

무선통신 자료를 활용한 통행발생량 분석 Trip Generation Analysis Using Mobile Phone Data

김경태 · 이인묵* · 민재홍 · 곽호찬
Kyoungtae Kim · Inmook Lee · Jae Hong Min · Ho-Chan Kwak

Abstract The recent trend in transportation planning information is to reduce traffic survey costs and enhance accuracy by using and converging various sources of external data. In Korea, mobile phone data can help generate useful transportation planning information, thanks to the universal use of mobile phones, which are present in a number greater than that of the population. This paper addresses measures to derive trip generation information from mobile phone data and verifies the value of the system for practical use by correlation analysis with KTDB trip generation data. The results show that trip generation information produced by mobile phone data correlates with existing (KTDB) trip generation data.

Keywords : Mobile phone data, Cell phone data, Trip generation, Trip production, Trip attraction

초 록 현재 통행발생량 산정 등 교통계획 정보의 생성 체계가 기존 조사 중심 체계에서 외부 데이터를 접목하여 조사 비용을 저감시키고 정확성을 높이는 방향으로 전환되고 있다. 우리나라는 인구보다 많은 휴대전화 가입자가 있기 때문에 이로부터 구축된 무선통신 자료는 교통계획에 매우 유용한 정보를 줄 수 있을 것이다. 본 연구에서는 이동통신사에서 제공하는 수도권 지역 성·연령별 유동인구 자료로부터 교통계획의 중요한 자료인 통행발생량을 산정하기 위한 방안을 제시하고 이를 KTDB의 통행발생량과 상관성 분석을 통하여 자료의 활용 가능성을 확인하였다. 그 결과 무선통신 자료를 이용한 통행발생량 추정은 기존의 KTDB에서 구축한 직접 조사 방식 기반에 의한 결과와 매우 높은 상관관계를 가지는 것으로 분석되었다.

주요어 : 무선통신 자료, 기지국자료, 통행발생, 통행생성, 통행유인

1. 서 론

국가교통데이터베이스(이하, KTDB)는 5년 단위 조사인 가구통행실태조사를 기반으로 기준년도의 OD를 생성하고 있으며, 중간년도는 사회경제지표, 교통량 조사자료, 교통카드 데이터, 수단수송실적 등 수단별 통계지표를 이용한 현행화 과정을 거쳐 OD를 보정하고 있다. 지속적인 연구를 통해 KTDB OD 자료의 신뢰성이 과거 대비 향상된 것으로 평가 되나, 조사기반의 정보생성체계의 한계인 표본수 및 조사항목 제약으로 인하여 여전히 신뢰성 향상을 위한 연구가 필요한 상황이다. 이러한 관점에서, 본 연구에서는 KTDB의 OD 생성 과정 중 첫 단계인 통행발생량 추정 방법론을 중심으로 기존의 방법론을 보완할 수 있는 새로운 방안을 제시하고, 이를 기존 KTDB 추정치와 비교하여 유의성을 검토하고자 한다.

통계청[5]에 따르면 우리나라 2014년 추계인구는 50,423,955명이다. 이에 비하여 이동전화서비스 가입자는 2014년 1월 54,823,782명에서 12월에는 57,207,957명으로 증가[6]하여 인구 1인당 이동전화서비스 가입자는 1.13으로서 인구를 상회하고 있는 실정이다. 즉, 국민 1인당 1대 이상의 이동전화를 보유하고 있어, 이동전화의 위치를 파악하면 그것이 바로 인구의 위치를 파악하는 것이 된다. 즉, 기지국 등 특정 위치에 대한 정보가 포함되는 무선통신 자료를 활용하면 별도의 조사 없이도 실제 인구의 이동을 파악할 수 있다는 점에 착안하여 이를 통행발생 추정에 활용하고자 하는 것이 본 연구의 목표이다. 개인정보 및 위치정보보호 문제로 이동전화서비스 가입자에 대한 직접적인 정보나 위치 정보는 파악할 수 없으나 기지국 단위 등으로 집합화 된 통신 자료를 활용하더라도 동단위 이상으로 세분화되는 분석에는 충분할 것이다.

각 이동통신사에는 엄청난 양의 무선통신 자료가 축적되고 있다. 이미 일부 이동통신사는 이를 활용한 빅데이터 분석을 통하여 관광지나 축제 등을 방문하는 고객들이 어느 곳에서 유입되는지에 대한 정보나 상권의 유동인구 등을 포함한 다양한 서비스를 제

*Corresponding author. Tel.: +82-31-460-5175, E-mail : mook79@krri.re.kr.

© 2015 The Korean Society for Railway. All rights reserved.

<http://dx.doi.org/10.7782/JKSR.2015.18.5.481>

공하고 있으며, 무선통신 자료 자체를 판매 또는 제공하고 있는 실정이다. 무선통신 자료는 여러 부문에서 매우 유용한 정보를 제공할 수 있을 것으로 판단되며, 통신 자료의 특성을 분석하고 활용 가능한 분야를 발굴하려는 노력이 필요한 시점이다. 그 일환으로서 본 연구에서는 이동통신사에서 제공하는 수도권 지역 성·연령별 유동인구 자료¹⁾로부터 교통계획의 중요한 자료인 통행발생량을 산정하는 방안에 대해서 제시하고 이를 KTDB의 통행발생량과의 상관성 분석을 통하여 자료의 활용 가능성을 확인하고자 한다. 다만, 2장의 자료구조에서 알 수 있듯이 셀(cell) 단위로 집합화 된 자료의 한계에 따라 수단통행의 추정에는 한계가 있어, 목적통행의 통행발생만을 범위로 설정하였으며, 셀에서 측정된 통신 기록만으로는 각 무선통신 이용자가 목적지에서 발생시킨 통신 인지 경유지(차량 등)에서 발생시킨 통신인지 판별할 수 없으므로 절대적 수치보다는 상관성 분석을 중심으로 연구를 진행하였다.

2. 본 론

2.1 현황 고찰

본 연구는 무선통신 자료를 이용한 기존(KTDB)의 통행발생량 추정의 보완 가능성을 중점적으로 다루고 있으므로 KTDB의 사례를 중심으로 현황을 고찰한다. 또한 실시간성이 있는 무선통신 자료의 활용성 검증에 있어 별도로 대조군에 해당하는 인구의 참값을 구하는 데에는 큰 한계가 있으므로, 현재 활용 가능한 자료 중 가장 시간적, 공간적으로 근사한 자료로서 KTDB를 대조군으로 설정하였다. KTDB에서 제공하는 수도권 OD의 현행화 과정은 목적통행 현행화, 수단통행 현행화, 검증의 단계로 구성된다. 목적통행 현행화 과정은 다시 3단계로 나뉘는데, 1단계로 인구변화를 반영한 동별 발생통행량을 보정하고, 2단계는 종사자수 변화를 반영한 통행유인량을 보정하며, 3단계는 수용학생수 변화를 반영한 통행유인량을 보정하는 단계로 구성된다. 이렇게 목적통행으로 구분하여 생성 및 유인통행량을 보정하는 것은 모형의 설명력을 높여 신뢰성을 제고하기 위한 것이다[1,2].

KTDB는 여객 OD 추정 과정에서 회귀분석 모형을 적용하고 있다. KTDB에서 통행목적별 발생(trip generation)모형 산정은 권역별로 목적 구분은 동일하나 모형에 반영된 독립 변수는 지역별 특성에 따라 일부 차이가 있다. 대표적으로 수도권의 존 단위 회귀분석 모형의 통행목적별 독립변수는 Table 1과 같다. Table 1에서 보면 통행생성(trip production, 통행유출) 모형과 통행유인(trip attraction, 통행유입) 모형으로 구분하고, 가정 기반 및 비가정 기반 모형으로 구분하여 제시하고 있으며, R² 값을 참고한다면 가정 기반 모형이 비가정 기반 모형에 비해서 상대적으로 설명력이 높은 것을 알 수 있다. 그러나 가정 기반 모형의 통근, 학원 및 기타 통행 이외에는 R² 값이 0.8 수준으로서 설명력이 높지 않은 것을 알 수 있다[2].

본 연구와 KTDB와의 차이점을 보면, 우선 조사방식에 있어 본 연구는 휴대전화 통신기록을 이용하기 때문에 이용 가능한 자료에 제한이 없으나 KTDB의 경우 5년 단위의 가구통행실태조사를 기본으로 하고 있기 때문에 2015년 현재 활용할 수 있는 기준년도 OD는 2012년으로서 시기적으로 차이가 있으며, 중간년도의 현행화 과정도 관련 통계(교통카드 데이터, TCS(Toll Collection System) 데이터 포함)가 작성이 완료된 이후에 진행할 수 있어 1년 이상의 시차를 두고 OD가 보정되고 있는 실정이다. 표본수도

Table 1. Trip production/attraction models of KTDB.

Trip purpose		Production model			Attraction model		
		Variables	%RMSE	R ²	Variables	%RMSE	R ²
Home based trip	Commute	Total employees	42.1	0.9672	Total employees	63.2	0.8894
	School	Pop. between 6 and 17	60.0	0.7989	Total students	62.8	0.8485
	Academy	Pop. between 18 and 21	60.7	0.8954	Employees related to academy	94.5	0.6880
	Shopping	Pop. between 6 and 21	74.8	0.7779	Employees in tertiary industry	133.8	0.5450
	Others	Pop. above 15	39.7	0.9190	Employees in tertiary industry, Total students	71.5	0.7808
Non-home based trip	Business	Total employees	79.2	0.8145	Total employees	93.5	0.6952
	Shopping	Employees in tertiary industry, Total students	92.8	0.7247	Employees in tertiary industry	165.3	0.4826
	Others	Employees in tertiary industry, Total students	70.6	0.8132	Employees in tertiary industry, Total students	90.1	0.6795

Source : 「2012 National transportation survey and developing database」 [2]

1) 본 연구에서 ‘유동인구’는 특정 기지국 범위에 존재하는 휴대전화의 수를 의미하며, 수도권의 주중 5일치(2014년 5월 12일~16일) 자료를 활용하였다.

본 연구는 전체 휴대전화 가입자의 약 50%를 대상으로 하지만, KTDB는 약 2.5% 수준의 표본조사를 활용한다.

또한 본 연구의 경우 전국 약 370만 개의 셀(수도권의 경우 약 93.9만 개) 단위로 자료생성이 가능하며, 이미 기록되고 있는 정보를 활용하므로 조사기간의 제한이 없어 요일별 패턴 분석 및 특정일에 대한 분석이 가능하고, 셀별 유동인구를 통계청 집계구, 법정동 등의 기준으로 분류할 수 있는 분석 상세도 측면의 장점이 있다[3]. 반면에 가구통행실태조사는 대표로 지정한 특정일(보통 일주일 정도)에 대한 조사로서 조사비용이 많이 소요[4]되며 인구가 많지 않은 지역의 경우 연 단위 전수화시 표본오차가 커질 가능성이 크다. 게다가 전술한 바와 같이 5년 단위 조사 시점 사이의 중간년도는 예측에 의한 방법으로 정보가 생성되므로 추가적 오차 발생 가능성이 존재한다.

2.2 무선통신자료의 구조 및 특성

2.2.1 자료 구조

통계청 자료[5]에 따르면 우리나라 2014년 추계인구는 50,423,955명²⁾이며, 휴대전화 가입자수(회선수 기준)는 2014년 12월 57,207,957명[6]으로 인구 1인당 휴대전화 가입자수는 1.13으로서 오히려 인구를 상회하고 있다. 통신사 별로는 SK텔레콤(이하, SKT)이 약 50%, KT, LG유플러스가 각각 약 30%, 약 20%의 점유율을 보이고 있고, 무선통신 자료는 국민 대다수가 이용하는 휴대전화 통신발생 기록으로부터 취득할 수 있는 매우 유용한 정보이며, 특히 교통 분야에서의 활용성이 크다고 할 수 있다. 개인정보 및 개인위치정보 보호 문제로 특정 개인의 통행 궤적을 추적할 수는 없으나 위치(기지국 등) 기반으로 집합화 된 통신 자료를 활용하는 것만으로도 충분히 여러 의미 있는 정보를 추출할 수 있다.

본 연구에서는 국내 통신사 중 하나인 SK텔레콤에서 제공하고 있는 무선통신 자료에서 추출한 성·연령별 유동인구 자료를 토대로 교통정보 관점에서 그 의미를 살펴보고, 이를 활용한 통행발생량 자료를 추출한 후 KTDB에서 제공하는 수도권 통행발생량과 비교를 통하여 그 활용성을 파악하고자 한다. SK텔레콤에서 제공하는 성·연령별 유동인구 자료의 구성은 Table 2와 같다.

Table 2. Structure of mobile phone data.

Columns	Contents	Types	Descriptions
1	Date	String	- 4-digit year, 2-digit month, 2-digit day (Example: 20140513)
2	X coordinate of cell ¹⁾	Number	- UTM-K coordinate system
3	Y coordinate of cell	Number	- UTM-K coordinate system
4~9	Male floating population by age	Number	- Daily number of mobile phone communicated in the cell (to 2 places of decimals) - For 6 age groups (10s, 20s, 30s, 40s, 50s, over 60s) by gender
10~15	Female floating population by age	Number	- Daily number of mobile phone communicated in the cell (to 2 places of decimals) - For 6 age groups (10s, 20s, 30s, 40s, 50s, over 60s) by gender
16	Origin code	String	- 8-digit code of "dong" unit

1): A cell covers 50m×50m area. The number of cells is 3,712,040 in national area, and 938,976 in Seoul metropolitan area.

셀의 X, Y 좌표는 UTM-K 좌표체계를 따르며 소수점 4자리까지 표시된다. 성·연령별 유동인구는 통행발생 중 통행유인량 산정의 핵심이 되는 항목으로, 1일 단위로 해당 셀에서 통신한 사람의 수³⁾를 나타낸다. 동일 날짜에서는 여러 번 통신한 기록이 있더라도 한 번으로 계산되며, 예를 들면, 해당 셀에서 하루에 100번의 통신 기록이 있더라도 한 번으로 산정되는 방식이다. 유동인구는 이동통신사에서 자료 제공 시 관측된 유동인구에 해당 이동통신사의 점유율을 나누어 전수화한 값이다.

마지막으로 유입지(Origin place) 코드는 해당 셀에서 통신한 사람이 전 달 심야시간대에 동일한 곳에 16일 이상 머문 곳을 법정동 기준으로 분류한 것으로서 전국을 5,050개 법정동으로 구분하였다. 위 기준에 부합하지 않는 경우는 불명으로 처리되며, 분석 자료 중 유입지 불명 자료는 평균 13,865,436건으로 전체 자료(72,636,670건)의 약 19.1% 수준이다.

2.2.2 분석대상 자료

본 연구에서는 2014년 5월 12일(월)부터 5월 16일(금)까지의 주중 5일치 자료를 이용하여 분석하였다. 통신 자료의 특성으로 인해 제공 자료의 셀이 370만 개를 상회하기 때문에 관련 자료도 매우 방대하게 생성되고 있다. 경기도 자료의 경우 1일치 제공 파일의 크기가 약 19.9~27.1GB(기가바이트) 수준이며, 서울의 경우 약 12.2~16.2GB, 인천은 약 3.1~4.2GB로서 자료의 파일 크기가

2) 통계청 국가통계포털(www.kosis.kr) 홈페이지 자료 인용[5].

3) 엄밀한 의미에서는 각 셀에서 통신(문자/통화 수발신, 데이터 통신 기록 기준이며, WiFi 통신은 제외됨)한 휴대전화 수임.

매우 크다. 가장 파일의 크기가 큰 경기도의 경우 2014년 5월 13일 자료의 수가 226,218,419개 행 수준으로서 전체 자료를 분석하기에는 일반적인 전산환경에서는 연산시간이 과다하게 소요되는 한계가 있다.

하지만 연령별 유동인구 값이 0이거나 매우 작아 무시할 만한 행이 많이 포함되어 있는 것에 착안하여 이러한 행들은 소거하여 연산시간을 적정하게 확보하였다. 불필요한 행을 소거하기 위해 전체 값을 검색하는 것 역시 고용량의 데이터인 상황에서는 연산시간 증가요소 이므로, 통신량이 가장 많은 값인 남자 40대와 남자 50대 두 칼럼의 값을 기준으로 연령별 유동인구 값이 0이 아닌 행을 선별하여 소거하였다(남자 40대와 50대 유동인구 값이 모두 0인 칼럼이 삭제된다). 그 결과 2014년 5월 13일 기준 경기도 자료에서 추출된 자료 기준으로 55,684,459개 행만 남음으로서 전체 자료의 약 24.6% 수준으로 행수를 축소하였다. 그럼에도 소거 전후의 유동인구의 합을 기준으로 하면 전체 자료의 합이 39,032,413이며, 추출된 자료는 34,234,286으로서 87.7% 수준이다. 따라서 추출된 자료만으로 분석하여도 충분히 의미 있는 값을 도출할 수 있을 것으로 판단된다. 본 연구에서 제시하는 유동인구 값은 불필요한 행을 소거하여 추출된 자료를 기준으로 하고 있다.

2.3 무선통신자료 기반의 통행발생량 산정 및 평가

2.3.1 통행발생량 산정

Table 2에서 설명한 자료를 기준으로 분석기간 주중 5일치에 대한 수도권의 시·군·구별로 통행발생량을 산정하였다. 통신자료의 특성에서 살펴본 바와 같이 성·연령별 유동인구는 공간적으로 분할된 수도권의 셀에서 통신한 자료를 근거로 도출하였기 때

Table 3. Estimation of trip attraction using mobile phone data.

Unit: trips

Metro region	Administrative region	Trip attraction	Metro region	Administrative region	Trip attraction	Metro region	Administrative region	Trip attraction
Gyeonggi	Gapyeong	156,991	Gyeonggi	Yangju	602,520	Seoul	Dongjak	1,073,688
	Deogyang, Goyang	1,016,354		Yangpyeong	212,226		Mapo	1,527,111
	Ilsandong, Goyang	912,794		Yeoju	263,956		Seodaemun	936,326
	Ilsanseo, Goyang	765,470		Yeoncheon	96,956		Seocho	2,632,375
	Gwacheon	238,543		Osan	512,629		Seongdong	1,200,828
	Gwangmyeong	894,844		Giheung, Yongin	875,500		Seongbuk	1,194,877
	Gwangju	703,235		Suji, Yongin	587,868		Songpa	2,218,426
	Guri	636,783		Cheoin, Yongin	662,707		Yangcheon	1,240,675
	Gunpo	682,558		Uiwang	454,085		Yeongdeungpo	2,121,641
	Gimpo	833,359		Uijeongbu	1,238,067		Yongsan	1,034,749
	Namyangju	1,425,441		Icheon	631,338		Eunpyeong	776,490
	Dongducheon	244,141		Paju	982,892		Jongno	1,355,798
	Sosa, Bucheon	553,893		Pyeongtaek	1,256,801		Jung	1,861,448
	Ojeong, Bucheon	497,916		Pocheon	467,423		Jungnang	1,011,018
	Wonmi, Bucheon	1,480,991		Hanam	455,148		Sum	36,279,329
	Bundang, Seongnam	1,780,138		Hwaseong	1,839,125		Ganghwa	97,611
	Sujeong, Seongnam	664,257		Sum	34,415,747		Gyeyang	876,315
	Jungwon, Seongnam	685,075		Gangnam	4,228,193		Nam	1,082,738
	Gwonseon, Suwon	836,411		Gangdong	1,224,799		Namdong	1,477,346
	Yeongtong, Suwon	905,590		Gangbuk	754,583		Dong	203,216
Jangan, Suwon	656,247	Gangseo	1,596,347	Bupyeong	1,363,451			
Paldal, Suwon	840,800	Gwanak	1,315,214	Seo	1,202,343			
Siheung	1,370,062	Gwangjin	1,285,944	Yeonsu	816,188			
Danwon, Ansan	1,201,600	Guro	1,350,125	Ongjin	25,469			
Sangrok, Ansan	884,256	Geumcheon	938,551	Jung	620,652			
Anseong	524,430	Nowon	1,420,699	Sum	7,765,330			
Dongan, Anyang	1,084,294	Dobong	725,706					
Manan, Anyang	800,033	Dondaemun	1,253,717	Total		78,460,406		

Note: Average trip attraction during 5 days

문에 수도권외의 통행발생(trip generation) 중 통행유인(trip attraction)량은 다음의 가정 하에 의해서 도출할 수 있다.

첫째, 휴대전화 가입자는 통행의 목적지에서 최소 한 번 이상의 통신을 하는 것으로 가정하였으며, 만약 통신을 하지 않은 경우에는 본 자료에서 제외된다. 둘째, 휴대전화로 통신한 지점은 통행의 목적지로 가정한다. 실질적으로는 통신을 한 지점이 통행의 최종 목적지가 아닌 중간 경유지 또는 이동 중일 수도 있다. 그러나 현 자료만 가지고는 이러한 자료의 구분이 불가하기 때문에 별도의 구분을 고려하지 않았다. 마지막으로, 통행인구는 휴대전화를 모두 소지하였다고 가정하여, 휴대전화를 소지하지 않은 경우의 통행은 고려하지 않았다.

상기 가정에 따라 수도권외의 각 셀에서의 성·연령별 유동인구는 합계는 해당 셀의 통행유인량과 같아지게 된다. 마지막으로, 각 셀을 분석하고자 하는 공간단위로 집합화하여 통행유인량이 산정된다. 이러한 가정 하에서 분석기간에 대해서 하루 단위로 통행유인량을 산정하여 평균값을 제시한 결과가 Table 3이다. 요일별 변동을 확인할 수 있도록 Table 4에는 일자별 총량을 표시하였다.

통행생성(trip production)량은 Table 2에서 설명한 자료의 구성 중 '유입지코드'를 활용하여 산정할 수 있다. 하지만 두 가지 측면에서 자료의 완결성 문제로 인해서 일부 제약이 따를 수 있다. 첫 번째는 자료의 특성에서 설명한 바와 같이 원칙적으로 본 연구에서 분석한 성·연령별 유동인구 자료는 수도권외의 공간적 범위로 하였기 때문에 수도권외를 벗어난 지역을 방문한 경우에는 본 자료에서는 제외된다. 그러나 이러한 문제는 수도권 이외의 지역에 대한 자료를 확보할 수 있다면 해결될 수 있는 부분이다.⁴⁾ 두 번째로, 유입지가 알려지지 않은 자료의 경우 총량은 알 수 있으나 유입지가 누락되어 있으므로 통행생성량 산정에는 제외한다. 전술한 바와 같이 이러한 경우는 전체 자료의 약 19.1%를 차지하고 있으며, 향후 이를 반영할 수 있는 방안에 대한 연구가 필요할 것으로 판단된다. 현 시점에서는 통신자료를 활용한 통행유인량 및 통행생성량의 산정 가능성을 분석하는 것에 중점을 두고 있으므로, 이에 대한 연구는 향후 연구로 남겨둔다.

통행유인량 산정 시의 가정이 적용되면, 통행유인량이 어느 곳으로부터 유입되었는지에 대한 정보(유입지 코드)가 있는 자료를 기준으로 하여 통행생성량을 산정할 수 있다. 이에 대한 결과는 다음의 Table 5와 Table 6이다.

2.3.2 통행발생량 자료의 평가

무선통신 자료로부터 산정된 통행발생량은 KTDB 통행발생량과 상관성 분석을 통하여 평가하였다. KTDB에서 제공하는 수도권외의 수단 OD 합계를 이용하여 수도권외의 시·군·구별로 통행유인량과 통행생성량을 도출하였다. KTDB와의 통행발생량 비교는 최근 조사 기준년도인 2012년 OD 및 무선통신 자료와 동일한 년도로서 2012년 기준으로 추정된 값인 2015년의 OD와의 비교를 시행하였다. 다만, 2015년 OD의 추정과정에서의 시점의 차이에 따른 오차가 발생할 수 있는 한계가 있다. KTDB에서 제공하고 있는 수도권외의 유인량의 총합은 2012년, 2015년 기준 각각 63,636,681통행, 59,467,281통행으로 산정되었으며, 생성량은 각각 63,576,731통행, 59,402,056통행으로 산정되었다.

총량을 비교해 보면 통행유인량은 통신자료에서 추정된 값이 더 큰 수준이며, 이는 목적지가 아닌 경유지나 이동 중 발생시킨 통신에 의해 초과 산정된 유인량 때문으로 추정된다. 통행생성량은 통신자료에서 유입지가 알려지지 않은 자료가 제외됨으로써 KTDB가 더 큰 것으로 분석되었다. 수도권외의 시·군·구는 모두 79개(경기도 44개, 서울시 25개, 인천시 10개)이며, KTDB의 유인 및 생성량과 본 연구에서 통신자료를 이용하여 통행발생의 유인량 및 생성량과 비교하였다. 먼저 상관관계를 살펴보면, 유인량의 상관관계는 높게 나타났다. 그러나 생성량은 상대적으로 낮게 나타나는데, 제공받은 자료의 특성상 해당 셀(지역)에 유입(attraction)된⁵⁾ 이용자는 기록에 남지만 유출(production)되는 이용자는 기록에 남지 않기 때문이다. 통신자료 만으로는 직접 생성량

Table 4. Estimation of daily trip attraction using mobile phone data.

Unit: 1,000 trips

Date	Day	Gyeonggi		Seoul		Incheon		Total		
		Male	Female	Male	Female	Male	Female	Male	Female	Sum
12 May, 2014	Mon.	19,709	14,232	19,941	15,527	4,346	3,365	43,995	33,124	77,119
13 May, 2014	Tue.	19,823	14,412	20,168	15,852	4,323	3,371	44,314	33,635	77,948
14 May, 2014	Wed.	18,692	13,761	19,122	15,163	4,059	3,205	41,873	32,130	74,002
15 May, 2014	Thu.	20,255	14,819	20,901	16,542	4,448	3,499	45,603	34,860	80,463
16 May, 2014	Fri.	21,167	15,210	21,418	16,764	4,628	3,581	47,214	35,555	82,769
Average		19,929	14,487	20,310	15,970	4,361	3,404	44,600	33,861	78,460

4) 실제 SK텔레콤에서 수도권 외의 지역의 자료를 구축하고 있어 논리적으로 문제는 없으나, 단순히 본 연구에 제공되지 않은 상황임.

5) 정확히는 해당 셀에서 통신을 발생시킨 인구에 대한 기록이지만, 이해를 돕기 위하여 해당 셀에 유입된 이용자는 해당 셀에서 통신을 발생시킨다는 가정 하에 '유입된'이라는 표현을 사용함.

Table 5. Estimation of trip production using mobile phone data.

Unit: trips

Metro region	Administrative region	Trip production	Metro region	Administrative region	Trip production	Metro region	Administrative region	Trip production	
Gyeonggi	Gapyeong	108,763	Gyeonggi	Yangju	388,863	Seoul	Dongjak	1,042,883	
	Deogyang, Goyang	891,768		Yangpyeong	178,871		Mapo	902,219	
	Ilsandong, Goyang	780,264		Yeoju	170,957		Seodaemun	746,804	
	Ilsanseo, Goyang	590,479		Yeoncheon	59,452		Seocho	1,485,878	
	Gwacheon	111,421		Osan	370,096		Seongdong	787,928	
	Gwangmyeong	983,390		Giheung, Yongin	788,585		Seongbuk	1,157,071	
	Gwangju	619,850		Suji, Yongin	623,091		Songpa	1,959,771	
	Guri	542,098		Cheoin, Yongin	483,519		Yangcheon	1,364,067	
	Gunpo	708,316		Uiwang	361,130		Yeongdeungpo	1,070,333	
	Gimpo	695,270		Uijeongbu	1,124,915		Yongsan	500,600	
	Namyangju	1,508,938		Icheon	424,894		Eunpyeong	1,011,071	
	Dongducheon	185,604		Paju	748,015		Jongno	326,651	
	Sosa, Bucheon	520,423		Pyeongtaek	852,376		Jung	470,988	
	Ojeong, Bucheon	419,201		Pocheon	309,456		Jungnang	1,133,701	
	Wonmi, Bucheon	1,224,446		Hanam	367,261		Sum	28,289,146	
	Bundang, Seongnam	1,043,305		Hwaseong	1,301,030		Ganghwa	68,904	
	Sujeong, Seongnam	416,613		Sum	27,834,284		Gyeyang	762,115	
	Jungwon, Seongnam	580,824		Seoul	Gangnam		2,212,119	Nam	1,071,301
	Gwonseon, Suwon	749,818			Gangdong		1,296,464	Namdong	1,340,696
	Yeongtong, Suwon	727,855			Gangbuk		934,853	Dong	119,441
	Jangan, Suwon	673,625			Gangseo		1,708,449	Bupyeong	1,352,519
	Paldal, Suwon	566,245			Gwanak		1,634,841	Seo	1,090,960
	Siheung	936,507			Gwangjin		1,062,240	Yeonsu	792,655
	Danwon, Ansan	851,878			Guro		1,239,124	Ongjin	9,843
	Sangrok, Ansan	827,554			Geumcheon		735,402	Jung	227,794
	Anseong	311,859			Nowon		1,534,069	Sum	6,836,229
	Dongan, Anyang	990,376			Dobong		945,980		
	Manan, Anyang	715,086			Dondaemun		1,025,642	Total	

Note: Average trip production during 5 days

Table 6. Estimation of daily trip production using mobile phone data.

Unit: 1,000 trips

Date	Day	Gyeonggi		Seoul		Incheon		Total		
		Male	Female	Male	Female	Male	Female	Male	Female	Sum
12 May, 2014	Mon.	15,852	11,479	15,392	12,171	3,819	2,933	35,064	26,583	61,647
13 May, 2014	Tue.	15,964	11,678	15,657	12,445	3,818	2,963	35,439	27,085	62,524
14 May, 2014	Wed.	15,068	11,165	14,870	11,929	3,601	2,816	33,539	25,910	59,449
15 May, 2014	Thu.	16,378	12,065	16,243	12,993	3,907	3,060	36,527	28,118	64,645
16 May, 2014	Fri.	17,113	12,410	16,611	13,134	4,109	3,156	37,833	28,700	66,533
Average		16,075	11,759	15,755	12,534	3,851	2,986	35,680	27,279	62,960

을 추정하지 않고 등록된 유입지 기준으로 추정할 수밖에 없기 때문에 수도권에 거주하는 사람이 수도권 이외의 지역에서 통신기록을 남기더라도 본 연구에서 입수한 자료에는 누락되기 때문이다. 앞서서도 언급한 바와 같이 이 문제는 수도권 이외의 통신기록을 확보할 수 있으면 해결될 수 있는 부분이다.

Table 7. Correlation Analysis between KTDB and this study.

Date	Day	Comparison with 2012 KTDB data [2]				Comparison with 2015 KTDB data			
		Trip attraction		Trip production		Trip attraction		Trip production	
		Correlation Coef.	Theil's U	Correlation Coef.	Theil's U	Correlation Coef.	Theil's U	Correlation Coef.	Theil's U
12 May, 2014	Mon.	0.9564	0.1349	0.8562	0.1318	0.9684	0.1624	0.8434	0.1380
13 May, 2014	Tue.	0.9544	0.1427	0.8609	0.1295	0.9667	0.1703	0.8480	0.1381
14 May, 2014	Wed.	0.9557	0.1223	0.8624	0.1323	0.9674	0.1480	0.8493	0.1343
15 May, 2014	Thu.	0.9576	0.1544	0.8633	0.1293	0.9675	0.1840	0.8489	0.1430
16 May, 2014	Fri.	0.9552	0.1669	0.8596	0.1324	0.9673	0.1963	0.8476	0.1480
Average		0.9561	0.1440	0.8606	0.1296	0.9677	0.1722	0.8476	0.1391

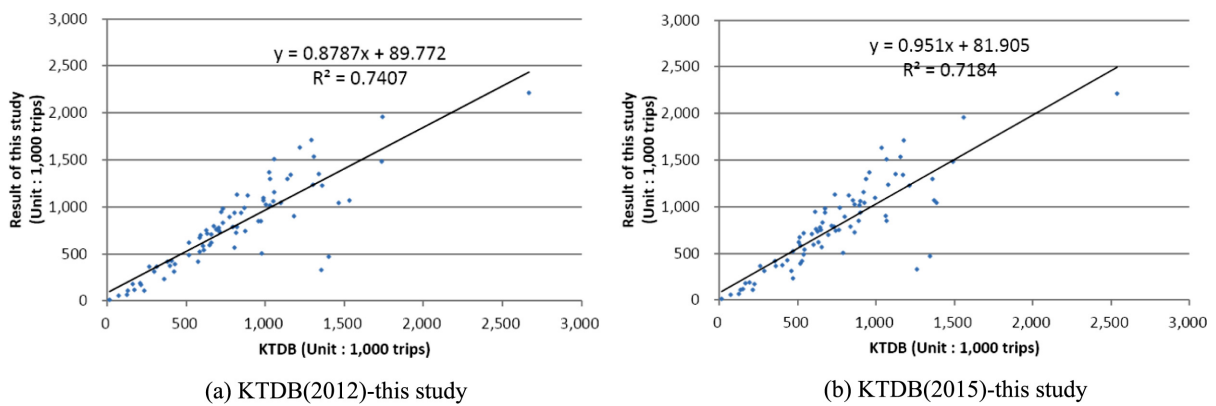


Fig. 1. Correlation analysis of trip production.

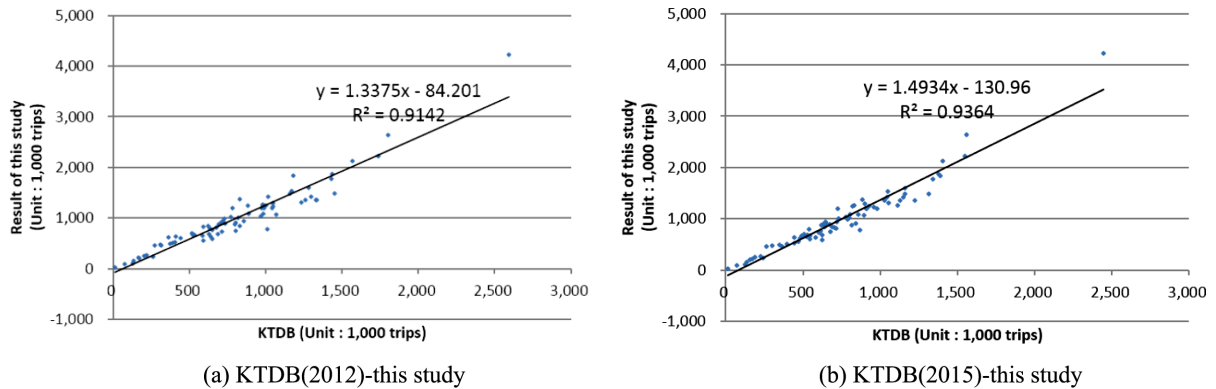


Fig. 2. Correlation analysis of trip attraction.

Table 7에 정리한 바와 같이 본 연구에서 제시한 무선통신 자료를 이용한 통행발생량 추정치는 기존의 KTDB에서 구축한 직접 조사 방식 기반의 통행발생량 추정에 의한 결과와 매우 높은 상관관계를 가지는 것으로 분석된다. 본 연구에서는 무선통신 자료의 교통계획 분야 활용가능성을 판단하기 위하여 별도의 전수화 과정을 포함하지는 않았지만, 높은 상관관계를 근거로 판단할 때에 향후 통행발생량 예측 시에 기존의 조사 방식에 의존하던 방식 이외에 통신자료를 이용한 새로운 방식도 가능할 수 있다는 것을 확인하였다.

3. 결 론

본 연구에서는 무선통신 자료를 이용하여 교통계획 자료의 기초가 되는 설문 또는 직접 조사에 의한 자료의 일부를 대체할 수 있는 가능성을 확인하였다. 이를 위해 무선통신자료의 일종인 셀 단위 통신 이용자의 연령, 성별 현황 정보를 이용하여 셀 단위 및 존

단위 통행유인(trip attraction)량을 추정하였으며, 이용자의 유입지 정보를 이용하여 존 단위 통행생성(trip production)량을 추정하였다.

본 연구에서 추정된 통행발생량을 2015년도 KTDB 통행발생량 추정치와 비교한 결과 통행유인량의 R^2 는 0.9364, 통행생성량의 R^2 는 0.7184로 산출되어 높은 상관관계가 있음을 확인할 수 있었다. 즉, 교통계획 기초자료 생성에 있어서 무선통신자료는 충분히 활용 가치가 있음을 확인하였다. 무선통신자료를 활용하여 기존의 교통조사 결과의 신뢰성을 유지하면서도 비용을 절감시키거나, 조사량을 유지하면서도 신뢰성 및 정확성을 향상시키는 방향으로 교통조사 및 DB구축을 발전시킬 수 있을 것이다.

본 연구에서 이용한 자료의 경우 통행유인량을 추정하는 데에는 매우 유용하지만, 상대적으로 통행생성량을 추정하는 데에는 한계가 따랐다. 이는 연구에서 활용한 자료의 한계로, 향후 무선통신자료 원본으로부터 가공하는 과정에서 유출되는 이용자 정보가 별도로 확보된다면 충분히 보완될 것으로 보인다. 더불어, 계층별, 연령별, 지역별, 성별 다양한 특성분석, 휴대전화를 소유하지 않은 인구 등 누락 정보의 보완을 통한 통행발생량 전수화에 대한 후속연구 및 현재 자료의 한계인 목적별 통행발생에 대한 보완방안에 대한 후속연구를 통하여 무선통신 자료의 대표성 및 신뢰성이 보완된다면 자료의 활용 가치가 보다 커질 것으로 보인다.

현재는 『위치정보의 보호 및 이용 등에 관한 법률』에 의하여 통신사에서 경로(셀 간 이동) 정보를 제공하지 않고 있으나, 개인 정보 및 개인위치정보를 침해하지 않는 선을 명확히 하고 그 범위 내에서 일부 경로 정보를 활용할 수 있다면 통행발생량뿐만 아니라 OD도 보다 정확히 추정할 수 있을 것이다. 그리고 무선통신 자료 이용성을 높이기 위하여 기지국 기반 데이터의 한계인 통신발생 정보에서 도착과 통과를 구분할 수 있는 방안, 일자별(주중/주말) 및 시간대별 분석을 통한 목적통행 추정의 보완, KTDB 읍명동 OD와 비교분석 등에 대한 연구가 향후 진행되어야 할 것이다. 무선통신 자료의 경우 현재 동단위 대비 공간적으로 세밀하기 때문에 후속연구가 성공적인 성과를 얻게 된다면 기존 KTDB에서 제시하는 OD 대비 더욱더 정교한 결과를 제시할 수 있을 것으로 기대된다.

이러한 일련의 후속연구의 진행을 위해서 아직은 활용 및 접근이 어려운 무선통신 자료를 지속적으로 연구개발에 활용할 수 있도록 자료 제공 및 유통체계 구축 등 이동통신사 등 민간부문과의 지속적 협조체계가 필요할 것이다.

후 기

본 연구는 한국철도기술연구원 주요사업의 연구비 지원으로 수행되었습니다.

References

- [1] KOTI (2013) 2013 Korea Transport Survey and Database, Ministry of Land, Infrastructure and Transport.
- [2] KOTI (2012) 2012 Korea Transport Survey and Database, Ministry of Land, Infrastructure and Transport.
- [3] K. Kim, I. Lee, H. Kwak, J. Min (2015) Application study of telecommunication record data in floating population estimation, *Seoul Studies* 16(3), pp. 181-191.
- [4] M. Byun, U. Seo (2011) How to measure daytime population in urban streets?: Case of Seoul pedestrian flow survey, *Survey Research*, 12(2), pp. 27-50.
- [5] www.kosis.kr (Accessed 15 March 2015).
- [6] www.kcc.go.kr (Accessed 15 March 2015).

(Received 22 June 2015; Revised 10 July 2015; Accepted 23 September 2015)

Kyoungtae Kim: ktim@krri.re.kr

Green Transport and Logistics Institute, Korea Railroad Research Institute, 176, Cheoldobangmulgwan-ro, Uiwang, Gyeonggi 16105, Korea

Inmook Lee: mook79@krri.re.kr

Green Transport and Logistics Institute, Korea Railroad Research Institute, 176, Cheoldobangmulgwan-ro, Uiwang, Gyeonggi 16105, Korea

Jac Hong Min: jhmin@krri.re.kr

Green Transport and Logistics Institute, Korea Railroad Research Institute, 176, Cheoldobangmulgwan-ro, Uiwang, Gyeonggi 16105, Korea

Ho-Chan Kwak: kwak01@krri.re.kr

Green Transport and Logistics Institute, Korea Railroad Research Institute, 176, Cheoldobangmulgwan-ro, Uiwang, Gyeonggi 16105, Korea