

모바일 통신 자료와 O/D 통행량의 상관성 분석 - 대구광역시 사례를

중심으로

김근욱* · 정연식**

Kim, Keun-uk*, Chung, Younshik**

Correlation Analysis Between O/D Trips and Call Detail Record: A Case Study of Daegu Metropolitan Area

ABSTRACT

Traditionally, travel demand forecasts have been conducted based on the data collected by a survey of individual travel behavior, and their limitations such as the accuracy of travel demand forecasts have been also raised. In recent, advancements in information and communication technologies are enabling new datasets in travel demand forecasting research. Such datasets include data from global positioning system (GPS) devices, data from mobile phone signalling, and data from call detail record (CDR), and they are used for reducing the errors in travel demand forecasts. Based on these background, the objective of this study is to assess the feasibility of CDR as a base data for travel demand forecasts. To perform this objective, CDR data collected for Daegu Metropolitan area for four days in April including weekdays and weekend days, 2017, were used. Based on these data, we analyzed the correlation between CDR and travel demand by travel survey data. The result showed that there exists the correlation and the correlation tends to be higher in discretionary trips such as non-home based business, non-home based shopping, and non-home based other trips.

Key words : Mobile phone big data, Call detail record (CDR), Origin-Destination (O/D), Correlation analysis, Paired t-test, Smart city

초 록

전통적으로 통행수요예측은 개별 면접조사를 통해 수집된 자료를 기반으로 수행되었으며, 통행수요 예측의 정확성은 이러한 문제로부터 지속적으로 제기되어 왔다. 최근 정보통신 기술의 발전에 따라 통행수요예측 연구에서 새로운 자료의 활용이 다루어지고 있다. 이러한 자료는 GPS기반 위치 자료, 휴대전화 신호의 자료, CDR (Call Detail Record) 자료 등으로 포함하며, 통행 수요예측의 오류를 줄이기 위한 활용과 관련한 연구가 진행되고 있다. 이를 바탕으로 본 연구의 목적은 통행수요예측의 기초자료로 CDR의 적용 가능성을 평가하는 것이다. 이를 위해 본 연구에서는 대구광역시 평일과 주말을 포함한 2017년 4월의 4일 동안 수집한 CDR 자료를 사용하였다. 즉, CDR 통신량과 기존 면접조사의 O/D 통행량 간의 상관성을 분석하였다. 그 결과, CDR과 전통적 방식에 의한 교통수요는 서로 상관성이 존재하는 것으로 나타났으며, 통행목적별 통행량 비교결과에서는 주말 첨두시 CDR이 비가정기반 쇼핑·여가 목적 통행량과 같은 선택적 통행에서 상관성이 높은 것으로 나타났다.

검색어 : 통신 빅 데이터, CDR, 기중점 통행량, 상관분석, 대응표본t-test, 스마트 도시

* 부산시의회 입법정책담당관실 연구위원 (Busan Metropolitan Council Legislation & Policy Office · kevinlim110@gmail.com)

** 중신회원 · 교신저자 · 영남대학교 도시공학과 조교수 (Corresponding Author · Yeungnam University · tpgist@yu.ac.kr)

Received August 12, 2019/ revised August 26, 2019/ accepted September 9, 2019

1. 서론

전통적인 방식을 적용한 교통수요 추정 및 예측 과정은 크게 2가지 한계를 가진다고 볼 수 있다. 우선, 교통수요 추정과 예측은 가구통행실태 조사와 같은 전통적 조사 방식에 의한 결과를 기반으로 진행되어, 매우 낮은 수준의 통행 표본을 기반으로 교통수요를 추정하게 되는 한계가 존재한다. 즉, 시간과 비용적 한계로 인해 국내의 경우 약 2 % 수준의 샘플을 표본으로 조사를 수행하지만, 정상적으로 수집되는 샘플은 약 1.0 %에서 1.2 % 수준에 머물며, 결국 이 자료를 기반으로 교통수요를 추정하게 되어 그 정확도에 대한 한계가 존재하게 된다. 이러한 한계는 샘플에 대한 비율을 높이면 극복 가능성이 존재하지만, 시간과 비용에 대한 제약조건을 해결해야 하는 문제와 직결된다. 또 다른 한계는 적용되는 방법론의 문제이다. 교통수요 예측을 위한 전통적 방법은 통행발생(trip generation), 분포(trip distribution), 수단선택(modal choice), 배정(traffic assignment)의 4단계 수요추정 기법을 적용하는 것이다. 그러나 이 방법론은 각 단계별 추정결과(Output)가 다음 단계의 투입요소(Input)가 되는 순차적 수요추정의 문제로 오차의 누적전이(error propagation)를 비롯하여 각 단계별 활용되는 모형에 대한 한계가 지속적으로 제기되어 왔다. 결과적으로 대부분의 선행 연구는 이러한 방법론적 한계 극복에 집중되어 왔다.

한편, 최근 정보통신 기술의 발전과 스마트 폰과 같은 개인 휴대용 기기 보급률이 증가하면서, 기존 교통수요 추정 및 예측 분야에 대한 새로운 도전이 시도되고 있다. 즉, 스마트 폰과 같은 개인 휴대용 기기의 신호 혹은 통화기록 정보를 활용하여 개인의 통행 기록을 추정하는 이른바 모바일 통신 빅 데이터 분야의 연구가 활발히 진행되고 있다(Kim et al., 2005). 특히, 국내의 경우 모바일 통신 가입자 수가 2017년 기준 63,658,688명²⁾으로 국내 전체 인구수가 5,136만명 수준임을 감안하면, 전체 인구수 보다 많은 가입자를 보유하고 있음을 알 수 있다. 다시 말해, 모바일 통신 자료를 활용하게 되면, 이론적으로 전 국민의 교통수요를 실시간적으로 파악할 수 있다고 가정할 수 있다.

이러한 배경 하에, 본 연구는 모바일 통신 빅 데이터를 활용하여 교통수요 추정에 사용되는 기종점 통행량(origin-destination: O/D) 자료의 정확성을 높이기 위한 방법론을 제시하고자 수행되었다. 즉, 본 연구에서는 통행 수요를 추정하는 데 있어 기초자료로서 모바일 빅 데이터의 활용가능성을 살펴보기 위해 기존 O/D 통행량과 모바일 자료의 양적(量的) 상관성을 비교·분석하였다. 이를

1) 국가교통DB의 국가교통조사 여객부분(2016년기준) 가구통행실태조사는 2010년 인구주택총조사 전국 총조사 가구의 1.15 % (202천 가구)로 조사(KTDB, 2016).
2) 과학기술정보통신부, 2017년 12월 무선통신서비스 가입자 통계 (Ministry of Science and ICT, 2018).

기반으로 교통수요 예측을 위한 기초자료로서 모바일 통신 빅 데이터의 기존 O/D 대체 혹은 보완 가능성 여부를 살펴보았다.

2. 선행연구 고찰

Table 1은 모바일 통신 자료를 활용하여 O/D와의 관계성 및 활용성에 대한 주요 선행연구를 요약한 것이다. Kim et al.(2015)는 이동통신사((주)SK-T)에서 제공하는 수도권 지역의 상·연령별 이동인구 자료로부터 통행발생량을 산정하기 위한 방법을 제시하고 해당 지역의 O/D 통행발생량의 상관성 분석을 통해 자료의 활용 가능성을 확인하였다. 그 결과 무선통신 자료를 이용한 통행발생량 추정은 기존의 국가교통 데이터베이스(Korea Transport Database: KTDB)에서 구축한 직접 조사 방식에 기반한 O/D 통행발생량 자료와 매우 높은 상관관계를 가지는 것으로 분석되었다. 그러나 이 연구는 중준(경기도 시군)을 공간적 단위로 설정하여 통행유출입(trip attraction & production)량을 도출하였다.

Yoon et al.(2015)는 현재의 교통수요 모형을 이용한 수요 추정의 경우 수요 변화요인을 신속하게 파악하고 분석하기 어려워 지역 간 교통수요예측에 한정하여 교통 분야 빅 데이터인 고속도로 요금징수시스템(toll collection system: TCS)에 의한 통행량자료, 이동 통신자료를 활용한 KTDB의 보완이 필요하며, 시계열적 변화를 보여주는 실측치를 활용한 검증이 필요하다고 주장하였다. 또한 Kim et al.(2016)의 연구에서는 휴대전화 이용자의 시간대별 기지국 위치정보를 기반으로 이용자의 이동경로를 추적하고, O/D를 추출한 가정기반 통행만을 대상으로 한 수도권 중심의 O/D를 구축하였다. 그 결과 행정동 단위로 변환한 이동통신자료 기반 O/D와 KTDB의 O/D가 높은 상관관계를 보이는 것으로 분석되었다.

Furletti et al.(2014)은 모바일 자료를 이용하여 도시 간 이동성에 대해 빅 데이터가 어느 정도 신뢰성 있고 시기적절한 추정이 가능한지에 대해 살펴본 결과 행정자료 소스의 통합을 통해 주거지와 근무지를 연계한 통근(통학)을 비롯한 O/D 매트릭스 자료의 구축이 가능함을 증명하였다. Pappalardo et al.(2015)은 인간의 이동성과 사회경제적 발전에 활용하는 차원에서 모바일 자료를 이용하여 개인의 이동량과 이동수단의 다양성을 추정하고 이를 사회경제적 지표와 상관성을 살펴보는 연구를 진행하였다. 그 결과 사회경제지표를 대체가능한 새로운 통계적 지표로서 이러한 빅 데이터의 활용이 가능하다고 주장하였다.

Alexander et al.(2015)은 이동통신 자료를 이용해 평균 일일 O/D통행을 추정하는 방법을 제시하고 이동통신 자료의 관측 빈도에 따라 집, 근무지 또는 다른 지점들을 주중과 시간대별로 추천하는 과정을 거쳤다. 이렇게 제시된 방법론은 일시적이고 공간적인 통행 분포를 지역적이고 국가적인 조사를 검증해줄 수 있는 자료로

Table 1. Summary of Previous Studies on Correlation Analysis of Two O/Ds from Mobile Phone Data and from Traditional Survey Data

Author	Research objective and findings	Data used
Kim et al.(2015)	- Introduce ways to derive trip generation information from mobile phone data and verify the value of the system for practical used by correlation analysis with KTDB trip generation data. - Trip generation information by mobile phone data correlated KTDB trip generation data.	Mobile phone data
Yoon et al.(2015)	- Find ways to initiate the use of big data collected by various public agencies and private companies for travel demand forecasting. - Big data such as TCS and mobile phone data with better spatiotemporal resolution than traditional survey data could enable more responsive travel demand forecasting.	TCS and mobile phone data
Kim et al.(2016)	- Analyze correlation between O/D from mobile phone data and O/D from KTDB. - Two O/Ds were highly correlated as 0.98 for the city level zone and 0.85 for the Dong level zone.	Mobile phone data for Seoul Metropolitan area
Furletti et al.(2014)	- Deploy the massive and constantly updated information carried by CDRs for estimating population statistics related to residence and mobility. - As a proxy of traditional O/D, a population and flow estimation based on mobile phone data were generally encouraging and very accurate in comparison to the traditional O/D.	Mobile phone data
Pappalardo et al.(2015)	- Identify the relations between human mobility patterns and socio-economic development. - Aggregated human mobility patterns were correlated with these socio-economic indicators. - The diversity of mobility exhibited the strongest correlation with the external socio-economic indicators. - The volume of mobility and the diversity of mobility show opposite correlations with the socioeconomic indicators.	Mobile phone data
Alexander et al.(2015)	- Present methods to estimate average daily O/D trips from triangulated mobile phone data. - The proposed methodology based on mobile phone data were analogous to the outputs of the trip generation and distribution steps of traditional four-step travel demand models.	Mobile phone data

Table 2. CDR Data Format in this Study

Fields	Field name	Data type	Description
1	Timecode	string	4-digit year, 2-digit month, 2-digit day
2	X_coord	Number	X coordinate of cellular station in UTM-K Coordinate system
3	Y_coord	Number	X coordinate of cellular station in UTM-K Coordinate system
4	Time_used	Number	call duration in hour unit

활용이 가능할 것으로 보였다. 이처럼 모바일 빅 데이터를 활용한 다양한 국내외 선행연구 검토 결과, 대체적으로 모바일 빅 데이터가 인간의 활동, 즉 ‘통행’을 추정하는데 매우 유의미한 자료임을 입증해주고 있으며 본 연구에서도 이러한 관점의 연장선상에서 모바일 빅 데이터가 향후 O/D를 보완·대체할 수 있는 자료로서 활용 가능성을 입증하기 위한 차원의 기초 연구로 대구광역시로 한정하여 모바일 자료와 기존 O/D자료와의 상관성을 분석해보는 것에 의의를 가진다.

3. 자료의 구성 및 연구방법

3.1 자료의 구성

본 연구에서 활용된 모바일 빅 데이터는 대구광역시 대상으로 2017년 4월 주중 2일(12일, 13일) 및 주말 2일(8일, 9일) 동안

(주)SK-T 이용자들이 의해 수집된 자료를 포함한다. 이 자료는 10세 이상의 이용자에게 대하여 10세 단위로 연령이 구분되어 있으며, 통신자료 취득 당시 해당 통신사의 가입자 수는 대한민국 총 인구 55,332,064명 기준 24,511,118명으로 총인구대비 44.3 %의 시장 점유율을 나타내었다.³⁾ 통신량 자료(call detail record: CDR)는 ArcGIS에서 사용 가능한 셰이프(shp)파일로 제공되어졌으며, 50 × 50 m 격자 형태의 셀(pCell)의 point로 각 연령대별 CDR이 속성 값으로 내포되어 있는 형태이다. 세부자료 설명은 Table 2와 같다. CDR 자료는 개별 이동통신 단말기에 관여하는 기지국 자료로부터 위치를 측정하여 정의된 pCell 중 가장 가까운 값에 사상(mapping)하여 생성된다. 개인식별 코드가 없이 각 셀에 누적

3) 과학기술정보통신부, 2017년 12월 무선통신서비스 가입자 통계 (Ministry of Science and ICT, 2018).

정보가 기록되는 방식이므로, 셀 간의 이동정보는 직접 추출할 수 없는 구조이다.

Kim et al.(2016)는 각 셀에 기록된 통신량은 개인의 통신기록으로서 음성(통화), 문자, 자료(인터넷)을 이용한 경우가 포함되며, 푸시(push) 서비스 등 수동적으로 발생하는 통신도 이에 포함이 된다. 이 통신량은 하루 중 해당 셀에서 이동통신 서비스를 이용한 기록이 있는 이용자의 통계수치라고 볼 수 있고 정확히는 회선에 대한 정보이지만 대체적으로 회선과 이용자가 동일하다는 가정 하에서 통신자료를 통행으로 다룰 수 있다고 주장하였다. 따라서 본 연구에서 기준한 통신량 자료는 1일 단위로 해당 셀에서 통신한 사람의 수로서 엄밀히 말해 각 셀에서 통신(문자, 통신 수·발신, 자료 통신 기록 기준이며, wifi통신은 제외)한 휴대전화의 수라고 볼 수 있다. 사실상 이러한 통신량 자료(CDR)는 통행의 방향성 즉, 기·종점을 의미하는 O/D로서 통행경로를 유추하기에는 다소 어려움이 있고 사실상 본 연구에서 다루고자 하는 통신량과 O/D 통행발생·유입량의 직접 비교를 통한 자료의 검증 목적과는 다소 부합되지 않는 영역이라고 판단된다. 이에 CDR 통신량과 O/D 통행발생·유입량을 기준으로 하여 통행량의 비교를 통한 활용가능성 분석에만 한정하여 활용하였다.

3.2 연구방법

CDR의 활용성 검증을 위해 O/D 자료와의 비교분석은 KTDB 제공 2016년도 대구광역시 행정동(139개 존) 여객 O/D의 통행목적별 통행발생·집중 통행량을 기준으로 CDR의 요일별·시간대별 자료를 추출하여 통행량에 대한 상관관계 분석을 시행하였다. 이후 동일 모집단의 사전·사후점수의 평균을 비교할 때 주로 사용되는 대응표본 t-test 분석을 실시하였다. 서로 다른 두 자료의 직접 비교를 위해 CDR을 기준 2일 단위 제공처에서 1일 통행량 자료로 변환하고 자료 취득 당시 (주)SK-T의 전국단위 44.3%의 점유율을 고려하여 보간법을 통해 행정동(소존) 단위의 통신량을 전수화시켜 도출하였다. 또한 통행 단위의 통일성을 기하기 위해 기준연도 O/D와 같이 일 단위 통행을 그대로 사용하되 통행목적별 통행량과

통행발생(production) 및 통행유입(attraction) 자료를 모두 활용하여 비교 분석하였다. 이는 CDR의 특성이 공간적 단위에서 집계되는 통신량이므로 이것이 통행발생인지 통행유입인지에 대한 정의가 모호함으로 이를 모두 적용시켜 O/D와 비교해 볼 필요가 있다고 판단하였다.

본 연구에서는 분석을 위해 통행목적 여객 O/D를 적용하였는데, 이는 CDR이 개인 모바일 특성에 기초한 수집 자료라는 점에 착안한 결과이다. 즉, 수단 O/D를 적용할 경우, 재차인원 또는 교통량 등을 고려한 지표이어야 하지만, 실제 CDR은 사람의 이동에 바탕에 둔 개인의 통신 자료라는 의미에서 목적 O/D의 적용이 보다 합리적인 것이라는 판단에 근거하였다. 또한, CDR은 개인의 모바일 통신에 의해 생성되어 매우 높은 샘플링으로 기록된 자료임을 고려할 때, 전수화 과정을 통해 도출된 O/D와의 비교 분석은 자료의 신뢰성 검증 차원에서 의미가 있을 것으로 판단된다. 한편, 본 연구에서 통신자료(2017년)와 O/D 자료(2016년)의 시점이 정확히 일치하지 않는다. 즉, 취득 시점에서 가장 근접한 시점인 2016년 O/D 자료를 활용하였다는 한계가 존재한다.

4. CDR 및 O/D 상관성 분석결과

4.1 비교자료 구성

CDR은 구조 상 그 자체만으로는 통행목적별을 알 수 없기 때문에 기존 O/D의 통행목적별 구분을 요일별, 시간대별 CDR과 비교함으로써, 행정동 기준으로 집계된 통행량(量)적인 측면에서 비교 분석하였다. 통행목적별 자료의 비교분석은 비교적 고정적이고 반복적인 통행 성격을 가진 통근, 통학과 같은 필수적 통행(mandatory trip)과 쇼핑, 여가, 기타 목적의 선택적 통행(discretionary trip)으로 구분하여 진행하였다. 즉, 평일 오전 및 오후 첨두시 CDR은 필수적 통행으로, 주말 오전 및 오후 첨두시 CDR은 선택적 통행으로 간주하였다. 또한, 통행 목적별 O/D의 경우 가정기반 통근(home-based work: HBW), 통학(home-based school: HBSc) 통행을 필수적 통행으로, 가정기반 쇼핑(home-based shopping:

Table 3. Definition of Comparison Datasets

Trip category	CDR	O/D	
Total trip	· Aggregated CDR (weekday AM peak + weekday PM peak)	Total production	
		Total attraction	
Mandatory trip	· CDR for weekday AM peak period (CDR AM peak)	Production	HBW, HBSc
	· CDR for weekday PM peak periods (CDR PM peak)	Attraction	
Discretionary trip	· CDR for weekend AM peak (CDR AM peak)	Production	HBS, HBO, NHBB, NHBS, NHBR, NHBO
	· CDR for weekend PM peak (CDR PM peak)	Attraction	

Note) Weekday AM peak: 07:00~09:00, weekday PM peak: 17:00~19:00, Weekend AM peak 09:00~12:00, and weekend PM peak: 14:00~16:00

HBS) 및 가정기타(home-based other: HBO), 비가정 기반 업무(non-home based business: NHBB), 쇼핑(non-home based shopping: NHBS), 여가(non-home based recreation: NHBR), 기타(non-home based other: NHBO) 통행은 선택적 통행으로 구분하여 CDR과 비교 분석하였다(Table 3).

4.2 분석결과

4.2.1 총통행량 기준

Table 4는 집계된 CDR과 총통행발생량(total production)과 총통행유입량(total attraction)의 비교결과를 제시한 것이다. 행정동 단위로 집계된 CDR과 총통행발생량 합계를 비교한 결과, CDR은 6,137,942통행, 총통행발생량은 5,994,680통행으로 약 2.3 % 차이를 보이는 것으로 나타났다. 중존인 구·군단위로 집계된 결과에서는 수성구, 남구, 중구의 경우 44.8~49.7 %의 높은 오차율을 보였으나, 북구와 동구, 달성군의 경우 10.1~19.4 %로 비교적 낮은 오차율로 나타났다. 구체적으로 행정동 단위로 살펴본 결과, 북구 국우동, 수성구 범물1·2동, 고산3동과 같은 경우에는 총통행발생량이 CDR보다 높았고, 남구 이천동, 대명2동의 경우 CDR이 총통행발생량보다 높게 나타났다. 즉, 대구광역시 중 대규모 공동주택단지 밀집한 지역일수록 총통행발생량이 CDR보다 높게, 반대로 공동주택이나 주거지 밀집이 낮은 지역인 경우 CDR이 총통행발생량보다 높게 나타났다.

CDR과 총통행유입량의 비교 결과, 수성구, 남구의 경우 44.4~48.5 %의 높은 오차율을 보였으나, 북구와 동구의 경우

7.2~14.1 %로 비교적 낮은 오차율로 나타났다. 특히 수성구 범물1·2동, 고산3동과 같은 경우에는 총통행유입량이 CDR보다 높은 것으로 나타났으나, 남구 이천동, 대명11동의 경우 CDR이 총통행유입량보다 높은 것으로 나타났다. 이러한 결과는 앞서 총통행발생량과 유사한 지역으로 대규모 공동주택단지 등 거주인구가 많은 지역의 경우, 총통행발생유입량이 CDR보다 높게 나타난 것을 알 수 있다. 이는 CDR과 O/D자료의 특성에 따른 이유로 판단된다. 즉, CDR은 이용자가 개인 모바일 기기의 접속 빈도(통신사 통신기기 기반 통화, 문자, 인터넷 접속 등)와 관련이 있다. 통행유입과 관련된 오후 침투시간(주중 오후 5사~7시 혹은 주말 오후 2사~4사)에는 식사를 하거나 업무 마감으로 인한 집중에 따른 낮은 개인 모바일 기기 접속으로 해석이 된다. 또한, 주거지에서는 대부분 고정된 wifi를 연결하여 개인 모바일 기기를 접속하여 통신사 CDR에는 통신 기록이 낮게 형성되었을 가능성이 존재한다. 그러나 이러한 해석 유추를 위한 보다 깊은 연구가 요구된다.

Table 5는 Table 4에 제시된 자료를 기반으로 CDR과 총통행발생 및 유입량에 대한 상관관계 분석 결과를 제시한 것이다. 분석 결과, 집계된 CDR과 총통행발생량은 0.666의 상관관계를 보였으며, 총통행유입량과는 0.673의 상관관계를 나타내었다. 한편, Table 6과 같이 집계된 CDR과 총통행발생 및 유입량의 평균 차이에 대한 대응표본 t-test 분석에서는 유의확률이 각각 0.679, 0.677로 나타나 차이가 없으므로 검증되었다. 즉, 집계된 CDR과 총통행발생 및 유입량의 두 집단 간 평균 차이는 없는 것으로 분석되었다.

Table 4. Comparison Result of Aggregated CDR and Total Production/Attraction

Fields	Total trip			Difference		Error rates	
	CDR (A)	Production (B)	Attraction (C)	A-B	A-C	((A-B/A)*100)	((A-C/A)*100)
Jung-Gu	969,670	535,173	585,259	434,497	384,411	44.8 %	39.6 %
Dong-Gu	664,352	743,576	758,183	-79,224	-93,831	-11.9 %	-14.1 %
Seo-Gu	555,189	421,704	423,874	133,485	131,316	24.0 %	23.7 %
Nam-Gu	694,123	358,814	357,356	335,309	336,767	48.3 %	48.5 %
Buk-Gu	1,029,911	1,134,357	1,103,968	-104,446	-74,057	-10.1 %	-7.2 %
Susung-Gu	690,000	1,033,275	996,183	-343,275	-306,184	-49.7 %	-44.4 %
Dalseo-Gu	1,036,247	1,366,165	1,356,134	-329,918	-319,887	-31.8 %	-30.9 %
Dalseong-Gun	498,449	401,616	413,723	96,833	84,726	19.4 %	17.0 %
total	6,137,942	5,994,680	5,994,680	143,262	143,262	2.3 %	2.3 %

Table 5. Correlation Analysis Results of Aggregated CDR and Total Production/Attraction

	Total production trips	Total attraction trips
Aggregated CDR	0.666**	0.673**

** Significant at the level 0.01 (2-tailed test)

Table 6. T-test Results of Aggregated CDR and Total Production/Attraction

Fields				t	df	Sig. (2-tailed)
	Mean	Std.Deviation	Std. Error Mean			
Aggregated CDR & Total production	1030.6	29288.6	2484.2	.415	138	.679
Aggregated CDR & Total attraction	1030.6	29148.1	2472.3	.417	138	.677

Table 7. Correlation Analysis Results of CDR for Weekday Peak Periods and Mandatory Production/Attraction

	Mandatory production	Mandatory attraction
CDR AM peak	.512**	.512**
CDR PM peak	.481**	.479**

** Significant at the level 0.01 (2-tailed test)

Table 8. T-test Results of CDR AM/PM Peak and Mandatory Production/Attraction

Category				t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean			
CDR AM peak & Mandatory production	-3465.7	15372.1	1303.8	-2.658	138	.009
CDR AM peak & Mandatory attraction	-3337.5	15450.6	1310.5	-2.547	138	.012
CDR PM peak & Mandatory production	3143.8	21144.6	1793.5	1.753	138	.082
CDR PM peak & Mandatory attraction	3272.1	21208.2	1798.9	1.819	138	.071

Table 9. Correlation Analysis Results of CDR for Weekend Peak Periods and Discretionary Production/Attraction

	Discretionary production	Discretionary attraction
CDR AM peak	.620**	.614**
CDR PM peak	.589**	.572**

** Significant at the level 0.01 (2-tailed test)

Table 10. T-test Results of CDR AM/PM Peak and Discretionary Production/Attraction

Fields				t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean			
CDR AM peak & Discretionary production	13127.7	23919.6	2028.8	6.471	138	.000
CDR PM peak & Discretionary attraction	13127.6	24070.6	2041.6	6.430	138	.000
CDR AM peak & Discretionary production	9322.1	26428.9	2241.7	4.159	138	.000
CDR PM peak & Discretionary attraction	9322.1	26804.3	2273.5	4.100	138	.000

4.2.2 필수적 통행량 기준

필수적 통행과 관련하여 평일 오전과 오후 첨두시간대 CDR과 가정기반 통근 및 통학에 대한 발생 및 유입 통행량을 비교하여 통행목적별 상관성을 평가하였다. Table 7에 제시된 바와 같이, 평일 오전 첨두시간대 CDR과 가정기반 통근 및 통학 통행(필수 통행) 발생·유입량은 모두 0.512의 상관관계를 나타내었다. 반면, 평일 오후 첨두시간대 CDR과 가정기반 통근 및 통학 통행(필수 통행) 발생·유입량은 각각 0.481과 0.479로 평일 오전첨두 시간대보다 낮은 상관관계를 보였다. 또한, Table 8의 대응표본 t-검정

결과, 평일 오전첨두시 CDR과 가정기반 통근·통학통행 발생·유입량의 유의확률은 유의수준 0.05보다 작은 0.009~0.012로 나타나 “두 통행량은 차이가 있음”으로 검증되었다. 반면, 평일 오후첨두시 CDR과 가정기반 통근·통학통행발생·유입량의 유의확률은 0.082~0.071로 나타나 “두 변수간 차이가 없음”으로 검증되었다.

4.2.3 선택적 통행량 기준

선택적 통행과 관련하여 주말 오전·오후첨두시 CDR과 비가정기반 업무, 쇼핑, 여가, 기타 통행발생 및 유입량을 비교하여 통행목

적별 상관성을 분석하였다. Table 9에서 각 자료별 상관관계를 분석한 결과, 주말 오전첨두시 CDR과 비가정기반 업무, 쇼핑, 여가, 기타 통행발생 및 유입량은 각각 0.620과 0.614의 높은 상관관계를 나타내었다. 반면, 주말 오후첨두시 CDR과 O/D 비가정기반 업무, 쇼핑, 여가, 기타 통행발생 및 유입량은 각각 0.589과 0.572로 주말 오전첨두시 CDR과 통행발생 및 유입량보다는 상대적으로 낮은 상관관계를 나타냈다. 또한, Table 10에서 대응표본 t-test 분석결과, 주말 오전·오후첨두시 CDR과 비가정기반 업무, 쇼핑, 여가, 기타 통행발생 및 유입량의 유의확률은 모두 0.000으로 나타나 “두 변수간 평균 차이가 있음”으로 검증되었다.

5. 결론

본 연구는 CDR 및 O/D 통행발생·유입량을 통행목적별로 집계하여 상관분석 및 대응표본 t-test를 통해 두 자료가 가진 차이점 및 상관성을 분석하였다. 먼저 지역적으로 CDR과 O/D의 통행량을 비교한 결과, 주거지가 밀집된 지역에서는 CDR보다 O/D 통행량이 상대적으로 많은 것으로 나타났으며, 도심지역과 같은 경우에는 CDR이 O/D 통행량보다 많은 것으로 나타났다. 이는 CDR이 실제 통행을 하고 있는 사람들의 활동을 나타내는 자료라는 점에서 기존 O/D가 일정 표본면접조사에 따른 가구 및 가구원 전수화계수를 통해 상대적으로 주거기능이 집중된 지역에서 통행량이 과대 추정되었음을 의미하는 것으로 해석된다. 또한, 각 통행목적별로 통행량을 구분하여 모형의 상관성을 나타내는 R^2 와 대응표본 t-test 유의확률을 비교하여 가장 유사한 대안을 비교하였다. 그 결과, CDR 통행량은 상대적으로 O/D 통행발생량보다는 O/D 통행유입량과의 관계가 높은 것으로 나타났으며, 통행목적별로는 필수적 통행에서 평일 오전첨두시 CDR과 가정기반 통근, 통학 통행발생 및 유입량과의 관계가 존재하는 것으로 나타났다. 선택적 통행에서는 주말 오전첨두시 CDR과 비가정기반 업무, 쇼핑, 여가, 기타 통행 발생량과 비교적 관계가 있는 것으로 나타났지만, 다른 통행량과는 특별한 상관성이 파악되지 않았다. 반면, 통행량의 목적별 비교에서 선택적 통행과 유사하다고 볼 수 있는 주말 CDR은 비가정기반 업무, 쇼핑, 여가, 기타 통행 발생량과 비교적 유사한 결과가 나타났다.

이러한 결과로부터 전통적 방식에 의한 통행수요 추정에서 쇼핑과 여가와 같은 불규칙적인 통행 수요 추정 시 면접 조사에 의한 수요 추정의 한계를 보완하기 위해 CDR 자료를 활용할 수 있는 가능성을 보여주었다. 즉, CDR은 O/D자료와 통행목적별로 통행량(量)적 상관성이 있는 것으로 나타났으며, 이를 통해 통행목적별 통행발생예측 모형 구축시 CDR을 주요 독립변수로 활용할 수 있을 것으로 판단된다. 또한, 본 연구의 결과는 기존 교통수요 관련 기반 정보가 존재하지 않는 지역에 대한 교통계획의 수립

시 활용 가능성을 제시하고 있다. 즉, 전통적 방식을 통해 교통수요를 추정하기에 한계가 존재하는 교통계획 관련 기술 낙후 국가나 지역의 경우에도 스마트폰은 상당부분 보급된 것으로 보도되고 있다. 이러한 지역에서 교통수요를 추정할 의 경우, CDR 자료의 활용은 매우 비용 효율적인 것으로 판단된다. 한편, 최근 이슈가 되고 있는 스마트 도시의 경우, 기존의 낙후된 도시 인프라에 스마트한 정보 통신이 가미하거나, 새로운 도시 건설시 처음부터 스마트한 도시로 계획하는 경우가 존재한다. 이러한 스마트한 정보 통신의 중심에 개인 스마트 기기가 중요한 역할을 담당하고 있으며, 결과적으로 개인 스마트 기기를 통한 CDR 자료의 활용 가능성을 지속적으로 증가할 것으로 판단되며, 이러한 추세를 비추어 볼 때, 본 연구의 결과는 향후 더 큰 관심을 받을 것으로 기대된다. 마지막으로 본 연구에서는 몇 가지 해결해야 하는 한계점도 드러났다. 특히, CDR 자료 활용 가능성 검증을 위해 사용된 O/D자료의 경우 서로 시간적으로 일치하지 않는 한계점이 존재했으며, CDR 자료와 O/D 자료와의 상관성 원인에 대한 구체적인 해석은 보다 과학적 혹은 객관적인 증거를 기반으로 제시해야 하는 한계점이 드러났다. 이러한 한계점은 향후 연구를 통해 해결될 것으로 기대한다.

감사의 글

본 성과는 2017년도 과학기술정보통신부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017R1A2B4008984).

References

- Alexander, L., Jiang, S., Murga, M. and González, M. (2015). “Origin-destination trips by purpose and time of day inferred from mobile phone data.” *Transportation Research Part C: Emerging Technologies*, Vol. 58, pp. 240-250.
- Furletti, B., Gabrielli, L., Garofalo, G., Giannotti, F., Milli, L., Nanni, M., Pedreschi, D. and Vivio, R. (2014). “Use of mobile phone data to estimate mobility flows. measuring urban population and inter-city mobility using big data in an integrated approach.” *Proceedings of the 47th Meeting of the Italian Statistical Society*, Cagliari, Italy.
- Kim, K. T., Lee, I. M., Min J. H. and Kwak, H. C. (2015). “Trip generation analysis using mobile phone data.” *Journal of the Korean Society for Railway*, KSR, Vol. 18, No. 5, pp. 481-488.
- Kim, K. T., Oh, D. K., Lee, I. M. and Min, J. H. (2016). “Home-based OD matrix production and analysis using mobile phone data.” *Journal of the Korean Society for Railway*, KSR, Vol. 19, No. 5, pp. 656-662.
- Kim, S., You, B. and Kang, S. (2005). “Origin-destination estimation based on cellular phone’s based station.” *Journal of Korean Society of Transportation*, KST, Vol. 23, No. 1, pp. 93-102.

KTDB (2016). *National Traffic Survey*, Available at: <https://www.ktdb.go.kr/www/contents.do?key=15> (Accessed: August 10, 2019).

Ministry of Science and ICT (2018). *December 2017 wireless service subscriber statistics*, Available at: <https://www.msit.go.kr/web/msipContents/contentsView.do?cateId=mssw11241&artId=1384069> (Accessed: May 25, 2019).

Pappalardo, L., Smoreda, Z., Pedreschi, D. and Giannotti, F. (2015).

“Using big data to study the link between human mobility and socio-economic development.” *2015 IEEE International Conference on Big Data (Big Data)*, IEEE, Santa Clara, CA, USA.

Yoon, S., Lee, C., Kim, H., Yook, D. and Kim, S. (2015). *A study on usability of big data to enhance reliability of regional travel demand forecasting*. Korea Research Institute for Human Settlements (in Korean).