

## 이동통신 자료를 활용한 가정기반 OD 구축 및 분석

### Home-based OD Matrix Production and Analysis Using Mobile Phone Data

김경태 · 오동규\* · 이인묵 · 민재홍

Kyoungtae Kim · Dongkyu Oh · Inmook Lee · Jae Hong Min

**Abstract** Based on time dependent location data of mobile phone users, users' ODs were produced after tracing their travel route and inducing their origins and destinations. System considered average signaling frequency, which means that the longer the travel length is the more frequent the signal is. This is a home-based OD and is limited to the Seoul Metropolitan area. The OD matrix from the mobile phone data which was aggregated to the cell and transformed to the 'Dong' area, was compared to the KTDB OD. The results can be analyzed and it was determined that they are highly correlated because individual coefficients are 0.98 and 0.85, the former between the OD of this study and the KTDB Si/Gun/Gu unit area OD and the latter between the OD of this study and the Dong unit area KTDB OD.

**Keywords** : Mobile phone data, OD production, KTDB, Trip distribution

**초 록** 본 연구에서는 휴대전화 이용자의 시간대별 기지국 위치정보를 기반으로 하여 이용자의 이동경로를 추적하고, 기/종점을 추출하여 OD를 구축하였다. OD 구축시 통행거리가 길어질수록 통신횟수가 많아짐을 고려하여 통행거리별 평균통신횟수를 산출/반영하였고, 가정기반 통행만을 대상으로 수도권 중심의 OD를 구축하였다. 셀 단위로 집계된 휴대전화 자료를 법정동 단위로 변환하고, 이를 행정동 단위로 변환하여 본 연구에서 구축한 OD를 KTDB OD와 비교분석 한 결과, 시/군/구 단위 OD와 동 단위 OD의 상관계수는 각각 0.98, 0.85로 나타나 상당한 연관성이 있는 것으로 분석되었다.

**주요어** : 이동통신자료, OD생성, KTDB, 통행분포

## 1. 서 론

미래창조과학부에 따르면 우리나라의 이동전화 가입현황은 2015년 12월 기준 총 58,935,081 회선이며[1], 이 중 일반 이용자용 휴대전화가입자 수는 53,616,057명에 이른다. 이는 동일시점 주민등록인구인 51,529,338명[2]과 유사하지만 약간 상회하는 수준으로, 1인당 1.04회선에 해당한다. 즉, 이동전화의 위치를 파악함으로써 인구 위치의 파악이 가능함을 의미하며, 이동전화 이용에 따라 생성되는 이동통신 자료를 활용하면 별도의 조사 없이도 실제 인구의 이동을 파악할 수 있음을 의미한다[3].

이미 2000년대 초반, Kim 등[4]은 이동전화 가입자가 폭발적으로 증가하면서 이동통신 자료를 이용한 OD 추정에 관한 연구를 진행한 바 있다. 당시 연구는 개별 회선의 시간별 기지국 정보(셀 트래킹)를 통해 해당 회선(이동통신 이용자)의 추론된 위치정보와 GPS에 의한 실측 위치정보를 비교함으로써, 이동통신 자료 중 기지국 정보에 의해 이동통신 이용자의 위치 파악이 교통계획 수준에서 유의하며 이를 통한 OD 추정의 가능성이 있음을 확인하였다. 하지만 특정 이용자(실험자, 영업차량 등)에 의해 생성된 일부 데이터를 이용한 개념적 연구로, 교통계획 자료 수준의 OD 통행량을 추정하는 데에는 미치지 못하였다.

이후 국내외에서 이동통신 자료 기반의 OD 추정 연구가 정체에 있다가, 2010년 이후 이동통신 사업자가 이동통신 통화, 문자, 데이터 등의 과금을 위해 저장하고 있는 이동통신 이용기록 상세자료인 CDR(Call Detail Records)을 공개하면서 다시금 관련 연구가 활성화 되고 있다. Calabrese 등[5], Wang 등[6]은 억단위 이상의 line으로 구성된 CDR 데이터를 이용하여 시간대별 OD 추정의 가능성을 확인하였다.

Iqbal 등[7]은 CDR에 마이크로 시뮬레이션(MITSIMLab) 및 제한적 교통조사 자료를 이용하여 시간대 OD 추정 정확성을 높이는 방안을 제시하였으며, Jiang 등[8], Schneider 등[9]은 CDR에 의한 통행사슬(trip chain) 및 경로가 조사기반의 가구통행조

\*Corresponding author. Tel.: +82-31-460-5846, E-mail: ohdong@krii.re.kr.

© 2016 The Korean Society for Railway. All rights reserved.

http://dx.doi.org/10.7782/JKSR.2016.19.5.656

사와 일관된다고 하였다.

국의 연구사례로 볼 때에 CDR은 정보의 품질 뿐만 아니라 별도의 조사 또는 데이터 수집과정 없이 이동통신 과금 과정의 부산물로서 얻어지는 장점도 있어, 향후 교통분야에 활용가능성이 확대될 것으로 판단된다. 하지만 국내의 경우는 국외와는 다르게, 「개인정보 보호법」 및 「위치정보의 보호 및 이용 등에 관한 법률」의 이유로 이동통신 사업자가 개인식별코드(ID)<sup>1)</sup>를 포함하는 데이터를 제3자에게 제공하지 않고 있는 실정이다. 대신 개인식별코드 없이 기지국 단위 등 지점(셀)을 기준으로 이용자 정보를 집계한 ‘제한적’ 이동통신 자료만을 제공하고 있다.

Kim 등[3]은 이러한 제한적 이동통신 자료를 이용하여 통행발생량을 추정하였으며, 기존 국가교통데이터베이스(이하, KTDB)에서 제공하는 통행발생량과 통계적으로 높은 상관관계가 있음을 검증함으로써 교통계획 기초자료 생성에 제한적 이동통신 자료의 활용 가능성을 확인하였다.

개인식별코드가 없이는 개별 통행자의 경로 추적 및 이동 파악에 한계가 있으므로, 자료 자체만으로는 OD 추정으로 발전시키는 데에 한계가 있다. 이러한 점을 보완하기 위하여 이동통신 사업자는 각 셀(cell)별 유입지 정보를 포함하여 이동통신 자료를 제공하고 있다. 유입지는 해당 셀에 집계된 이동통신 이용자들이 어느 지역(법정동 기준)들로부터 유입되었는지를 나타내는 항목으로, 각 이동통신 이용자가 전달 심야시간대에 동일한 곳에 16일 이상 머문 곳이 유입지로 판별되며 16일 이상 머문 곳이 없는 경우 유입지 불명으로 분류된다.

본 연구에서는 개인 식별코드가 없이 지점 기준으로 구축된 제한적 이동통신 자료를 활용하여 OD 통행량을 추정하는 방안 및 그 결과를 제시하였다. 또한 OD 통행량 추정에 활용 가능성 확인을 위하여 추정 결과를 기존 KTDB의 OD 통행량과 비교하여 상관성을 검증하였다.

## 2. 본 론

### 2.1 데이터 구축

#### 2.1.1 자료 구조

국내 이동통신 사업자별 회선 점유율은 2015년 12월 현재 SKT, KT, LG U+, MVNO(가상 이동통신망 사업자) 각각 약 45.3%, 25.6%, 19.1%, 9.9%이며, MVNO를 제외하면 SKT, KT, LGU+가 각각 50.3%, 28.5%, 21.2%를 차지하고 있다[1]. 이동통신 사업자별 회선수(점유율)는 월단위로 미래창조과학부를 통해 제공되고 있으며, 이 정보는 이동통신 자료 기반의 통행량 산출시 전수화에 필요한 정보이므로 확인할 필요가 있다.

본 연구에서는 SKT에서 제공하고 있는 제한적 이동통신 자료를 활용하였다. 서론에서 밝혔다시피 국외에서는 개별 이동통신 이용자별 ID(개인식별코드)가 있는 CDR을 제공하는 반면에, 국내의 경우 관련 법령의 제약에 따라 지점 등을 기준으로 집계하여 ID를 제거한 형태의 자료만을 제공하고 있다. 본 연구에서 활용한 이동통신 자료의 구조는 Table 1과 같다.

자료는 SKT에서 설정한 50m×50m 단위의 셀(cell)인 pCELL 기준으로 구성되어 있다. pCELL은 개별 이동통신 단말기에 관

Table 1. Structure of mobile phone data [1].

Columns	Contents	Types	Descriptions
1	Date	String	- 4-digit year, 2-digit month, 2-digit day (Example: 20140513)
2	X coordinate of cell <sup>1)</sup>	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)
10~15	Female floating population by age	Number	- For 6 age groups (10s, 20s, 30s, 40s, 50s, over 60s) by gender - 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.  
 Data: Kim K.t. etc., trip generation analysis using mobile phone data, journal of the Korean society for railway

1) 암호화하여 직접적 식별표시나 숫자를 알 수 없는 경우도 포함함.

여하는 기지국 데이터로 부터 위치를 측정하여 정의된 pCELL 중 가장 가까운 값에 사상(mapping)하여 생성된다. 개인식별코드가 없이 각 셀에 누적 정보가 기록되는 방식이므로, 셀 간의 이동 정보는 직접 추출할 수 없는 구조이다. 자료에서 2~3번째 칼럼의 좌표는 pCELL의 좌표이다.

4~15번째 칼럼은 기간(본 연구에서는 1일 단위) 중 성별, 연령별 해당 셀에서 통신기록이 있는 이용자의 수이며, 중복해서 통신한 기록이 있어도 하나의 셀에는 한번만 계산된다. 즉, 하루 중 해당 셀에서 이동통신 서비스를 이용한 기록이 있는 이용자의 성별, 연령별 통계수치이다. 이용자 개념은 정확히 회선에 대한 정보이지만, 본 연구에서는 회선과 이용자가 동일하다 가정하였다. 여기서 통신기록이라 함은 음성(통화), 문자, 데이터(인터넷)를 이용한 경우가 포함이 되며, 푸시(push) 서비스 등 수동적으로 발생하는 통신도 포함이 된다. 다만, 단순히 기지국에 연결되어 있는 대기상태는 제외된다. 이용자 속성인 성별, 연령별 정보는 가입자 정보로부터 산출되었다.

마지막으로 유입지(origin place) 코드는 OD 생성의 핵심이 되는 정보로, 해당 이용자의 추정된 거주지이다. 전(前) 달의 심야 시간에 16일 이상 동일한 지역(법정동 단위)의 셀에 머문 기록이 있는 경우 거주지로 판정되며<sup>2)</sup>, 그 외의 경우에는 불명으로 처리된다. 분석대상 자료 중 약 19.1%가 불명으로 처리되었다[3].

### 2.1.2 분석대상 자료

본 연구에서는 2014년 5월 10일(토)부터 5월 16일(금)까지 SKT 통신사의 수도권 자료를 대상으로 추출하였다<sup>3)</sup>. 자료의 형식은 선행연구[3]에서 제시한 형태와 동일한 형태이며, 선행연구에서는 전제자료에서 일부 데이터를 추출하여 분석하였으나, 본 분석에서는 모든 자료를 활용하여 분석하였다. 다만 거주지를 파악할 수 없는 불명 자료의 경우에는 제외하였다.

## 2.2 분석 방법론

본 연구에서는 일반적으로 교통 분야에서 사용하는 모형을 사용하지 않고, 통신자료의 특성을 고려하여 OD를 구축하였다. 전체 통행의 약 20% 미만을 차지하고 있는 비가정 기반 통행의 경우에는 통신자료를 활용하여 반영하기는 현실적으로 어려운 측면이 있다. 셀 단위의 성·연령별 유동인구 자료 중에서 중복 산정된 자료를 제외하여 법정동 기준의 출발지-도착지 OD를 생성하며, 중간 경유지에 대한 분석이 곤란하므로 도착지의 유동인구는 다시 출발지로 회귀한다는 가정 하에 OD를 생성하였다. KTDB와 비교하기 위해서 법정동 기준의 OD를 행정동 기준으로 변환하였으며, 이 때 변환비율을 적용한 기준은 인구를 기준으로 하였다.

여기서 가장 중요한 부분은 과연 중복 산정된 자료를 어떠한 방식으로 제거할 것인가 하는 것이다. Fig. 1에서 보는 바와 같

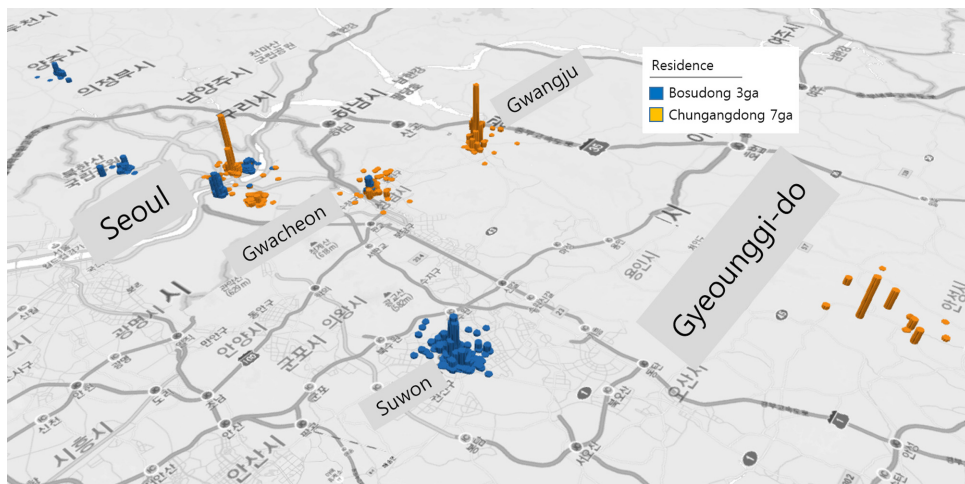


Fig. 1. Trip distribution from Bosudong 3ga and Chungangdong 7ga, Busan to Seoul metropolitan area.

2) 휴대전화 ID와 개인 주소를 매치시키면 거주지를 정확히 판별할 수 있지만 개인정보보호법 때문에 불가함. 16일을 거주지 판단 기준으로 정한 것은 특별한 기준은 없으나, 자료제공통신사인 SKT에서 1개월 중 절반 이상 야간에 머문 곳을 거주지로 보고 기초자료를 작성하여 제공하였기 때문에 본 연구에서도 이에 따름.  
 3) 휴대폰 명의자와 사용자가 다른 경우, 개인이 2대 이상의 휴대폰을 이용하는 경우, 휴대폰이 없는 경우는 소수이므로 데이터에서 별도로 구분하지 않음.

**Table 2.** Frequency of signalizing according to travel length.

Span of travel length	below 1km	1-5km	5-10km	10-30km	30-100km	Over 100km
Frequency	1.1880	1.9846	2.1839	2.9898	4.0721	6.7708

이 부산광역시 보수동3가와 중앙동7가 거주자가 특정일에 수도권에서 통신한 기록을 기준으로 공간적 분포를 나타낸 것이다. 즉 통신자료를 이용하여 특정 지역에 거주지를 둔 통행자가 수도권 지역을 방문하였다는 것은 알 수 있으나, 과연 어디가 최종 목적지인지에 대한 정보와 특정 1명이 몇 번의 통신기록을 남겼는지에 대한 기록은 알 수 없다. 따라서 방문지역에서의 중복 카운팅을 제외하고, 최종목적지를 파악할 수 있다면 정확한 통행기록(거주지→방문지)을 확인할 수 있을 것이다. 본 연구에서는 현 시점에서 취득할 수 있는 정보의 한계로 인해서 거주지와 방문지 간의 거리를 이용하여 거리대별로 중복횟수를 산정하였다. 거리대는 전체 6개 구간으로 분류하고, 2012년 KTDB의 거리대별 통행량과 비교하여 추출한 비율은 Table 2와 같다. 향후 중복횟수를 산정하는 부분은 충분한 자료가 확보된다면 조정 가능성이 있는 부분으로 사료된다. OD를 구축하는 과정은 다음과 같다.

- ① 특정 1일간 특정 지역(행정동)에서 통신이 발생된 총량 데이터 수집
- ② 해당 통신량의 유입지(거주지)별 분류
- ③ 유입지-특정 지역(행정동)간 거리 산출
- ④ 통신발생량을 중복횟수로 나누어 유입지-해당지역 OD 산출

비교 대상인 KTDB자료를 2012년자료로 설정한 이유는, 2012년 OD가 가구통행실태조사 결과를 이용하여 구축된 OD이기 때문이다. 본 연구의 중점이 가구통행실태조사를 통해 제작하는 OD와 통신자료를 이용하여 제작하는 OD를 비교하는 것이므로 2012년 OD와 통행패턴을 비교하였고, 총량 차이변화를 비교하기 위해 2015년 예측 KTDB OD도 병행하여 비교하였다.

## 2.3 분석 결과

### 2.3.1 분석결과 및 검증

제시된 절차에 따라서 일자별로 OD를 구축하였으며, Table 3은 행정동 단위로 구축된 OD를 시도 단위로 취합하여 평일평균 값을 제시한 것이다. 2012년 KTDB자료는 본 연구 결과에 비해 총량 기준 6% 작은 값을 가지고, 2015년 KTDB자료는 본 연구 결과에 비해 12% 작은 값을 가지는 것으로 분석되었다. 특히, 수도권 내부통행의 경우 인천→경기간 통행을 제외하면 2012년 KTDB자료는 본 연구 결과의 값에 비해 83%~99%의 값을 가지는 것으로 분석되었다. 인천↔경기간 통행은 KTDB자료가 본 연구결과에 비해 56%~60% 수준인 것으로 나타났는데, 그 원인은 2012년 KTDB의 행정동간 OD matrix에서 값을 가지는 쌍의 수가 583,121건에 불과하지만, 본 연구에서 수집한 통신데이터를 OD로 구축한 쌍에서 값을 가지는 수는 1,488,961건에 달하여 나타난 결과이다. 즉, KTDB 구축시 가구통행실태조사에서 인천, 경기 지역의 넓은 지역에 조사가 누락된 행정동이 많아 OD가 과소추정될 가능성("0"셀이 통신데이터보다 많이 존재)이 있는 반면, 본 연구의 통신데이터는 모든 존에서 이동 정보가 수집되기 때문에 누락이 거의 없어 나타난 현상이라 할 수 있다.

또한 Table 4에서는 통신자료를 이용하여 구축된 OD 자료를 이용하여 KTDB 자료와 비교하였다. 그 결과를 보면 시군구 단위에서는 2012년 KTDB(2011년 기준년도로 구축)와 매우 높은 상관관계를 보이고 있으며(Theil의 부등계수가 0에 가까울수록 일치성이 높음), 2015년 KTDB(예측자료로서 2014년 값을 제시하고 있지 않아 2015년 값과 비교하였음)도 기준년도와의 비교 보다는 다소 낮지만 매우 높은 상관관계를 보이고 있다. 행정동 단위의 비교에서는 상관계수가 0.1 정도 떨어지지만 모두 0.85 이상을 보이고 있는 수준이다. 따라서 통신자료를 이용하여 구축한 OD의 경우에도 어느 정도의 활용성은 충분히 있는 것으로 보인다.

### 2.3.2 통신자료 기반 OD의 지표 분석

통신데이터를 활용하여 OD를 구축할 경우 통신데이터의 특성을 살려 다양한 지표를 산출할 수 있다. 본 연구에서는 분석 사례로 주말계수, 즉, 토요일, 일요일의 통행량이 평일평균에 비해 얼마나 큰 가중치를 가지는지를 분석하였고, 성별, 연령별로 구분하여 산출하였다. Table 5에서는 산출 결과를 수도권에 해당하는 값만을 정리하여 제시한 결과이다. 분석 결과 수도권에서 주말의 통행량은 평일평균에 비해 전체적으로는 낮은 것으로 나타났고, 연령대별로 뚜렷한 특성을 보이는데, 특히 10대, 20대의 경우 토요일 통행량이 평일평균 통행량보다 큰 것으로 나타났다. 이는 평일의 경우 집-학교(직장)-집 2통행만 발생하다가 주말에는 집-친교-집, 집-식사-집 등 통행패턴이 변화하는 것으로 추정된다.

**Table 3.** Compared results of aggregated OD matrix.

Data	Conurbation	Seoul	Incheon	Gyeonggi	Others	Sum	
2014 This study	Seoul	27,616,222	440,533	3,676,203	284,908	32,017,867	
	Incheon	440,533	5,617,782	952,302	47,185	7,057,803	
	Gyeonggi	3,676,203	952,302	23,864,191	362,850	28,855,546	
	others	284,908	47,185	362,850	-	694,943	
	Sum	32,017,867	7,057,803	28,855,546	694,943	68,626,158	
2012 KTDB	Seoul	25,746,563	434,368	3,216,666	321,666	29,719,264	
		93%	99%	87%	113%	93%	
	Incheon	424,140	5,452,298	570,734	54,342	6,501,514	
		96%	97%	60%	115%	92%	
	Gyeonggi	3,064,428	536,655	23,403,855	351,016	27,355,953	
		83%	56%	98%	97%	95%	
	Others	333,590	68,709	384,674	-	1,050,900	
		117%	146%	106%	-	151%	
	Sum	29,568,721	6,492,030	27,575,930	990,950	64,627,631	
		92%	92%	96%	143%	94%	
	2015 KTDB	Seoul	22,144,540	474,275	3,432,787	324,937	26,376,539
			80%	108%	93%	114%	82%
Incheon		454,663	5,142,353	570,806	64,501	6,232,324	
		103%	92%	60%	137%	88%	
Gyeonggi		3,232,082	538,653	22,650,030	372,429	26,793,193	
		88%	57%	95%	103%	93%	
Others		336,617	81,016	409,459	-	827,092	
		118%	172%	113%	-	119%	
Sum		26,167,902	6,236,297	27,063,082	761,867	60,229,148	
		82%	88%	94%	110%	88%	

**Table 4.** Correlation analysis between KTDB and this study.

Date	Day	Comparison to KTDB OD - 'Si', 'Gun' & 'Gu' area unit				Comparison to KTDB OD - 'Dong' area unit			
		2012 KTDB		2015 KTDB		2012 KTDB		2015 KTDB	
		Correlation Coef.	Theil's U	Correlation Coef.	Theil's U	Correlation Coef.	Theil's U	Correlation Coef.	Theil's U
10 May, 2014	Sat.	0.9843	0.0895	0.9705	0.1254	0.8594	0.2693	0.8502	0.2776
11 May, 2014	Sun.	0.9826	0.1573	0.9680	0.1499	0.8603	0.3034	0.8521	0.3076
12 May, 2014	Mon.	0.9854	0.0851	0.9732	0.1248	0.8618	0.2647	0.8557	0.2707
13 May, 2014	Tue.	0.9854	0.0853	0.9730	0.1262	0.8610	0.2651	0.8547	0.2713
14 May, 2014	Wed.	0.9855	0.0856	0.9732	0.1205	0.8605	0.2682	0.8539	0.2742
15 May, 2014	Thr.	0.9850	0.0894	0.9717	0.1375	0.8600	0.2637	0.8534	0.2706
16 May, 2014	Fri.	0.9856	0.0909	0.9736	0.1394	0.8650	0.2583	0.8588	0.2652
Average (all week days)		0.9854	0.0859	0.9726	0.1219	0.8635	0.2655	0.8565	0.2719
Average (business days)		0.9855	0.0855	0.9731	0.1288	0.8638	0.2618	0.8574	0.2683

Table 5. Results of calculating weekend weight factor by gender and ages

Date	From	To	Men						Women						Total
			10~	20~	30~	40~	50~	60~	10~	20~	30~	40~	50~	60~	
SAT.	Seoul	Seoul	1.064	1.038	0.885	0.824	0.855	0.846	0.927	0.987	0.888	0.877	0.888	0.872	0.896
		Incheon	1.228	1.002	0.705	0.656	0.766	0.798	1.159	0.949	0.921	0.880	0.938	0.896	0.819
		Gyeonggi	1.225	1.000	0.722	0.670	0.784	0.825	1.153	0.980	0.916	0.879	0.912	0.893	0.826
	Incheon	Seoul	1.228	1.002	0.705	0.656	0.766	0.798	1.159	0.949	0.921	0.880	0.938	0.896	0.819
		Incheon	1.079	1.047	0.940	0.903	0.901	0.892	0.938	1.029	0.916	0.903	0.907	0.907	0.935
		Gyeonggi	1.174	0.991	0.780	0.740	0.796	0.819	1.041	1.012	0.975	0.902	0.903	0.895	0.847
	Gyeonggi	Seoul	1.225	1.000	0.722	0.670	0.784	0.825	1.153	0.980	0.916	0.879	0.912	0.893	0.826
		Incheon	1.174	0.991	0.780	0.740	0.796	0.819	1.041	1.012	0.975	0.902	0.903	0.895	0.847
		Gyeonggi	1.096	1.029	0.901	0.878	0.887	0.876	0.952	1.013	0.905	0.900	0.903	0.903	0.920
SUN.	Seoul	Seoul	0.840	0.836	0.640	0.576	0.624	0.634	0.712	0.784	0.649	0.644	0.685	0.682	0.671
		Incheon	0.934	0.701	0.435	0.375	0.465	0.523	0.819	0.640	0.593	0.575	0.663	0.635	0.528
		Gyeonggi	0.944	0.715	0.446	0.393	0.510	0.574	0.839	0.667	0.607	0.587	0.672	0.665	0.551
	Incheon	Seoul	0.934	0.701	0.435	0.375	0.465	0.523	0.819	0.640	0.593	0.575	0.663	0.635	0.528
		Incheon	0.819	0.856	0.685	0.630	0.674	0.693	0.683	0.843	0.680	0.675	0.721	0.730	0.708
		Gyeonggi	0.881	0.728	0.473	0.428	0.512	0.568	0.767	0.759	0.656	0.616	0.669	0.664	0.564
	Gyeonggi	Seoul	0.944	0.715	0.446	0.393	0.510	0.574	0.839	0.667	0.607	0.587	0.672	0.665	0.551
		Incheon	0.881	0.728	0.473	0.428	0.512	0.568	0.767	0.759	0.656	0.616	0.669	0.664	0.564
		Gyeonggi	0.864	0.851	0.649	0.620	0.657	0.679	0.724	0.838	0.679	0.680	0.722	0.729	0.700

### 3. 결 론

본 연구에서는 통신자료를 이용하여 수도권의 OD를 구축하였고, 그 결과를 KTDB OD와 비교하였으며, 두 OD 사이에 높은 상관관계가 있음을 확인하였다. 그러나, 비교적 단순한 절차에 의해서 생성되었으므로, 여러 가지 한계는 가지고 있다. 첫째, 비가정기반 통행은 반영할 수 없다는 한계를 가지고 있으며, 둘째, 수도권에 들어온 통행은 반영할 수 있지만 수도권에서 타 지역으로의 통행은 누락되었다. 그러나 이 부분은 지역적 범위를 전국으로 확대하여 분석한다면 해결될 수 있는 부분이다. 셋째, 장거리 통행의 경우에는 해당 지역에서 숙박하는 경우가 발생하지만, 본 분석에서는 숙박 통행은 없는 것으로 가정하여 당일 모두 거주지로 돌아가는 것으로 하여 OD를 구축하였다. 이 외에도 중복비용의 제거 방식 등 다양한 한계는 존재할 수 있겠으나, 본 연구에서 구축한 OD가 기존의 KTDB에서 구축하고 있는 OD와 상관성이 매우 높은 점을 고려한다면 여러 가지 시사점을 줄 수 있을 것으로 판단된다.

특히, 기존 KTDB OD 구축방법인 표본조사에 의한 방법뿐만 아니라 본 연구에서 제시한 통신자료를 이용한 방법도 OD 구축이 가능함을 보여주었고, 기존 OD 구축 방법에 비해 다양하고 세분화된 OD를 구축할 수 있다는 데에 연구의 의의가 있다.

본 OD를 활용하여 현재의 조사방식으로는 통행수요가 많지 않은 OD쌍은 조사에서 누락되어 값이 존재하지 않는 부분을 해결할 수 있을 것이다. 기존년도 KTDB에 비해서 값이 존재하는 OD쌍의 수가 본 연구에서는 2배 이상인 점이 그 증거가 될 수 있다. 또한 다양한 연령대, 성별로 OD의 구축이 가능하다는 점, 주말 계수를 굳이 구축할 필요없이 바로 산정할 수 있다는 점, 요일별 OD를 구축할 수 있다는 점 등은 통신자료가 가지는 매우 유용한 장점이 될 것이다. 또한, 이러한 다양한 OD 구축은 철도정책 수립 및 철도교통계획에서 적절히 활용할 수 있을 것이다.

향후 통신자료가 교통분야에 보다 유용한 정보가 되기 위해서는 개인이력정보를 확인할 수 있는 방향으로 자료가 구축될 필요성이 있다. 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] <http://www.msip.go.kr> (Accessed 18 April 2016).
- [2] <http://kosis.kr> (Accessed 18 April 2016).
- [3] K. Kim, I. Lee, J.H. Min, H. Kwak (2015) Trip generation analysis using mobile phone data, *Journal of the Korean Society for Railway*, 19(5), pp. 481-488.
- [4] S. Kim, B. Yoo, S. Kang (2005) Origin-Destination estimation based on cellular phone's base station, *Journal of Korean Society of Transportation*, 23(1), pp. 93-102.
- [5] F. Calabrese, G.D. Lorenzo, L. Liu, C. Ratti (2011) Estimating origin-destination flows using mobile phone location data, *IEEE CS Pervasive Computing*, October-December 2011, 4(10), pp. 36-44.
- [6] P. Wang, T. Hunter, A.M. Bayan, K. Schectner, M.C. Gonzalez (2012) Understanding road usage patterns in urban areas, *Scientific Reports*, 2, pp. 1-6.
- [7] M.S. Iqbal, C.F. Choudhury, P. Wang, M.C. Gonzalez (2014) Development of origin-destination matrices using mobile phone call data, *Transportation Research Part C*, 40, pp. 63-74.
- [8] S. Jiang, Y. Yang, G. Fiore, E.Jr.J.F. Frazzoli, M. Gonzalez (2013) A review of urban computing for mobile phone traces: current methods, challenges and opportunities. *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, Chicago, Illinois, pp. 1-9.
- [9] H.S. Schneider, S.V. Ukkusuri, M.C. Gonzalez (2013) Spatiotemporal patterns of urban human mobility, *Journal of Statistical Physics*, 151(304), pp. 151.

(Received 18 July 2016; Revised 16 August 2016; Accepted 22 August 2016)

---

**Kyoungtae Kim:** [ktim@krri.re.kr](mailto:ktim@krri.re.kr)

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

**Dongkyu Oh:** [ohdong@krri.re.kr](mailto:ohdong@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](mailto:mook79@krri.re.kr)

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

**Jae Hong Min:** [jhmin@krri.re.kr](mailto:jhmin@krri.re.kr)

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