

스마트카드 자료를 활용한 대중교통 승객의 통행목적 추정*

전인우¹ · 이민혁¹ · 전철민¹*

Estimating the Trip Purposes of Public Transport Passengers Using Smartcard Data*

In-Woo JEON¹ · Min-Hyuck LEE¹ · Chul-Min JUN¹*

요 약

스마트카드 자료에는 개별 승객의 대중교통 이용기록이 저장되고, 이를 활용하면 정류장별, 시간대별 통행수요를 분석할 수 있다. 다만 스마트카드 자료에는 통행목적이 기록되어 있지 않기 때문에 통근, 통학, 여가 등의 목적별 수요는 설문조사 자료를 기반으로 추정되고 있다. 하지만 설문조사 자료에는 일부 표본의 통행만 기록되어 있어 전반적인 대중교통 통행수요를 추정하는데 한계가 있다. 만약 스마트카드 자료에서 통행목적을 추정할 수 있다면, 전수조사에 가까운 통행목적별 대중교통 수요에 대한 분석이 가능하다. 이에 본 연구에서는 스마트카드 자료에 기록된 승객의 O-D 통행빈도, 체류 시간, 출발 시각 등을 고려하여 통근, 통학, 귀가의 통행목적을 추정하는 방법론을 제시한다. 결과적으로 제시한 방법론을 적용하여 승객 중 근로자와 대학생을 분류하였다. 제시한 방법론의 검증으로는 가구통행실태조사 자료의 목적별 통행패턴과 본 연구를 통해 추정한 목적별 통행패턴을 비교하였다.

주요어 : 스마트카드, 대중교통, 통행목적, 통행패턴

ABSTRACT

The smart card data stores the transit usage records of individual passengers. By using this, it is possible to analyze the traffic demand by station and time. However, since the purpose of the trip is not recorded in the smart card data, the demand for each purpose such as commuting, school, and leisure is estimated based on the survey data. Since survey data includes only some samples, it is difficult to predict public transport demand for each purpose close to the complete enumeration survey. In this study, we estimates

2019년 03월 04일 접수 Received on March 04, 2019 / 2019년 03월 15일 수정 Revised on March 15, 2019 / 2019년 03월 18일 심사완료 Accepted on March 18, 2019

* 본 연구는 국토교통부 국토교통기술촉진연구사업의 연구비지원(18CTAP-C133228-02)에 의해 수행되었습니다.

1 서울시립대학교 공간정보공학과, Dept. of Geoinformatics, University of Seoul

* Corresponding Author E-mail : cmjun@uos.ac.kr

the purposes of trip for individual passengers using the smart card data corresponding to the complete enumeration survey of public transportation. We estimated trip purposes such as commute, school(university) considering frequency of O-D, duration, and departure time of a passenger. Based on this, the passengers are classified as workers and university students. In order to verify our methodology, we compared the estimation results of our study with the patterns of the survey data.

KEYWORDS : *smartcard, public transportation, trip purposes, travel pattern*

서 론

스마트카드 자료에는 대중교통 이용객들의 승·하차 정류장 및 시간, 이용 노선, 환승 등의 통행정보가 기록된다. 서울시 및 수도권의 하루 평균 대중교통 이용객 수는 천만 명이 넘고, 이 중 98%가 스마트카드를 사용한다(Bin *et al.*, 2011). 즉, 스마트카드 자료는 대표적인 교통 빅데이터이며, 대중교통 이용실태에 대한 전수 조사 자료로써 활용 가치가 높다. 이로 인해 시간대별/수단별 수요 및 정류장 단위의 O-D (Origin-Destination)별 수요 등 대중교통 수요에 대한 다양한 분석이 스마트카드 자료를 바탕으로 이루어졌다(Han *et al.*, 2015; Yang *et al.*, 2018).

하지만 통행목적별 수요를 파악하기 위해 스마트카드 자료를 활용하는 것에는 한계가 있다. 스마트카드 자료에는 승객의 통행목적이 기록되지 않기 때문이다. 통행목적별 수요를 분석한 기존 연구에서는 가구통행실태조사와 같은 설문조사 자료를 주로 이용하였다. 가구통행실태조사에는 거주지, 성별, 연령, 직업 등 개인의 인구 사회적 속성과 통행목적이 포함된 통행정보가 기록되어 있다. 이로 인해 인구 속성 및 통행목적과 연계한 대중교통 이용패턴에 대한 분석이 가능하다는 장점이 있지만, 개인정보 노출 우려로 인해 거시적인 교통존 단위의 통행정보가 기록되어 있고, 설문조사 방식이기 때문에 전수조사에 한계가 있다.

본 연구에서는 스마트카드 자료에 나타난 통행정보를 활용하여 승객의 통행 목적을 추정하는 방법론을 제시하고자 한다. 전수에 가까운 대중교통 통행수요를 알 수 있는 스마트카드 자료에서 통행 목적을 추정할 수 있다면, 설문조사 자료를 이용하는 것보다 정확한 통행목적별 수요를 추정할 수 있다. 이를 위해 본 연구에서는 승객의 승·하차 정류장 정보를 이용하여 거주지 및 활동지를 추정하고, 거주지와 활동지를 오가는 통행의 빈도, 출발 시각 및 체류 시간 등을 분석하여 통근, 통학 등의 통행 목적을 분류하는 방법론을 개발하였다. 제안하는 방법론의 검증은 스마트카드 자료에서 추정한 통행목적별 패턴과 가구통행실태조사 자료에서의 통행목적별 패턴을 비교·분석하여 수행하였다.

관련 연구 분석

스마트카드 자료를 바탕으로 대중교통 이용객의 통행 목적을 추정한 관련 연구에서는 승객의 거주지 및 활동지를 분석하고, 분석된 거주지와 활동지를 오가는 통행정보를 이용하여 통행 목적을 추정하고 있다(Barry *et al.*, 2002; Lee and Hickman, 2014; Amaya *et al.*, 2018).

Jun and Dongyuan(2013)은 스마트카드에서 근로자의 거주지를 추정하는 방법론을 제안하였다. 이 연구에서는 오전 침두시에 대중교통을 이용한 승객을 전부 근로자로 가정하고 이들의 오전 침두시 승차 정류장을 거주 지역, 오후 침두시의 승차 정류장을 직장으로 가정하였다. Zou *et al.*,(2018)은 하루 중 첫 번째 통행의

승차 정류장과 마지막 통행의 하차 정류장의 거리가 500m 이내일 경우, 해당 정류장들을 거주지에서 이용한 정류장으로 보았다. Sari *et al.*, (2018)은 런던의 대중교통을 이용한 승객들을 대상으로 거주지 및 활동지를 추정하는 휴리스틱 알고리즘을 제안하였다. 하루 중 첫 번째 통행의 승차 정류장과 마지막 통행의 하차 정류장의 거리가 500m 이내이고, 정류장 이용 횟수가 임계값을 초과하면, 해당 정류장들을 거주지로 결정하였다. 활동지 역시, 동일한 방법으로 추정하였다.

Medina(2018)은 싱가포르의 가구통행실태조사 자료와 스마트카드 자료를 이용하여 통근과 통학 통행을 분류하는 방법론을 제시하였다. 가구통행실태조사 자료를 이용하여 근로자와 학생을 분류하는 이항 선택 모형을 만들고, 이를 스마트카드 자료에 적용하여 통근, 통학, 귀가, 기타로 구분하여 통행목적을 추정하였다. DBSCAN (Density-based Spatial Clustering of Application with Noise) 알고리즘을 이용하여 통행패턴이 유사한 군집을 생성하였고, 이를 통행목적별로 나누어 통행패턴의 차이를 비교하였다. Alsger *et al.*, (2018)도 스마트카드에서 승객의 통행목적들을 추정하기 위해 가구통행실태조사 자료를 활용하였다. 우선, 가구통행실태자료에서 통행목적별 통행량을 교통존 단위로 집계하여, 교통존별 통행목적 비율을 산정한다. 그리고 스마트카드 자료에 나타난 승객의 통행에 대하여, 하차 정류장이 위치한 교통존의 통행목적

비율과 출발시각을 고려하여 통근, 통학, 쇼핑, 여가, 귀가 등의 통행목적들을 추정하였다.

방법론

본 연구에서는 스마트카드 자료에 나타난 승객의 거주지에서 활동지로의 통행빈도, 활동지에서의 체류시간, 활동지로의 출발시각 범위 등을 고려하여 승객의 통행목적들을 추정하는 방법론을 제안한다. 본 연구에서 추정하는 통행목적은 정기적인 통행에 해당하는 통근과 통학이며, 통학은 대학교로의 통행만을 대상으로 한다. 즉, 본 연구를 통해 승객의 통행목적이 추정되면, 해당 승객을 근로자와 대학생, 그 외 유형으로 분류하는 것이 가능하다.

1. 분석 대상

본 연구에서는 통근 및 통학 통행을 추정하기 위해 2015년 10월 12일 ~ 2015년 10월 18일 동안 이용한 스마트카드 자료를 활용하였고, 서울시 및 수도권에 위치한 대중교통 정류장을 이용한 승객을 대상으로 하였다. 이 중 5일간 10회 이상의 통행기록을 보유한 성인(Adult) 승객 데이터만 추출하여 분석에 사용하였다. 하루의 통행은 오전 4시부터 시작한다고 가정하여 오전 0시~3시까지의 통행은 이전 날짜 통행의 연장으로 간주하였다. 표 1은 본 연구에서 활용한 스마트카드 데이터의 구조로, 개별 이용객의 대중교통 이용정보(승·하차 정류장 및 시간,

TABLE 1. An example of smartcard data

PersonID	RoutelD	Departure stationID	Arrival stationID	Departure Time	Arrival Time	Person type
1	3121	71543	801493	2015-10-12 08:30:11	2015-10-12 09:28:47	Adult
1	3121	801511	71530	2015-10-12 21:07:05	2015-10-12 22:18:19	Adult
1	Line 1	151	332	2015-10-13 08:10:05	2015-10-13 09:15:51	Adult
1	Line 3	332	151	2015-10-13 22:50:15	2015-10-13 00:07:32	Adult
2	3324	9001874	9010311	2015-10-15 12:10:35	2015-10-15 14:35:11	Adult

TABLE 2. The variables in the algorithm

Variable	Description
S^R	Set of stations in residential area
S^A	Set of stations in activity(work, university) area
k	Travel date(day)
$S_{f,d}^k$	Departure station of first trip on the k-th day
$S_{f,a}^k$	Arrival station of first trip on the k-th day
$S_{l,d}^k$	Departure station of last trip on the k-th day
$S_{l,a}^k$	Arrival station of last trip on the k-th day
$Z(S)$	Set of administrative zones adjacent to station

이용 노선, 승객 유형 등)가 기록되어 있다.

2. 거주지 및 활동지 추정

본 연구에서 승객의 거주지란, 하루 첫 통행의 출발지역이면서 마지막 통행의 도착지역으로, 정류장 집합을 의미한다. 활동지는 거주지에서 출발하여 도착한 지역으로, 거주지와 마찬가지로 정류장 집합으로 산출된다.

거주지 및 활동지 추정 알고리즘은 다음과 같으며, 알고리즘에 사용된 변수들에 대한 설명은 표 2와 같다.

Step. 1 $S^R = \emptyset$, $S^A = \emptyset$, k

Step. 2 만약, $Z(S_{f,d}^k) \cap Z(S_{l,a}^k) \neq \emptyset$ 라면, S^R 에 $S_{f,d}^k$ 추가, S^R 에 이미 $S_{f,d}^k$ 가 포함되어 있다면 Step 3 진행

Step. 3 만약, $Z(S_{f,a}^k) \cap Z(S_{l,d}^k) \neq \emptyset$ 라면, S^A

에 $S_{f,a}^k$ 추가, S^A 에 이미 $S_{f,a}^k$ 가 포함되어 있다면 Step 4 진행

Step. 4 k 를 1 증가시켜 Step 2 진행, k 를 더 이상 증가시킬 수 없다면, 알고리즘 종료

그림 1은 표 1의 ID가 1번인 승객의 거주지 및 활동지를 추정하는 과정을 설명하기 위한 예시이다. 표 1에서 승객의 12일의 $S_{f,d}^k$ 는 71543번 정류장이고 $S_{f,a}^k$ 는 801493번 정류장, $S_{l,d}^k$ 는 801511번 정류장, $S_{l,a}^k$ 는 71530번 정류장이다. Step 2를 진행하여 $Z(S_{f,d}^k)$ 를 계산해보면 71543번 정류장과 인접한 {서초4동, 서초2동, 반포1동}이 계산되고, $Z(S_{l,a}^k)$ 는 71530번 정류장과 인접한 {반포1동, 서초1동, 서초2동}으로 계산된다. 이를 이용하여 $Z(S_{f,d}^k)$ 와 $Z(S_{l,a}^k)$ 에 공통으로 속한 행정구역이 있는지 확인

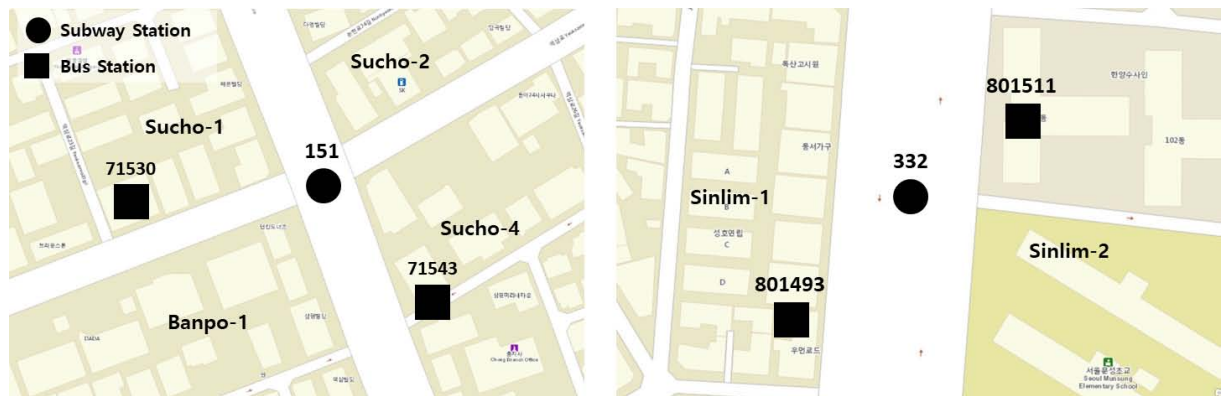


FIGURE 1. Estimating residence zones and activity zones of passengers

해 보면 {반포1동, 서초2동}이 있고 이로 인해 71543번 정류장이 S^R 에 저장된다. 동일한 방법으로 Step 3을 진행하여 $Z(S_{f,a}^k)$ 와 $Z(S_{t,d}^k)$ 를 확인하면 {신림1동, 신림2동}이 공통으로 속한 행정구역으로 계산되고 S^A 에 801493번 정류장을 저장한다. 이를 알고리즘이 종료될 때까지 진행해보면 ID가 1번인 승객의 거주지(S^R)에는 {71530, 151}이 속하게 되고, 활동지(S^A)에는 {801493,332}가 속하게 된다.

3. 통행목적 추정

승객의 스마트카드 데이터에 앞서 분석된 거주지에서 활동지로 3일 이상 통행한 기록이 포함되어 있을 경우, 해당 승객은 통근 및 통학 목적의 정기적인 통행을 한다고 가정하였다. 통행목적은 활동지의 유형(대학교 존(zone)), 활동지에서의 체류시간, 활동지로의 출발시각 범위를 고려하여 통근과 통학으로 분류하였고 거주지로의 통행은 귀가로 분류하였다. 표 3은 통행목적 분류 기준을 나타낸 것이다. 표 3에서 O는 승객의 거주지이며, D는 승객의 활동지를 의미한다.

대학교 존은 그림 2와 같이 버스 정류장은 500m, 지하철역은 700m 버퍼 분석을 통해 결정하였다. 버스 정류장이 대학교 필지에서 500m 내에 있거나 지하철역이 700m 내에 있다면, 해당하는 정류장들은 대학교 존으로 분류하였다.

체류시간 기준은 가구통행실태조사의 근로자 및 대학생 활동시간 통계치를 바탕으로 결정하였다. 그림 3은 가구통행실태조사에 나타난 근로자 및 대학생의 활동시간을 누적분포함수로 나타낸 것이다. 통근 통행 이후의 활동시간은 8시간 이상에서, 통학 통행(성인 기준) 이후의 활동시간은 4시간 이상에서 누적분포가 급격히 증가하는 것을 알 수 있다. 이에 본 연구에서는 활동지에서의 체류시간이 8시간 이상일 경우, 통근 목적으로, 활동지가 대학교 존이면서 체류시간이 4시간 이상일 경우, 통학 목적으로 승객의 통행을 분류하였다.

또한 활동지로의 출발시각 범위는 최대 출발시각과 최소 출발시각의 차이를 의미하여 차이가 1시간 이내일 경우, 통근으로, 1시간 초과일 경우, 통학으로 분류하였다. 즉, 승객의 활동지

TABLE 3. Rules for estimating trip purposes

Trip purpose	Number of O-D trips(day)	Destination	Duration(h)	Range of departure time(h)	Passenger type
Commute	$N \geq 3$	-	$D_t \geq 8$	$\lambda \leq 1$	Adult
University		University zone	$D_t \geq 4$	$\lambda > 1$	Adult
Home		All trips where the destinations are residential areas			

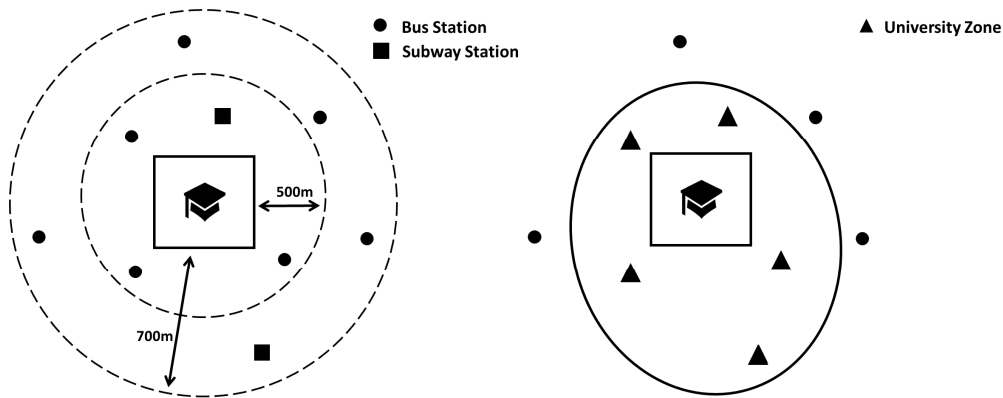


FIGURE 2. Determining university zones

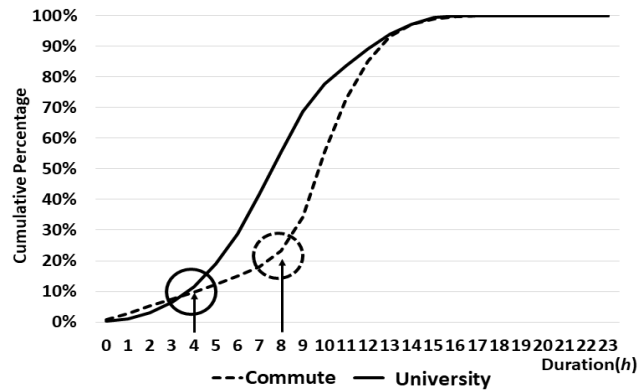


FIGURE 3. The cumulative distribution of duration for commute and school in survey data

로의 통행 데이터가 있을 때, 출발시각의 범위가 1시간 이내이면 근로자로, 1시간을 초과할 경우, 대학생으로 분류한 것이다. 이는 출근 시간이 비교적 일관적인 근로자와 수강 시간표에 따라 통학 시간이 자유로운 대학생을 구분하기 위함이다.

이 중 거주지와 활동지가 추정된 승객은 약 86%이며, 근로자로 분류된 승객은 약 150만 명, 대학생으로 분류된 승객은 약 15만 명으로 나타났다. 표 4는 본 연구의 방법론을 통해 표 1에 나타난 승객의 통행 목적을 추정한 예시이다.

실험 및 결과

본 연구에서는 2015년 10월 12일 ~ 2015년 10월 18일에 스마트카드를 통해 서울시 및 수도권권의 대중교통 정류장을 이용한 승객의 통행 목적을 추정하였다. 스마트카드 자료에 있는 약 1억 건의 대중교통 이용기록 중 10회 이상 통행 기록이 있고 승객 유형이 '성인' 인 약 574만 명의 승객을 연구대상으로 설정하였다.

1. 거주지 및 활동지의 공간분포

그림 4는 본 연구를 통해 근로자 및 대학생으로 분류된 승객들의 거주지 및 활동지 분포를 시각화한 것이다. 붉은 지역일수록 더 많은 승객들이 거주하거나, 활동하는 지역임을 의미한다. (a)는 거주지 분포에 해당하며, 서울시의 주거중심지 역할을 하는 강북 지역(도봉구, 강북구 등)과 강남, 강서지역(관악구, 양천구 등)을

TABLE 4. Smartcard data with trip purposes estimated

PersonID	Departure stationID	Arrival stationID	Departure Time	Arrival Time	Person type	Trip purpose
1	71543	801493	2015-10-12 08:30:11	2015-10-12 09:28:47	Adult	Commute
1	801511	71530	2015-10-12 21:07:05	2015-10-12 22:18:19	Adult	Home
1	151	332	2015-10-13 08:10:05	2015-10-13 09:15:51	Adult	Commute
1	332	151	2015-10-13 22:50:15	2015-10-13 00:07:32	Adult	Home
2	9001874	9010311	2015-10-15 12:10:35	2015-10-15 14:35:11	Adult	University

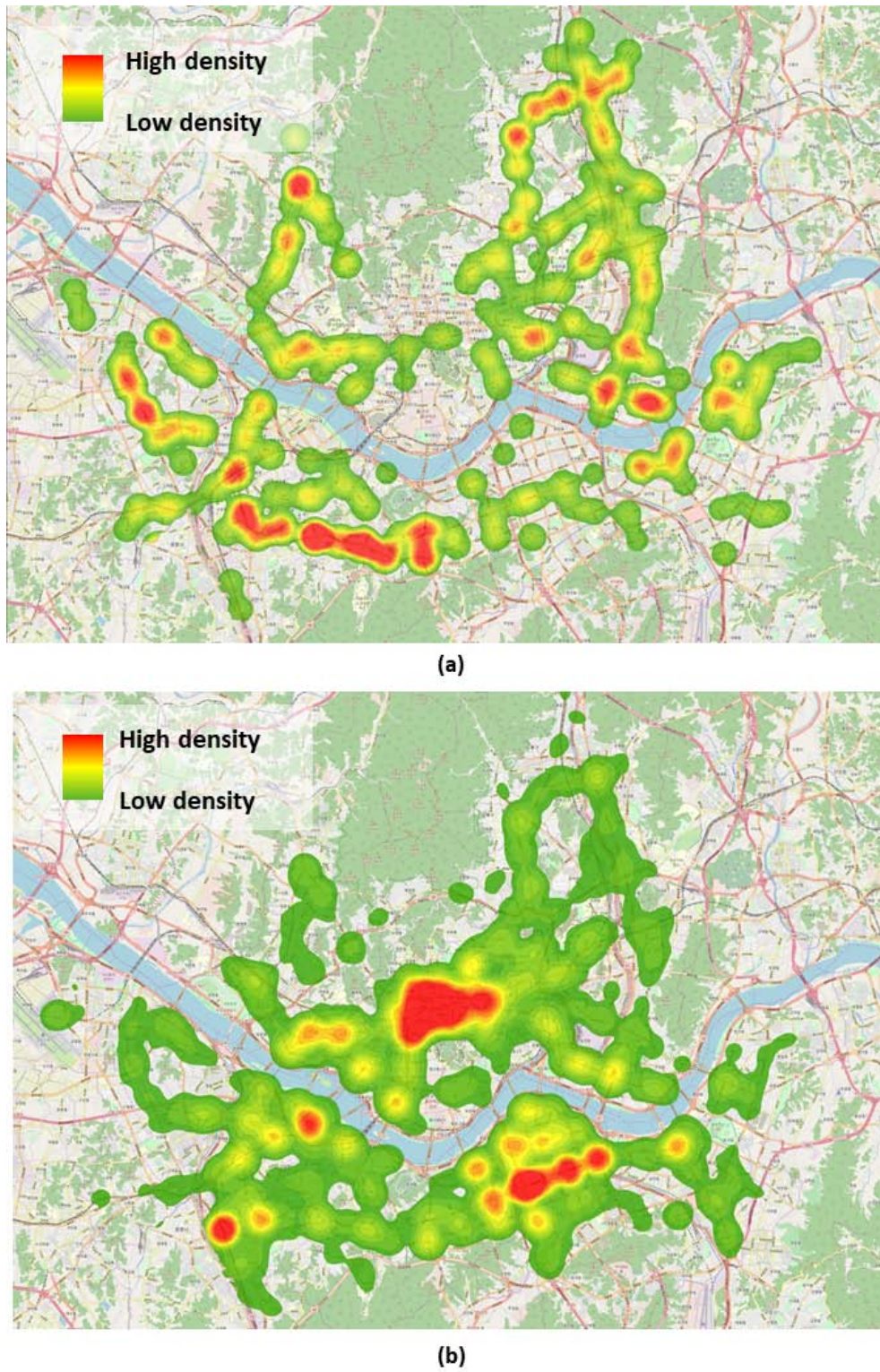


FIGURE 4. Spatial distribution of estimated population of residence zones (a) and activity zones (b)

포함한 서울 외곽 지역에 승객의 거주지가 대부분 추정된 것을 확인할 수 있었다. (b)는 활동지 분포로, 서울 중심부(종로구, 중구 등)와 강남(강남구, 서초구 등), 여의도, 가산디지털단지 등 주요 4대 고용 중심지가 활동지로 분석되었다. 이 외에도 대학 밀집 지역에 해당하는 신촌 일대가 활동지로 나타났다. 이를 통해 본 연구에서 제시한 방법론으로 추정한 승객의 거주지 및 활동지가 실제 토지이용패턴과 유사함을 확인할 수 있었고 기존 대중교통 이용 통계에서 나타난 지역별 수요가 정성적으로 드러난 것을 확인할 수 있었다.

2. 가구통행실태조사와의 통행목적별 패턴

비교

스마트카드 자료에서 분석한 통행목적별 패턴이 실제 통행목적별 패턴과 유사한지 검증하기 위해 2010년 가구통행실태조사 자료를 활용하였다. 표 5는 본 연구를 통해 분석한 스마트카드 자료의 목적별 통행비율과 가구통행실태조사 자료에 나타난 대중교통 이용객의 목적별 통행비율을 비교한 것이다. 통근 통행과 귀가 통행의 비율은 거의 유사하게 추정되었지만 스마트카드 자료에서 통행 목적을 추정하지 못한 기타 통행의 비율이 가구통행실태조사에 비해 높게 나타난 것을 확인할 수 있었다.

그림 5는 통행목적별 출발시각의 분포를 비교한 것이다. 그림 5의 (a)는 통근 통행, (b)는 통근 통행을 한 승객의 귀가 통행, (c)는 통학 통행, (d)는 통학 통행을 한 승객의 귀가 통행의 출발시각별 통행비율을 나타낸 것이다. 통근 통행의 경우, 오전첨두시(7시~9시)에 대부분의 통행이 분포하는 형태를, 본 연구의 결과 및 가

구통행실태조사에서 모두 확인할 수 있었다. 근로자의 귀가 통행 역시, 본 연구의 결과와 가구통행실태조사에 나타난 패턴이 유사함을 보였다. 통학 통행의 경우, 가구통행실태조사에서는 오전 첨두시(7시~9시)에 대부분의 통행이 발생하지만, 본 연구의 결과에서는 10시 이후에도 지속적으로 통학 통행이 발생하는 것을 확인하였다. 대학생의 귀가 통행은, 가구통행실태조사와 본 연구 결과 모두 근로자의 귀가 통행에 비해 첨두시에 집중되는 현상이 덜한 것으로 나타났다.

그림 6은 목적별 통행시간의 분포를 30분 간격으로 나누어 비교한 것이다. 본 연구에서 제안한 방법론을 통해 스마트카드에서 추정한 통근/통학별 통행시간과 가구통행실태조사 자료에 나타난 통근/통학별 통행시간은 매우 유사하게 나타났다. 통근/통학 통행 모두 30분 이내 급간에서 약간의 차이를 보이지만, 전체적으로 모든 급간에서 통행비율이 비슷하게 나타나는 것을 확인하였다.

본 연구를 통해 추정한 스마트카드의 통행목적별 출발시각 및 통행시간 분포를 가구통행실태조사 자료와 비교해본 결과, 전체적으로 유사한 패턴을 보임을 확인할 수 있었다. 대학생으로 분류된 승객들의 통학 및 귀가 통행에 대한 출발시각이 가구통행실태조사와는 다소 상이한 모습을 보였는데, 이는 가구통행실태조사 자료와 본 연구에서 분석한 스마트카드의 시점이 다르기 때문에 발생한 문제로 판단된다. 동일한 시점의 설문조사 자료와 스마트카드 자료를 비교한다면, 본 연구를 통해 추정한 대학생들의 통행패턴을 보다 정교하게 검증할 수 있을 것이라 사료된다.

TABLE 5. Comparison of travel ratio by trip purpose between smartcard data and survey data

Trip purpose	Trips of survey data	Percentage	Trips of smartcard data	Percentage
commute	46,663	39%	1,299,358	39%
school	10,487	9%	110,576	3%
others	6,251	5%	428,823	13%
home	56,419	47%	1,466,100	44%

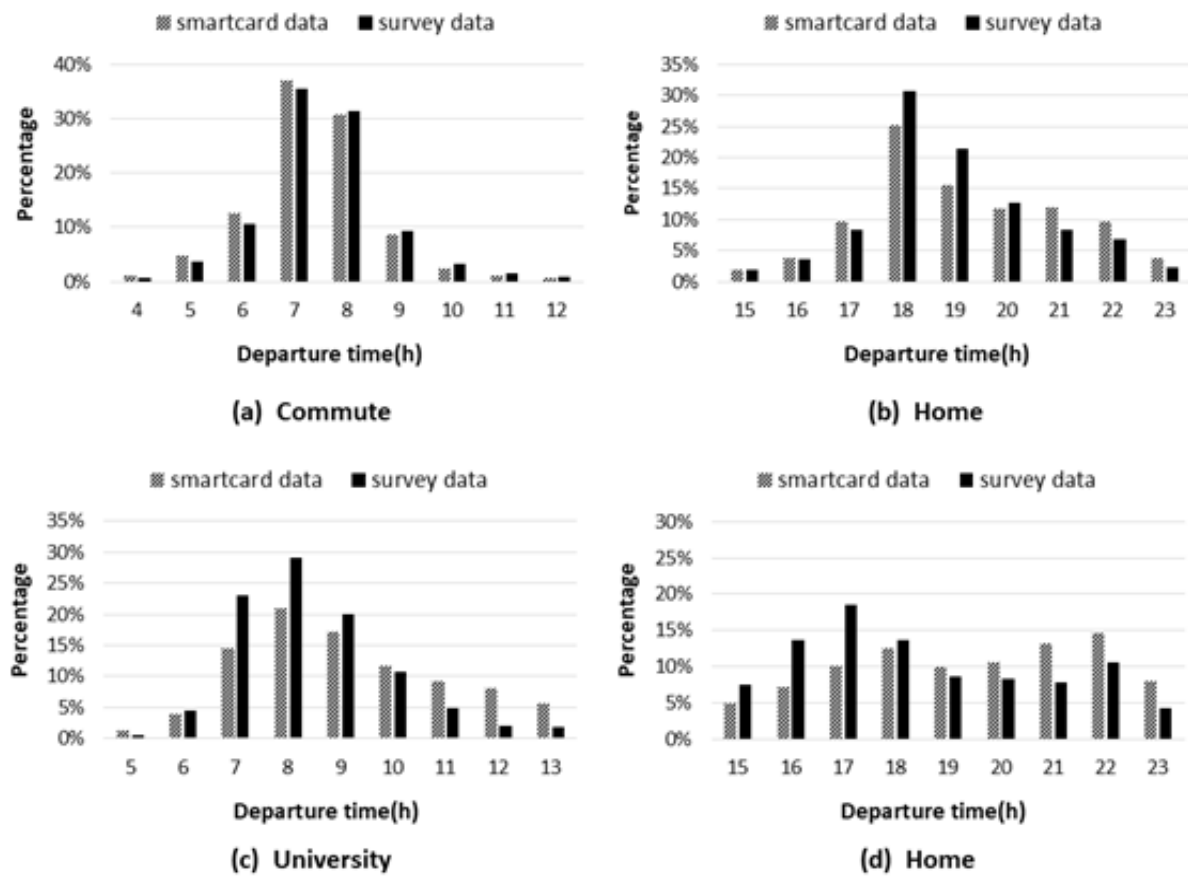


FIGURE 5. Comparing departure times of smartcard data and survey data

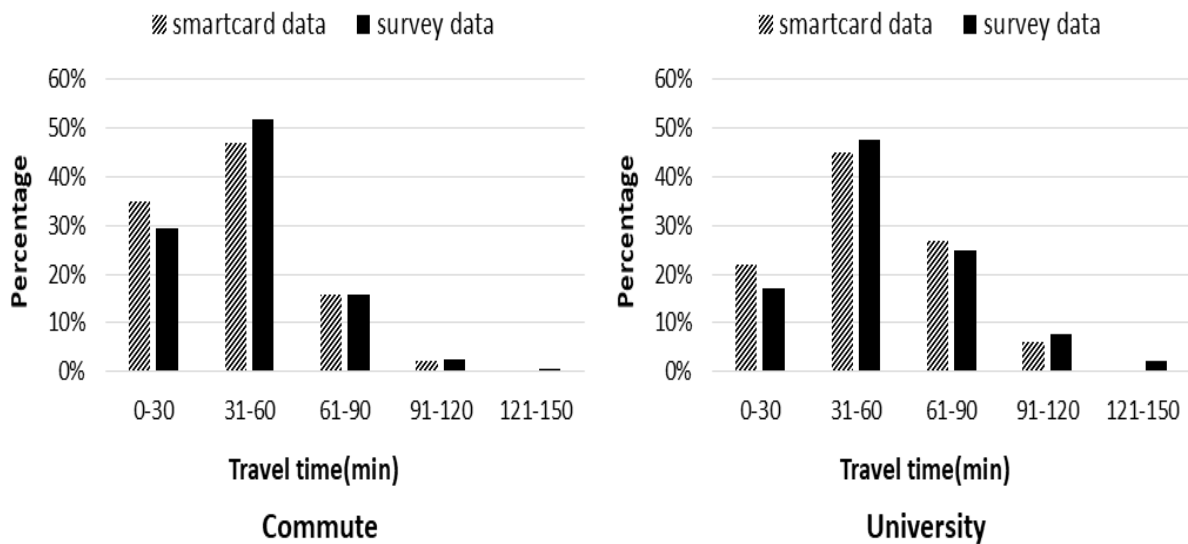


FIGURE 6. Comparing the travel times of between smartcard data and survey data

결 론

본 연구에서는 스마트카드 자료에서 승객의 통행목적별을 추정하는 방법론을 제안하였다. 승객의 통행목적별을 추정하기 위해 승·하차 정류장 정보를 이용하여 승객별 거주지 및 활동지를 탐색하였고, 근로자와 대학생을 구분하기 위해 활동지에 대학교 존을 고려하였다. 이 후 승객의 거주지에서 활동지까지의 통행빈도, 체류시간, 출발시각의 범위 등의 통행정보를 이용하여 통행목적별을 통근, 통학, 귀가로 나누어 추정하였다.

스마트카드 자료에서 추정한 승객들의 거주지 및 활동지는 실제 토지이용패턴과 유사한 모습을 보였다. 스마트카드 자료로부터 추정한 통행목적별 패턴은 가구통행실태조사 자료를 이용하여 목적별 출발시각 및 통행시간 분포 측면에서 비교·분석하였다. 각 지표에 대해서 본 연구의 결과와 가구통행실태조사 자료의 결과가 전반적으로 동일한 패턴을 보였다. 다만, 보다 정교한 검증은 위해서는 동일한 시점에 조사된 스마트카드 및 가구통행실태조사 자료를 이용해야 할 것이라 판단된다.

본 연구는 스마트카드 자료 기반의 통행목적별 수요 분석을 가능하게 한다는 점에서 학술적 의의가 있고, 대학생과 근로자를 분류한다는 점에서 유사 연구와 차별성을 가진다. 최근, 전동킥보드와 같은 개인용 모빌리티 및 카셰어링과 같은 모빌리티 서비스에 대한 관심도가 높아지고 있고, 이에 대한 주요 수요층으로 대학생들이 지목되고 있다. 따라서 스마트카드에서 대학생들을 분류하고, 그들의 대중교통 이용패턴을 분석할 수 있다는 점은 모빌리티 서비스 수요 분석 측면에서 매우 활용가치가 높다고 할 수 있다. 또한 쇼핑, 여가활동 등 비정기적인 목적의 대중교통 통행에 대한 추정 및 도보, 택시 등 타수단과 연계한 분석이 수행된다면, 보다 미시적으로 시민들의 대중교통 이용패턴을 파악할 수 있으리라 사료된다. **KAGIS**

REFERENCES

- Alsger, A., A. Tavassoli, M. Mesbah, L. Ferreira and M. Hickman. 2018. Public transport trip purpose inference using smart card fare data. *Transportation Research Part C: Emerging Technologies*, 87:123–137.
- Amaya, M., R. Cruzat and M. A. Munizaga. 2018. Estimating the residence zone of frequent public transport users to make travel pattern and time use analysis. *Journal of Transport Geography*, 66:330–339.
- Barry, J., R. Newhouser, A. Rahbee and S. Sayeda. 2002. Origin and destination estimation in New York City with automated fare system data. *Transportation Research Record: Journal of the Transportation Research Board*, 1817: 183–187.
- Bin, M.Y., S.J. Lee and E.S. Jung. 2011. A Study on Analysis Travle Patterns and Applications Using Gyeonggi-Do Traffic Card Data. Gyeonggi Research Institute, pp.1–148 (빈미영, 이수진, 정의석. 2011. 경기도 교통카드자료를 이용한 통행패턴 분석과 활용방안 연구. 정책연구, pp.1–148)
- Han, S.U., H.Y. Kang and M.H. Lee. 2015. Residential Distribution of Public Trasport Commuter by Smart Card through the Use of the Major Sub way Influence Area : Fous on Metropolitan, *Journal of Korea Planning Association*, 50(4):103–117 (한상욱, 강희용, 이명훈. 2015. 교통카드 데이터를 활용한 주요 역세권별 대중교통 이용 통근통행자의 주거지 분포. 국토계획, 50(4):103–117)
- Jun, C. and Y. Dongyuan. 2013. Estimating smart card commuters origin–destination

- distribution based on APTS data. *Journal of Transportation Systems Engineering and Information Technology*, 13(4):47–53.
- Lee, S.G. and M. Hickman. 2014. Trip purpose inference using automated fare collection data. *Public Transport*, 6(1–2): 1–20.
- Medina, S.A.O. 2018. Inferring weekly primary activity patterns using public transport smart card data and a household travel survey. *Travel Behaviour and Society* 12:93–101.
- Sari Aslam, N., T. Cheng and J. Cheshire. 2018. A high-precision heuristic model to detect home and work locations from smart card data. *Geo-spatial Information Science*, pp. 1–11
- Yang, H.J., H.W. Nam and C.M. Jun. 2018. Analysing Potential Improvement of Public Transit Services in OD Level Using Time-Distance Accessibility and Smartcard Traffic Volume, *The Korean Association of Geographic Information Studies*, 21(2):80–93 (양현재, 남현우, 전철민. 2018. 시간거리 접근성과 교통카드 기반 통행량을 이용한 OD 별 잠재적 대중교통 서비스 개선량 분석. *한국지리정보학회지*, 21(2):80–93)
- Zou, Q., X. Yao, P. Zhao, H. Wei and H. Ren. 2018. Detecting home location and trip purposes for cardholders by mining smart card transaction data in Beijing subway. *Transportation* 45(3):919–944.