 단계에 있어 더욱 신뢰도 있고 간결한 방법론이 개발될 수 있을 것이라 판단된다.

REFERENCES

- Alam, M. G. R., Haque, M., Hassan, M. R., Huda, S., Hassan, M. M., Strickland, F. L. and AlQahtani, S. A.(2023), "Feature cloning and feature fusion based transportation mode detection using convolutional neural network", *IEEE Transactions on Intelligent Transportation Systems*, vol.

- 24, no. 4, pp.4671-4681.
- Anda, C., Erath, A. and Fourie, P. J.(2017), “Transport modelling in the age of big data”, *International Journal of Urban Sciences*, vol. 21, no. sup1, pp.19-42.
- Andersson, A., Engelson, L., Börjesson, M., Daly, A. and Kristoffersson, I.(2022), “Long-distance mode choice model estimation using mobile phone network data”, *Journal of Choice Modelling*, vol. 42, 100337.
- Bandyopadhyay, S., Datta, A., Ramakrishnan, R. K. and Pal, A.(2024), “Generalizable journey mode detection using unsupervised representation learning”, *IEEE Transactions on Intelligent Transportation Systems*.
- Brändle, N.(2021), “Inferring modal split from mobile phones: Principles”, *Issues and Policy Recommendations*.
- Breyer, N., Gundlegård, D. and Rydergren, C.(2021), “Travel mode classification of intercity trips using cellular network data”, *Transportation Research Procedia*, vol. 52, pp.211-218.
- Bwambale, A., Choudhury, C. F., Hess, S. and Iqbal, M.(2021), “Getting the best of both worlds: A framework for combining disaggregate travel survey data and aggregate mobile phone data for trip generation modelling”, *Transportation*, vol. 48, no. 5, pp.2287-2314.
- Chen, C., Bian, L. and Ma, J.(2014), “From traces to trajectories: how well can we guess activity locations from mobile phone traces?”, *Transportation Research Part C: Emerging Technologies*, vol. 46, pp.326-337.
- Chen, C., Ma, J., Susilo, Y., Liu, Y. and Wang, M.(2016), “The promises of big data and small data for travel behavior (aka human mobility) analysis”, *Transportation Research Part C: Emerging Technologies*, vol. 68, pp.285-299.
- Cho, S. H., Seo, Y. H., Kho, S. Y. and Rhee, S. M.(2017), “Estimation of interregional mode choice models and value of travel time accommodating taste variation of individuals”, *Journal of the Korean Society for Railway*, vol. 20, no. 2, pp.288-298.
- Currans, K. M. and Clifton, K. J.(2015), “Using household travel surveys to adjust ITE trip generation rates”, *Journal of Transport and Land Use*, vol. 8, no. 1, pp.85-119.
- Doyle, J., Hung, P., Kelly, D., McLoone, S. F. and Farrell, R.(2011), *Utilising mobile phone billing records for travel mode discovery*.
- Engelson, L.(2021), *Long-distance mode choice model estimation using mobile phone network data*.
- Eromietse, E. J. and Joseph, O. O.(2019), “Comparative assessment of radial basis function neural network and multiple linear regression application to trip generation modelling in Akure Nigeria”, *International Journal for Traffic & Transport Engineering*, vol. 9, no. 2, pp.163-176.
- Etu, J. E. and Oyedepo, O. J.(2018), “Forecasting trip generation for high density residential zones of Akure, Nigeria: Comparability of artificial neural network and regression models”, *Journal of Civil Engineering, Science and Technology*, vol. 9, no. 2, pp.76-86.
- Fourez, T., Verstaavel, N., Migeon, F., Schettini, F. and Amblard, F.(2023), *Transport mode detection on GPS and accelerometer data: A temporality based workflow*.
- Jiang, S., Ferreira, J. and Gonzalez, M. C.(2017), “Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore”, *IEEE Transactions on Big Data*, vol. 3, no. 2, pp.208-219.

- Jun, S. H.(2011), “A new statistical sampling method for reducing computing time of machine learning algorithms”, *Journal of The Korean Institute of Intelligent Systems*, vol. 21, no. 2, pp.171-177.
- Kalatian, A. and Shafahi, Y.(2016), “Travel mode detection exploiting cellular network data”, *In MATEC Web of Conferences, EDP Sciences*, vol. 81, 03008.
- Kim, J. Y., Kim, S. J., Lee, G. J. and Choo, S. H.(2021a) “Estimating a mode choice model considering shared E-scooter service - Focused on access travel and neighborhood travel”, *The Journal of Korea Institute of Intelligent Transport Systems*, vol. 20, no. 1, pp.22-39.
- Kim, K. H., Lee, D. Y., Kim, D. H., Won, M. S., Hong, S. M. and Song, T. J.(2021b), “A study on the classification and understanding of travel boundary by city population scale based on mobile travel data”, *Journal of Korean Society of Transportation*, vol. 39, no. 5, pp.662-680.
- Lee, G. J. and Choe, G. J.(2010), “A bike mode share estimation model and analysis of the bike demand factor effects”, *Journal of Korean Society of Transportation*, vol. 28, no. 3, pp.145-155.
- Phithakkitnukoon, S., Sukhvilul, T., Demissie, M., Smoreda, Z., Natwichai, J. and Bento, C.(2017), “Inferring social influence in transport mode choice using mobile phone data”, *EPJ Data Science*, vol. 6, pp.1-29.
- Qu, Y., Gong, H. and Wang, P.(2015), “Transportation mode split with mobile phone data”, *In 2015 IEEE 18th international conference on intelligent transportation systems*, IEEE, September, pp.285-289.
- Ryu, S. K., Rho, J. H. and Kim, J. E.(2012), “A study on the modal split model using zonal data”, *Journal of Korean Society of Transportation*, vol. 30, no. 1, pp.113-123.
- Sadeghvaziri, E., Rojas IV, M. B. and Jin, X.(2016), “Exploring the potential of mobile phone data in travel pattern analysis”, *Transportation Research Record*, vol. 2594, no. 1, pp.27-34.
- Shin, K.(2014), “Improving methods for estimating transportation mode choice model in Busan-Ulsan metropolitan area”, *Journal of the Korea Academia-Industrial cooperation Society*, vol. 15, no. 7, pp.4580-4587.
- Toole, J. L., Ulm, M., González, M. C. and Bauer, D.(2012), “Inferring land use from mobile phone activity”, *In Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, August, pp.1-8.
- Wang, H., Calabrese, F., Di Lorenzo, G. and Ratti, C.(2010), “Transportation mode inference from anonymized and aggregated mobile phone call detail records”, *In 13th International IEEE Conference on Intelligent Transportation Systems*, IEEE, September, pp.318-323.
- Won, M., Choi, J., Lee, H. and Kim, J.(2021), “Development of an algorithm to analyze home-based work trips using mobile phone data”, *Journal of Korean Society of Transportation*, vol. 39, no. 3, pp.383-398.
- Xu, D., Song, G., Gao, P., Cao, R., Nie, X. and Xie, K.(2011), “Transportation modes identification from mobile phone data using probabilistic models”, *In International Conference on Advanced Data Mining and Applications*, Springer, Berlin, Heidelberg, December, pp.359-371.

APPENDIX

Pseudo code of regional trip modes classification methodology

```

1   S: dataframe of all user's home-based travel trajectories
2   serial_num: list of all user's serial number
3   air_travels: arrival and departure pairs of air travel
4   railroad_travels: arrival and departure pairs of air travel
5   bus_travels: arrival and departure pairs of air travel
6   N1 is a maximum travel time of the airport travel
7   N2 is a minimum travel time of the railroad travel
8   N3 is a minimum dwell time of the departure bus terminal
9   N4 is a minimum travel time of the bus travel
10  N5 is a maximum travel time of the bus travel
11  airport_cnt ← 0      /* # of air travel */
12  non_airport_cnt ← 0  /* # of non-air travel */
13  railroad_cnt ← 0     /* # of railroad travel */
14  non_railroad_cnt ← 0 /* # of non-railroad travel */
15  bus_cnt ← 0          /* # of bus travel */
16  non_bus_cnt ← 0     /* # of non-bus travel */
17  other_mode_cnt ← 0  /* # of other mode travel */
18
19  For i ← 0 to len(serial_num)
20      tmp ← extract data from S which satisfying S['serial_number'] = serial_num[i]
21      If len(the ploygon including airport) >= 2 in tmp Then
22          departure_data ← extract data from tmp corresponding to the initial travel
23                          to the polygon including the airport
24          departure_data_sorted ← sort airport_data by time
25          departure_data_final ← drop data with non-existent pre-arrival trips in departure_data_sorted
26
27          For i ← 0 to len(departure_data_sorted-1)
28              If i != len(departure_data_sorted-1) Then
29                  arrival_data_final ← extract data from tmp which is polygon including the airport
30                                      between departure_data_final[i].index and departure_data_final[i+1].index
31              Else
32                  arrival_data_final ← extract data from tmp which is polygon including
33                                      the airport after departure_data_final[i].index
34              End if
35          End for
36          air_travels ← merge departure_data_final and arrival_data_final
37          air_travels ← sort air_travel by time
38      End for
39      For j ← 0 to len(air_travels_final - 1)

```

```

36         timedelta ← air_travels[2 * j + 1]['in_time'] - air_travels[2 * j]['out_time']
37     If timedelta <= N1 Then
38         airport_cnt ← airport_cnt + 1
39     End if
40 End for
41 non_airport_cnt ← non_airport_cnt + 1
42 Else if len(the ploygon including railroad) >= 2 in tmp Then
43     railroad_data ← extract data from tmp corresponding to the initial travel to the polygon including the airport
44     For i ← 0 to len(raiload_data)
45         If data with non-exist pre- and post-arrival trip at the railroad_data[i].index in tmp
46             drop data
47         End if
48     End for
49     For i ← 0 to len(raiload_data-2)
50         railroad_data['delta'] ← railroad_data[i+1]['station_code'] - railroad_data[i]['station_code']
51     End for
52     Initialize railroad_travels
53     tmp_travel ← []
54     For i ← 0 to len(raiload_data-2)
55         If railroad_data['delta'][i] * railroad_data['delta'][i+1] > 0
56             append data to tmp_travel
57         Else
58             append tmp_travel to railroad_travels
59             tmp_travel ← []
60         End if
61     End for
62     For i ← 0 to len(railroad_travels)
63         timedelta ← railroad_travels[i][len(railroad_travels[i]-1)]['in_time'] - railroad_travels[i][0]['out_time']
64         If timedelta >= N2 Then
65             railroad_cnt ← railroad_cnt + 1
66         End if
67     End for
68     non_railroad_cnt ← non_railroad_cnt + 1
69 Else if len(the ploygon including bus terminal) >= 2 in tmp Then
70     terminal_data ← extract data from tmp corresponding to the initial travel to the polygon including the busstop
71     departure_data ← []
72     For i ← 0 to len(terminal_data)
73         If terminal dwell time at terminal_data[i]['terminal_code'] >= N3 Then
74             append data to departure_data
75         End if
76     End for

```



```

77     Initialize bus_travels
78     tmp_travel ← []
79     For i ← 0 to len(terminal_data-1)
80         departure ← terminal_data[i]
81         append data to tmp_travel
82         For j ← i+1 to len(terminal_data-1)
83             arrival ← terminal_data[j]
84             timedelta ← arrival['in_time'] - departure['out_time']
85             If timedelta is between  $N_4$  and  $N_5$  Then
86                 append data to tmp_travel
87                 append tmp_travel to bus_travels
88                 i ← j+1
89                 break
90             End if
91         End for
92     End for
93     For i ← 0 to (len(bus_travels) // 2 - 1)
94         If there is post-arrival trips in bus_travels[2 * i + 1] Then
95             bus_cnt ← bus_cnt + 1
96         End if
97     End for
98     non_bus_cnt ← non_bus_cnt + 1
99     Else
100         other_mode_cnt = other_mode_cnt + 1
101     End for

```