

빅데이터와 교통정책의 연계 방향

유경상 이 창 윤규근

A Study on Application of Big Data for Transportation Policy Development and Improvement



서울연구원
The Seoul Institute

**빅데이터와
교통정책의
연계 방향**

연구책임

유경상

교통시스템연구실 연구위원

연구진

이 창

글로벌미래연구센터 연구위원

윤규근

교통시스템연구실 연구원

이 보고서의 내용은 연구진의 견해로서
서울특별시의 정책과는 다를 수도 있습니다.

요약

빅데이터 활용성 높이고 민간자료 확보해 서울시, 다양한 교통정책서비스 발굴해야

교통부문, 빅데이터 활용해 새로운 정책·서비스 개발 요구 날로 커져

정보통신기술의 발달로 전 세계적으로 생산 및 소비되고 있는 정보량은 과거 어느 때보다 많다. 데이터 처리 능력에 대한 사용자들의 욕구와 실제 기기의 성능 향상이 서로 상승 작용을 일으키고 있어 이러한 증가 추세는 물론 정보의 유형도 매우 다양해질 것으로 보이며, 사물인터넷과 인공지능으로 대표되는 4차 산업혁명이 도래하면 이러한 현상은 더욱 가속화될 것으로 전망된다.

빅데이터는 일반적인 데이터베이스 도구로 수집, 저장, 관리, 분석이 어려울 정도의 규모와 특성을 가지며, 이러한 유형의 빅데이터를 처리하기 위한 하드웨어 및 소프트웨어 기술이 계속 발전하고 있다. 가까운 미래에는 이러한 기술을 이용하여 빅데이터 활용이 획기적으로 늘어날 것으로 전망된다.

교통 부문에서도 교통카드 기록을 비롯하여 통행자의 이동 과정과 교통운영 과정에서 다양한 데이터가 끊임없이 생성되어 축적되고 있으며, 향후 자율주행차량의 등장 및 사물인터넷의 도입확산으로 그 양은 폭발적으로 늘어날 것으로 판단된다. 이와 더불어 교통 부문은 이용자 집단의 규모가 크고 실시간 정보 수집 및 제공의 효용이 높다는 특징이 있기 때문에, 빅데이터를 활용한 다양한 정책 및 서비스 개발 요구가 점차 증대되고 있다.

해외 각국, 장거리 통행패턴 분석 등의 분야에서 빅데이터 적극 활용

이러한 요구에 대응하여 국내외에서도 다양한 빅데이터를 활용하여 교통정책이나 새로운 서비스 개발을 계속 시도하는 추세이다. 미국, 영국, 싱가포르 등의 해외 각국에서는 도시 간 장거

리 통행패턴 분석, 가구통행조사 보완, 대중교통 운영 분석 등에 고속도로 요금소 진출입 기록, 교통카드 이용기록, 이동통신망 기지국 연결기록, 소셜미디어 이용기록 등 여러 분야에서 생성되는 다양한 빅데이터를 활용하고 있다.

국내에서도 차량내비게이션 표본 자료, 통신사 제공 유동인구 자료, 택시 운행기록(DTG), 교통카드 기록, 앱택시 요청 및 연결 자료 등 다양한 빅데이터를 활용하여 심야시간대 버스 운행 노선 계획, 택시 운행 현황 분석 및 개선점 도출, 장애인 콜택시 운영개선, 대중교통 소외지역 파악, 통행 O/D 추정 등 교통의 여러 분야에서 적극적인 활용이 시도되고 있는 것으로 파악되었다.

교통 부문은 정보를 얻고자 하는 대상이 고정되어 있지 않고 계속 움직이고 있는 경우가 많아 대규모의 표본 확보가 어려운 대표적인 분야로, 모집단 규모에 가깝게 수집·축적되고 있는 빅데이터와 관련 기술의 활용으로 이러한 한계를 다소 완화할 가능성이 크다.

향후 교통 부문에서 빅데이터 활용 및 분석이 본격적으로 이루어진다면 대규모의 비용과 시간이 소요되는 가구통행실태조사를 보완하거나 대체할 잠재력이 있을 것으로 생각되며, 그동안 데이터가 없어 정교한 정책이나 서비스 개발이 어려웠던 부문에 큰 도움이 될 것으로 판단된다. 또한, 빅데이터를 활용한 새로운 정책과 서비스 제공을 통해 이용자 편의 증진 등 사회적 편익의 증가도 클 것으로 전망된다.

공공역할 큰 교통부문, 빅데이터 활용이 정교한 정책 마련에 ‘큰 도움’

그러나 교통 부문의 정책 또는 서비스 개발에 필요한 데이터이지만 개인정보를 포함하고 있는 빅데이터 중에서 개인정보 노출에 대한 우려로 매우 제한적으로 공개되고 있거나 아예 공개되지 않은 경우가 많다. 공개된 데이터라도 특정 개인을 식별할 수 없도록 집계 또는 가공되어 제공되고 있으며 대부분 기관에서 여러 가지 이유로 원자료(Raw Data)를 제공하지 않는 것이 일반적이다.

이러한 현실적인 제약으로 인하여 원자료의 형태를 직접 확인하고 특성을 파악하는 것은 매우 어려운 실정이다. 그러나 이 연구에서는 공개된 집계 또는 통계 데이터의 내용 및 속성, 데이터의 수집 장치 및 수집방식, 일상생활에서의 경험과 데이터 보유기관 담당자와의 면담 등을 토

대로 원자료의 형태, 주요 포함 정보 등을 최대한 파악하고자 노력하였다.

이 연구에서는 이러한 노력을 통해 교통 부문과 밀접하게 연관되어 있다고 판단되는 총 18종의 빅데이터 원자료의 주요 내용을 정리하여 제시하였으며, 각각에 대하여 주요 포함 정보, 수집 원, 수집 및 데이터 적재주기, 저장형태, 수집·보유기관, 자료 특성 등을 상세히 파악하여 제시하였다.

한편, 교통 분야는 민간보다는 공공의 역할이 더 크기 때문에 빅데이터 분석 및 활용이 행정 당국의 업무를 효율화하고 더욱 정교한 정책 마련 및 시행에 도움을 줄 수 있을 것으로 판단되어, 서울시 교통 관련 실무부서 업무와 조사된 빅데이터와의 연관성을 분석하여 제시하였다.

업무에 가장 빈번하게 사용될 수 있는 빅데이터는 대중교통 통행 행태 및 통행패턴 분석, 노선 조정, 심야버스 노선 개발·운영, 버스운영업체 관리 등에 활용이 가능한 교통카드 기록과 버스 운영시스템(BMS) 데이터로 나타났다. 이 외에도 장애인 콜택시 서비스 개선, 도심 수요관리, 교통사고 예방 등 관련 부서별로 다양한 업무에 조사된 빅데이터가 활용될 수 있을 것으로 기대된다.

빅데이터 활용 가능한 11개 교통정책·서비스 도출... 7개 실행가능성 커

이 연구에서 조사된 빅데이터 원자료 현황을 바탕으로 관련 전문가 인터뷰, 연구진 회의, 외부 자문회의 등 다양한 방법을 이용해 기존 빅데이터 활용 정책·서비스 외에 총 11가지의 새로운 교통 관련 정책과 서비스를 도출하였다. 또한, 도출한 11가지 정책·서비스 중에서 비교적 실행 가능성이 큰 7가지는 향후 더욱 심층적인 추가 연구를 위해 관련 빅데이터 활용 구상안을 세부적으로 제시하였다.

하지만 이 연구에서는 실제 데이터를 적용하여 도출된 정책·서비스의 효용성을 검증하지 못하였고, 개략적인 빅데이터 활용방안을 구상하여 제시한 수준으로 한계가 있다. 따라서 향후 실제 빅데이터를 확보하여 구체화한 후, 이를 관련 교통정책에 얼마나 활용할 수 있고 그 효과가 어느 정도 발생할지에 대한 심층적인 추가 연구가 필요하다.

검증된 빅데이터 현장 활용으로 교통현상 이해시민편의 증진 ‘일석이조’

먼저, 차량내비게이션 애플리케이션 기록을 활용하여 기존 승용차 O/D를 보완할 수 있을 것으로 추정하였다. 차량내비게이션 애플리케이션 기록은 승용차 통행의 출도착지 정보와 이동 경로 정보를 실시간으로 수집할 수 있고 표본의 크기가 크기 때문에, 기존 가구통행실태조사에서 누락될 수 있는 소수의 통행을 보완하는 데에 활용될 가능성이 높다고 판단하였다. 이와 함께 내비게이션의 경로 재탐색이 빈번하게 실행될 것으로 예상되는 도로 위험구간 파악에도 활용될 수 있을 것으로 기대하였다.

다음으로, 기존에 정교한 파악이 쉽지 않다고 알려져 있는 지하철 노선 간 환승 통행 비율을 추정하는 데에 통신사 기지국 접속 기록을 이용해 볼 수 있을 것으로 추정하였다. 일부 제한사항을 해결할 방안을 찾는다면 환승객 규모를 지금보다 정확한 수준으로 조사할 수 있을 뿐 아니라 운송수익 분담 등에도 활용될 수 있을 것으로 예상하였다.

또한, 실제 보행 네트워크 기반으로 산출되는 포털지도 길 찾기 검색기록을 기반으로 하여 대중교통 거점 보행 접근 거리를 추정해 볼 수 있다. 기록을 분석하면 지역별·대중교통 거점별로 접근 통행에 대한 평균 접근 거리 등을 산출할 수 있으며, 이 외에도 대중교통 소외지역 등 다양한 정보를 얻을 수 있을 것으로 기대하였다.

그리고 교통카드 기록과 통신사 기지국 접속 기록을 분석하여 교통약자의 심야 보행안전 개선을 위한 지점을 선별할 수 있을 것으로 판단하였다. 교통카드 기록을 이용하여 심야시간대에 하차 인원수가 많은 지하철역과 버스정류장을 선별한 다음, 통신사 기지국 접속 기록을 활용하여 분석된 통행 인구를 이용하여 가로별 유동인구를 산출할 수 있을 것이다. 이어서 성별 및 연령대별 통행 비율을 구하여 교통약자로 분류될 수 있는 집단과 여성의 비중이 높은 지역과 이동 경로를 탐색한 뒤, 해당 구간에 심야 보행안전을 위한 다양한 조치를 시행할 수 있을 것으로 예상하였다.

이 외에도 상용차 운행 데이터·통신사 기지국 접속 기록·교통사고 정보 연계 분석을 통한 위험물 수송차량 운행행태와 도로 위험요소 식별, 상용차 운행 데이터에 기반을 둔 전세버스 통행 O/D 및 통행패턴 분석 등을 실현 가능할 활용 방안으로 소개하였다. 빅데이터 활용 예시로 들었던 정책과 서비스들에 대한 검증 이후 현장에 적용된다면 관련 연구자들의 실제 교통현상에 대한 이해도와 일반 시민들의 편의가 동시에 높아질 것으로 기대된다.

미래 교통 빅데이터 시대 대비한 사전연구 활성화 등 선제대응 바람직

이처럼 교통 부문에서 활용성이 클 것으로 전망되는 빅데이터는 개인정보 보호, 기업의 영업비밀, 데이터의 불완전성 등 여러 가지 법적·제도적 문제와 정보에 대한 인식의 차이로 실제 정책 수립에 활용하기가 매우 어려운 실정이다.

서울시도 빅데이터의 중요성을 인식하고 관련 조직 정비, 민간기관과의 제휴 확대, 빅데이터캠퍼스 설립, 중앙정부 법·제도 개선 건의 등 많은 노력을 해오고 있으며, 실제 빅데이터를 활용한 다양한 정책수립 지원 사례를 제시하고 있다. 이러한 노력에도 불구하고 다양한 이유로 인해 기대만큼 활용되지 못하고 있는 빅데이터의 활용성 제고, 더 많은 민간 빅데이터에 대한 확보 노력이 계속되어야 한다.

그러려면 빅데이터 보유기관과의 업무 제휴 강화, 중앙정부 및 입법부 건의를 통한 법적·행정적 제도 개선, 서울시 자체 보유 데이터의 체계적 관리 및 개방이 필요하며, 나아가 빅데이터를 활용한 새로운 가치 창출에 민간의 창의성과 효율성이 접목될 수 있는 연구 환경 조성 등을 위해 노력할 필요가 있다.

또한, 자율주행차량의 등장과 사물인터넷의 보편화로 미래교통 부문에서 대량으로 생성·축적될 것으로 전망되는 빅데이터의 효과적 활용을 위해 자료수집체계, 가공 및 관리방안, 데이터 간 연계처리 및 통합방안에 관한 사전연구가 활발히 이루어져야 한다. 이러한 과정들을 거치면서 정비된 체계를 추진동력으로 삼아 빅데이터 기반의 새로운 정책서비스 발굴을 계속해 나가는 것이 바람직할 것이다.

목차

01 빅데이터 시대의 도래	2
1_빅데이터 시대는 현재진행형	2
2_교통 부문도 이러한 변화의 물결 속에 있어	9
3_교통 관련 빅데이터 현황과 활용 가능성 파악 필요	10
02 국내외 교통 부문에서의 빅데이터 활용	14
1_해외에서는 통행패턴 분석 시 활발하게 활용	14
2_국내 사례는 교통 운영 측면에서의 활용에 중점	25
3_교통 부문에서 빅데이터의 활용가치 클 것으로 기대	36
03 교통 부문에서 활용 가능한 빅데이터	40
1_교통 관련 빅데이터들의 현황 검토	40
2_빅데이터마다 활용성은 서로 달라	66
04 빅데이터와 교통정책 연계 구상	74
1_교통정책 담당부서의 실무와 연계 가능성	74
2_빅데이터와 공간정보 연계 시 시너지 효과 기대	77
3_빅데이터를 활용한 다양한 교통정책 구상 가능	78
05 결론	94
참고문헌	97
Abstract	99

표

[표 2-1] 이지패스 트랜잭션의 통행 변환 기준	15
[표 3-1] 교통 관련 빅데이터 원자료 주요 내용	41
[표 3-2] 교통카드 기록 세부내용	43
[표 3-3] 포털지도 경로 검색 기록 세부내용	44
[표 3-4] 앱택시 요청 및 연결 자료 세부내용	46
[표 3-5] 대리운전 요청 및 연결 자료 세부내용	47
[표 3-6] 카드매출 기록의 세부내용	48
[표 3-7] 통신사 기지국 접속 기록 세부내용	50
[표 3-8] 버스운영시스템(BMS) 데이터 세부내용	51
[표 3-9] 택시 운행 데이터 세부내용	52
[표 3-10] 상용차 운행 데이터 세부내용	54
[표 3-11] 내비게이션 애플리케이션 기록 세부내용	55
[표 3-12] 공영주차장 출입 기록 세부내용	57
[표 3-13] 유료도로 요금소 이용 기록 세부내용	58
[표 3-14] 검지기 자료 세부내용	59
[표 3-15] 교통정보 CCTV 자료 세부내용	61
[표 3-16] 교통사고 정보 세부내용	62
[표 3-17] 불법 주정차 단속 정보 세부내용	63
[표 3-18] 속도위반 단속 기록 세부내용	64
[표 3-19] 택배 물품 추적 기록 세부내용	65

[표 3-20] 제공 및 연관 데이터셋별 주요 특성	68
[표 4-1] 주요 공간정보 기초자료	77



그림

[그림 1-1] 연간 생성 및 복제 데이터 규모 추산	2
[그림 1-2] 빅데이터의 특성	5
[그림 1-3] 교통 부문 빅데이터의 범위	6
[그림 1-4] 연구의 흐름 및 주요 내용	11
[그림 2-1] 북동축 내 상위 10개 통행량 가종점	16
[그림 2-2] AirSage의 이동통신 신호 가공 및 활용 과정	17
[그림 2-3] AirSage 데이터와 가구통행조사로 추정된 가로별 통행 분포	18
[그림 2-4] 조사 기간 내 특정 시점의 로스앤젤레스 일대 소셜 미디어 사용자 위치	19
[그림 2-5] 시간대별 대상 지역 내 통행량 분포	20
[그림 2-6] 런던 교통부에서 보유 및 수집 중인 빅데이터	22
[그림 2-7] BSEP 시행을 통한 버스 서비스 공급 수요 격차 해소	24
[그림 2-8] T map for Biz.의 주요 활용 분야	25
[그림 2-9] 서울시 심야버스 노선도	27
[그림 2-10] 택시 운행빈도 분석 예시	29
[그림 2-11] 서울 시내 175,448개 셀 분포도	30
[그림 2-12] 장애인 콜택시 운영 분석 결과 활용 방안	32
[그림 2-13] 대중교통 소외 지역과 카카오택시 이용 빈도가 높은 지역	34
[그림 2-14] 출근시간대 택시타기 힘든 지역(서울숲푸르지오, 강남 및 남부순환로 일대)	35
[그림 2-15] 퇴근시간대 택시타기 힘든 지역(용산역, 종로, 강남, 여의도 일대)	35
[그림 3-1] 정보 생성 경로별 빅데이터 분류	42

[그림 3-2] 교통카드 기록 수집 과정	44
[그림 3-3] 포털지도 경로 검색 결과 예시(네이버 지도)	45
[그림 3-4] 앱택시 배정 과정 예시(카카오택시)	46
[그림 3-5] BMS 데이터 수집 과정 예시(서울시)	52
[그림 3-6] 택시 운행 데이터 수집 과정 예시(서울시)	53
[그림 3-7] 내비게이션 애플리케이션 작동 시 정보의 흐름 예시	56
[그림 3-8] 공영주차장 출입 기록 수집 및 가공 과정	57
[그림 3-9] 루프 검지기 자료 생성 및 가공 과정	60
[그림 3-10] 원자료와 제공 데이터셋 간의 관계	67
[그림 4-1] 빅데이터와 현재 서울시 업무 간 연관성	75
[그림 4-2] 빅데이터의 활용이 가능한 정책 예시	78
[그림 4-3] 내비게이션 기록 기반 승용차 통행 O/D의 보완 예시	80
[그림 4-4] 내비게이션 기록 기반 도로 위험구간 파악 예시	82
[그림 4-5] 지하철 노선 간 환승 통행 비율 추정 예시	84
[그림 4-6] 포털지도 길 찾기 안내 시 제공되는 도보 시간 관련 정보 예시	86
[그림 4-7] 다양한 기준에 따른 도보 접근 거리 산출 예시	86
[그림 4-8] 새로운 역세권 개념 정의 예시	87
[그림 4-9] 심야시간대 안전한 귀가를 위한 정책 예시	88
[그림 4-10] 심야 보행 안전 지원 과정 예시	89
[그림 4-11] 위험물 수송차량 통행 관리 예시	90
[그림 4-12] 전세버스 통행 O/D 파악 및 행태 분석 예시	92

01

빅데이터 시대의 도래

- 1_빅데이터 시대는 현재진행형
- 2_교통 부문도 이러한 변화의 물결 속에 있어
- 3_교통 관련 빅데이터 현황과 활용 가능성 파악 필요

01 | 빅데이터 시대의 도래

1_빅데이터 시대는 현재진행형

1) IT 기술의 발달로 정보 처리 능력과 생산량이 폭발적으로 증가

20세기 후반 정보통신기술의 발달로 인해 인류 문명에 정보화 시대가 도래하였으며, 개인 컴퓨터와 휴대전화 등 다양한 정보통신기기가 빠른 속도로 보급되어 보편화되었다. 오늘날 전 세계적으로 생산·소비되고 있는 정보량은 과거의 어느 시점보다도 많은 수준이지만 사물인터넷(IoT, Internet of Things)의 확산으로 정보 생성량의 증가속도는 더욱 빨라질 전망이다.

글로벌 정보 인프라 솔루션 기업인 EMC가 IT시장조사기관인 IDC에 의뢰하여 2014년 4월에 발표한 보고서에 의하면 2013년 한 해 동안 전 세계에서 생성 및 복제된 데이터의 크기는 4.4ZB(제타바이트, 10^{21} 바이트)에 달하고 2020년에는 44ZB까지 증가할 것이라고 예측되었다. 이를 [그림 1-1]과 같이 저장용량이 128GB(기가바이트)인 태블릿 PC에 저장하여 쌓는다고 가정하면 지구와 달 사이 거리의 6.6배에 이를 것으로 추정되었다.



자료: EMC(2014)

[그림 1-1] 연간 생성 및 복제 데이터 규모 추산

이렇게 방대한 규모의 데이터가 생성될 수 있게 된 것은 사용자들의 데이터 처리 속도 및 저장용량 증가에 대한 욕구와 실제 기기의 데이터 처리 능력 향상이 서로 상승 작용을 일으키며 점점 더 높은 수준으로 향상되고 있기 때문인 것으로 판단된다. 사용자들은 지금보다 더 빠른 기기와 양질의 서비스를 이용하고 싶어하며, 제조업체와 개발사들은 이러한 수요에 대응하여 더 나은 성능의 장치와 서비스를 개발해 보급하기 때문이다.

데이터 저장장치의 용량은 2000년대 초반까지 사용되었던 3.5인치 플로피 디스켓이 1.44MB 수준이었으나 현재 시중에서 판매되고 있는 휴대용 하드드라이브는 10TB까지 용량이 증가하였다. 20년이 채 되지 않았지만 약 700만 배의 증가율을 보인 것이다. 저장 용량이 늘어남에 따라 개별 데이터의 크기도 점점 커지는 추세에 있다. 데이터 처리용량 또한 18개월마다 마이크로칩의 처리 능력이 2배로 증가한다는 무어의 법칙이 성립할 정도로 기하급수적으로 증가해왔다.

데이터를 전송하는 기술도 발달하여, 세계 각국에서 주도권을 놓고 경쟁 중인 5G는 현재 상용화된 4G LTE보다 20배가량 빠른 20Gbps까지 전송속도가 향상될 것으로 기대된다.¹⁾ 이러한 통신망에 접속하게 될 이용자들은 지금보다 더 큰 용량의 콘텐츠를 안정적인 속도로 이용할 수 있게 될 것이다.

교환되는 정보의 유형뿐 아니라 이를 전달하는 데에 사용되는 매체도 다양해졌다. 이에 따라 전통적인 우편제도와 신문, 전기적 신호를 활용하는 유무선 전신과 전화, 대중 매체로 분류될 수 있는 라디오와 텔레비전 등이 예전만큼의 영향력을 유지하지 못하고 있으며, 온라인으로 각종 정보를 제공하는 포털 사이트와 전자우편 및 메신저들, 다양한 소셜 미디어가 정보 교류에서 차지하는 비중이 점차 확대되고 있다.

2) 수년 전부터 빅데이터에 관한 관심 증대

정보화 시대의 이전과 현재를 비교해보면 인류 사회의 일상생활도 크게 변화했다는 것을

1) 이코노미스트 커버스토리, 「미래를 건 표준전쟁」 표준 잡는 자 천하를 얻는다, 2017년 2월 27일자, <http://jmagazine.joins.com/economist/view/315452>

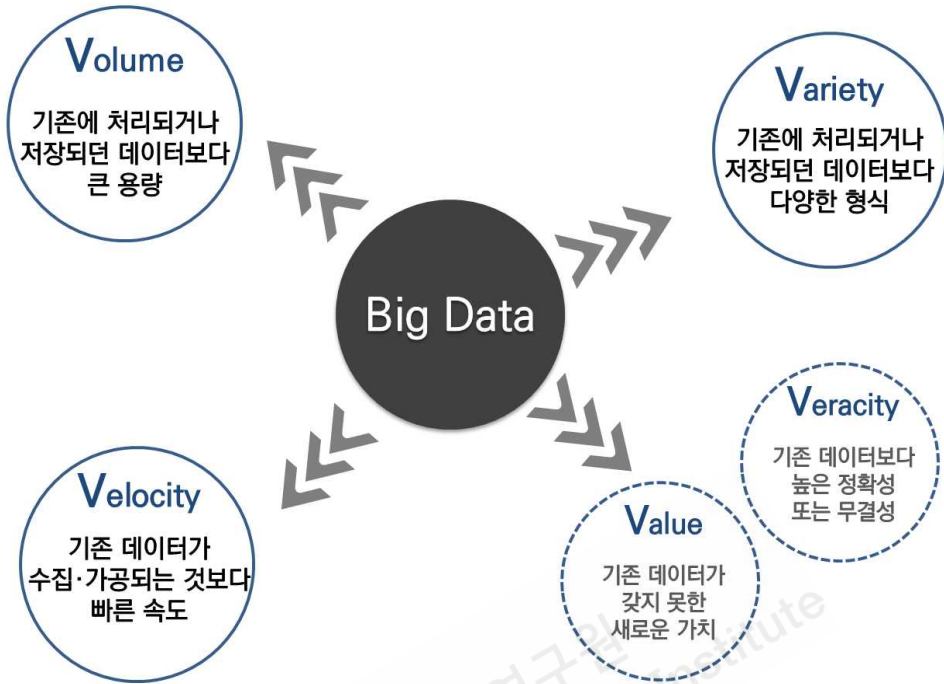
알 수 있다. 이전 기술보다 방대한 규모의 정보를 처리할 수 있는 환경이 조성되면서, 이용자들의 정보 생산 및 소비 행태도 변화하였다. 예를 들어, 이동통신기술은 휴대용 단말기를 이용하여 장소에 무관하게 통화할 수 있게 하였고, 직접 상대방과 얼굴을 마주하고 대화를 나눌 수 있는 영상통화까지도 가능하게 되었다. 불과 이십여 년 전 숫자 몇 개를 보낼 수 있었던 호출기와 이후 등장한 단문메시지서비스를 거쳐, 최근에는 문자, 사진, 영상, 파일 등을 사용자 간에 전송할 수 있게 되었다.

사람들은 유선을 이용한 음성통화와 짧은 문자로도 충분히 소통했었지만, 현재는 스마트폰을 통해 다양한 형태의 정보를 실시간으로 교환하면서 더욱 활발한 생산자 또는 소비자의 역할을 보여주고 있다. 그와 함께 엄청난 양의 데이터들이 생산되면서 이들에 관한 관심도 높아지게 되었다. 생산된 데이터를 들여다봄으로써 이용자들의 기호와 습관 등을 파악할 가능성이 열렸음에도 불구하고, 기존의 방법이나 도구를 이용해서는 분석할 수 없을 정도로 데이터의 규모가 방대했기 때문에 한계가 존재했었다.

하지만 데이터 처리용량 및 기술들은 계속해서 발달하여 가까운 미래에 실현될 것이라는 공감대가 일찍부터 형성되어 있었기 때문에, 기존의 기술로는 처리가 힘든 방대한 규모의 데이터 분석과 관련된 논의가 활발하게 진행되었다. 이러한 논의는 빅데이터 분석이 본격적으로 정착되어 활용되기에 앞서 빅데이터 자체의 정의에서부터 출발하였다.

3) 빅데이터의 정의는 다양하게 이루어지고 있어

빅데이터는 일반적인 데이터베이스 소프트웨어로는 수집, 저장, 관리, 분석이 어려울 정도의 규모를 가진 데이터셋으로 정의될 수 있다(McKinsey Global Institute, 2011). 2001년에 컨설팅업체 메타그룹(META Group Inc.)은 기존의 데이터 관리가 데이터의 양(Volume), 데이터 생성 및 처리의 속도(Velocity), 데이터 종류의 다양성(Variety) 등의 측면에서 새로운 전기를 맞이하였다는 내용의 자료를 작성하여 배포하였다(META Group, 2001). 이 자료에서 언급된 '3V'는 현재 빅데이터의 기본적인 특성으로 받아들여지고 있다. 메타그룹은 전자상거래(E-Commerce)가 도입되어 정착된 이후로 다음과 같은 문제가 초래되었다고 배포자료에서 기술하고 있다.



[그림 1-2] 빅데이터의 특성

(1) 데이터의 양 증가

전자상거래는 판매자와 구매자 사이에서 발생하는 매매 활동을 처리하기 위하여 하나의 거래를 여러 개의 트랜잭션(transaction)으로 구분하여 처리한다. 이로 인해 저장될 수 있는 데이터의 양은 최대 10배까지 증가할 수 있는데, 해당 기업이 이 정보들을 가치 있는 자산으로 인식하게 되면 폐기를 주저하게 될 것이며 결국 관리해야 할 데이터의 규모가 방대해질 것이다. 역설적으로 단위 데이터당 가치는 상대적으로 하락하며, 이는 신규 데이터 저장 공간에 대한 투자 가치 저하로 이어진다.

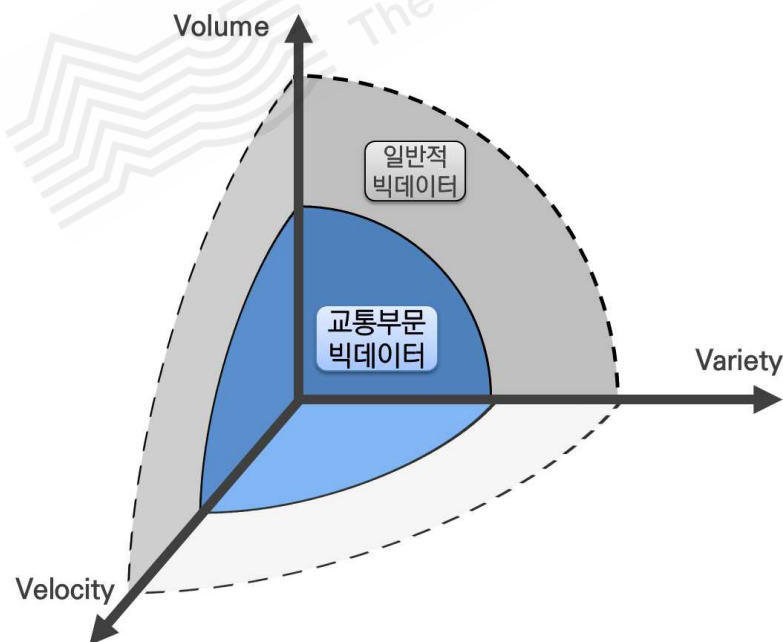
(2) 데이터 생성 및 처리 속도의 증가

전자상거래를 통한 거래에서는 거래의 처리 속도가 빠를수록 해당 서비스의 품질이 좋다고 인식됨에 따라 웹사이트 반응속도, 재고량 조회, 트랜잭션 처리, 주문 추적, 배송 등의 처리 속도를 향상시켜야 한다. 기업들은 많은 양의 데이터를 처리하여 원하는 결과를 빠르게 도출해 내기 위해서 다양한 물리적, 논리적 방안들을 마련해야 한다.

(3) 데이터 종류의 다양성 증가

적합하지 않은 데이터형식(format), 정렬되지 않은 데이터구조(structure), 의미(semantic)가 일관적이지 않은 데이터들은 체계적인 데이터 관리에 있어 큰 장애물로 작용한다. 현재 확인된 데이터 형태들을 새롭게 데이터베이스 유형에 추가한다고 해도 또다시 새로운 유형들이 등장할 것이기 때문에 데이터의 다양화에 대응하는 것은 매우 어려운 작업이 될 수 있다.

메타그룹이 언급한 상기 (1)~(3)의 특성 외에도 데이터의 정확성(Veracity)이나 가치(Value)가 빅데이터의 특성에 포함되어야 한다고 제안되기도 하는 등 여러 연구자가 다양한 방식으로 빅데이터를 정의하려고 시도하였다. 하지만 기술이 발전함에 따라 수년 전의 빅데이터가 오늘날에는 평범한 데이터로 취급될 수 있으므로 빅데이터의 정의 또는 기준은 정량적으로 고정되어 있지 않고 기술의 수준이나 시대에 따라 달라질 수 있다. 또한, 산업 부문에 따라 주로 활용되는 소프트웨어나 일반적인 데이터의 크기에 따라 빅데이터의 정의가 변동될 수 있다고 간주될 수 있다(McKinsey Global Institute, 2011).



[그림 1-3] 교통 부문 빅데이터의 범위

이에 따라, 이 연구에서는 아래와 같이 빅데이터의 본래 특성에서 다소 벗어나는 데이터들도 빅데이터로 취급하기로 한다. 교통 부문에서 수집되고 있는 데이터 중 앞에서 소개된 특성들을 모두 보유한, 엄밀한 의미의 빅데이터라고 할 수 있는 데이터는 많지 않기 때문이다. 예를 들어, 수도권 지역의 교통카드 기록은 하루 2~3GB로 전 세계적인 범위에서 생성되고 있는 각종 웹 검색 기록과 비교할 때 데이터 규모 측면에서 아주 작은 축에 속한다고 할 수 있다.

[이 연구에서의 빅데이터 정의]

- ① 데이터의 규모가 해당 분야의 기존 관련 자료들과 비교하여 상대적으로 큰 데이터
- ② 실시간으로 수집되지는 않지만 오랜 기간 축적되어 규모가 방대해진 데이터
- ③ 정형화된 형태로 수집된 대규모 데이터

4) 빅데이터는 분야를 막론하고 혁신적 변화를 가져오는 중

빅데이터는 그에 대한 기본적인 개념이 어느 정도 정립된 이후 다방면으로 활용되고 있다. 빅데이터 활용의 가장 큰 강점은 사실상의 모집단으로 간주할 수 있는 방대한 표본을 이용하여 보다 통계적으로 신뢰할 수 있는 분석을 수행할 수 있다는 것이다. 기존에 시간 및 비용 등의 문제로 인해 수집이나 가공이 어려웠던 데이터에 대해서 빅데이터 관련 기술들이 많이 접목되고 있다.

먼저, 지난 2016년 11월에 있었던 미국 45대 대통령선거에서는 수많은 여론조사 결과가 힐러리 클린턴 민주당 후보가 당선될 것이라고 예견했었다. 그러나 실제로는 도널드 트럼프 공화당 후보가 당선되어 대통령에 취임하였는데, 대선 관련 구글 검색 추이를 이용한 빅데이터 분석 결과가 이와 적중하자 많은 이들이 충격을 받았었다.

다음으로, 트위터와 페이스북 등의 소셜 미디어에 게시되는 수억 건의 글에서 가장 많이 언급된 단어를 분석하거나 특정 키워드에 대해 언급한 글의 수 등을 파악하는 텍스트 마이닝(Text mining)은 소셜 미디어 이용자들 간의 최근 화제에 대하여 분석함으로써 신규 마케팅이나 정책 개발의 근거를 마련해 주고 있다. 온라인 상거래 업체의 물품 추천과

인터넷 포털들의 검색어 추천은 자신을 포함한 같은 성별 내 유사 연령대 집단의 검색 및 페이지 방문기록을 분석하여 이루어지기도 한다.

아울러, 공공 부문에서도 서울시의 ‘우리마을가게 상권분석서비스’와 같이 공공의 편익을 위해 빅데이터를 활용한 정책들을 속속 시행하거나 구상하고 있다. 우리마을가게 상권분석서비스는 상권정보와 관련된 32종의 빅데이터 및 일반 데이터를 활용하여, 자영업자가 가장 많이 창업하고 있는 43개의 생활밀착업종을 선별하여 업종별로 다양한 정보를 상권 단위로 제공하고 있다.²⁾

위와 같은 사례들로 미루어 봤을 때 빅데이터의 활용을 통해 기존에는 알기 어려웠던 전체 구성원들의 행동 패턴이나 시간에 따른 각종 추세를 파악하는 것이 과거보다 쉬워졌음을 알 수 있다. 또한, 빅데이터 분석을 통해 제공되는 서비스나 산출물은 운영자와 이용자에게 다양한 편익을 가져다줌으로써 분야별로 커다란 변화가 도래하고 있는 것도 확인된다. 이러한 사례들은 나날이 증가하는 정보량에 적절히 대응하고 더욱 편리한 서비스를 개발하여 제공하기 위해서는 빅데이터에 대한 체계적인 접근이 필요하다는 것을 체감할 수 있게 해준다.

²⁾ 서울시 우리마을가게 상권분석서비스(http://golmok.seoul.go.kr/sgmc/cmmn/get_service_info.do?tab=2)

2_교통 부문도 이러한 변화의 물결 속에 있어

사람과 물품의 이동은 사회 구성원들의 기본 생활과 경제 활동에 있어 필수불가결하며, 국민의 이동권은 국가에 의해 보장되어야 한다는 점에서 교통정책은 언제나 높은 관심의 대상이다. 도로, 철도, 항만, 공항, 대중교통 등 사회간접자본의 상당수가 교통 관련 시설에 해당하며, 건설 시에는 운영 이익(Profit)보다 사회적 편익(Benefit)이 더 강조된다. 앞서 언급되었던 빅데이터 활용사례들을 보면 공공성이 강한 교통 부문에도 관련 빅데이터의 활용성이 충분할 것으로 판단된다.

먼저, 정보 수집의 대상이 될 수 있는 통행자 집단의 규모가 크기 때문에 빅데이터의 활용이 적절한 교통정책을 수립하는 데 있어 큰 역할을 할 수 있을 것으로 보인다. 개인별로 출발지에서 목적지까지 통행하는 사람이라면 모두 교통시스템의 이용자라고 할 수 있다. 승용차, 택시, 버스, 도시철도 등 어떤 수단을 이용하더라도 모두가 교통시스템의 범위 내에 포함된다. 사실상 분석자의 관심 대상이 되는 시간적·공간적 범위 내에서 움직이고 있는 모든 사람이 여기에 해당할 것이다.

다음으로, 교통과 관련된 정보의 실시간 제공을 원하는 수요 집단이 존재하기 때문에 빅데이터 플랫폼을 이용하여 즉각적인 분석을 달성할 수 있을 것이다. 예를 들면, 자신의 통행에 대해 미리 계획하는 때도 있지만 이동 중의 돌발 상황에 대처하기 위해 새로운 경로를 재탐색할 필요가 있을 때가 있다. 이와 같은 상황에서는 이동 중에 전방의 교통정보가 즉시 필요하다. 이러한 교통정보는 이미 수집되어 축적된 빅데이터를 적절히 활용하여 예측해 낼 수 있다. 이렇게 빅데이터를 교통에 활용함으로써 통행자에게 더욱 질 좋은 교통서비스를 제공할 수 있을 것으로 판단된다.

마지막으로, 미래 교통시스템에서 일부를 담당할 것으로 예견되는 자율주행차량의 운행은 교통 부문에서의 빅데이터 분석 및 활용을 더욱 촉진할 것으로 보인다. 자율주행차량은 무인주행을 위하여 차량 위치 정보, 주변 사물 정보 등을 실시간으로 수집하고 이에 기초한 연산과정을 거쳐 차량의 이동을 결정한다. 연산에 활용된 데이터가 모든 자율주행차량으로부터 수집된다면 도로 구간별 속도, 교통량, 돌발 상황 발생 여부 등의 추정에 충분히 활용될 수 있을 것이다.

3_교통 관련 빅데이터 현황과 활용 가능성 파악 필요

앞서 살펴본 바와 같이 교통 분야에서 빅데이터의 활용은 그동안 시도하지 못했던 새로운 접근법으로 교통현상을 해석하고, 새로운 서비스를 발굴하거나 더욱 세밀한 교통정책을 수립·시행하는 데 이바지할 수 있을 것으로 기대된다. 그러므로 국가 또는 지방자치단체는 정책적 차원에서 더욱 적극적으로 빅데이터를 활용할 필요가 있다. 이 연구에서는 교통 분야에서 빅데이터의 활용을 더욱 촉진하기 위해 관련 빅데이터의 기본적인 현황을 분석하고, 이들 데이터를 이용하여 새로운 가치를 창출할 수 있을 가능성을 검토해 보는 것을 주요한 연구목적으로 설정하였다.

1) 교통 관련 빅데이터 활용사례 및 빅데이터 현황 파악

교통 분야에서는 도로에 설치된 각종 검지기 및 CCTV, 택시와 버스 등의 상용차에 장착된 디지털운행기록계(DTG; Digital Tacho Graph), 대중교통 이용자들의 교통카드 등으로부터 통행에 관련된 데이터들이 대량으로 수집되고 있으며, 각 부문을 관장하고 있는 다양한 기관에서 저장·관리되고 있다. 각 기관은 수집된 원자료(Raw Data)를 내부적으로 가공하여 운영 방안 마련이나 정책수립에 활용하기도 하며, 처리 과정에서 산출될 수 있는 통계들을 보고서 형태로 정리하여 일정 주기 또는 수시로 공개하기도 한다.

하지만 최근 들어 빅데이터 분석에 대한 학계, 연구계 및 다양한 기관의 시도가 활발하게 이루어지고 있는 가운데, 분석에 필요한 원자료 자체의 획득을 위해서는 데이터 보유기관의 협조가 필요한 것이 대부분이다. 이러한 형태의 빅데이터 공유도 개인정보 보호 문제로 인하여 극히 제한적으로 이뤄지고 있다. 민감한 개인정보 부분을 제외한 나머지 데이터들을 적극적으로 공개하는 분위기가 사회적으로 형성될 수 있다면 빅데이터를 바라보는 시각의 전환 또는 다른 데이터와의 융합을 통해서 새로운 부가가치를 창출하게 될 가능성도 한층 커질 것으로 기대된다.

이 연구에서는 빅데이터의 공개 및 활용에 대한 사회적 분위기 형성과 국가 및 지방자치단체의 빅데이터 활용을 촉진하는 데 일조할 목적으로 기존 교통 분야에서 빅데이터 활용사례를 검토하고 교통 관련 부문에서 생성되고 있는 빅데이터의 목록과 저장 위치, 형태 등

현황을 우선 파악하고자 하였다. 활용하고자 하는 빅데이터가 어떤 기관에 어떤 형태로 존재하는지 사전에 알고 있어야 교통정책 수립 시 빅데이터의 활용 가능성을 가늠해볼 수 있기 때문이다.

2) 파악된 빅데이터 현황을 토대로 교통정책 수립 시 활용 가능성 검토

교통 및 관련 부문에서 현재 수집·축적되고 있는 빅데이터의 현황과 특성을 파악한 후에는 교통정책 수립 시 어떻게 활용할 수 있을지에 대한 고민이 필요하다. 이 연구에서는 파악된 빅데이터들과 교통계획, 교통운영, 대중교통 등 세부 교통정책 분야 간 연관성을 검토하고 각 분야에서 활용 가능성이 큰 데이터들을 선별하고자 하였다. 이러한 과정을 거친 뒤 최종적으로 교통 부문에서 새롭게 도출될 수 있을 정책 또는 서비스들을 구상해 보고자 하였다.

이렇게 함으로써 기존 교통정책을 더욱 정교하게 하고 새로운 정책을 개발하는 데 빅데이터 활용의 중요성을 인식시킬 수 있을 것으로 판단된다. 또한, 향후 수집이 필요할 것으로 생각되는 데이터 후보군을 미리 제시하여 관련 데이터를 취급하고 있는 기관이나 조직에 신규 자료를 수집·관리할 동기를 부여할 수도 있을 것이다.

궁극적으로는 이 연구를 통해서 교통 관련 종사자들의 빅데이터 활용에 관한 관심을 증대시키고 새로운 교통정책 수립 시에 빅데이터 활용을 더욱 촉진시킬 수 있는 계기가 마련될 수 있기를 기대한다.



[그림 1-4] 연구의 흐름 및 주요 내용

02

국내외 교통 부문에서의 빅데이터 활용

- 1_해외에서는 통행패턴 분석 시 활발하게 활용
- 2_국내 사례는 교통 운영 측면에서의 활용에 중점
- 3_교통 부문에서 빅데이터의 활용가치 클 것으로 기대

02 국내외 교통 부문에서의 빅데이터 활용

1_해외에서는 통행패턴 분석 시 활발하게 활용

1) 미국 동부의 도시 간 장거리 통행패턴 분석

미국 컨설팅회사 RSG(Resource Systems Group)는 미국의 수도인 워싱턴 DC와 미국 동부 대도시인 보스턴을 연결하는 북동축(Northeast Corridor) 고속도로를 이용하는 도시 간 장거리 차량 통행패턴에 관해 빅데이터를 이용하여 분석하고자 하였다(RSG, 2015). 일반적으로 항공 및 철도의 수요는 노선별 항공권이나 승차권 발권 수를 집계할 수 있어 비교적 자료 수집이 쉬운 데 반해, 승용차 통행 수요에 대해서는 알 수 있는 부분이 상대적으로 제한적이다. 따라서 이러한 분석은 현재의 차량 통행패턴을 파악하여 단기적 교통운영 정책수립에 참고할 수 있을 뿐만 아니라 장기적 관점에서 해당 지역 내에 철도와 같은 대안 수단 인프라 투자를 검토할 때도 잠재적인 통행 수요의 규모를 가늠해 볼 수 있는 기준이 될 수 있다.

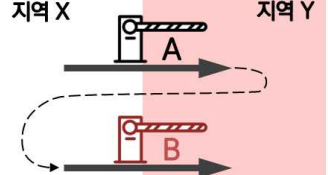
RSG의 도시 간 장거리 차량 통행패턴 분석은 미국 고속도로의 자동요금수납시스템인 이지패스(E-Z Pass)를 통해 수집된 통행료 결제기록에 기초하여 수행되었다. 활용된 데이터에서는 이지패스 결제정보가 통행 당사자의 연락처 정보와 연계되어 개인 및 통행 특성에 대한 사후 추가 조사가 가능하다는 큰 장점이 있었다. 그러나 코네티컷주(州) 등 일부 지역에서는 고속도로 무료 통행으로 인해 이러한 자료가 수집될 수 없었고 고속도로축의 말단 지역으로 갈수록 요금소의 간격이 벌어져 통행 가중점의 정확도가 떨어진다는 단점이 있었다.

이에 차량 번호판 인식(License Plate Capture)을 통해 차량의 진출입 기록을 보완하였으며, 통행자와 통행의 특성 수집을 위해 해당 구간 이지패스 이용자들을 대상으로 한 온라인 설문조사도 함께 수행하였다.

이지패스를 채택한 8개 고속도로 운영 주체들이 2013년 4월 27~29일(토~월) 동안 약 750만 건의 통행료 결제 트랜잭션을 수집하였으며, 데이터에는 개인 태그 번호, 징수 지점

및 시간 정보가 기본적으로 포함되었다. RSG는 트랜잭션을 통행으로 변환하기 위하여 3가지의 기준을 [표 2-1]과 같이 설정하였다.

[표 2-1] 이지패스 트랜잭션의 통행 변환 기준

	1) 동일 요금소 징수 기록이 반복되는 경우: 두 번째 통행이 복귀 통행일 가능성이 크므로 별도 통행으로 간주
	2) 시간적으로 연속된 요금소 징수 기록의 간격이 일정 시간 보다 길 때 장시간 정차한 것으로 추정되므로 별도 통행으로 간주
	3) 연속된 요금소 징수 기록이 한 번의 통행에 포함될 수 없는 경우: 예를 들어, 특정 지역으로 진입하는 서로 다른 요금소를 경유했다면, 별개의 진입 통행으로 간주

위와 같은 기준으로 이지패스 결제기록을 통행으로 변환한 결과, 750만 건의 트랜잭션이 500만 통행으로 줄어들었으며, 이 중 185,000통행은 진출입 요금소가 모두 분석 대상인 북동축에 위치하여 관련성이 매우 큰 것으로 추정되었다. 여기에 차량 번호판 인식으로 수집된 17,000통행의 차량 소유주와 뉴저지 턴파이크(Turnpike; 유료 고속도로) 이용자 전체를 대상으로 설문조사 참여를 요청하였다. 약 15,000명이 이에 응했으나 통행 방향, 통행 가종점, 통행 거리 등에 관한 기준에 맞지 않는 응답을 제외한 5,577건의 응답만을 분석에 활용하였다.

데이터의 출처가 다양함에 따라 집단별로 표본 비율에도 차이가 있었기 때문에 자료별로 별도의 가중치를 부여하는 과정을 거쳤으며, 이 과정에서 현금 결제 이용자에 대한 고려도 함께 진행되었다. 최종 결과에 따르면 뉴욕~필라델피아 간 통행량이 6,630만 통행으로 압도적으로 많은 것으로 확인되어 해당 구간이 대안 수단 도입의 잠재력이 가장 큰 것으로 추정되었다. 상위 5개 가종점 쌍 중 뉴욕이 포함된 곳이 4개로 나타나 뉴욕이 북동축의 핵심 도시임이 확인되었다. 또한, 설문조사로부터 통행자들의 상세한 가종점, 통행 목적 및 빈도, 동승자, 정차 횟수, 복귀 통행 정보, 인적 특성 등을 수집하여 정리하였다.



출처: Resource Systems Group(2015)

[그림 2-1] 북동측 내 상위 10개 통행량 가중점

통행요금을 부과받은 차량들의 통행만을 추출했기 때문에, 통행요금을 내지 않은 차량에 대해서는 대표성을 갖지 못한다는 점이 분석데이터의 한계로 언급되었다. 그 외에도 북동측 말단에서의 낮은 요금소 밀도로 인해 통행량이 과소 추정되었다는 점과 표본 수집 방식이 균등하지 못하였다는 점도 한계점으로 언급되었다.

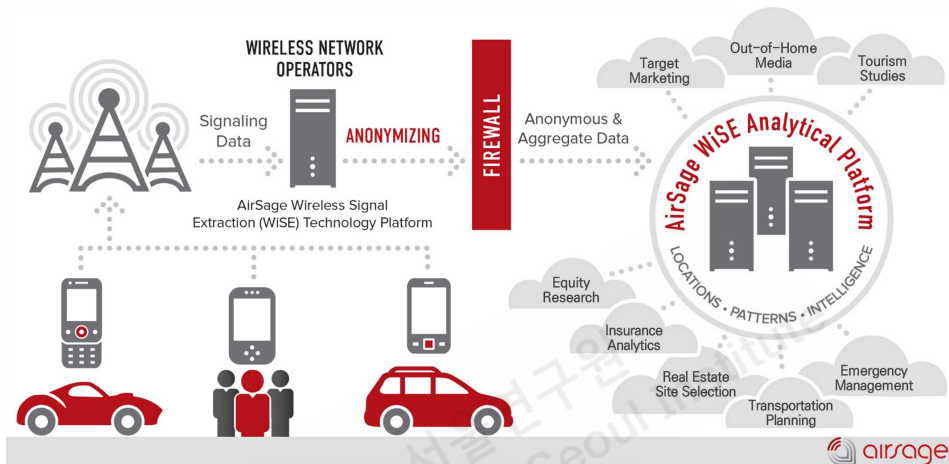
우리나라도 모든 고속도로에 설치운영되고 있는 하이패스 시스템으로부터 수집된 통행료 결제기록을 활용하여 지역 간 통행에 대하여 시간대별, 일자별, 요일별 통행패턴 등을 분석하고자 한 시도(윤서연 외, 2015)가 있었으며, 이러한 시도는 앞으로도 계속될 것으로 전망된다.

2) 이동통신데이터를 활용하여 통행을 분석하는 AirSage

AirSage는 미국 통신사들의 이동통신데이터를 이용하여 통행 정보를 생성하는 업체이며, 2000년에 설립된 이후로 정밀 위치 정보, 시간대별 통행패턴 분석 등을 수행하였다.³⁾ 이

³⁾ AirSage 홈페이지(<http://www.airsage.com/>)

동통신 빅데이터 분석과 연관된 기술이 축적되면서 관련 특허를 취득하였고 위치기반 서비스 공급자 중에서 가장 넓은 서비스 지역을 보유하고 있다. 하루 약 150억 건에 이르는 위치 정보를 휴대폰, 태블릿PC, PDA 등 이동통신 기기로부터 수집하여 분석하는 것이 가능하며, 이는 미국 전체 인구의 3분의 1가량을 분석할 수 있는 규모이다. 수집된 모든 정보는 익명 처리되어 프라이버시 문제를 사전에 방지하고 있다.



출처: AirSage 홈페이지(<http://www.airsage.com/Technology/How-it-works/>)

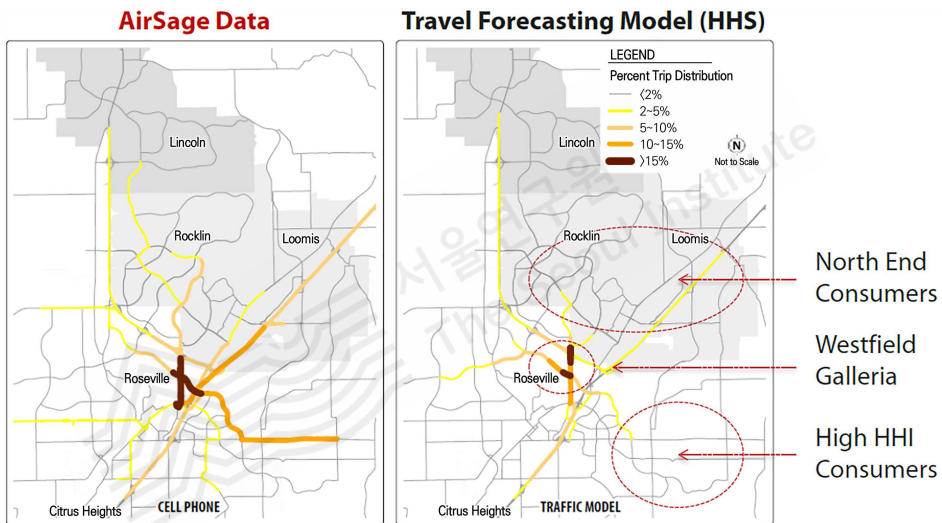
[그림 2-2] AirSage의 이동통신 신호 가공 및 활용 과정

AirSage가 분석하여 산출할 수 있는 결과물에는 특정 장소 2개 간을 이동하는 통행의 수와 유형을 정리한 통행 행렬(Trip Matrix), 특정 위치에 도착하거나 출발하는 통행의 수와 각 통행의 기·종점을 분석한 선택 구역 분석(Select Zone Analysis), 시간대에 따라서 특정 지역에 도착하거나 출발하는 통행 수를 분석한 출·도착 분석(Arrivals and Departures), 특정 지역에서 거주하거나 근무하는 사람들의 숫자를 분석하는 가정-직장 행렬(Home-Work Matrix) 등이 있다.

분석 결과의 정확도는 다양한 방법으로 검증되었는데, 그중 하나는 AirSage가 분석한 특정 교량의 통과 교통량을 현장의 검지기로 측정한 값과 비교하는 방법이었다. AirSage의 결과는 실제 측정량의 3% 오차 이내로 산출되었으며, 이 정도의 오차는 검지기 자체에서도 발생할 수 있는 수준이었기 때문에 AirSage의 분석기법이 일반적으로 신뢰할 만하다

고 알려지게 되었다.

가구통행조사(Household Survey; HHS)에 대비한 AirSage 이동통신데이터의 장점은 표본의 불균등 분포로 인한 음영지역 발생이 줄어든다는 것이다. [그림 2-3]은 캘리포니아 Rocklin에 위치한 Westfield Galleria 방문객들의 통행 분포를 분석한 것이다. AirSage 데이터를 이용하여 추정된 통행 분포는 우측의 HHS 기반 통행예측모형과 달리 북부지역과 동부 부촌으로의 통행들을 표현할 수 있었다. 해당 지역에서의 HHS에 대한 응답률이 낮았다면 충분히 일어날 수 있는 일이지만 이동통신데이터에서는 이러한 누락이 상당 부분 방지된다는 점에서 향후 활용 가능성이 크다고 판단된다.



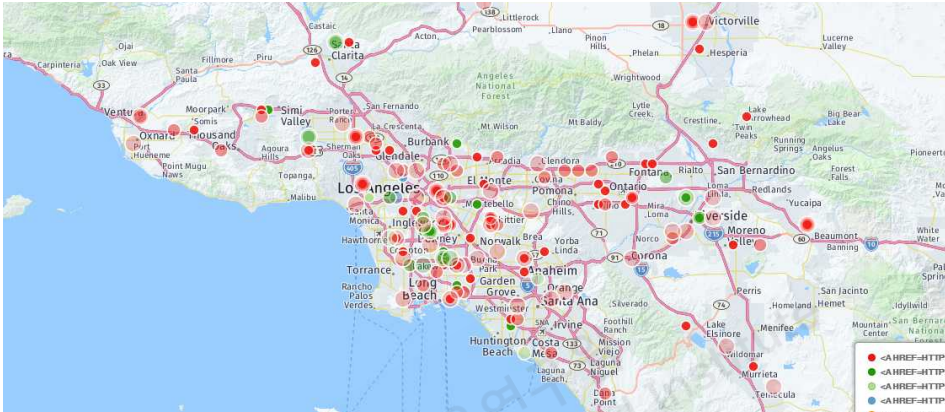
출처: Ryan Kinskey(2014)

[그림 2-3] AirSage 데이터와 가구통행조사로 추정된 가로별 통행 분포

3) 소셜 미디어 이용 기록을 활용한 지역 간 O/D(기종점 통행량) 분석

Gao 등은 미국 캘리포니아주의 로스앤젤레스와 인접 카운티를 포함한 지역에 대해서 1개월간 작성된 680만 건의 위치 정보를 포함한 SNS 트윗을 분석하였다(Gao et al., 2014). 이를 가구통행조사와 비교하여 빅데이터 분석이 고비용의 가구통행조사를 보조하거나 대신할 수 있는지 그 가능성을 판단하고자 하였다.

캘리포니아 대학교 샌타바버라 캠퍼스의 STKO(Space and Time for Knowledge Organization) 연구실과 샌디에이고 주립대학교 연구진은 미국 센서스(Census) 조사에 투입되는 비용이 1억 7,000만 달러에 달하는 데에 비해, 전체 가구의 1%만을 표본으로 추출하여 조사한다는 점이 상당히 비효율적이라고 생각하였다. 또한, 1%의 가구가 전체 집단을 대표할 수 있는지에 대한 의문도 제기하였다.



출처: LA Tweets by Gao(http://songgaogeo.carto.com/viz/55088584-bb57-11e3-8469-0e73339ffa50/public_map)

[그림 2-4] 조사 기간 내 특정 시점의 로스앤젤레스 일대 소셜 미디어 사용자 위치

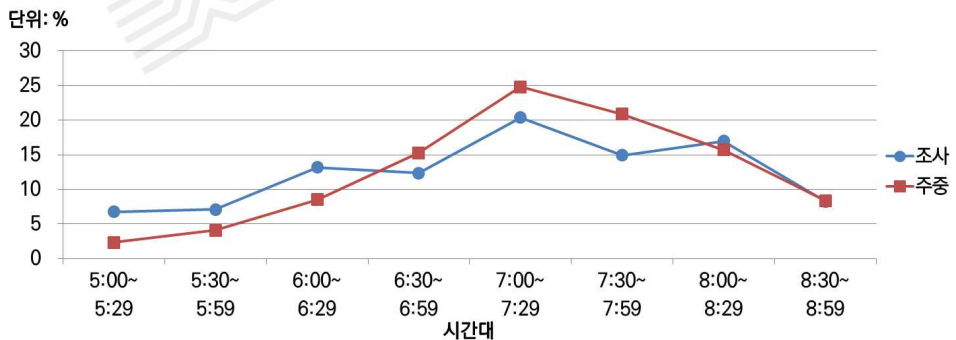
연구진의 판단에 따르면 로스앤젤레스 카운티로 통근하는 직장인이 471,000명 수준이라는 센서스 당국의 보고서에 대한 신뢰도 저하될 수밖에 없었다. 한편, 트위터나 인스타그램과 같은 소셜 미디어에서는 이용자가 특정 게시물을 등록할 때에 위치 정보도 함께 표시할 수 있는 기능을 제공하며, 이로부터 다른 이용자들이 어느 장소에서 어떤 글이나 사진을 올렸는지 확인할 수 있다. 연구진은 이 점에 착안하여 트윗 이용자들의 출발지와 도착지를 소셜 미디어로부터 취합하여 생성한 O/D 통행과 센서스를 통해 얻은 O/D 통행을 비교하였다.

연구진은 트위터 스트리밍 API를 이용하여 2013년 12월 7일부터 이듬해 1월 7일까지 로스앤젤레스 및 인접 4개 카운티(Orange, Riverside, San Bernardino, Ventura)에서 위치 정보가 등록된 트윗 680만 건을 수집하였다. 트윗에서 식별된 개별 이용자는 110,868명

이었고 이 중 10% 수준인 약 11,000명은 하루에 위치 정보가 등록된 트윗을 5개 이상 게시하였다. 개인별 트윗 게시의 간격은 평균 126분, 중앙값 79분으로 나타났는데, 약 80%의 이용자는 190분 미만의 간격을 두고 새로운 트윗을 게시하였다.

연구진은 개인 기반 통행 궤적 추적(individual-based trajectory detection)과 지역 기반 통행 집계(place-based trip aggregation) 등 두 단계에 걸쳐서 O/D 추정을 수행하였다. 1단계는 트윗으로부터 얻을 수 있는 시·공간 정보로부터 이용자의 통행 궤적을 추출하는 과정이었다. 개인의 트윗을 시간적 순서에 따라 나열한 뒤 각 트윗이 작성된 위치를 교통존(Traffic Zone)에 대응시킴으로써 센서스의 O/D와 동일한 공간적 기준으로 통일시켰다. 이후에 특정 시간 동안 같은 교통존 내에서 트윗이 작성되었다면 이를 통합하여 복수의 통행으로 집계되지 않도록 하였다. 2단계에서는 분석하고자 하는 시간대에 대하여 교통존 간 출·도착 통행량을 집계함으로써, 지역 내 O/D 통행을 추정하였다.

추정된 O/D 통행을 2008~2012년의 ACS(American Community Survey) 결과와 비교하기 위해, 오전 5시부터 9시까지 30분 간격으로 집계한 8개 시간대에 대해 통행량 분포 비율을 산출하여 상관계수를 분석하였다. ACS 결과와 O/D 추정 결과 간 상관계수는 주중 0.91, 주말 0.69로 높았고, 평균 통행시간도 각각 32분과 29분으로 큰 차이가 없어 연구진의 트윗 분석을 통해 도출된 O/D가 어느 정도 일관성을 가지는 것을 확인하였다.



참고: Gao, Song, et al.(2014)의 표 'Table 1. The comparison of morning peak-hour trips between the survey and the results detected from geotagged tweets'를 재구성

[그림 2-5] 시간대별 대상 지역 내 통행량 분포

또한, 5개 카운티 간의 일일 통행 분포에서도 로스앤젤레스-Orange 간 통행량이 가장 많고 로스앤젤레스-San Bernardino, 로스앤젤레스-Ventura 순으로 이어지는 등, 통행량 순위가 ACS 결과와 동일하게 집계되었다. 특기할 만한 부분은 ACS에서 조사되지 않은 Riverside와 나머지 4개 카운티 간의 통행이 트윗 기반 O/D 추정에서는 수집되었다는 것이다. 이는 앞에서 살펴본 AirSage 데이터와 마찬가지로 센서스의 표본 추출 과정에서 누락될 수 있는 소수 집단의 통행 행태까지 반영할 수 있다는 것을 의미한다. 연구진은 여기에서 더 나아가 트윗의 메시지 내에서 어떤 활동을 하고 있는지를 알아내기 위한 텍스트 마이닝에 관한 연구도 진행할 것이라고 언급하였다.

4) 런던 교통부의 대중교통 운영 개선

Transport for London(2015)에 의하면 런던에서는 하루 약 3,000만 통행이 도로와 대중교통을 통해 발생하고 있는데, 예산 규모 면에서 세계에서 가장 큰 조직 중 하나인 런던 교통부(TfL; Transport for London)가 런던 일대에서 운영 중인 교통 관련 시설 및 시스템들을 총괄하고 있다. 혼잡통행료 징수구역(Congestion Charge Zone)과 공해배출 제한구역(Low Emission Zone)의 설정운영 등 선진적인 교통 수요관리 정책의 도입과 시행을 선도하고 있으며, 이러한 정책을 통해 2000년 이후로 총 통행의 10.6%가 차량에서 대중교통, 자전거, 보행 등 친환경수단으로 전환된 것으로 추정되고 있다.

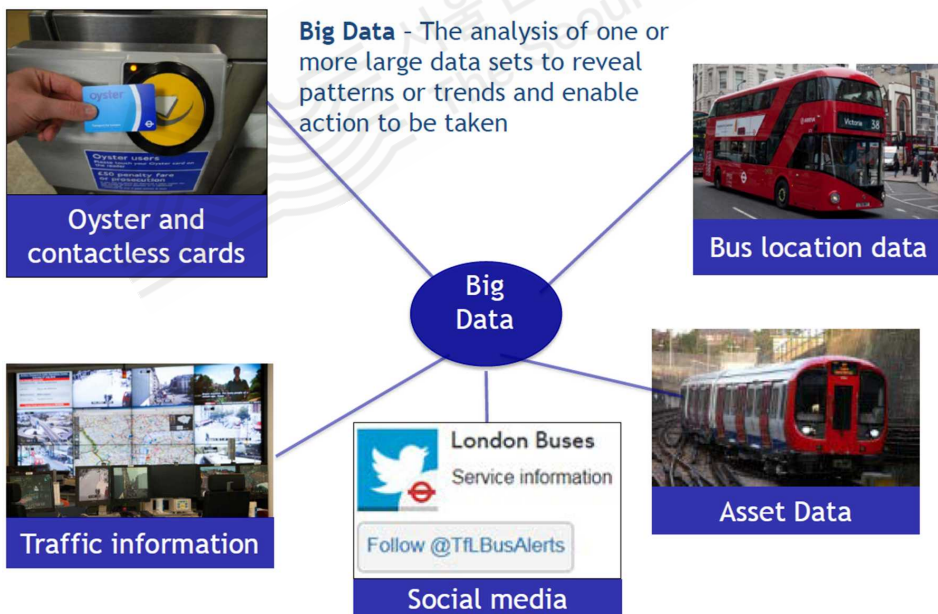
런던 교통부에서 제공하고 있는 교통 관련 오픈 데이터들은 무료로 활용 가능하며, 이들은 460여 개의 통행 정보 관련 애플리케이션에서 활용되고 있다. 또한, 런던 교통부에서 운영하는 소셜 미디어 계정은 290만 명의 팔로워가 구독 중이고, 홈페이지(<http://tfl.gov.uk>)에는 월간 1,200만 명이 방문하며, 2014년에는 2억 6,000만 건의 이메일을 이용자에게 발송하였다. 이처럼 런던 교통부에서는 다양한 경로를 통하여 통행 관련 정보들을 제공함으로써 시민들의 편의를 높이고 있다.

한편, 런던 교통부에서 수집하고 있는 자료 중에는 [그림 2-6]과 같이 오이스터 카드(Oyster Card: Greater London 일대에서 통용되는 교통카드) 태그 자료, 버스 위치 정보, 교통상황 정보, 보유 자산 정보, 소셜 미디어 기록 등이 빅데이터로 인식되고 있으며,

이들 데이터는 런던 교통부의 여러 정책에서 적극적으로 활용되고 있다.

먼저, 런던 교통부는 오이스터 카드로 버스를 이용할 때에는 승차 시에만 카드를 태그하기 때문에 정확한 하차지점을 알 수 없다는 것이 단점이라고 인식하였다. 따라서 현재 버스노선 승차 및 다음 수단 승차 시 카드 태그 위치와 시간으로부터 하차지점을 추론하는 알고리즘을 개발하여 승객들의 하차지점을 추정해 냈다.

또한, 교량의 보강 작업을 위한 차량 통행 제한과 대응 방안 마련을 위해 오이스터 카드 자료를 이용하여 해당 교량을 지나는 버스노선의 탑승객들이 어느 정도 되는지 파악하였다. 분석 결과, 일주일 동안 4만 명이 111,000통행을 하였으나 이 중 절반은 교량 근처에서 통행을 시작하는 것으로 확인되어 별도의 조치가 필요하지 않은 것으로 나타났다. 나머지 절반의 통행에 대해서는 요금이 중복되어 징수되지 않도록 환승 관련 장치를 설치하는 한편, 교량 통제로 영향을 받을 것으로 보이는 이용객들만을 대상으로 대안 경로 안내 이메일을 발송하였다.



출처: Transport for London(2015)

[그림 2-6] 런던 교통부에서 보유 및 수집 중인 빅데이터

이외에도 런던 교통부는 지하철역 이용객의 특성을 분석하거나 첨두 시간 수요를 분산하기 위한 혼잡시간대 정보 안내 등에도 이러한 빅데이터들을 활용하여 이용객들의 편의를 높이기 위해 노력하고 있다.

5) 싱가포르 교통부의 교통운영 개선

싱가포르 교통부에서는 다양한 종류의 센서들로부터 수집되는 교통 데이터들을 활용하여 기존 교통시스템의 효율을 높이기 위해 노력해 왔다(Lui, 2014).

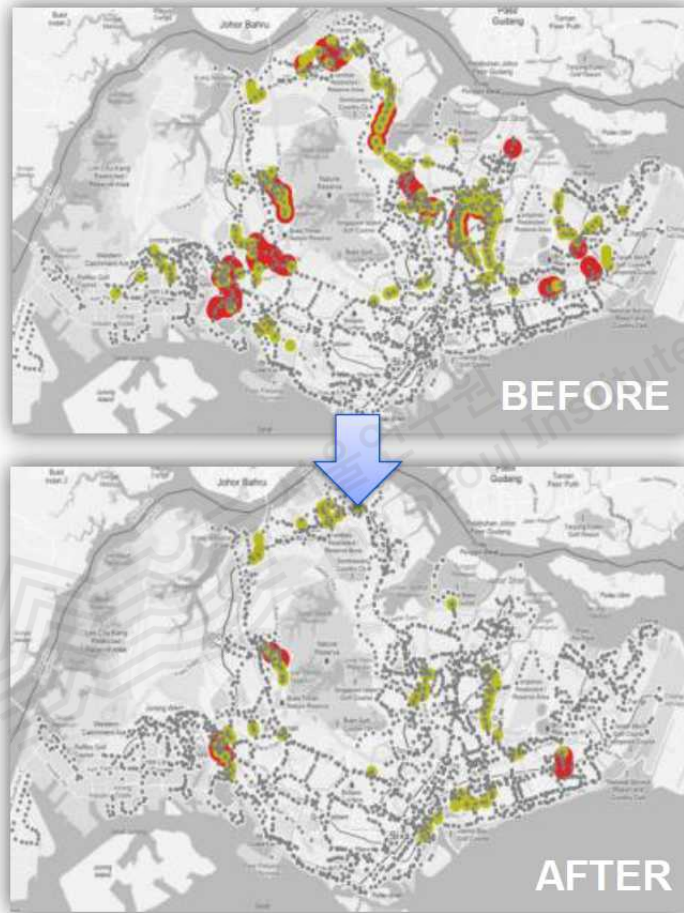
먼저 버스 서비스 증진 프로그램(Bus Service Enhancement Programme; BSEP)의 하나로, 버스 데이터를 이용해서 버스 서비스의 수요와 공급이 불일치되는 지역을 [그림 2-7]과 같이 선별하여 개선함으로써 서비스 수준을 개선하였다.⁴⁾ 해당 지역과 도심지를 직결하는 노선을 신설하거나 첨두 시간에만 배차되는 셔틀 노선을 도입하는 등 수요와 공급을 일치시키는 방안들을 마련하여 시행하였다. 또한, 배차 간격과 대기시간 감소를 위하여 운전기사에게 내비게이션을 제공하고 실시간으로 버스노선을 재조정하는 등 전반적인 버스 서비스 개선도 함께 추진하였다.

다음으로, 택시의 운행 효율을 높이기 위하여 택시 승하차 자료를 활용하였다. 택시들이 승객을 태우지 않고 시가지를 배회하는 시간을 줄이기 위해, 택시들의 승객 탑승 여부를 열지도(Heat map) 형태로 표현하여 승객들이 택시를 필요로 하는 지역과 택시들이 승객을 찾아서 운행하는 지역을 시간대별로 판별하였다. 이후 택시기사들과 택시 이용자들에게 이러한 정보를 제공하여 공급자와 수요자가 적절히 연결될 수 있도록 유도하였으며, 추가로 택시 운영업체들에 보유 차량의 운영 효율을 높여야 한다는 규정을 적용함으로써 정책 추진의 실효성을 확보하였다.

마지막으로, Wi-Fi AP를 지하철 플랫폼에 설치하여 승객 밀도를 균등하게 분산시키고자 하였다. Wi-Fi AP에 접속하고 있는 승객들만을 대상으로 알림이나 메시지 송신이 가능하

⁴⁾ Singapore Ministry of Transportation, 「How your bus rides are better under the BSEP」, <https://www.mot.gov.sg/Transport-Matters/Public-Transport/How-your-bus-rides-are-better-under-the-BSEP/>

므로 역사 내의 승객들에게 혼잡 정보를 제공하는 것이 가능하다는 점에 착안한 것이다. 또한, AP 접속 기록을 이동통신 접속 기록으로 보완하여 인접 카페, ATM, 택시 승차대 등의 위치를 안내함으로써 이용자들이 필요로 하는 서비스를 적재적소에서 받을 수 있도록 하고 있다.



출처: Lui(2014)

[그림 2-7] BSEP 시행을 통한 버스 서비스 공급 수요 격차 해소

2_국내 사례는 교통 운영 측면에서의 활용에 중점

1) SK텔레콤의 티맵(T map)

SK텔레콤의 티맵(T map)은 국내에서 최초로 GPS 방식을 이용하여 교통정보와 최적 경로 안내를 제공하는 스마트폰 애플리케이션으로 출시되었다. 택시, 배송차량, 버스 등 교통 정보 수집차량에서 이동체적 정보가 최대 5분 주기로 전송되면 교통정보센터에서 이를 수집하여 맵 매칭 작업을 통해 가공한 후, 서비스 이용자들에게 실시간 교통상황, 길 안내, 도착예정시간 등을 제공하게 된다.



출처: SK텔레콤 티맵 홈페이지(<https://www.tmap.co.kr/>)

[그림 2-8] T map for Biz의 주요 활용 분야

2016년 7월부터는 SK텔레콤 가입자가 아니더라도 티맵 서비스를 활용할 수 있도록 개방하였으며, 이로 인해 더 많은 이용자가 유입됨에 따라 수집되는 정보량도 증가하여 교통 정보의 신뢰성도 더욱 높아지게 되었다. 2016년 12월을 기준으로 한 월간 이용자 수는 약 820만 명으로 집계되어 여러 내비게이션 애플리케이션 중 월간 이용자가 가장 많은 것으로 나타났다.⁵⁾

SK텔레콤에서는 배경지도 데이터, 업체·매장·지번 및 지적 등의 지점 정보, 실시간 및 예측 교통정보 등 티맵을 구동하는 데에 필요한 다양한 데이터들을 활용하여 [그림 2-8]과 같은 서비스들을 제공하거나 계획 중에 있다. 또한, 이들 데이터를 외부에 유료로 개방함으로써 부가적인 수익도 창출하고 있다.

2) 서울시 심야버스 노선 개발 및 검증

대중교통 운영시간이 종료된 밤늦은 시간까지 활동하는 시민들은 이용 가능한 통행수단에 많은 제약이 따른다. 택시를 이용하는 것이 가장 일반적으로 고려되는 방법인데, 일부 택시기사의 승차 거부나 난폭운전, 안전에 대한 우려 등으로 시민들이 체감하는 서비스의 질이 주간 시간대보다 떨어지게 된다. 이로 인해 발생하는 시민들의 민원도 상당하며 2015년 서울시에서 민원 7,760건이 접수된 것으로 나타났다.⁶⁾

이에 서울시는 심야시간대에 귀가하는 시민들을 대상으로 대중교통 서비스를 공급하기 위하여, 런던이나 홍콩 등지에서 시행되고 있는 심야버스 도입·운행을 추진하였다(서울특별시, 2017a). 기관사와 역사 및 차량까지 운용 인력 등 많은 노동력의 투입이 필요한 지하철에 비하여, 버스의 운영은 상대적으로 비용이 저렴하다. 무엇보다도 버스는 쉽게 운영 노선을 변경할 수 있어 기존 노선과 다른 형태의 노선으로 서비스 공급이 가능하다는 장점이 있으므로 심야시간대 대중교통수단으로 우선 검토되었다.

5) 한국경제, 「“국민 내비 T맵에 도전한다”...길안내는 기본, 편의는 ‘덤’」, 2017년 1월 27일자, <http://news.hankyung.com/article/201701264809g?nv=o>

월간 이용자는 조사 기간 해당 서비스를 한 번 이상 쓴 이용자 기준. 2위 카카오내비 220만 명, 3위 KT 내비 200만 명.

6) 이투데이, 「서울시 “택시 승차거부 방지 대책 논의 중... 신고건수 50% 감소”」, 2016년 10월 10일자, <http://www.eteday.co.kr/news/section/newsview.php?idxno=1395253>

심야시간대 버스운행 경로 결정을 위해서는 통행의 출발지와 도착지를 우선 파악할 필요가 있었기 때문에, 서울시에서는 기구축된 구 단위 택시 O/D 자료를 활용하여 심야버스 초기 노선 안을 도출하였다. 하지만 해당 자료는 전체 시민이 아닌 택시 승객만의 밀집도만을 표현하여, 실제 이용 수요 반영 여부의 검증이 필요하였다.

따라서 KT의 유동인구 빅데이터로부터 심야시간대 통화 시에 연결된 기지국 위치를 출발지로, 통신요금 청구서가 발송되는 주소인 거주지를 도착지로 간주하여 검증에 활용하였다. 또한 택시 DTG 정보로부터 추출한 승차 및 하차 위치 정보도 함께 이용하여 정확도를 높이고자 하였다. 이에 기초하여 서울시 전역을 모서리 길이가 500m인 육각형 존(Zone)으로 덮은 후 존별 유동인구 밀집도를 분석하여 초기 노선 안을 검증하고 보완하는 한편, 잠정적으로 결정된 노선에 대하여 예상되는 통행 수요를 분석하였다.



출처: 서울대중교통 홈페이지(<http://bus.go.kr/>)

[그림 2-9] 서울시 심야버스 노선도

서울시는 분석 결과를 반영하여 초기 노선 안을 일부 수정한 뒤 기존에 시범적으로 운영을 시행한 2개 노선을 9개로 확대하였으며, 노선 통폐합 및 신규 운영 등을 거쳐 2017년 4월 현재 총 9개 노선을 심야시간대에 운영하고 있다. 하루 평균 이용승객은 6~8천 명의 수준을 유지하고 있고 시민들의 만족도도 높아 성공적으로 정착된 것으로 판단된다.

3) 택시 운행 데이터 분석

택시는 2014년 기준 하루 평균 약 222만 통행을 처리하여 6.8%의 수단분담률을 담당하고 있다.⁷⁾ 버스정류장이나 지하철역과 같이 특정 지점에서만 서비스가 제공되지 않고 승객이 원하는 때와 장소에 이용할 수 있다는 점에서 일반적인 대중교통과는 차이점이 있다. 하지만 운행시간이나 배차 간격 등의 운영 특성이 고정되어 있기 때문에 택시와 승객 간의 연결이 잘 이뤄지지 않을 때도 있다.

서울시에서는 이러한 수요와 공급의 불일치로 인해 발생하는 택시 운행의 비효율과 승객들의 불만을 줄이기 위해서 택시정보시스템에서 수집하고 있는 택시 운행정보 빅데이터를 분석한 바 있다(서울특별시, 2017b). 서울시 택시정보시스템에서는 택시에 부착된 통합형 디지털운행기록계(IDTG; Integrated Digital Tacho Graph)에 10초 단위로 기록된 위치 정보, 고도, 방향 및 속도, 승객 탑승 여부를 실시간으로 수집하고 있다. 서울시에서는 수집된 데이터를 이용해 요일, 시간대, 날씨에 따라 빈 택시들의 분포가 어떻게 달라지는지를 분석하였다.

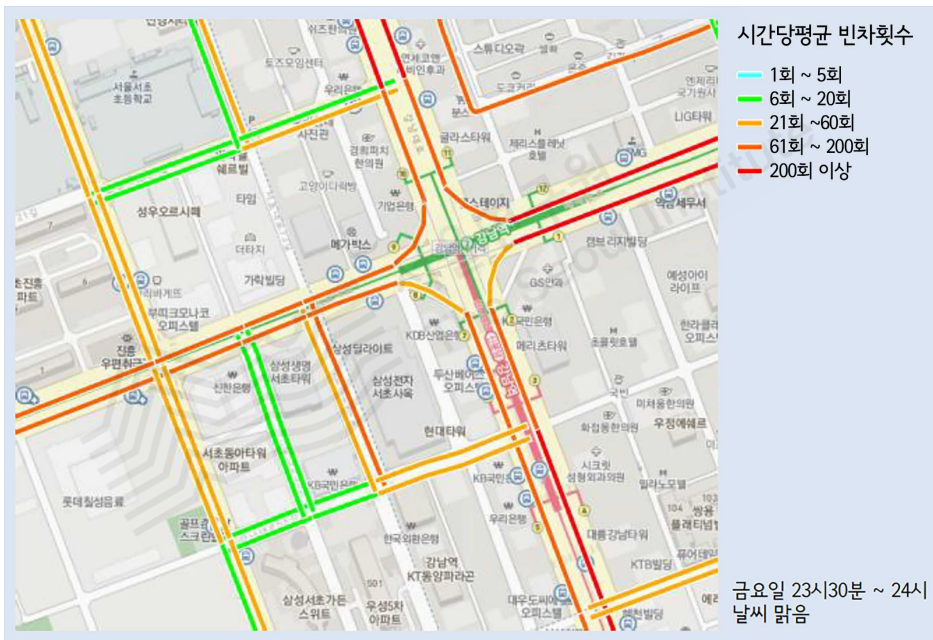
먼저, 택시들의 위치 정보를 도로 표준노드링크 네트워크의 링크상에 표현하였다. 차량 진행방향을 고려하여 상·하행을 구분하였으며, 이 과정을 통해서 GPS 오차로 인해 도로 외의 지점에서 관측된 택시들을 도로 네트워크 위에 재배치하였다. 분석에 활용된 링크는 50m 단위로 분할되어 있어 보다 세부적인 분석이 가능하였다.

이후에는 특정 일자의 해당 링크상에서 발생하는 승차 또는 하차 시점 데이터를 추출하고 각 데이터와 쌍을 이루는 하차 또는 승차 시점 데이터를 추출하여 분석 대상 시간 동안에

⁷⁾ 서울통계(<http://stat.seoul.go.kr/>), 1일 교통수단별 통행현황(분담률)

택시 승객들의 움직임을 나타내었다. 또한, 지하철 역사 인근에서 시간 및 요일별로 발생하는 택시 승하차 정보를 분석하여 어디에서 지하철역으로 왔는지, 지하철역에서 탑승하여 어디로 이동했는지 등을 분석하였다.

또한, 택시 운행정보를 통해 심야시간대에 택시 탑승이 활발하게 일어나는 지역을 중심으로 택시 승차가 집중되는 지점과 비교적 한산한 지점을 분석하였다. 이를 통해 늦은 밤에 택시 잡기 어렵기로 소문난 지역들이지만, 조금 걸어서 이동할 수 있는 거리에 상대적으로 빈 택시가 많은 구간이 어디인지 파악하였다. 이러한 정보를 잠재 승객들에게 제공한다면 택시 수요와 공급의 불일치가 어느 정도 해소될 것으로 기대하였다.



출처: 서울특별시(2015)

[그림 2-10] 택시 운행빈도 분석 예시

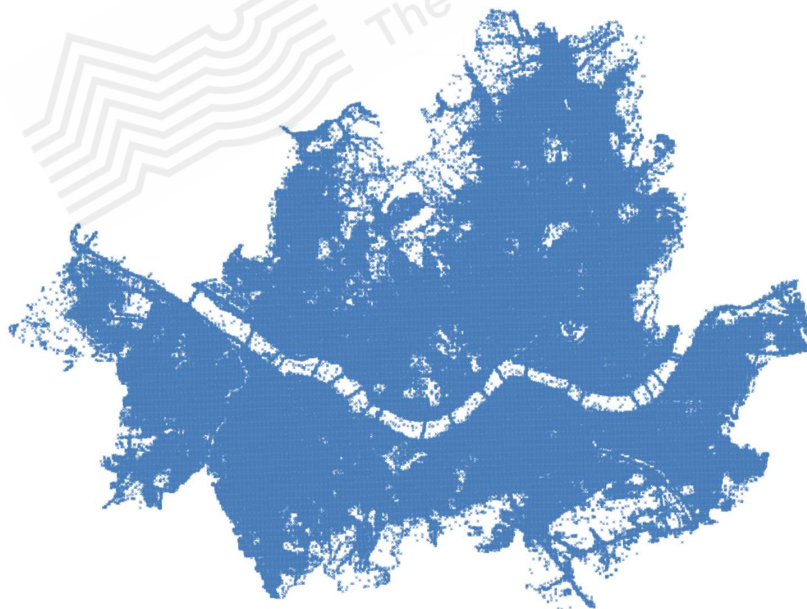
한편, 택시 승차 건수가 많은 지하철역과 열차 운행의 종착역들을 대상으로 지하철 출구로부터 반경 100m 내 지역에서의 택시 승차 건수를 서로 비교하였으며, 하루 평균 승차 건수와 요일·시간대별 승차 건수 등을 분석하여 택시 승차 특성을 분석하였다. 이를 통해 금요일과 토요일에는 유흥상권 인근 역사 주변, 일요일에는 고속터미널역, 서울역, 용산

역, 신용산역 등 지역 간을 운행하는 교통수단을 이용할 수 있는 지점에서 택시 승차가 많은 것을 확인하였다. 시간대별 승차 건수 분포 분석을 통해서도 지역의 특성에 따라 유흥상권의 외봉형 패턴, 유흥·오피스 혼재 상권의 쌍봉형 패턴, 교통환승지역의 일자형 패턴 등을 확인할 수 있었다.

서울시의 이러한 분석은 시민들의 교통수단 중 하나인 택시로부터 수집될 수 있는 빅데이터를 활용하여 정책적 개선방안을 마련할 수 있다는 가능성을 보여줬다는 점에서 의의를 가진다. 향후 서울시뿐 아니라 인접한 경기도나 인천광역시 택시에 대해서도 함께 분석한다면 수도권 지역에서의 통합적인 택시 정책 마련에도 도움이 될 것으로 판단된다.

4) 통신자료 기반 수도권 O/D 추출

해외에서 이동통신데이터를 이용하여 가중점 통행을 분석했던 사례를 앞서 기술하였는데, 국내에서도 이와 유사한 사례가 존재한다. 한국철도기술연구원에서는 SK텔레콤의 성별·연령별 유동인구 자료를 이용해 수도권과 50개 지방 도시의 O/D를 추정하였다.



출처: 김경태(2014)

[그림 2-11] 서울 시내 175,448개 셀 분포도

SK텔레콤에서는 전국을 50m×50m의 셀로 구분하여, 기지국과 이용자의 단말기 간 통화, 문자, 데이터 통신 등의 기지국 통신연결기록 정보와 이용자의 통신가입 등록정보로부터 성별과 연령 등을 수집하여 집단별로 셀별 집계된 유동인구 자료를 집계한다. 유동인구 자료에는 일자, 셀의 중심점 좌표, 성별(남/여) 및 연령대(10대~60대 이상)별 유동인구 수, 법인·외국인 여부, 법정동 기준의 유입지 코드가 포함되어 있다.

한국철도기술연구원의 연구에서는 셀 단위로 제공된 자료를 법정동 기준으로 취합하여 기존 O/D에서 활용된 지리적 구분과 동일하게 하였다. 유동인구 자료 중에서 중간 경유지로 판단되는 자료를 제외하면서 법정동 기준의 출·도착지 O/D를 생성하였다. 이후 도착지를 기준으로 다시 출발지로 복귀하는 O/D를 생성하여 추가하고, 이를 행정동 단위로 변환하여 최종적인 O/D를 추정하였다.

추정된 O/D와 현재 KTDB에서 제공되고 있는 O/D의 시군구 단위 도착 및 발생통행량 간의 상관계수를 분석하였는데, 도착통행량은 0.96, 발생통행량은 0.86 수준으로 나타나 높은 상관관계가 있는 것을 확인하였다. 또한, 유동인구 자료 기반으로 추정된 O/D와 KTDB O/D를 이용하여 회귀모형을 구축한 결과, 결정계수(R^2)가 도착통행량은 0.94, 발생통행량은 0.71로 계산되었다. 비교적 높은 수준의 상관관계는 통신연결기록 기반 유동인구 자료를 이용한 O/D의 추정이 큰 비용이 투입되는 가구통행조사를 일부 보완하는 역할을 할 수도 있음을 시사한다.

그러나 통신요금 청구서 주소지 외의 지점에서 출발하는 비가정 기반 통행의 반영이 현실적으로 어렵다는 점과 어떤 수단을 이용하여 통행했는지 파악하기 어렵다는 점 때문에 가구통행조사의 즉각적이고 완전한 대체는 힘들 것으로 판단된다.

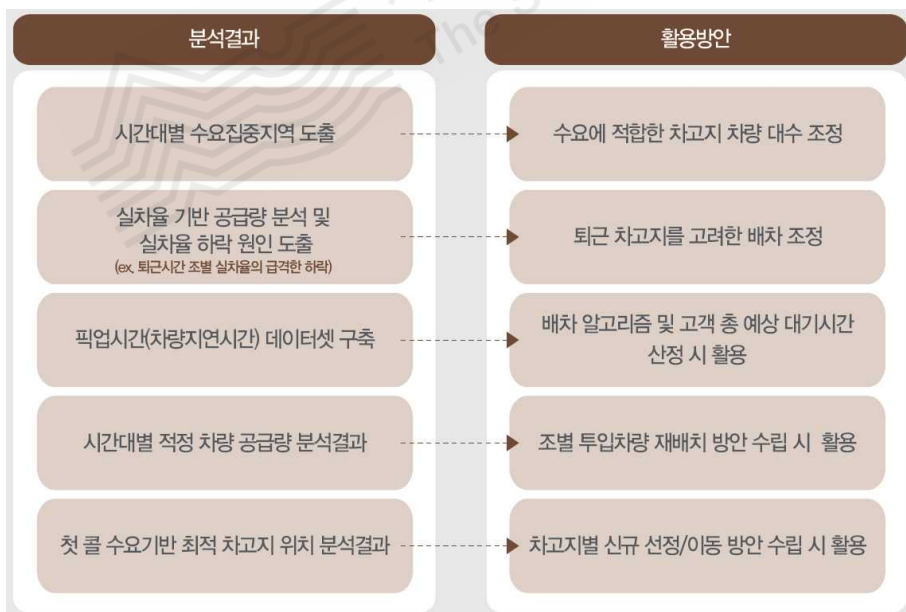
5) 장애인 콜택시 운영 분석

서울시에서는 장애인들의 이동권 보장을 위하여 장애인 콜택시를 운영하고 있으나 이를 필요로 하는 장애인 수에 비해 여전히 공급이 부족한 실정이다. 저녁이나 심야시간대에는 콜택시 운영 대수가 줄어들어 이용이 불편해지는 한편, 지역별로 수요와 공급이 편중되어 일부 지역에서 지나치게 대기시간이 증가하는 상황도 발생하고 있다. 따라서 서울시는 장

장애인 콜택시 운행 자료를 분석함으로써 정책적으로 조정 가능한 부분을 찾아내고 이를 실제 정책에 반영하여 수요와 공급 불일치를 해소하는 등 서비스를 개선하고자 하였다.

장애인 콜택시 운영 시에 수집되어 저장되는 데이터에는 장애인 콜택시의 운행정보, 수요자의 콜 내역, 고객·운전자·차량·상담원 등의 운영자원 정보가 있으며, 여기에 가로교통량 정보와 기상정보 등을 연계하여 분석을 수행하였다. 먼저 현재 보유 정보로부터 수요·공급 패턴과 차량 지연시간 등 운영현황을 알 수 있는 지표들을 산출하였다. 요일·시간대·자차구에 따른 승차희망 건수를 산출하여 수요가 집중되는 시공간적 지점을 파악하고 시간대별 투입차량과 콜처리 소요시간을 이용하여 공급계수, 실차율, 휴게율 등의 운영지표를 분석하였다.

또한, 차량 지연이 집중되는 시간대 및 지역을 도출하여 서비스 질 저하가 많이 발생하는 시간과 장소를 분석하였으며, 이를 토대로 차량 지연시간 통계 데이터모형과 최적배차 알고리즘을 개발하고 적정차량 공급량과 차고지 최적 위치를 선정하여 콜택시의 운영 효율을 높이는 방안들을 마련하여 시행하고자 하였다.



출처: 서울특별시(2017c)

[그림 2-12] 장애인 콜택시 운영 분석 결과 활용 방안

장애인 콜택시를 운영하는 주체인 서울시설공단에서는 차고지 위치나 시간대별 차량 투입 대수를 재조정하고 자동배차시스템을 도입하는 등 운영 측면에서 변화를 줌으로써, 이용객들의 평균 대기시간 감소를 기대하였다. 이처럼 운영과정에서 축적된 빅데이터를 활용하여 대대적인 증차 및 인력 충원 없이도 더 나은 서비스를 공급할 수 있음을 알 수 있다.

6) 카카오택시 운영 결과 분석

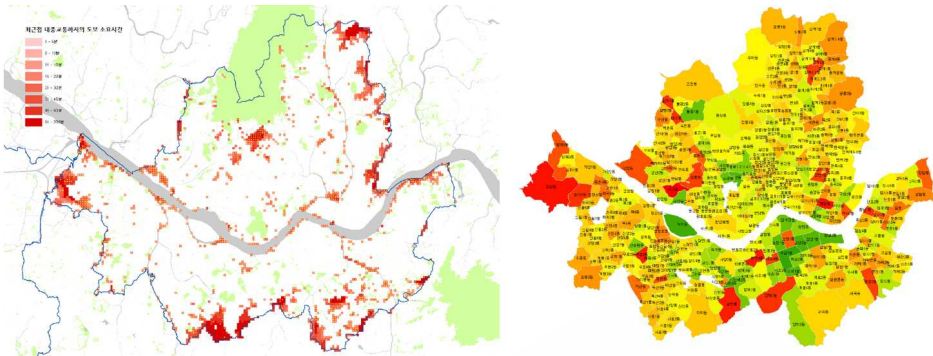
(주)카카오에서 2015년 3월 출시한 카카오택시는 택시 승차를 원하는 승객들이 자신이 원하는 출발지와 목적지를 입력하여 배차를 요청하면 해당 승객을 탑승시키고자 하는 택시기사가 이를 수락함으로써 택시 서비스에 대한 수요와 공급이 연결되도록 보조하는 스마트폰 애플리케이션이다. 앞의 택시 운행정보 분석 사례에서도 언급했듯이 택시의 수요와 공급의 불일치를 최소화하는 서비스로 이용자들 사이에서 인기가 높다.

유무선 전화를 이용하는 기존의 콜택시 서비스와는 다르게, 카카오택시는 승객이나 택시 기사로부터 별도의 비용을 받지 않고 있어 콜택시 시장의 상당 부분을 차지하면서 영향력이 확대되고 있다. 출시된 지 약 1년 5개월 만에 누적 호출 건수가 2억 건을 초과하는 등 이용객들의 호응도 상당하다. 관련 연구 결과,⁸⁾ 앱택시 이용 경험이 있는 설문조사 응답자 중에서 전반적인 서비스가 불만족스럽다는 응답은 1.6%에 불과할 정도로 시장을 파고들고 있다.

운영 주체인 (주)카카오에서는 승객들의 카카오택시 이용 정보를 취합하여 현재의 택시 이용행태에 대한 다양한 분석을 수행하였다(카카오, 2017). 먼저, 승객들이 택시를 잡기 위해 길거리에서 손을 드는 모습이 다소 줄어든 것으로 확인되었다. 2016년 8월 기준으로 광화문 주변에서 발생한 557만여 건의 택시 콜 중에서 약 33%에 해당하는 183만여 건이 실내에서 일어난 것으로 분석되었다. 이는 택시가 승객을 태우기 위해 돌아다니고 있는 도로변에 나가지 않고도 실내에서 간단한 조작으로 택시를 호출할 수 있는 앱택시의 장점을 이용자들이 잘 활용하고 있다는 것을 의미한다.

⁸⁾ 서울연구원 내부자료

다음으로, 버스와 지하철 등의 대중교통 이용이 어려운 지역에서 애플리케이션의 이용 빈도가 높았던 것으로 나타났다. [그림 2-13]과 같이, 가장 가까운 버스정류장 또는 지하철역에서 도보로 이동했을 때의 소요시간이 15분을 넘는 지역들과 카카오택시 이용 빈도가 높았던 지역 사이에 어느 정도의 상관관계가 존재한다는 것을 짐작할 수 있다.



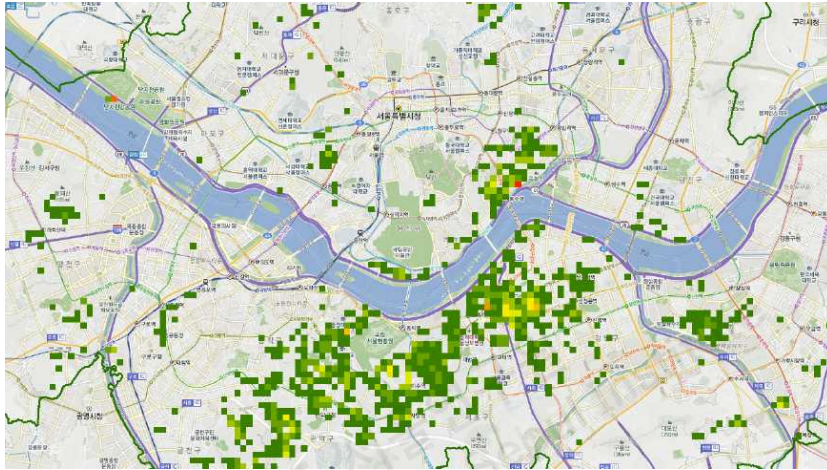
출처: 카카오(2017)

[그림 2-13] 대중교통 소외 지역과 카카오택시 이용 빈도가 높은 지역

또한, 시간대 및 지역에 따라 택시 콜이 잘 성사되지 않는 지역과 택시 탑승 횟수가 많은 지역을 분석하였으며, 그 결과 택시 탑승 수요와 공급 간의 시공간적 불일치가 확인되었다([그림 2-14] 및 [그림 2-15] 참조). 이러한 정보가 승객들과 기사들에게 적절히 제공된다면, 이와 같은 불균형을 해소하는 데 큰 도움이 될 것으로 보인다. 그 밖에도 택시 콜 요청의 38%가 5km 이하, 65%가 10km 이하라는 점에서 승객들이 콜택시를 이용하는 통행 거리가 대부분 단거리라는 것이 확인되었다. 단거리 승객에 대한 승차 거부가 빈번한 것이 기사의 황포에만 기인하는 것이 아니라, 수요가 많기 때문일 수도 있다는 점을 시사한다.

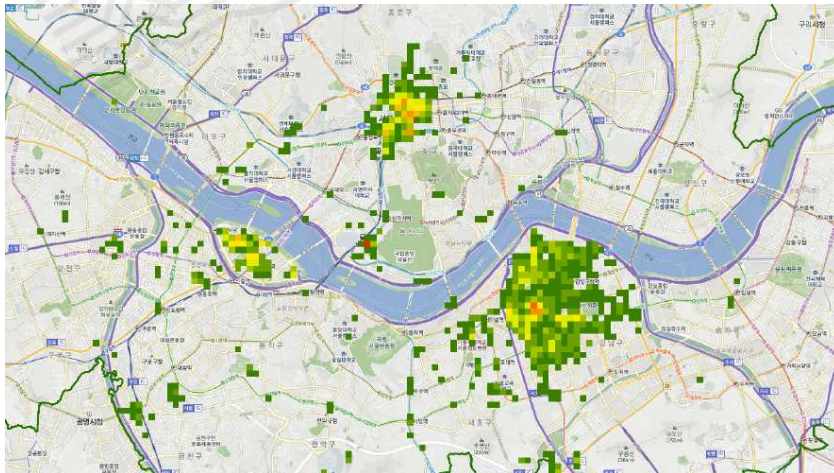
아울러 (주)카카오는 자율주행차량의 도입 및 확산, 카쉐어링 등의 공유 경제 정착 등의 변화에 대응하기 위해서, 카카오택시 운영을 통해 수집할 수 있게 된 빅데이터들로부터 새로운 서비스를 개발하려는 시도를 하고 있다. 자율주행차량의 운행비용이 저렴해지고 개인차량 소유가 감소하게 되면 대중교통보다는 택시와 같이 개인화된 통행수단에 대한 수요가 증가할 것으로 전망되기 때문이다.

한편, 현재 카카오택시 이용자 중에서 출발지와 도착지가 유사한 경로상에 있는 경우를 분석한 결과, 하루 평균 8만 건의 콜 중에서 2만 건은 출발지와 도착지가 5분 거리 내에 있는 것으로 나타났다. 이는 유사한 경로로 통행할 두 명의 승객이 합승하여 이동할 수 있다고 가정하면, 차량운행 횟수가 8만 회에서 6만 회로 25% 감소될 수 있으며 대기환경 개선과 교통혼잡 완화에 이바지할 수 있음을 의미한다.



출처: 카카오(2017)

[그림 2-14] 출근시간대 택시타기 힘든 지역(서울숲푸르지오, 강남 및 남부순환로 일대)



출처: 카카오(2017)

[그림 2-15] 퇴근시간대 택시타기 힘든 지역(용산역, 종로, 강남, 여의도 일대)

3_교통 부문에서 빅데이터의 활용가치 클 것으로 기대

자료의 방대한 규모는 빅데이터가 보유하고 있는 가장 큰 특성이며, 이로부터 얻을 수 있는 주요한 이점은 표본 데이터의 통계적 대표성 강화라고 할 수 있다. 이러한 이유로, 빅데이터는 기존에 시간과 비용의 문제로 인해 충분한 표본을 확보하지 못한 분야에 우선 접목되고 있음을 알 수 있었다.

교통 부문은 많은 수의 표본 자료 수집이 어려운 대표적인 분야이다. 정보를 얻고자 하는 대상이 제자리에 머물러 있지 않고 계속 움직이고 있는 경우가 많으므로 자료 수집을 위해서는 조사자 또는 장치가 표본과 함께 이동하거나 이동 경로상에 촘촘히 배치되어야 할 때도 있다. 그렇지 않으면 특정 시간 및 장소에서 수집되는 자료에만 의존하여 분석을 수행해야 하는 한계가 있다.

예를 들어, 도로에 매설된 루프 검지기에서 수집되는 차량 속도 및 교통량 정보는 검지기를 통과하는 차량에 한해 조사가 가능하며, 검지기를 통과하는 지점속도만을 수집할 수 있다. 택시 등 위치 정보 수집이 가능한 일부 차량을 동원하여 시간대별 구간별 통행속도 데이터도 수집하지만, 전체 집단의 크기에 비해 표본 규모가 작을 수밖에 없다.

또한, 고속도로 요금소 출발·도착 정보는 고속도로 내부 구간에 대해서만 대략적인 경로 추정이 가능할 뿐이며, 특정 지점의 통과 여부가 포함된 세부 경로 정보나 휴게소, 졸음쉼터 등의 시설 이용 여부는 확인이 어렵다. 더욱이, 출발 요금소 이전 및 도착 요금소 이후의 통행 구간에 대해서는 정보 수집이 쉽지 않은 실정이다.

한편, 가중점 여객 O/D 구축을 위해 5년마다 주기적으로 실시하는 가구통행실태조사에서 2016년 약 400만의 서울 내 가구 중 4만여 가구만을 표본으로 가구원들의 통행 정보를 조사한 바 있다. 모집단의 1%대에 불과한 표본으로 전체 집단을 대표해야 하는 통계적 대표성 및 많은 시간과 비용 투입의 필요성 등에 대한 논란이 벌어지기도 한다.

그러나 이러한 논란은 이 연구에서 기술한 국내외 빅데이터 활용사례를 살펴보면 상당 부분 완화될 수 있다는 가능성을 엿볼 수 있다. 각 사례에서는 교통카드 기록과 실시간 차량 위치 정보, 통신기록 등 빅데이터를 이용하여 전체 모집단에 가까운 사람들이 어떤

방식으로 통행하는지를 분석하고 있음을 알 수 있다. 이제 연구자들은 빅데이터를 활용한 통행자 정보 수집 방법론 개발과 이를 활용한 교통정책 수립이나 신규 서비스 도출은 물론, 교통환경의 변화로 인한 영향도 분석해 볼 기회를 가질 수 있게 되었다.

사례들로부터 알 수 있는 빅데이터 활용의 효용은 크게 자료 수집 비용 및 시간의 절감과 신규 정책 및 서비스 도입으로 인한 사회적 편익의 증대로 요약할 수 있다. 비교적 긴 주기로 대규모 조사를 통해 수집하던 정보들을 비교적 쉽게 획득할 수 있으므로 기존에 투입되던 인력과 재원을 다른 분야에 유용하게 활용할 수 있게 될 것으로 기대된다. 그뿐만 아니라, 빅데이터 분석 결과를 토대로 새로운 인프라, 통행수단, 운영시스템 도입 등을 통해 통행의 이동성, 접근성, 편의성, 쾌적성 등을 정교하게 개선할 수 있게 될 것으로 기대된다.

이러한 점들이 이 연구에서 목표로 하는 빅데이터와 교통정책 간의 연계를 통해 기대되는 사항들이다. 3장에서부터는 교통과 관련된 빅데이터가 현재 어떤 것들이 있는지 파악하고, 이러한 빅데이터가 효과적으로 활용될 수 있을 것으로 생각되는 교통정책 또는 서비스 구상안을 도출해 보고자 한다.

03

교통 부문에서 활용 가능한 빅데이터

- 1_교통 관련 빅데이터들의 현황 검토
- 2_빅데이터마다 활용성은 서로 달라

03 | 교통 부문에서 활용 가능한 빅데이터

1_교통 관련 빅데이터들의 현황 검토

1) 가공자료로부터 원자료(Raw Data)의 특성 추정

2장에서 기술한 국내외의 빅데이터 활용사례에서 사용된 빅데이터 중 대부분은 원자료에서 개인정보를 제거하고 가공한 집계 또는 통계 데이터들이다. 원자료는 다양한 장치들을 통해 수집저장되고 있으나 대부분의 원자료는 개인정보 노출이나 기업비밀 유지 등과 같은 다양한 이유로 인해 제공되지 않는 것이 일반적이다.

특정 개인에 대한 식별이 가능하게 되면 개인의 사적인 활동이 공개되어 프라이버시 침해와 관련된 문제가 발생할 수 있다. 다만 개인정보를 식별할 수 없는 일부 데이터는 공식적인 업무협약 및 계약 체결 등을 통해 원자료 보유기관과의 협업, 사전에 규정된 절차를 거친 후 무료공개 또는 구매 등을 통해 연구자들이 필요한 자료를 받을 수 있는 것으로 확인되었다.

이러한 현실적인 제약으로 인하여 원자료의 형태를 직접 확인하고 특성을 파악하는 것은 매우 어려운 실정이다. 그러나 이 연구에서는 공개된 집계 또는 통계 데이터의 내용 및 속성, 데이터의 수집 장치 및 수집방식, 일상생활에서의 경험과 데이터 보유기관 담당자와의 면담 등을 토대로 원자료의 형태, 주요 포함 정보 등을 최대한 파악하고자 노력하였다.

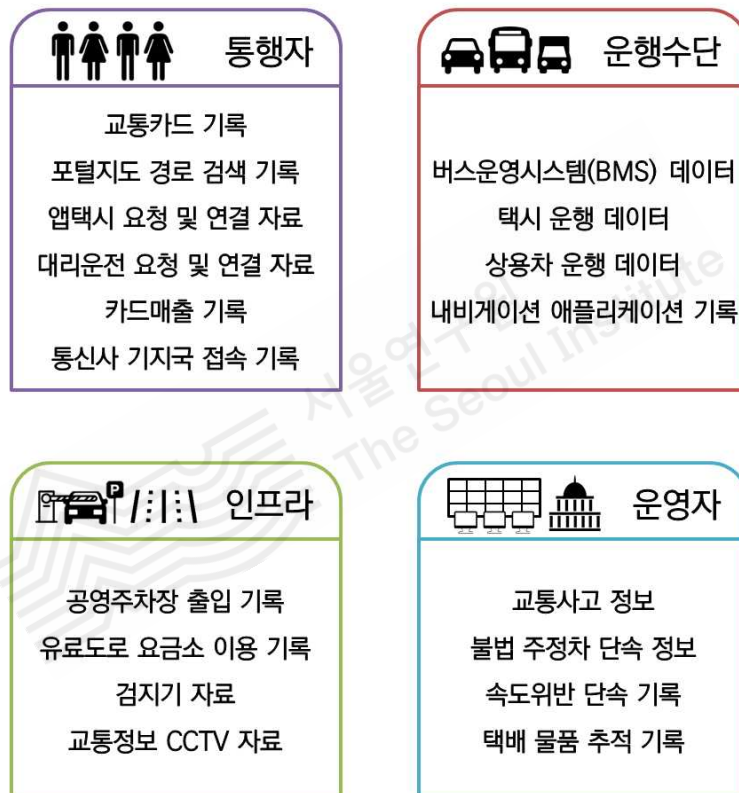
이러한 노력을 통해 다음 [표 3-1]과 같이 교통과 밀접하게 연관되어 있다고 판단되는 총 18종의 빅데이터 원자료의 주요 내용을 정리하였다. 이 외에도 다양한 빅데이터가 있으나 그 범위가 너무 넓어 이 연구의 대상에서는 제외하기로 한다.

[표 3-1]의 자료들은 이용자 단말기, 차량 내 장치, 인프라상에 구축된 장비 등 다양한 경로를 통해 수집되어 집계되고 있으며, 원자료별 주요 포함 정보, 수집원, 수집주기, 저장 형태, 수집기관, 자료 특성 등에 대해서는 별도로 상세하게 정리하였다.

[표 3-1] 교통 관련 빅데이터 원자료 주요 내용

구분	자료명	주요 포함 정보	보유기관
1	교통카드 기록	교통카드 태그 일시, 승하차 정류장, 태그 단말기 정보, 적용요금 구분(어린이, 일반 등), 추가 요금 정보	교통카드사
2	포털지도 경로 검색 기록	출·도착지 검색기록, 수단별 경로 정보 등	포털지도 운영업체
3	앱택시 요청 및 연결 자료	승객 위치, 요청 목적지, 요청 시각, 배차차량 정보, 콜 수락 시각, 승하차 시각, 결제 요금	애플리케이션 운영업체
4	대리운전 요청 및 연결 자료	차량 출·도착지, 배차 기사, 이용시간, 요금 및 결제 방식	대리운전업체
5	카드매출 기록	카드 사용 일시, 카드 사용자 정보, 가맹점 정보, 거래 금액, 결제 유형(할부 여부 등)	카드사
6	통신사 기지국 접속 기록	이용자 정보, 단말기 정보, 기지국 정보, 기지국 연결 일시, 연결 시간	이동통신사
7	버스운영시스템 (BMS) 데이터	버스 및 업체 관련 정보, 정류장·교차로별 차량 도착 정보	TOPIS
8	택시 운행 데이터	택시 등록정보, 운전자 정보, 시각별 차량 위치, 속도, 진행방향, 공차 여부, 감·가속 정보 등	한국 스마트카드사
9	상용차 운행 데이터	차량·운전자 정보, 운송사업자 등록번호, 시각별 차량 위치, 속도, 진행방향, 브레이크 신호, 가속도 등	교통안전공단
10	내비게이션 애플리케이션 기록	이용자 정보, 출·도착지 검색기록, 시각별 차량 위치, 이동 속도	내비게이션 운영업체
11	공영주차장 출입 기록	주차장명 및 위치, 차량번호, 입·출차 시각, 총 주차시간, 정산요금, 정기권 여부	주차장 운영 주체
12	유료도로 요금소 이용 기록	출발·도착 일시 및 요금소, 차량 정보(번호판), 차종 구분, 요금 정보	한국도로공사 민간운영업체
13	검지기 자료	검지기 위치, 검지기 도달·통과 시점, 검지기 점유시간, 통행속도	경찰청 국토교통부
14	교통정보 CCTV 자료	카메라 설치 위치, 녹화일시	서울시설공단, 경찰청 등
15	교통사고 정보	사고 발생일시 및 장소, 사고내용, 피해자가해자 정보, 상해 정도	도로교통공단
16	불법 주정차 단속 정보	소속 자치단체, 차량번호, 단속 차종, 단속 구역, 단속 위치, 위반 일시, 단속 구분	지방자치단체
17	속도위반 단속 기록	차량번호, 단속 일시, 단속지점, 단속 시 통과속도, 운전자 정보 등	경찰청
18	택배 물품 추적 기록	발신자, 수신자, 물품 정보, 접수·집화·도착시간, 수송차량 및 담당자 정보	택배업체

[표 3-1]의 원자료들은 생성되는 시점과 경로가 매우 다양하게 나타났는데, 먼저 1~6은 통행자가 움직임에 따라 생성되는 것으로 각 개인의 통행이 어떻게 이뤄졌는지 추정하는 데에 활용될 가능성이 크다. 7~10은 승객 또는 화물을 수송하는 교통수단들의 움직임과 관련된 정보들이며, 11~14는 교통수단들이 교통인프라를 이용하면서 생성되는 기록들이다. 마지막으로 15~18은 정부, 공공기관, 민간업체 등의 운영자가 시스템 관라운영 차원에서 수집하게 되는 정보들이며, [그림 3-1]과 같이 정리할 수 있다.



[그림 3-1] 정보 생성 경로별 빅데이터 분류

2) 원자료별 세부내용 및 특성

(1) 교통카드 기록

[표 3-2] 교통카드 기록 세부내용

항 목	내 용
자료명	교통카드 기록
주요 포함 정보	교통카드 태그 일시, 승하차 정류장, 태그 단말기 정보, 적용요금 구분(어린이, 일반 등), 추가 요금 정보 등
수집원	교통카드 단말기
수집 시점 및 데이터 적재주기	실시간(교통카드 태그 시)
저장형태	데이터베이스
크기	2~3GB/일(1,500만~2,000만 통행)
수집기관	교통카드사

교통카드는 대중교통을 편리하게 이용하기 위한 필수품으로 정착된 지 오래 되었으며, 하나의 카드로 전국의 교통수단을 이용하는 수준에 이르렀다. 특히, 대중교통을 이용하여 환승이 필요한 통행을 할 때 교통카드를 이용하지 않으면 환승할인을 받을 수 없다. 이런 이유로 대중교통 통행 중 버스는 약 98.7%, 지하철은 100.0%에 이르는 통행에서 교통카드가 이용되고 있으며,⁹⁾ 이로부터 교통카드 기록만을 분석하는 것으로도 사실상 대중교통을 이용하는 시민 전체의 통행 특성을 알아낼 수 있다.

교통카드 정보는 버스 또는 지하철 개찰구에 장착된 교통카드 단말기에 카드를 접촉할 때 기록되며, 일정한 주기마다 서버로 전송되어 적재된다. 교통카드 기록에는 통행자가 대중교통에 승하차한 시각과 정류장/지하철역사, 단말기 정보 등을 포함하고 있어 개별 통행자가 이용한 교통수단과 이동 경로를 확인할 수 있다.

교통카드 기록은 대중교통 통행과 관련된 많은 정보가 포함되어 있음에도 카드 ID가 암호화되어 특정 개인을 식별해내는 것이 사실상 불가능하다. 따라서 다른 빅데이터에 비해

⁹⁾ 서울시 내부 통계 자료

개인정보 노출의 우려가 비교적 적어 많은 연구에서 활용되고 있는 편이다.

교통카드 기록은 대중교통 통행 O/D 추정, 대중교통 노선 및 운영 조정, 노선별 승객 수송 특성 분석 등 이미 다양한 부분에서 널리 이용되고 있다. 교통카드가 본격적으로 보급·활용된 지 10여 년이 지나 연구자와 정책입안자들에게 비교적 친숙한 데이터임과 동시에, 거의 모집단 전체를 설명하는 데이터이므로 분석 결과에 대한 신뢰도가 매우 높다. 교통카드 기록은 향후 각종 공간데이터 등과의 통합·분석을 통해 대중교통 개선에 있어 더 많은 가치가 창출될 수 있을 것으로 판단된다.



[그림 3-2] 교통카드 기록 수집 과정

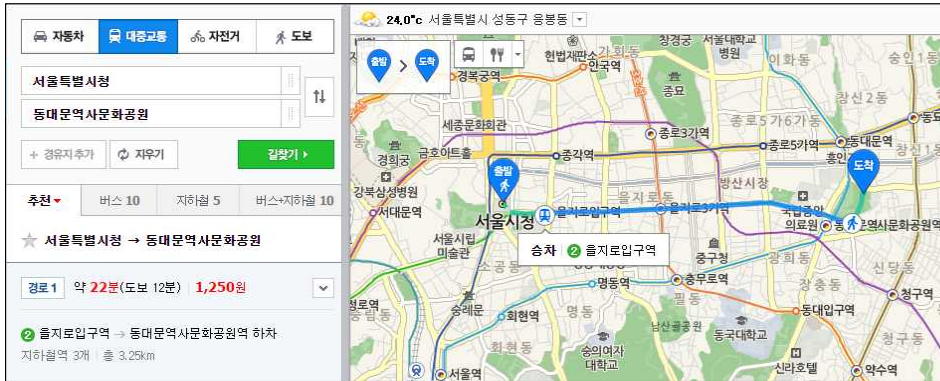
(2) 포털지도 경로 검색 기록

[표 3-3] 포털지도 경로 검색 기록 세부내용

항 목	내 용
자료명	포털지도 경로 검색 기록
주요 포함 정보	출·도착지 검색기록, 수단별 경로 정보, 통행비용 정보 등
수집원	포털지도 및 애플리케이션
수집 시점 및 데이터 적재주기	실시간(검색 시)
저장형태	파일
수집기관	포털지도 운영업체

포털지도 서비스는 네이버, 다음, 구글 등 인터넷 포털 업체의 대표 서비스 중 하나이며, 인터넷에 접속 가능한 컴퓨터만으로 관심 지역의 공간지리 정보를 무료로 얻을 수 있다는

점에서 기존의 종이지도에 대한 수요를 상당 부분 대체했을 뿐 아니라 포털 사이트의 접속자 수를 증가시키는 데에도 큰 역할을 한 것으로 알려져 있다.



[그림 3-3] 포털지도 경로 검색 결과 예시(네이버 지도)

최근의 포털지도에서는 거의 모든 도로의 주변 환경을 촬영한 사진을 제공하고 있으며, 승용차·대중교통·자전거·도보 등 다양한 교통수단을 이용할 때의 최적 경로도 자체 알고리즘을 통해 산출하여 제공하고 있다. 이러한 편리함에 힘입어 이동 방법을 잘 모르는 목적지로 통행하고자 할 때 포털지도를 통해 사전에 최적의 경로와 수단정보를 탐색하는 사람들이 많아졌다.

일부 포털지도 서비스에서는 이용자들이 요청한 출발지와 도착지 정보를 기반으로 교통수단별 경로 정보, 통행 거리 및 통행비용 관련 정보 등을 탐색하여 이용자에게 제공하는 동시에 이 정보를 로그 파일 형태로 서버에 저장·축적하고 있다. 포털지도 검색 기록은 이용자들이 실제로 해당 장소로 통행을 했는지가 확실하지 않다는 한계가 있지만, 가구통행실태조사에서 수집하는 통행 정보와는 다르게 평소에 자주 방문하지 않는 장소들을 주로 검색한다는 점에서 가구통행실태조사에서 빠질 수 있는 통행을 보완할 수 있는 잠재력이 있다. 또한, 포털지도에서 제공되는 정보에는 대중교통 이용 시 도보로 이동하는 거리와 시간 정보도 있어 대중교통 이용이 불편한 지역 등을 정교하게 찾아내는 데에도 도움을 줄 수 있을 것으로 보인다.

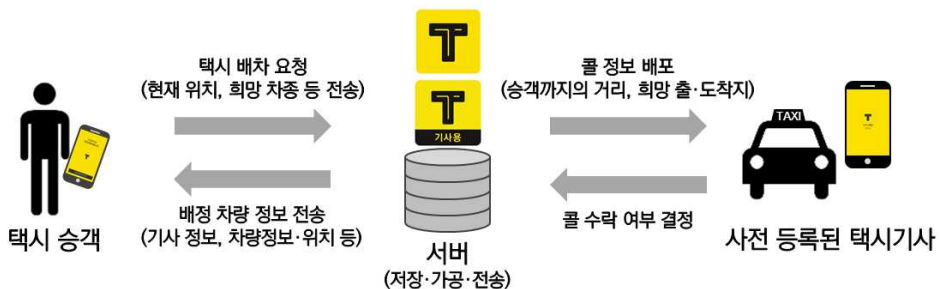
(3) 앱택시 요청 및 연결 자료

이 자료의 활용은 앞에서 국내 빅데이터 사례 중 하나로 기술하였으며, 앱택시의 원활한 서비스를 위하여 수집되거나 생성되는 정보들을 토대로 택시 운행형태 및 통행 특성에 대한 상세한 해석이 가능함을 알 수 있었다. 출시된 이후로 택시 관련 업계에 큰 영향을 미치고 있는 앱택시의 영향력은 당분간 계속될 것으로 보이며, 애플리케이션을 통해 수집되는 데이터의 양도 기하급수적으로 늘어날 것으로 전망된다.

[표 3-4] 앱택시 요청 및 연결 자료 세부내용

항 목	내 용
자료명	앱택시 요청 및 연결 자료
주요 포함 정보	승객 위치, 요청 목적지, 요청 시각, 배차차량 정보, 콜 수락 시각, 승하차 시각, 결제 요금 등
수집원	스마트폰 애플리케이션(승객용, 기사용)
수집 시점 및 데이터 적재주기	실시간(콜 요청 및 수락 시)
저장형태	N/A
수집기관	애플리케이션 운영업체

앱택시 이용을 위한 스마트폰 애플리케이션에서는 고객들이 원하는 택시 승차 지점 및 목적지 정보를 수집하며, 사전에 차량번호 및 운전자 정보를 등록한 택시기사가 배차받기를 요청하면 해당 정보를 이용자에게 보내 정확한 연결이 이뤄질 수 있도록 한다.



[그림 3-4] 앱택시 배정 과정 예시(카카오택시)

또한, 배정된 택시가 어디까지 왔는지 이용자가 직접 확인할 수 있도록 택시의 실시간 위치를 애플리케이션 내의 지도에도 표시해 준다.

이 애플리케이션에서 수집되는 정보는 서울시의 기존 택시정보시스템에서는 확인이 어려운 이용 승객의 가입정보까지 포함하고 있어 택시 이용자의 통행 행태를 보다 심층적으로 분석할 수 있게 하는 가능성을 가지고 있다. 즉, 택시 운행정보(DTG 데이터)를 통해서는 택시 차량의 승객탑승 여부 등 제한적인 정보만을 확인할 수 있지만, 애플리케이션을 통해 수집되는 정보는 이용자의 가입정보를 포함하고 있으므로 연령대별, 성별 주요 이용 시간대 및 출·도착지 등과 연계한 다양한 분석이 가능할 것으로 추정된다.

(4) 대리운전 요청 및 연결 자료

[표 3-5] 대리운전 요청 및 연결 자료 세부내용

항 목	내 용
자료명	대리운전 요청 및 연결 자료
주요 포함 정보	차량 출·도착지, 배차 기사, 이용시간, 요금 및 결제 방식 등
수집원	대리운전 콜센터
수집 시점 및 데이터 적재주기	실시간(이벤트 발생 시)
저장형태	N/A
수집기관	대리운전업체

대리운전 서비스가 본격적으로 이루어진 것은 2000년대 초반이며, 음주운전에 대한 사회적 경각심 증대와 처벌 강화 등이 서비스 보급 및 확산의 촉매 역할을 하였다. 음주 상태이지만 자가용을 집까지 가지고 가고자 하는 이용자가 대리운전업체에 대리기사 배정을 요청하면 콜센터가 이 요청을 소속 대리기사에게 전달하고 가장 먼저 응답하는 대리기사가 요청자의 위치로 이동하여 차량을 대신 운전해 목적지까지 통행하도록 돕는 것이 일반적인 서비스 흐름이다.

대리운전업체들이 직접 운영하거나 계약을 체결한 콜센터에서는 고객의 연락처와 현재 위치, 목적지 등의 정보를 수집하여 가지고 있으며, 대리운전기사의 배정 시간과 등록정보

등도 함께 보유하고 있을 것으로 보인다. 또한, 대리운전기사가 새로운 콜을 받기 위해서는 기존에 배정되었던 대리운전이 종료되었다는 정보를 업체에 등록해야 하므로, 서비스 종료시각도 일부 수집되어 보유하고 있을 것으로 추정된다.

대리운전 요청 및 연결 자료는 자가용을 이용하는 시민들의 심야시간대 활동이 어느 지역에서 많이 발생하고, 주요 목적지가 어디인지를 파악할 수 있는 정보원으로 활용될 수 있다. 이러한 정보는 심야 대중교통수단의 잠재적인 수요를 판단하여 노선을 계획하는데 활용할 수 있을 뿐만 아니라 주요 출발지와 목적지를 연결하는 경로가 음주운전자들의 잠재적인 이동 경로일 가능성이 크므로 음주단속 및 교통사고 예방을 위해 관련 인력 및 장비를 적재적소에 배치하는 데 도움을 줄 수 있을 것으로 판단된다.

(5) 카드매출 기록

[표 3-6] 카드매출 기록의 세부내용

항 목	내 용
자료명	카드매출 기록
주요 포함 정보	카드 사용 일시, 카드 사용자 정보, 가맹점 정보, 거래 금액, 결제 유형(할부 여부 등) 등
수집원	가맹점 카드단말기
수집 시점 및 데이터 적재주기	실시간(카드결제 시)
저장형태	데이터베이스
수집기관	카드사

최근에는 소규모 점포나 노점에서도 신용카드나 체크카드로 결제가 이루어질 만큼 카드의 이용이 활성화되어 있다. 한국은행의 분석에 따르면 2016년 현재 1인당 보유 신용카드는 1.98장이었으며, 결제 건수를 기준으로 2건 중 1건은 신용카드를 이용하는 것으로 나타나 현금이나 계좌이체 등의 다른 결제 방식보다 압도적이었다(한국은행, 2016). 더욱이, 한국은행에서는 현금으로 결제한 후 거슬러 받는 동전을 없애는 대신 전자화폐로 돌려주는 ‘동전 없는 사회’를 구상하고 있을 정도로 국민의 경제 활동에서 카드결제가 차지하는 비중이 커지고 있다.

개인이 카드를 특정 매장에서 사용하게 되면 카드번호, 카드 사용일시 및 장소, 금액 등이 결제단말기를 통해 VAN(Value Added Network; 부가 가치 통신망)사로 전송되고, 이 정보는 다시 카드회사로 전송되어 이용 한도 초과, 분실 신고 여부 등을 확인한 후 결제가 승인되고 영수증이 발급된다. 카드사에서는 결제 일시 및 가맹점 정보 등을 집계하여 개인의 소비패턴이나 특정 집단의 최근 소비 동향 등을 파악하여 새로운 마케팅 전략을 수립하거나 판촉활동을 벌이는 데 사용하는 등 수집된 정보를 적극적으로 활용하고 있다.

카드사뿐만 아니라 서울시의 유동인구 분석에서도 가로 주변의 경제 활성화 정도를 추정하는 데 통신사 기지국 접속 기록과 함께 카드매출 기록을 서로 연계하여 활용한 바 있다. 또한, 1장에서 기술한 서울시의 ‘우리마을상권분석서비스’에서도 카드사의 매출기록이 활용되고 있고, 최근에는 연세로 대중교통전용지구 조성의 효과를 측정하는 데에도 활용되는 등 카드매출 기록 정보는 가계의 경제 활동 중 지출의 상당 부분을 포함하고 있다는 점에서 그 활용성이 매우 크다고 할 수 있다.

(6) 통신사 기지국 접속 기록

이동통신사에서는 음성통화, 단문 문자, 3G·4G 데이터 등의 원활한 서비스를 제공하기 위해 전국적으로 통신 기지국과 중계기들을 설치하여 운영하고 있다. 통신사들은 기지국과 연결이 되지 않는 음영 지역을 최소화하고 이동 시 끊김이 발생하지 않도록 촘촘한 통신네트워크를 구축해 왔다. 우리나라 이동전화 가입 건수는 2016년 6월 기준 6천만 건을 넘어서 우리나라 인구수의 1.2배에 달할 정도이며, 이들의 통신기록을 이용할 수 있다면 사실상 우리나라 전 국민에 대한 위치 및 관련 분석이 가능할 것으로 생각된다.

이용자가 휴대전화 단말기를 이용하여 통신네트워크 접속이 필요한 활동을 하면 단말기와 기지국 간의 연결이 발생하며, 기지국에서는 단말기 접속이 시작된 시각과 종료된 시각을 기록하게 된다. 단말기의 사전 가입정보로부터 이용자 식별이 가능하며, 기지국 위치 정보를 토대로 통신사에서는 단말기(또는 개인)의 세부 이동 경로를 추정해 낼 수 있다.

[표 3-7] 통신사 기지국 접속 기록 세부내용

항 목	내 용
자료명	통신사 기지국 접속 기록
주요 포함 정보	이용자 정보, 단말기 정보, 기지국 정보, 기지국 연결 일시, 연결 시간
수집원	기지국
수집 시점 및 데이터 적재주기	실시간(통신망 접속 시)
저장형태	데이터베이스
크기	12.2~16.2GB/일
수집기관	통신사

또한, 통신사 기지국 접속 기록을 지역·시간대별로 집계하여 특정 지역의 유동인구나 체류 인구 등을 파악할 수 있으며, 집계 방식에 따라 다양한 분석이 가능할 것으로 판단된다. 통신사가 보유한 기지국 접속 기록은 개인의 인적정보를 가장 많이 포함하고 있다는 점에서 교통을 비롯한 여러 분야에서 다양하게 활용될 수 있는 잠재력이 있으나, 개인 프라이버시 침해에 대한 우려로 개인을 식별하지 못하도록 가공된 집계 자료만 제한적으로 활용할 수 있는 상태이다. 향후 통신사들의 적극적인 협조하에 개인정보가 노출되지 않는 선에서 필요에 따라 다양한 형태로 집계·가공하여 다른 빅데이터와 함께 활용한다면 많은 분야에서 더욱 큰 부가가치를 창출할 수 있을 것으로 판단된다.

예를 들어, 대중교통 이용자의 교통카드 기록, 승용차 이용자의 스마트폰 내비게이션 자료 등과 서로 연계하여 활용할 수 있다면 개인 통행의 특성을 보다 세부적으로 분석해 낼 수 있을 것으로 기대된다.

(7) 버스운영시스템(BMS) 데이터

서울시는 2004년 버스체계 개편과 함께 BMS(Bus Management System; 버스운영시스템)를 도입하여 운영하고 있다. 이 시스템을 통해 정류장별 노선별 실시간 버스 도착정보가 ARS, 문자, BIT(Bus Information Terminal; 버스정보안내기), 인터넷 등을 통해 이용자들에게 제공될 수 있었고 버스 운영의 체계적인 관리가 가능하게 되면서 이용자들의 서비스 만족도가 크게 향상되었다.

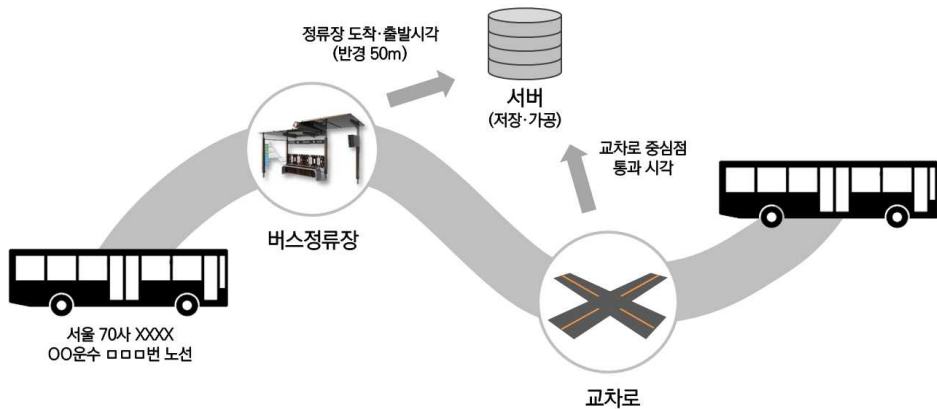
[표 3-8] 버스운영시스템(BMS) 데이터 세부내용

항 목	내 용
자료명	버스운영시스템(BMS) 데이터
주요 포함 정보	차량 및 업체 정보, 정류장·교차로별 차량 도착정보 등
수집원	버스운행 단말기
수집 시점 및 데이터 적재주기	수집: 정류장 출·도착, 교차로 통과 등 이벤트 발생 시(1초) 적재: 실시간(20초)
저장형태	데이터베이스
크기	약 1.2~1.3GB/주
수집기관	TOPIS

BMS가 도입·운영됨에 따라 현장 조사를 거치지 않고도 서울시에서 운행하는 모든 버스에 대하여 운행 현황을 실시간으로 파악할 수 있게 되었으며, 버스 서비스에 대한 공정한 평가도 가능해졌다. 또한, 버스 운영업체와 기사의 재량에 맡겼던 배차시간 준수 여부도 수시로 확인할 수 있게 되어 버스운행 효율성의 체계적인 진단이 가능해졌다.

BMS 데이터는 버스에 장착된 단말기에서 20초마다 주기적으로 센터로 전송되며, 정류 소별 출발 및 도착, 교차로 중간 지점 통과 등의 이벤트가 발생하는 시점에도 수시로 운행정보가 센터로 전송된다. 차량의 실시간 위치 정보인 GPS 좌표는 별도로 저장되지는 않지만 앞에서 설명된 특정 지점에 접근되었는지 확인하는 과정에서 위치 정보가 활용된다.

BMS 데이터는 버스 도착정보를 안내해주는 스마트폰 애플리케이션 서비스 제공의 초석이 되었고, 교통카드 정보와 함께 대중교통 이용실태와 서비스 질을 분석하는 데 있어 가장 많이 활용되고 있는 데이터라고 할 수 있다. 또한, 향후 지하철의 운행정보와 연계된다면 수도권 대중교통 이용자들의 수단 간 환승에 큰 편의를 가져다줄 것으로 판단된다.



[그림 3-5] BMS 데이터 수집 과정 예시(서울시)

(8) 택시 운행 데이터

[표 3-9] 택시 운행 데이터 세부내용

항 목	내 용
자료명	택시 운행 데이터
주요 포함 정보	택시 등록정보, 운전자 정보, 시각별 차량 위치, 속도, 진행방향, 공차 여부, 감·가속정보 등
수집원	디지털운행기록계(DTG)
수집 시점 및 데이터 적재주기	수집: 10초 단위 적재: 실시간(2.5분)
저장형태	데이터베이스
수집기관	한국스마트카드사

버스와 지하철은 고정된 노선과 승하차 지점 및 운영시간이 정해져 있지만 택시는 요금 체계가 균일하게 적용된다는 점 외에는 일반적인 대중교통과는 다른 수단이다. 그러므로 대중교통수단과는 달리 운영 측면에서 행정당국의 적극적인 규제가 어려운 것이 현실이다. 하지만 택시의 승차 거부, 난폭운전 등에 대해 행정당국에 항의하는 택시 서비스 관련 민원이 끊임없이 발생하고 있어, 이에 적극적으로 대처하고 택시 서비스의 질을 개선할 수 있는 적절한 관리수단이 필요하다.

이에 서울시는 개별 차량에 장착된 통합형 디지털운행기록계(IDTG)로부터 택시 운행정보를 실시간으로 수집하기 시작하면서 2012년 말 택시운행정보 통합관리시스템을 구축하여 운영하고 있다. 이 시스템을 통해 간편하게 택시 운행 실태에 대한 파악이 가능하게 됨으로써 민원에 적극적으로 대처할 수 있게 되었으며, 택시 운행 및 관리와 서비스의 개선을 도모할 수 있게 되었다.



[그림 3-6] 택시 운행 데이터 수집 과정 예시(서울시)

디지털운행기록계를 통해 수집되어 서버에 저장되는 정보에는 택시 등록정보, 택시운전자 정보, 승객탑승 여부 등과 같은 택시 서비스 관련 정보와 시각별 차량 위치, 속도, 진행방향 등 차량의 움직임 관련 정보가 포함되어 있다. 이러한 정보들은 10초마다 차내 장치에 기록되며, 2분 30초마다 서버로 전송되어 저장·축적된다.

택시를 이용하는 승객들은 대중교통과 달리 자신이 원하는 출·도착지와 가장 가까운 지점에서 승하차하기 때문에, 택시 운행 데이터를 이용하면 대중교통의 교통카드 기록보다 이용자들의 통행을 더 정확하게 파악할 수 있다. 또한, 대중교통을 이용하기 어려운 시간과 장소에서 택시를 이용할 가능성이 크기 때문에 대중교통 서비스의 소외지역 파악과 심야버스 노선 개발과 같은 대중교통 서비스 개선에도 많이 활용되고 있다.

(9) 상용차 운행 데이터

화물차, 버스 등 상용차는 일반 승용차보다 단위 시간당 주행거리가 길고 영리활동이 주된 목적인 점을 고려하여 운행에 대한 감독과 관리도 더 엄격하게 시행하고 있다. 2011년부터 사업용 자동차에 대해서는 신규 등록 시 DTG 장착을 법적으로 강제하고 있으며,

DTG에 기록된 차량 운행정보를 교통안전공단의 eTAS(Tachograph Analysis System; 운행기록분석시스템)에 전송하여 저장·관리토록 하고 있다.¹⁰⁾

[표 3-10] 상용차 운행 데이터 세부내용

항 목	내 용
자료명	상용차 운행 데이터
주요 포함 정보	차량 및 운전자정보, 운송사업자 등록번호, 시각별 차량위치, 속도, 진행방향, 브레이크 신호, 가속도 등
수집원	디지털운행기록계
수집 시점 및 데이터 적재주기	수집: 10초 단위 적재: USB-비정기/통신-실시간(180초)
저장형태	데이터베이스
수집기관	교통안전공단

상용차 DTG를 통해 수집되고 있는 정보는 국토교통부 고시인 「자동차 운행기록 및 장치에 관한 관리지침」의 별표 2에서 법적으로 규정되어 있다. [표 3-10]에 기술한 정보 외에도 일일 주행거리, 누적 주행거리, 분당 엔진회전수(RPM) 등도 운행기록으로 저장되도록 명시되어 있으며, 연료소모량과 연비 관련 장치가 장착된 차량에서는 해당 정보도 함께 수집할 수 있도록 하고 있다.

교통안전공단에서는 이와 같은 운행기록 빅데이터를 활용하여 사업용 자동차 운전자의 위험운전행동을 분석하고 교통안전지도를 작성하는 등 도로교통의 안전 향상을 도모하고 있다. 또한, 위험운전행동이나 실제 사고가 자주 일어난 지점을 분석하여 제공하고 있으며, 운전자와 사업체별로 특화된 다양한 분석 결과를 열람할 수 있도록 함으로써 위험운전에 대한 경각심을 일깨우는 데에 일조하고 있다.

상용차 운행 데이터는 위험물 수송차량이나 관광버스의 운행기록도 함께 포함하고 있어 향후 이 차량들이 주로 운행하는 지역이나 경로를 파악하여 관련 정책을 수립하는 데에도 활용할 수 있을 것으로 판단된다.

¹⁰⁾ eTAS 운행기록분석시스템(<http://etas.ts2020.kr/>)

(10) 내비게이션 애플리케이션 기록

[표 3-11] 내비게이션 애플리케이션 기록 세부내용

항 목	내 용
자료명	내비게이션 애플리케이션 기록
주요 포함 정보	이용자 정보, 출·도착지 검색기록, 시각별 차량 위치, 이동 속도 등
수집원	내비게이션 또는 애플리케이션
수집 시점 및 데이터 적재주기	수집: 출·도착지 검색 시 / 1~3초 적재: 실시간(10~30초)
저장형태	파일 또는 데이터베이스
수집기관	내비게이션 운영업체

초기 자동차 내비게이션은 차량에 장착된 전용 단말기 내 저장된 지도상에 GPS로부터 수신한 현재 위치를 표현하여 경로를 안내하는 방식이었다. 이용자들은 저장장치를 분리하거나 이동식 저장 매체를 이용하여 오프라인에서 지도 정보를 갱신해줘야 최신 지도를 참조하면서 주행할 수 있었다. 그뿐만 아니라, 도로의 교통상황이 제대로 반영되지 못한 경로를 안내하는 단말기도 상당히 많이 존재하였다.

하지만 정보통신기술의 발달로 실시간 교통정보를 수집하여 이를 반영한 최적의 경로를 안내해줄 수 있게 되었으며, 최근에는 스마트폰에서 내비게이션이 애플리케이션 형태로 보급되어 이용되고 있다. SK의 티맵, KT의 올레내비, (주)카카오의 카카오내비 등이 대표적인 스마트폰 애플리케이션들은 공공으로부터 교통정보를 공급받거나 내부에서 수집한 공한 실시간 교통정보들을 기반으로 하여 최적 경로를 안내하고 있다.

이용자가 스마트폰에서 내비게이션 애플리케이션을 작동하면 내비게이션을 실행한 단말기의 정보가 전송되며 단말기에 고유 ID가 부여된다. 이용자가 입력한 출발지와 도착지 정보를 토대로 최적의 경로를 탐색하여 제공하고 이용자가 이 경로를 따라 운행 시 실시간으로 단말기 위치, 이동 속도 등이 스마트폰에 저장되었다가 일정 주기마다 서버로 전송된다. 서버에서는 최적 경로를 5~10분의 간격으로 업데이트하여 개별 단말기로 전달하

며, 현재 안내한 이동 경로보다 소요시간이 적은 경로가 탐색되면 이용자에게 새로운 경로를 안내하게 된다.



[그림 3-기] 내비게이션 애플리케이션 작동 시 정보의 흐름 예시

수많은 차량의 내비게이션 애플리케이션을 통해 서버로 동시에 수집되는 차량의 위치 정보, 속도정보, 출도착지 정보, 이동 경로 정보 등을 활용할 수 있다면 도로의 교통상황을 더욱 세밀하게 파악할 수 있고, 차량을 이용한 개인의 통행 특성을 보다 정교하게 알아낼 수 있을 것으로 보여 관련 정책수립에 큰 도움을 줄 것으로 기대된다.

또한, 교통카드 기록을 이용하여 통행의 가종점 파악이 가능한 대중교통 통행과는 달리, 개인 차량을 이용하는 통행의 가종점 파악은 매우 어려운 실정이었다. 하지만 향후 내비게이션으로부터 수집되는 정보를 활용하면 어느 정도 파악이 가능할 것으로 생각된다. 이렇게 되면, 현재 가구통행실태조사를 통해 구축하고 있는 승용차 O/D의 검증 또는 보완에도 활용할 수 있을 것으로 판단된다.

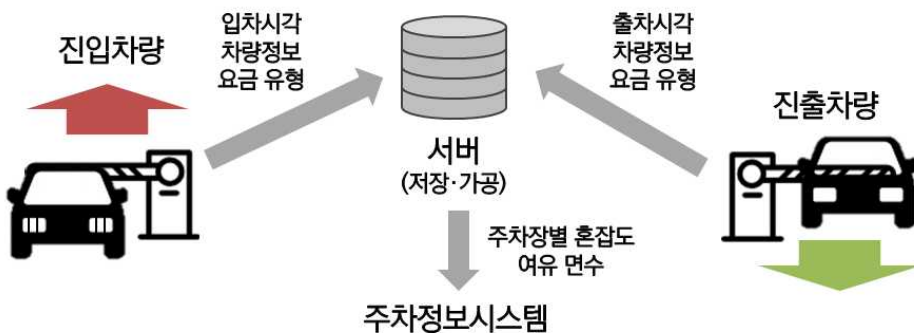
(11) 공영주차장 출입 기록

개인 차량을 이용한 통행의 기점과 종점에서는 항상 주차와 관련된 사항들이 중요한 부분을 차지한다. 주요 교통수단 중 가장 개인화되었다고 할 수 있는 승용차를 이용하여 이동할 때에는 편리함을 느껴도, 막상 도착지에서 운행하지 않는 차량을 보관하기 위한 공간을 찾는 것은 항상 어려운 일이다. 일반적으로 대형건물에서는 부설주차장을 제공하지만 소규모 점포들이 밀집된 지역에서는 주차공간을 찾기가 쉽지 않다. 이에 지방자치단체 등 행정당국에서는 시민의 주차 불편을 조금이라도 완화하기 위해 공영주차장을 곳곳에 설치하여 운영하고 있다.

[표 3-12] 공영주차장 출입 기록 세부내용

항 목	내 용
자료명	공영주차장 출입 기록
주요 포함 정보	주차장 명 및 위치, 차량번호, 입·출차 시각, 총 주차시간, 정산요금, 정기권 여부 등
수집원	무인 주차요금 정산기 또는 직원
수집 시점 및 데이터 적재주기	실시간(차량 출입 시)
저장형태	파일
수집기관	주차장 운영 주체

공영주차장에서는 주차 차량의 요금 징수 및 정산을 위하여 차량의 출입 기록을 무인주차 시스템 또는 운영직원의 단말기를 통해 수집관리하고 있다. 입차 시각과 출차 시각 간의 차이를 이용하여 총 주차시간을 산정하고, 주차장별 기본요금과 주차시간에 따른 추가 요금을 적용하여 총 주차요금을 산출한다. 요금 부과는 주차 정기권 등록 여부에 따라 결정되고 있으며, 대상 차량의 종류에 따라 일부 감면 또는 면제 등의 혜택도 주어진다. 현재 서울시에서는 공영주차장에서 수집되는 정보를 토대로 주차안내시스템을 구축하여 주차정보를 시민에게 제공하고 있으며, 민간에서도 이를 이용한 서비스들을 제공하는 중이다.



[그림 3-8] 공영주차장 출입 기록 수집 및 가공 과정

서울시에서 공영주차장을 이용하는 차량 대수는 하루 평균 약 12,000대(2013~2015년) 수준으로, 전체 승용차 통행 규모에 비하면 크지 않은 편이다. 하지만 공영주차장 출입 기록으로부터 총 주차대수, 주차장 회전을 등을 산출하여 인근 지역의 주차수요 수준을 추정하거나 관련 주차정책을 수립하는 데 활용하기도 한다. 또한, 이 정보는 신규 공영주차장을 건설하거나 주차수요에 따른 탄력적인 요금제 시행 검토 등에도 활용될 수 있을 것으로 보인다.

(12) 유료도로 요금소 이용 기록

[표 3-13] 유료도로 요금소 이용 기록 세부내용

항 목	내 용
자료명	유료도로 요금소 이용 기록
주요 포함 정보	출도착 일시 및 요금소, 차량정보(번호판), 차종 구분, 요금
수집원	하이패스 및 요금소 단말기
수집 시점 및 데이터 적재주기	실시간(요금소 통과 시)
저장형태	N/A
수집기관	한국도로공사 및 민간운영업체

한국도로공사에서 전국적으로 건설·운영하는 고속도로는 대표적인 유료도로이다. 2015년 현재 총 영업 거리는 31개 노선 3,872km에 달하며, 민간자본이 투자된 민자고속도로를 포함하면 33개 노선 4,194km로 늘어나게 된다. 한국도로공사가 운영하는 고속도로의 하루 평균 이용 차량 대수는 약 408만 대로, 차종별로는 소형차 87.2%, 중형차 3.2%, 대형차 9.6%이다(한국도로공사, 2016).

이뿐 아니라 일부 교량과 터널에서도 별도의 통행요금을 징수하는 곳이 있다. 요금 징수 방식은 진입 시 입구 영업소에서 통행권을 발급하고 진출 시 출구 영업소에서 이용 거리에 따른 통행료를 징수하는 폐쇄식과 차량의 진출입 지점과 관계없이 영업소를 통과하는 차량에 대해 균일요금을 징수하는 개방식으로 나뉜다.

교통시설에 따라 달라질 수 있으나, 통행요금은 비접촉식인 하이패스를 사용하여 자동으로 결제하거나 영업소 직원과 대면하여 현금이나 전자카드 등으로 결제하는 방식이 있다. 하이패스는 사전에 등록된 차내 단말기와 가입정보로부터 고유의 식별 ID를 부여하여 진출입 요금소가 어디인지를 파악할 수 있다. 하이패스가 도입되기 전에는 모든 차량이 마그네틱 통행권을 발급받아 고속도로에 진입하였다가 도착지에서 통행권을 제시하고 요금을 결제하는 방식이었으나 2016년 9월 기준 하이패스 이용률이 74.8%에 이르고, 차내 단말기 장착률이 69.5%에 이를 정도로¹¹⁾ 통행료 자동징수시스템이 정착되었다.

2장의 빅데이터 활용사례에서도 볼 수 있듯이 고속도로 요금소 통과 기록은 지역 간 통행 패턴을 파악하거나 신규 인프라 건설 검토에 활용될 수 있으며, 특히 가구통행실태조사의 지역 간 통행 O/D를 검증 또는 보완하는 용도로도 활용될 수 있을 것으로 판단된다.

(13) 검지기 자료

[표 3-14] 검지기 자료 세부내용

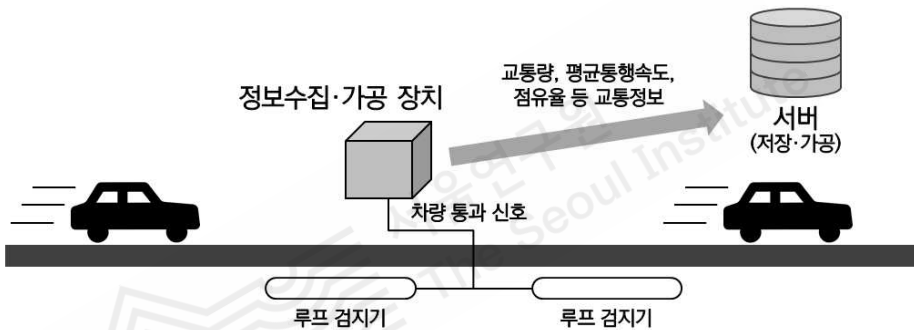
항 목	내 용
자료명	검지기 자료
주요 포함 정보	검지기 위치, 검지기 도달/통과 시점, 검지기 점유시간, 통행속도 등
수집원	검지기
수집 시점 및 데이터 적재주기	실시간(5분)
저장형태	데이터베이스
수집기관	TOPIS, 한국도로공사

전국적으로 구축된 지능형교통체계(Intelligent Transportation System; ITS)에서는 실시간 교통상황을 파악하기 위해서 도로상에 루프 검지기나 영상 검지기 등을 설치하여

¹¹⁾ 파이낸셜뉴스, 「도로공사, 작고 가벼운 '심카드형 하이패스 단말기' 첫 선」, 2016년 10월 12일자, <http://www.fnnews.com/news/201610120841141454>

교통량, 속도 등 교통정보를 수집하고 있다. 이렇게 수집된 정보는 VMS, 교통방송, 차량 내비게이션, 포털지도 등 다양한 채널을 통해 필요한 곳에 제공되고 있으며, 수집되는 정보들을 집계하여 통계정보로도 제공되고 있다.

통상적인 루프 검지기의 경우, 도로에 매설된 검지기를 차량이 통과할 때 변화하는 자기장으로부터 유도되는 전기 신호를 감지한 후 가로변의 제어장치에서 이 신호를 해석하여 교통정보를 생성한다. 교통량, 통행속도, 점유율 등을 포함한 교통정보는 통신망을 통해 교통정보센터로 전송되며, 센터에서는 이를 가공하여 도로 이용자 등 교통정보가 필요한 곳에 다양한 채널을 통해 제공된다. 한편, 비교적 최근에 설치되고 있는 영상 검지기는 현장에서 촬영한 화상을 센터로 전송하여 필요한 정보를 분석하는 형태를 취하고 있다.



[그림 3-9] 루프 검지기 자료 생성 및 가공 과정

하지만 ITS를 위해 현장에 설치된 검지기는 유지보수가 어렵고 정확도 측면에서 오류 가능성을 항상 내포하고 있을 뿐만 아니라, 최근에는 차내의 설치된 단말기의 GPS 위치 정보 등을 활용하여 소통 정보들을 획득할 수 있어 그 효용이 다소 낮아진 것도 사실이다. 그러나 개인을 식별할 수 있는 정보를 수집하지 않기 때문에 개인정보 노출에 대한 우려가 스마트폰 등 다른 장치보다 적고, 수집된 데이터의 활용성이 높다는 장점이 있다.

(14) 교통정보 CCTV 자료

CCTV(Closed Circuit Television; 폐쇄회로 화면) 카메라는 관측하고자 하는 특정 지점의 현장 상황을 실시간 영상으로 확인할 수 있는 장비이며, 차량의 흐름 및 교통상황 파

악, 범죄예방 및 추적, 불법주차 단속 등 다양한 용도로 사용되고 있다. 교통량 측정을 위한 영상 검지기는 촬영된 영상을 바탕으로 이미지 프로세싱을 통해 화면의 임의 지점을 통과한 차량을 계수하는 방식으로 필요한 정보를 추출한다. 범죄의 용의 차량 추적과 불법주차 단속을 위한 CCTV 카메라 영상에는 차량 번호판을 자동으로 식별하는 기술이 적용되고 있다.

[표 3-15] 교통정보 CCTV 자료 세부내용

항 목	내 용
자료명	교통정보 CCTV 자료
주요 포함 정보	카메라 설치 위치, 녹화일시, 녹화 영상
수집원	폐쇄회로 카메라
수집 시점 및 데이터 적재주기	실시간(상시)
저장형태	N/A
수집기관	경찰청, 서울시설공단 등

CCTV 카메라 영상은 다른 자료와 비교하면 데이터의 규모가 매우 방대하고 특정 개인의 위치 및 이동 동선 파악에 악용될 수 있다는 점 때문에 오랜 기간 자료를 보관하는 데 한계가 있고 데이터의 활용성 측면에서도 기대에 못 미치고 있다. 그러나 최근 발전하고 있는 영상 분석 및 인공지능 기술을 고려한다면 향후 수집된 데이터의 활용 잠재력이 매우 클 것으로 전망된다.

(15) 교통사고 정보

2015년 우리나라에서 1년간 일어난 교통사고 건수는 약 23만 건에 달하며, 같은 기간 서울시에서 발생한 교통사고 사망자 수는 376명이었다. 2013년을 기준으로 자동차 1만 대당 사망자 수는 2.2명으로 집계되었는데, 칠레 6.7명과 터키 3.3명에 이어 OECD 가입국 중 우리나라는 3위에 해당하는 것으로 나타나, 우리나라의 교통사고 심각성은 다른 국가들의 교통사고 통계와 비교하면 매우 큰 것으로 나타났다.¹²⁾

[표 3-16] 교통사고 정보 세부내용

항 목	내 용
자료명	교통사고 정보
주요 포함 정보	사고 발생일시 및 장소, 사고 상세내용, 피해자 및 가해자 정보, 상해 정도 등
수집원	경찰청 및 보험사 접수기록
수집 시점 및 데이터 적재주기	1년(매년 4월경)
저장형태	데이터베이스
수집기관	도로교통공단

경찰청훈령인 「교통사고조사규칙」의 제40조 제3항에 의하면, 교통사고 조사가 종결된 지 3일 이내에 교통사고 통계 원표에 표시된 항목을 도로교통공단의 교통사고관리시스템(TCS)에 입력하도록 규정되어 있다. 교통사고 통계 원표에는 발생한 교통사고에 대해 총 68가지에 달하는 항목을 조사하여 입력하게 되어 있으며, 사고 발생일시 및 장소, 관할 경찰서, 사망자·부상자 수, 사고유형 및 원인, 사고 당사자 특징 등이 포함되어 있다.

경찰과 보험사에서는 교통사고 후속 처리를 위해 관련 기록을 보관하고 있으며, 도로교통공단에서는 이러한 정보들을 1년에 1회씩 관계기관들로부터 수집하여 통계 처리와 교통사고분석시스템(Traffic Accident Analysis System; TAAS)을 통한 사고분석, 교통사고 발생 수준 등을 모니터링하고 있다. TAAS 시스템 내의 GIS 분석 기능을 이용하여 지역별, 사고유형별 교통사고 발생 현황 등을 확인할 수 있다.

교통사고 정보는 사고 발생 원인 분석을 통해 사고를 예방하는 등 교통사고 감소를 위한 관련 정책 개발에 활용되고 있다. 예를 들면, 사고가 자주 발생하는 도로 구간의 선형을 개량하거나 안전시설물을 보강하고, 사고를 자주 일으키는 특정 집단에 대한 안전 교육을 강화하는 등의 대책을 마련하는 데 사용될 수 있다. 이외에도 다른 유형의 정보와 연계되어 도로교통 안전 증진에 이바지할 수 있는 정책들을 도출할 수 있을 것으로 기대된다.

12) 도로교통공단 교통사고분석시스템(<http://taas.koroad.or.kr>)

(16) 불법 주정차 단속 정보

[표 3-17] 불법 주정차 단속 정보 세부내용

항 목	내 용
자료명	불법 주정차 단속 정보
주요 포함 정보	소속 자치단체, 차량번호, 단속 차종, 단속 구역, 단속 위치, 위반 일시, 단속 구분 등
수집원	주차단속카메라 및 단속원
수집 시점 및 데이터적재주기	실시간(적발 시)
저장형태	데이터베이스
수집기관	지방자치단체

불법 주정차는 주차 가능 구역으로 지정되지 않은 장소에 차량을 주차하거나 정차하는 행위를 일컫는다. 불법 주정차 차량은 다른 차량의 주행을 방해하고, 심지어는 긴급차량의 출동을 지연시켜 2차 피해 발생에 대한 원인을 제공하기도 한다. 그러므로 불법 주정차 단속 권한을 가진 행정 당국에서는 주차단속원과 고정식 카메라, 단속 차량 등 다양한 수단을 이용하여 불법 주정차 차량을 단속하고 있다. 또한, 주차 민원 접수 부서나 최근에 등장한 스마트폰 애플리케이션을 통해 신고를 접수하고, 단속원이 직접 현장을 방문하여 확인 후 계도 또는 과태료를 부과하고 있다.

단속 시에는 차량번호 조회를 통해 차량 등록정보 및 소유주 정보를 확인할 때도 있으며, 불법 주정차를 적발한 장소와 시간을 입력하여 추후 발생할 수 있는 이의 제기에 대응한다. 이러한 과정을 거쳐 서울시에서만 하루 약 9천 대의 차량이 적발되고 있는데, 인력과 장비의 부족 등으로 불법 주정차 차량 전체를 단속하는 데 한계가 있어 실제 불법 주정차 규모는 훨씬 클 것으로 보인다.

불법 주정차 단속 정보는 불법 주정차가 빈번한 지역을 선별하고 이를 예방하기 위한 계도활동, 주차장 공급 확대 등 주차정책 마련에 활용될 수 있다. 예를 들면 불법 주정차 발생 빈도가 높은 지역에 대해 사후 단속이 아닌 주차단속원 수시 순찰, 고정식 주차단속

CCTV 설치 등의 선제 조치를 함으로써 불법 주정차 발생 가능성을 원천적으로 차단하는데 활용될 수도 있다.

(17) 속도위반 단속 기록

[표 3-18] 속도위반 단속 기록 세부내용

항 목	내 용
자료명	속도위반 단속 기록
주요 포함 정보	차량번호, 단속 일시, 단속지점, 단속 시 통과속도, 운전자 정보 등
수집원	과속 단속 카메라
수집 시점 및 데이터 적재주기	실시간(적발 시)
저장형태	N/A
수집기관	경찰청

과속은 2015년 서울에서만 660,997건,¹³⁾ 전국에서는 8,472,983건(경찰청, 2016)이 적발되었을 정도로 우리나라 국민이 가장 흔히 위반하는 교통법규에 해당한다. 과속의 우려가 있는 도로 구간에 고정식 또는 이동식 단속 장비를 설치한 후 규정된 제한속도를 초과하여 주행하는 차량을 단속한다. 전방에 매설된 검지기를 통해 단속지점을 통과하는 모든 차량의 속도를 측정한 뒤 제한속도를 초과하는 위반 차량을 식별하여 카메라로 차량 앞부분을 촬영한다. 따라서 카메라로 촬영되어 단속 기록이 저장되는 차량은 모두 과태료 부과 대상이며, 초과한 속도의 크기에 따라 벌점 부여 대상이 되기도 한다. 수집되는 정보에는 적발 일시 및 위치, 차량번호, 적발 당시의 통과속도, 단속지점의 제한속도 및 초과 속도 등이 있으며, 자동차 등록원부의 차량 소유자 정보와 연계되어 위반 사실 통지서가 등록된 주소지로 발송된다.

13) 서울연구원 인포그래픽스, 「서울의 교통법규 위반 단속현황은?」, 2016년 6월 20일자, <https://www.si.re.kr/node/55356>

(18) 택배 물품 추적 기록

[표 3-19] 택배 물품 추적 기록 세부내용

항 목	내 용
자료명	택배 물품 추적 기록
주요 포함 정보	발신자, 수신자, 물품 정보, 접수·집화·도착시간, 수송차량 및 담당자 정보 등
수집원	택배전산시스템 및 단말기
수집 시점 및 데이터 적재주기	실시간(이벤트 발생 시)
저장형태	N/A
수집기관	택배업체

택배는 물품을 원하는 주소로 수송해 주는 서비스이며, 물품의 크기 및 무게 등에 의해 이용 요금이 달라진다. 온라인쇼핑몰의 활성화와 함께 관련 시장이 지속해서 성장해왔으며, 2016년 한 해 동안 처리된 택배 물량은 20억 5천만 건으로 추정되고 있다.¹⁴⁾ 택배사들은 물품의 효율적인 집화 및 배송을 위하여 교통 거점에 물류센터를 공동 또는 개별로 운영함과 동시에 간·지선 배송 체계를 구축하여 배송 효율성을 극대화하고 있다. 물품 배송을 위해 택배사는 발신자와 수신자의 이름, 주소, 연락처 등 개인정보를 수집하고 있으며, 배송물품의 위치를 추적하여 고객에게 서비스를 제공하고 있다.

택배 물품의 정보 및 추적 정보는 공공에서 파악이 어려운 택배 물동량과 물류의 가·종점을 추정하는 데 활용할 수 있으며, 물류의 거점과 물류시설의 입지를 선정하는 등 관련 물류 정책수립에 유용하게 사용될 수 있을 것으로 판단된다.

¹⁴⁾ 물류신문, 「[르포] 2017년 택배 힘겨운 노동현실, '바뀌는 중'」, 2017년 4월 5일자, <http://www.klnews.co.kr/news/articleView.html?idxno=115216>

2_빅데이터마다 활용성은 서로 달라

지금까지 정리한 18가지 종류의 빅데이터들은 크기, 수집 주기, 수집 경로, 정보의 종류 등의 그 특성이 매우 다양하며, 데이터 취득 및 활용 가능성에도 차이가 있다. 대외적으로 공개되고 있는 데이터도 있지만, 아래와 같이 주로 개인정보 보호나 기업의 영업비밀 등을 이유로 아직도 공개되지 못하는 데이터들이 대부분이다.

1) 개인정보 보호

전체 집단에서 특정 성격을 가지는 표본의 비율이나 일정 기간 발생한 특정 사건의 빈도와 같은 집계 또는 통계 수치들은 개인을 식별할 수 있는 정보를 포함하고 있지 않으므로, 단순히 전체 집단의 특성이나 동향을 파악하는 데 널리 활용되고 있다. 하지만 개인을 식별할 수 있는 빅데이터의 원자료는 개인정보 보호를 위해 법적으로 공개가 금지되어 있으며, 공개 시에는 개인을 식별할 수 없도록 집계 가공하여야 한다.

데이터를 보유하고 있는 기관에서는 개인정보 유출로 인해 발생할 수 있는 피해를 방지해야 할 의무가 있다. 그런데 정보제공을 요청받았을 때 개인별 ID 및 식별코드, 연락처, 주소 등의 정보는 삭제 처리하고 가공해야 하는 인적물적 부담이 발생하게 되므로 데이터 보유 기관 입장에서는 데이터 공개를 꺼리게 된다. 데이터를 사용하고자 하는 기관이나 연구자 관점에서는 연구목적에 따라 주소지 등 일부 개인정보가 필요할 때가 있으나, 이러한 정보가 제공되지 않으면 데이터의 활용성이 크게 저하되는 이슈가 항상 존재한다.

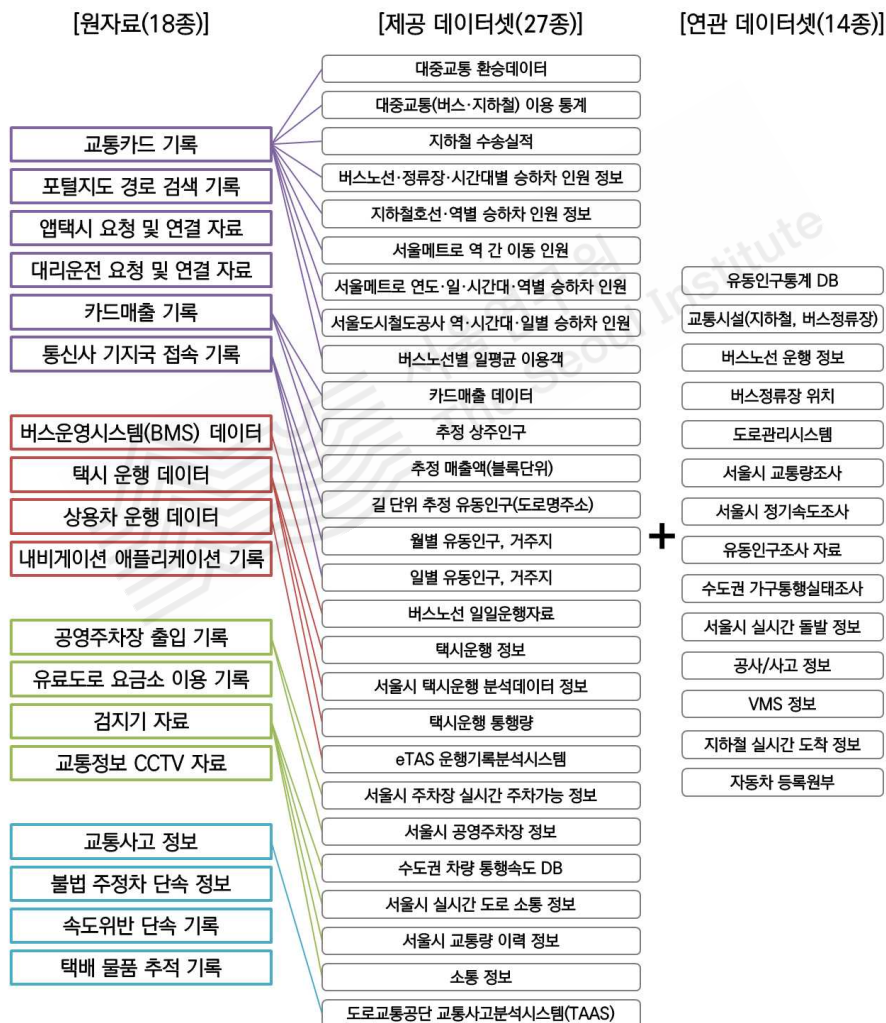
2) 기업의 영업비밀

민간업체들은 원활한 서비스를 제공하기 위하여 고객에 대한 대면 확인, 연락처, 위치, 카드 종류 및 결제 한도 등 자사의 서비스에 필요한 개인정보를 부득이하게 수집해야 한다. 이를 위해 사전에 고객으로부터 개인정보 수집에 대한 동의를 얻는 절차를 거쳐야 할 뿐 아니라, 정보 수집 및 저장관리를 위한 각종 설비를 구축하고 수집된 정보가 외부로 유출되어 고객들이 피해를 보지 않도록 보안 대책도 마련하여 시행해야 한다.

여기에는 막대한 비용이 투입되므로 기업들은 수집된 정보를 마케팅 등 이윤을 극대화하

는 방향으로 활용하고자 하며, 이러한 정보가 자사의 영업 비밀에 해당할 뿐만 아니라 잠재적인 자산으로 인식하기 때문에 일반에 공개하는 것을 극도로 회피하는 경향이 크다.

위와 같은 이유로 빅데이터 원자료 제공은 거의 이루어지지 않고 있으나, 공공기관을 중심으로 일부 민간업체에서 개인정보를 모두 제거한 1차 가공자료를 [그림 3-10]과 같이 제공하고 있는 것으로 나타났다. [표 3-20]에서는 현재 제공되고 있는 데이터셋 27종과 연관 데이터셋 14종에 대한 세부사항을 기술하였다.



[그림 3-10] 원자료와 제공 데이터셋 간의 관계

[표 3-20] 제공 및 연관 데이터셋별 주요 특성

제공/연관 데이터셋	제공기관	자료 출처	시간범위	수집 주기	형태	공개 여부	주요 항목
카드매출 데이터	서울시 빅데이터캠퍼스	KB카드	2014.1.~현재	월	csv	공개	가맹점수, 법인/남성/여성 매출액, 연령/요일/시간별 매출액
월별 유동인구, 거주지	서울시 빅데이터캠퍼스	SKT	2016.1.~현재	월	csv	공개	기준년월, 50m×50m 단위 셀 X/Y 좌표, 성별, 연령대, 요일, 시간대, 유동인구 수
일별 유동인구, 거주지	서울시 빅데이터캠퍼스	SKT	2015.8.~현재	월	csv	공개	기준일, 50m×50m 단위 셀 X/Y 좌표, 성별, 연령대, 요일, 시간대, 유동인구 수
추정 상주인구	서울시 빅데이터캠퍼스	오픈메이트	2014.하반기~현재	반기	csv, shp	공개	기준년월, 지번, 행정동, X/Y 좌표, 연령대별/성별 인구
유동인구통계 DB	서울시 빅데이터캠퍼스	KT	2015.9.	1회	csv, shp	공개	서울시 동단위 시간별 유동인구 현황(2015년 9월 한 달간 시간대별 유동인구 통계 데이터)
대중교통 환승데이터 (교통카드 거래내역)	서울시 빅데이터캠퍼스	한국스마트카드	2014.4.~현재	월	csv	공개	카드번호, 운행출발일시, 교통수단, 환승횟수, 버스노선ID, 교통사업자ID, 차량ID, 사용자 구분, 승하차일시, 승하차정류장ID
대중교통 이용 통계 (버스/지하철)	서울시 빅데이터캠퍼스	한국스마트카드	2014.1.~2015.10.	월	csv	공개	운행일자, 노선/호선, 정류장/역, 승차시간 구분, 시간구간ID, 승차 총승객수, 하차 총승객수
교통시설 (지하철, 버스정류장)	서울시 빅데이터캠퍼스	서울시	2014~현재	필요시	csv	공개	버스노선, 버스정류장, X/Y좌표, ARS번호, 자치구명, 지하철역, X/Y 좌표
택시운행 정보	서울시 빅데이터캠퍼스	한국스마트카드	2013.1.~현재	월	csv		택시ID, GPS X/Y/Z좌표, 시간, 방향, 속도, 상태, 승차/공차 여부
추정 매출액(블록단위)	서울시 빅데이터캠퍼스	서울시	2013.1.~현재	월	csv, shp	공개	기준년월, 블록코드, 서비스업종, 평균영업개월수, 기간별 생존율, 요일/시간/연령/성별 매출비율/금액/건수
길 단위 추정 유동인구 (도로명주소 도로링크)	서울시 빅데이터캠퍼스	서울시	2015.1.~현재	월	csv, shp	공개	기준년월, 도로링크ID, 시군구코드, 행정동코드, 요일, 연령대, 시간대, 유동인구 수

[표 3-20 계속] 제공 및 연관 데이터셋별 주요 특성

제공/연관 데이터셋	제공기관	자료 출처	시간범위	수집 주기	형태	공개 여부	주요 항목
택시운행 통행량 (택시운행본석데이터)	서울시 빅데이터캠퍼스	서울시	2015.1.~현재	월	csv	공개	링크ID, 요일, 시간, 날씨, 목적지, 승하차건수, 공차건수
버스노선 운행 정보	서울연구원	서울시 교통정보센터	2007~2012	실시간	xls	공개	업체명, 업체ID, 노선번호, 노선ID, 기종점, 인가/운행/여비대수, 인가거리, 운행시간, 대당 횟수, 총운행횟수, 배차간격, 첫차/막차시간
버스노선 일일운행자료	서울연구원	서울시 교통정보센터	2006	실시간	xls	공개	버스ID, 도착/출발 구분, 정류장별 출/도착 시간
버스노선별 일평균 이용객	서울연구원	서울시 교통정보센터	2005~2012	실시간	xls	공개	버스노선, 업체명, 기종점, 이용객수, 대당 이용객수
버스정류장 위치	서울연구원	서울시 교통정보센터	2007~2014	실시간	shp	공개	노선번호, 정류소명, X/Y 좌표, 경위도
도로관리시스템	서울연구원	서울시 도로관리시스템	2002, 2004, 2006, 2010, 2013, 2015	1년	shp	공개	도로기능, 면적, 연장, 노선명, 폭원, 규모, 시종점, 차로수 등
서울시 교통량조사	서울연구원	서울시	2005~2011	1년	xls	공개	지점명/위치, 도로명, 차로수, 방향, 요일/시간대별 구분, 시간당 교통량, 일일교통량
서울시 정기속도조사	서울연구원	서울시	2005~2010	1년	xls	공개	조사지점, 조사일시, 차종별 평균통행속도 등
수도권 차량 통행속도 DB	서울연구원	서울시	2001~2010	실시간	txt	공개	수집년월일, 수집시간, 링크시종점노드, 통행속도, 시도/시군구
유동인구조사 자료	서울연구원	서울시	2009, 2012~2014	비정기	xlsx, dbf, shp	공개	조사지점 공간 특성, 유동인구수, 유동인구 관찰조사(연령, 소지품, 복장 등), 유동인구 면접조사(거주지, 통행 주목적, 방문횟수 등)
지하철 수송실적	서울연구원	서울시	2000~2012	실시간	xls	공개	노선, 역명, 승하차 구분, 월별합계
수도권 가구통행실태조사	서울연구원	서울연구원	1996, 2002, 2006, 2010	5년	txt	공개	기준년도 목적 및 수단 O/D, 장래 목적 및 수단 O/D
서울시 실시간 돌발 정보	서울 열린데이터 광장	서울교통정보센터	N/A	수시	API	공개	돌발ID, 발생일시, 종로예정일시, 돌발유형, 링크ID, X/Y 좌표, 돌발내용

[표 3-20 계속] 제공 및 연관 데이터셋별 주요 특성

제공/연관 데이터셋	제공기관	자료 출처	시간범위	수집 주기	형태	공개 여부	주요 항목
서울시 실시간 도로 소통 정보	서울 열린데이터 광장	서울교통정보센터	N/A	수시	API	공개	링크ID, 속도, 여행시간
서울시 교통량 이력 정보	서울 열린데이터 광장	서울교통정보센터	2006.1.~현재	수시	API	공개	지점번호, 연월일, 시간, 유입유출 구분, 차로번호, 교통량
서울시 주차장 실시간 주차가능 정보	서울 열린데이터 광장	서울시 주차정보안내시스템	N/A	수시	API	공개	주차장코드, 주차장명, 위/경도, 배정코드, 배정코드명
서울시 공영주차장 정보	서울 열린데이터 광장	서울시 주차정보안내시스템	N/A	비정기	csv, API	공개	주차장코드, 주차장명, 주소, 주차장종류, 운영구분, 주차현황 정보 제공 여부, 주차면수, 현재 주차대수, 유·무료 구분 등
서울시 버스노선별·정류장별·시간대별 승하차 인원 정보	서울 열린데이터 광장	교통카드 정산시스템	2015.1.~현재	월, 일	csv, API	공개	사용년월, 노선번호, 노선명, 정류장ID, 정류장명, 시간대별 총승하차 승객수
서울시 지하철호선별·역별 승하차 인원 정보	서울 열린데이터 광장	교통카드 정산시스템	2015.1.~현재	일	csv, API	공개	사용일자, 호선명, 역명, 승차 총승객수, 하차 총승객수
서울메트로 역 간 이동 인원	서울 열린데이터 광장	서울메트로	2013.10.~2015.8.	비정기	csv	공개	사용년월, 승차역, 하차역, 사용자 구분, 이용인원
서울메트로 연도별·일별·시간대별·역별 승하차 인원	서울 열린데이터 광장	서울메트로	2010~2016.11.	비정기	csv	공개	역명, 날짜, 승하차 구분, 시간대별 승하차 인원수
서울도시철도공사 역별·시간대별(일) 승하차 인원	서울 열린데이터 광장	서울도시철도공사	2014	비정기	xlsx	공개	일자, 호선명, 역명, 승차 총승객수, 하차 총승객수

[표 3-20 계속] 제공 및 연관 데이터셋별 주요 특성

제공/연관 데이터셋	제공기관	자료 출처	시간범위	수집 주기	형태	공개 여부	주요 항목
서울시 지하철 실시간 도착 정보	서울 열린데이터 광장	서울시 교통정보과	N/A	수시	API	공개	지하철호선, 상/하행, 도착지 방면, 내리는 문 방향, 이전/다음역, 지하철역, 환승노선수, 연계호선ID, 연계지하철역ID, 열차종류, 도착예정시간, 열차번호, 종착역, 도착정보 생성시간
서울시 택시운행 분석데이터 정보	서울 열린데이터 광장	서울시 빅데이터시스템	2015~	월	file, API	공개	지점 X/Y 좌표, 요일, 시간, 날씨, 목적지, 승하차건수, 공차건수
소통 정보	국토교통부 교통정보 공개서비스	ITS 국가교통정보센터	2015~	5분	csv	공개	집계일자, 링크ID, 속도, 교통량, 밀도, 통행시간, 지체시간, 차량행렬 대기길이, 검시기점유율
공사/사고 정보	국토교통부 교통정보 공개서비스	ITS 국가교통정보센터	2015~	1시간	csv	공개	생성일, 사고ID, 발생시간, 상태, 분류코드, 분류명, 메시지, 지점 X/Y 좌표
VMS 정보	국토교통부 교통정보 공개서비스	ITS 국가교통정보센터	2015~	1시간	csv	공개	생성시간, 기관코드, VMS기기 정보, 표출시간, 내용
자동차 등록원부	자동차민원 대국민포털	국토교통부	-	수시	N/A	비공개	등록번호, 차대번호, 차명, 사용본거지, 자동차 소유자, 원동기형식, 차종, 용도, 세부유형, 구조장치 변경사항, 검사유효기간 등
eTAS 운행기록분석시스템	교통안전공단	교통안전공단	2009~	수시	xls	비공개	연도·분기·월·요일·시간대·운전자·차량별 위험운전행동 통계, 평균 운행속도, 총운행시간·거리 등
도로교통공단 교통사고분석시스템 (TAAS)	도로교통공단	도로교통공단	2007~	1년	xls	공개	교통사고 발생 추세, 지역·사상자·사고유형·운전자·차종·도로환경별 교통사고 발생 현황 및 통계

04

빅데이터와 교통정책 연계 구상

- 1_교통정책 담당부서의 실무와 연계 가능성
- 2_빅데이터와 공간정보 연계 시 시너지 효과 기대
- 3_빅데이터를 활용한 다양한 교통정책 구상 가능

04 빅데이터와 교통정책 연계 구상

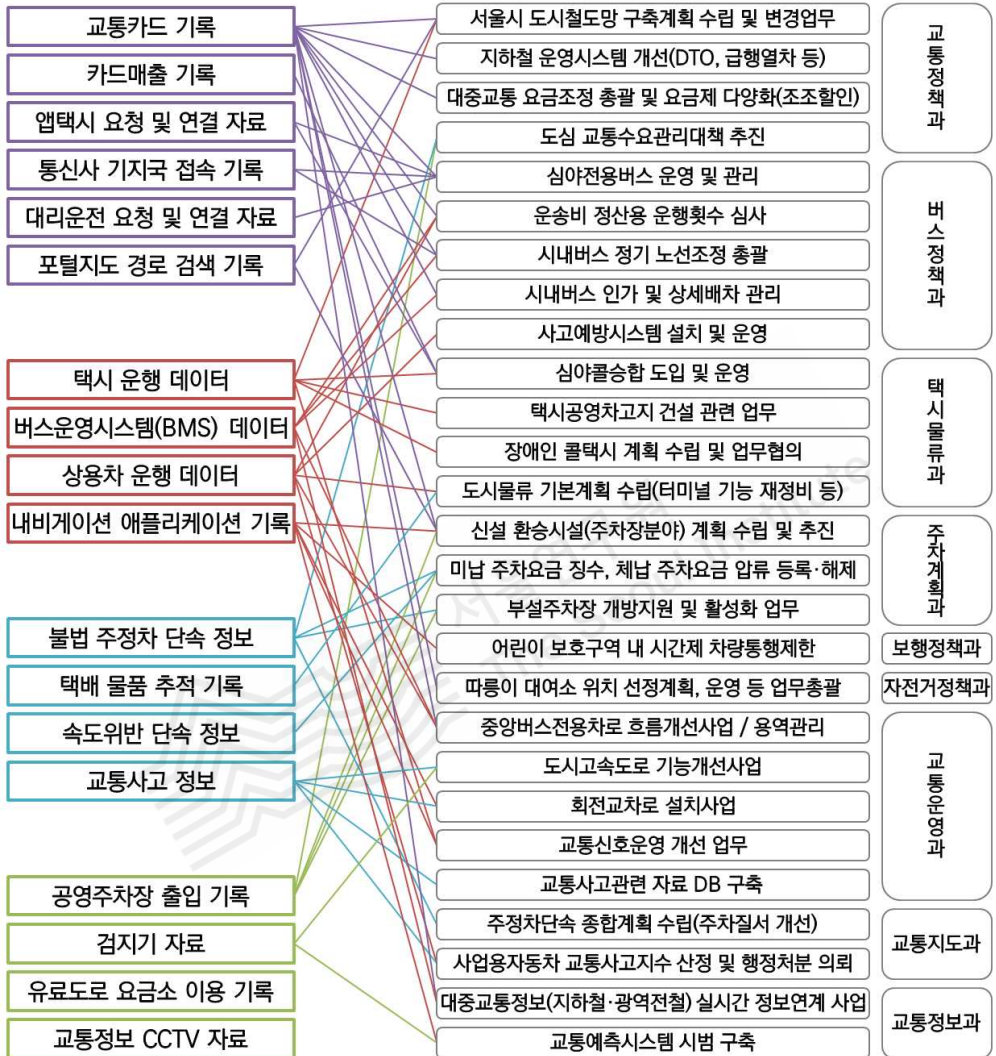
1_교통정책 담당부서의 실무와 연계 가능성

교통 분야는 민간 영역보다는 공공 영역의 역할이 더 크기 때문에 빅데이터 분석 및 활용이 행정 당국의 업무를 효율화하고 더욱 정교한 정책 시행에 큰 도움을 줄 수 있을 것이다. 이 연구에서는 행정당국의 빅데이터 활용 촉진을 위해 3장에서 파악한 빅데이터와 서울시 교통 관련 부서 업무와의 연관성을 개략적으로 파악하여 [그림 4-1]과 같이 정리해 보았다. 빅데이터의 활용이 가능할 것으로 판단되는 서울시 관련 부서의 업무는 서울시 공식 홈페이지(<http://www.seoul.go.kr>)의 조직도에 나타난 내용을 근거로 하였다.

[그림 4-1]을 살펴보면 가장 많은 업무와 연관성이 있고 실제로 가장 많이 사용되는 빅데이터는 교통카드 기록과 BMS 데이터이다. 교통카드 기록은 거의 전체 대중교통 이용자로 부터 수집되는 정보로 BMS 데이터와 연계하여 대중교통 통행 행태 및 통행패턴 파악, 노선 조정, 운영업체 관리감독 등 많은 분야에서 적극적으로 활용되고 있다. 택시 및 상용차 운행 데이터, 내비게이션 애플리케이션 기록 등은 현재 여러 가지 이유로 적극적인 활용이 이루어지지 못하고 있지만 향후 다양한 업무에서 활용될 수 있다면 그 효과가 상당할 것으로 판단된다.

시민의 호평을 받은 심야 노선버스의 노선은 택시 운행 데이터와 통신사 기지국 접속 기록을 활용하여 계획된 사례이다. 지금은 성공적으로 정착되어 심야 노선버스 인지도와 이용률이 높아진 상황으로, 노선 추가 신설 및 증차 등에 대한 요구가 여전히 존재하고 있다. 이러한 요구에 대응하기 위해서는 향후 택시 운행 데이터와 통신사 기지국 접속 기록 외에도 앱택시 및 대리운전 요청 및 연결 자료 등을 추가하여 분석하면 더욱 정교한 노선개발이 이루어질 수 있을 것으로 기대된다. 앱택시 및 대리운전 요청 자료에는 실제 택시에 탑승하지 못한 승객이나 대리운전기사를 배정받지 못한 고객의 정보를 포함하고 있으므로 심야 노선버스의 잠재수요를 좀 더 정교하게 파악하여 대응할 수 있을 것이다. 이러한

자료들은 심야 노선버스와 유사한 심야 콜 승합 도입 및 운영 업무에도 활용될 수 있을 것으로 생각된다.



[그림 4-1] 빅데이터와 현재 서울시 업무 간 연관성

한편, 시외 지역과 서울 간 통행자들을 대상으로 파크 앤 라이드(Park and Ride)¹⁵⁾ 이용을 촉진하기 위해 추진하고 있는 환승주차장 신설계획 수립 관련 업무에는 포털지도 경로 검색 기록과 내비게이션 기록을 활용할 수 있을 것으로 생각된다. 포털지도 경로 검색 기록을 토대로 주요 입지 후보지들에 대하여 승용차를 이용하는 시외 구간과 대중교통을 이용한 시내 구간에서의 총 통행시간을 알 수 있을 것으로 판단되며, 내비게이션 기록에서는 후보 지점을 경유하여 통행한 차량의 출·도착지를 파악하여 시설의 잠재적 이용수요를 가늠해 볼 수 있을 것으로 판단된다.

이외에도 교통 관련 부서의 다양한 업무에서 빅데이터를 활용하여 업무를 효과적으로 처리하고, 시민들이 체감하고 감동할 수 있는 정교한 정책을 마련·시행할 수 있을 것으로 기대된다.



¹⁵⁾ 대도시 외곽에 거주하는 이들이 도심으로 가는 대중교통수단을 이용하기 위해 자신의 차를 세워두는 주차시설
출처: 두산백과(<http://terms.naver.com/entry.nhn?docId=1977680&cid=40942&categoryId=32191>)

2_빅데이터와 공간정보 연계 시 시너지 효과 기대

이 연구에서 살펴본 빅데이터 대부분의 원자료는 GPS 좌표나 주소 등 데이터가 생성된 장소에 대한 정보를 포함하고 있다. 이는 빅데이터를 공간정보와 연계하여 다양한 분석이 가능하다는 것을 시사한다. 공간정보는 고도·수계·경사·생태 등의 자연지형 정보, 도로·철도·교량·터널 등의 교통시설 정보, 상하수도·전력망·송유관·가스관 등의 인프라 정보, 토지용도·건축물·공시지가·실거래가 등의 부동산 정보 등으로 매우 다양하며, 빅데이터 분석의 목적에 따라 적절한 정보와 연계가 필요하다. 간단한 예를 들면, 교통카드 기록과 토지 용도를 연계하여 대중교통 통행이 많이 일어나는 지역의 토지이용 특성을 알아낼 수 있을 것이다.

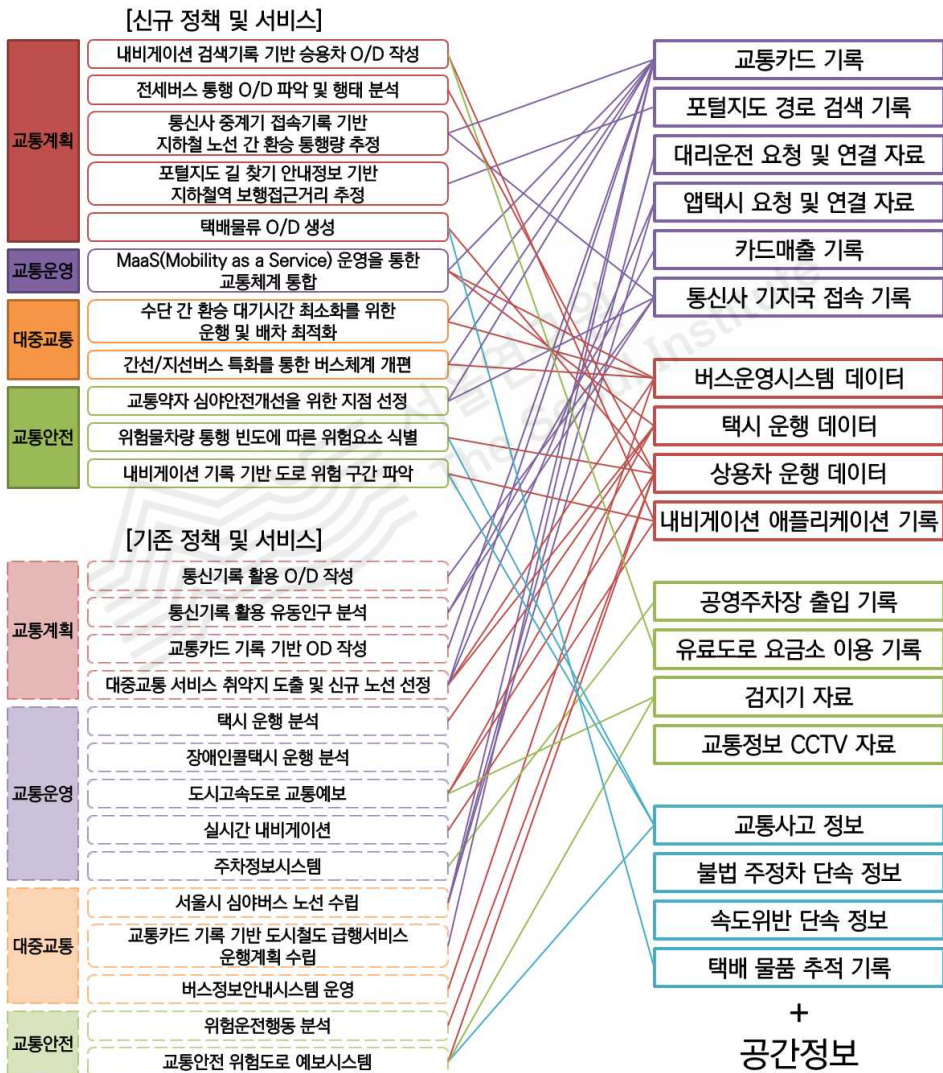
교통 부문에서 수집되는 빅데이터 원자료와 연계할 수 있을 것으로 생각되는 대표적인 공간정보는 [표 4-1]과 같다.

[표 4-1] 주요 공간정보 기초자료

분야	주요 공간정보 기초자료
교통	KT Olleh Map 네트워크 DB, KTDB 교통네트워크 GIS DB, ITS 전국표준노드링크
공간경계	통계청 기초단위구, 통계청 집계구, 통계청 도시화지역, 통계청 행정경계(읍면동), 행정안전부 주민등록 행정기관코드, 안전행정부 국가기초구역
건축	건축물대장, 건축물현황도, 재산세 과세자료(주택, 건축분), 주택공시가격자료, 주택(매매 및 전월세) 실거래가 자료
토지	지적도, 토지(임야)대장, 개별공시지가, 재산세 과세자료(토지분), 토지특성도
도시계획	서울시 도시계획 용도지역지구도, 서울시 도시계획 시설, 서울시 도시계획사업, 국토교통부 용도지역지구도
환경	도시생태현황도, 환경부 국토환경성평가지도, 환경부 생태자연도, 환경부 토지피복도, 산림청 맞춤형 조림지도, 산림청 산림입지도, 산림청 산사태위험지관리도, 산림청 수치임상도, 산림청 산지구분도, 광역토지이용현황도, 서울시 토지이용현황도, 수치지질도, 정밀토양도, 기상청 기상자료, 대기오염측정자료(시간별, 일별), 하수관망도
지형	수치지형도, 종이지형도, 수치표고모델, 위성영상, 항공사진, 행정안전부 새주소시스템 기본도
복지문화	문화재 보존관리 지도

3_빅데이터를 활용한 다양한 교통정책 구상 가능

3장에서 살펴본 빅데이터는 2장에서 기술한 국내외 사례를 비롯하여 그동안 교통의 여러 분야에서 부분적으로 활용이 이루어졌다. 예를 들면, [그림 4-2]의 왼쪽 아래 ‘기존 정책 및 서비스’와 같이, 통행 O/D 추정, 위험운전행동 분석, 고속도로 교통예보, 심야버스 노선 수립 등 다양한 분야에서 빅데이터 활용이 시도되고 있음을 알 수 있다.



[그림 4-2] 빅데이터의 활용이 가능한 정책 예시

이 연구에서는 조사된 빅데이터 원자료를 바탕으로 관련 전문가 인터뷰, 내부 연구진 회의, 외부 자문회의 등 다양한 방법을 통해 기존 빅데이터 활용 정책서비스 외에 [그림 4-2]의 왼쪽 위에 기술한 총 11가지의 새로운 정책 또는 서비스를 도출하였다. 또한, 이들 정책 및 서비스 개발에 필요할 것으로 판단되는 빅데이터를 [그림 4-2]의 오른쪽에 표시하여 활용성이 높을 것으로 예상되는 데이터를 정리하였다.

도출된 11가지의 정책서비스 중에서 비교적 실행 가능성이 큰 7가지는 향후 더욱 심층적인 추가 연구를 위해 관련 빅데이터 활용방안을 다음과 같이 구상하였다.

1) 내비게이션 애플리케이션 기록 기반 승용차 O/D 보완

서울시 통계에 의하면 서울시 통행 중에서 승용차가 차지하는 수단분담률은 2014년 기준 22.8%로, 전체 통행의 5분의 1 이상이 승용차를 이용해 이뤄지는 것으로 나타났다. 이는 서울시의 교통체계에서 승용차가 차지하는 비중이 여전히 높다는 것을 의미하며, 향후 교통정책의 방향 설정을 위해서는 이들 승용차 통행에 대한 세부 현황 파악이 매우 중요할 것으로 생각된다.

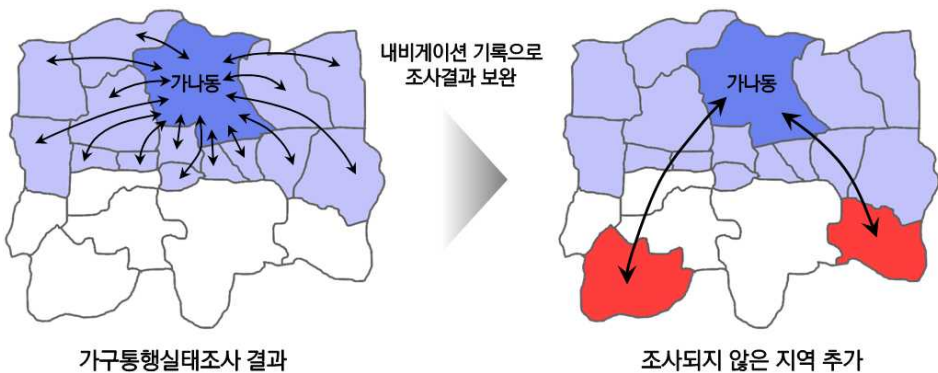
승용차를 제외한 다른 수단들의 통행에 대해서는 다양한 데이터를 이용하여 비교적 세부적으로 현황 분석이 이뤄지고 있다. 우선, 지하철과 버스를 아우르는 대중교통을 이용하는 통행은 교통카드 기록 및 버스운행 기록으로부터 100%에 가까운 통행 표본을 확보하여 사실상의 전수 분석이 가능하다. 또한, 택시에 대해서도 운행 DTG 자료를 토대로 세부적인 통행 행태를 분석할 수 있으며, 이미 심야버스 노선 개발과 같은 실제 정책수립에서도 활용되고 있다.

반면, 도보 및 자전거로 대표되는 기타 통행은 상대적으로 짧은 거리의 통행이라는 점과 수단분담률이 5% 미만이라는 점으로 인해, 그동안 교통정책에서는 크게 다루어지지 못한 것이 사실이다. 여기에는 도보와 자전거가 완전히 개인화된 수단이기 때문에 통행 정보 수집이 쉽지 않다는 점도 작용하였다. 대표적인 개인통행수단인 승용차도 도보 및 자전거와 마찬가지로 통행의 가종점, 통행의 목적 등 통행 관련 정보 수집이 매우 어렵다는 공통점이 있다.

승용차 통행에 대한 정보가 가장 상세하게 수집되는 경로는 5년마다 국가 단위로 이루어지고 있는 가구통행실태조사라고 할 수 있다. 통행의 가종점뿐만 아니라 통행 목적이나 응답자 개인정보도 함께 조사하기 때문에 이를 토대로 다양한 분석이 가능하다. 그러나 이 조사는 소요인력 및 비용이 대규모이고 전체 모집단 크기 대비 표본의 크기가 작아 통계적 대표성에 대한 이슈가 항상 존재하고 있다. 또한, 작은 표본 데이터를 이용하여 승용차 통행 전수 O/D를 만들고 있으므로 조사를 통해서 수집되지 않는 지역의 일부 통행이 누락될 가능성에 대한 우려도 있다.

이러한 문제를 극복하기 위해서는 조사 표본 규모를 최대한 확대해야 하지만 시간과 비용 문제로 쉽지 않으므로 가능하면 별도 자료를 활용하여 보완하는 과정이 필요할 것으로 판단된다. 여기에 사용할 수 있는 자료에는 통신사 접속 기록 등 여러 가지가 있을 수 있으나 최근 많이 이용되고 있는 차량내비게이션 자료도 활용될 수 있을 것으로 보인다.

차량내비게이션 애플리케이션 기록은 3장에서 기술하였듯이 출·도착지 정보와 이동 경로 정보를 실시간으로 수집할 수 있고, 표본의 크기가 매우 커 가구통행실태조사를 대체할 수 있는 잠재력을 지니고 있다. 그러나 내비게이션 기록도 해당 장치 또는 애플리케이션을 이용하는 승용차만을 대상으로 통행 정보를 수집할 수 있어 표본이 모집단의 다양한 특성을 제대로 설명할 수 없다는 단점이 있으므로 단기적으로는 가구통행실태조사를 대체하는 것이 어려울 것으로 판단된다. 따라서 내비게이션 기록은 가구통행실태조사의 표본에서 누락될 수 있는 지역의 O/D를 보완하는 방향으로 활용하는 것이 바람직하다.



[그림 4-3] 내비게이션 기록 기반 승용차 통행 O/D의 보완 예시

2) 내비게이션 기록 기반 도로 위험구간 파악

내비게이션의 경로 안내를 받으며 운전하는 도중에 안내된 경로를 벗어나 내비게이션이 최적 경로를 재탐색하는 경우가 종종 있다. 내비게이션이 안내하는 경로를 이탈하는 이유로는 운전자의 부주의로 인한 실수, 운전자가 고의로 자신이 생각한 경로를 이용하고자 하는 경우, 내비게이션 정보와 실제 도로 환경의 불일치, 도로의 기하 구조 및 시야 불량 등이 있을 수 있다.

첫 번째는 운전자가 내비게이션의 안내에 집중하고 있지 않다가 경로를 벗어나는 경우이다. 라디오를 듣거나 휴대전화를 보는 등 운전 외의 행동을 하는 도중에 내비게이션이 제공한 정보를 인지하지 못하고 잘못된 경로를 택할 수 있다.

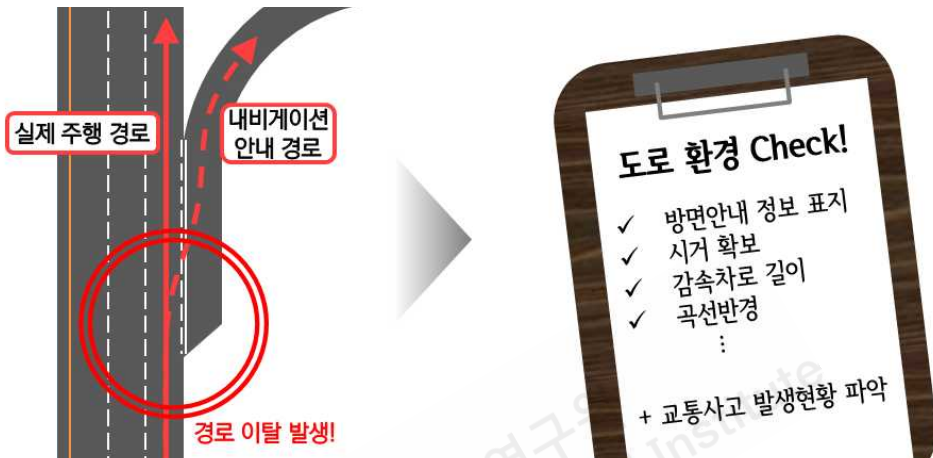
두 번째는 운전자가 해당 경로에 대한 정보를 평소 잘 알고 있어, 본인의 판단에 따라 경로를 선택하는 것이 더 낫다고 생각하는 경우이다. 이러한 운전자는 내비게이션이 안내한 경로보다 오랜 기간 축적된 경험을 토대로 고의로 다른 경로를 이용하며, 내비게이션 이용 목적이 경로 정보보다는 속도위반 단속용 CCTV 위치 파악이나 실시간 소통정보를 얻기 위한 경우가 많다.

마지막 내비게이션 정보와 실제 도로 환경의 불일치, 도로의 기하 구조 및 시야 불량은 자칫 교통사고 유발 등 심각한 문제로 이어질 수 있다. 내비게이션의 지도 정보가 최신 정보로 갱신되지 않아 도로 환경이 바뀐 것을 모르는 채 안내할 때도 있지만, 도로의 기하 구조가 운전자의 혼란을 일으키거나 도로의 각종 시설물이나 지형적 조건이 운전자 시야를 가려 안내된 경로를 벗어날 수 있다. 또한, 실제 도로 환경에서 운전자가 파악 가능한 정보가 불완전하여 안내된 경로를 이탈할 수 있다.

내비게이션의 안내 경로를 벗어나 경로 재탐색이 자주 일어나는 곳은 아마도 도로의 기하 구조가 너무 복잡한 나들목이나 램프 분합류 지점, 지형적 조건 등으로 시야 확보가 쉽지 않은 지점 등이 될 것으로 추정되며, 이러한 지점은 다른 곳보다 교통사고 발생의 위험성이 클 것으로 판단된다.

따라서 내비게이션 기록 분석을 통해 경로 재탐색 횟수나 비율이 높은 지점을 선별하여

잠재적 위험지역으로 설정하고, 교통사고 자료 및 공간정보와 연계 분석하거나 현장 조사 등을 통해 그 원인을 파악하여 대처할 필요가 있다. 만일, 도로 환경이 좋지 못하여 경로 재탐색이 빈번한 경우라면 교통사고 예방을 위해 도로 기하 구조를 개선하거나 시야를 방해하는 도로 시설물을 이전하는 등의 대책을 마련·시행하여야 한다.



[그림 4-4] 내비게이션 기록 기반 도로 위험구간 파악 예시

3) 통신사 기지국 접속 기록 기반 지하철 노선 간 환승 통행 비율 추정

수도권 지역에는 현재 서울지하철 1~9호선, 분당선, 경의·중앙선, 경춘·경강선, 인천지하철 1·2호선, 신분당선, 공항철도, 의정부경전철, 용인경전철 등의 노선이 운영되고 있으며, 개통을 앞둔 우이신설 경전철, 착공 단계인 신림 경전철까지 건설되면 지금보다 더욱 복잡한 도시철도 네트워크가 구축될 것으로 전망된다.

서로 다른 노선들이 겹치는 환승역에서는 통행자들이 쉽게 노선 간 환승을 할 수 있도록 환승 통로나 무빙워크 등의 시설들이 설치되고, 환승 저항을 줄여 이용자들이 다른 노선을 하나의 시스템으로 인식하고 편리하게 이용할 수 있도록 하고 있다. 특히 수도권 대중교통 통합요금제 시행으로 환승 할인이 이루어지고 있으며, 수익금 공동관리를 통해 도시철도 운영사 간 운임 수익금을 정산하고 있다.

도시철도 운영사 간 운임 수익금의 원활한 정산을 위해서는 환승역에서의 노선 간 환승

통행량 추정이 필수적이다. 서울지하철 9호선과 신분당선 등의 민자 노선에서는 역사 내 환승 게이트를 설치하여 환승 통행량을 정확하게 집계하고 있다. 따라서 민자 노선 운영자들은 특정 기간 타 노선에서 해당 노선으로 유입되거나 타 노선으로 유출된 환승 통행자 수를 비교적 쉽게 파악할 수 있다. 하지만 환승 게이트가 설치되지 않은 일반 환승역에서는 노선 간의 승객 이동 규모가 어느 정도 수준인지 확인이 쉽지 않다. 이로 인해 운임 수익금 정산 시 도시철도 운영사 간 법적 분쟁이나 마찰이 발생하는 경우가 종종 있었다.

지금까지는 각 도시철도 운영사가 환승 통로나 역사 내 특정 지점에서 환승 통행량을 직접 계수하거나, 모형을 통해 환승량을 시뮬레이션하여 추정한 후 이를 기반으로 수익금을 정산하고 있다. 하지만 이러한 방법을 통해 추정된 환승 통행량을 검증할 방법이 마땅히 없어 도시철도 운영사 간에 이견이 여전히 발생할 우려가 있다. 일부에서는 환승 통로 등에 CCTV 카메라를 설치하여 환승 통행량을 계수하는 시도도 있었지만, 카메라 설치 위치 및 프라이버시 침해, 환승 여부 판단의 어려움, 과도한 역사 내 이용자 등으로 실행에 많은 어려움을 겪고 있다.

이처럼 환승 게이트가 설치되지 않은 역사에서의 노선 간 환승 통행량 추정은 매우 어렵지만, 지하철 이용자의 통신사 기지국 접속 기록을 활용할 수 있다면 노선 간 환승 비율 추정이 어느 정도 가능할 것으로 보이며, 이 비율은 모형을 사용해 추정한 환승 통행량의 검증에 사용될 수 있을 것으로 판단된다.

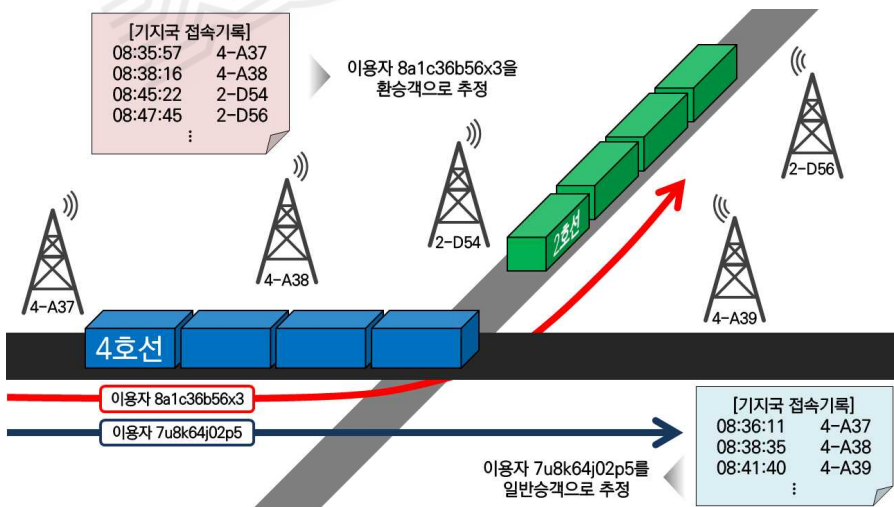
4장에서 기술한 통신사 기지국 접속 기록은 지하철로 대표되는 지하 공간의 이용자에 대해서도 개략적인 위치의 파악이 가능하다. 이동통신사들은 지하철 내에서도 끊임 없는 통신 서비스를 위하여 노선을 따라 일정 간격으로 중계기를 설치하였으며, 지하철에서 통신 기기를 이용하는 사람들은 중계기를 통해 지상의 가장 가까운 기지국과 접속하게 된다. 지상의 기지국은 통상적으로 도시부에서 100~300m 간격으로 설치되어 있어 지하철 이용자의 이동 경로를 개략적으로 파악할 수 있다.

만일 특정 지하철 이용자의 단말기가 노선 1에 인접한 기지국에서 접속되었다가 노선 2에 인접한 기지국과 접속되었다면, 중간의 노선 1과 노선 2가 겹치는 환승역에서 환승을 했을 것이라고 짐작할 수 있다. 이러한 일련의 기지국 접속 기록들을 집계할 수 있다면, 특정

환승역에서의 노선 간 하루 평균 환승 비율, 요일·시간대별 환승 비율 등의 추정이 가능할 것이다. 또한, 통신사 가입정보와 연계하여 성별, 연령대별, 지역별로 구분하여 환승 비율을 추정하는 등 다양한 분석도 가능할 것으로 판단된다.

하지만, 이러한 방식으로는 지상 기지국에는 지하철 이용자뿐만 아니라 인근의 모든 수단 이용자들이 접속하기 때문에 어느 이용자가 지하철을 이용했는지 확인이 어려운 한계가 있다. 다른 수단의 이용자가 혼재된 자료를 여과 없이 활용하게 되면 큰 오차가 발생할 우려가 크기 때문에, 일련의 기지국 접속 기록을 토대로 이동 경로와 이동 속도, 도로 및 도시철도 네트워크 등을 고려하여 승용차나 버스 등 공로수단을 이용하는 통행자들을 우선 선별해 내는 과정이 필요할 것으로 생각된다.

또한, 통신사의 기지국 접속 기록이 지하철 승객 모두를 대상으로 수집되지는 않기 때문에 모집단 전체에 대한 분석은 어려울 것으로 생각된다. 그러나 모집단 전체의 환승 통행량 추정은 어렵더라도 노선 간 환승 통행 비율 추정을 통해 그동안 정확히 알려지지 않았던 지하철 승객들의 이동 경로나 통행 행태를 어느 정도 유추해 볼 수 있을 것이다. 지하철 운영사는 이를 토대로 수익금 정산뿐만 아니라 환승시설 개선이 우선 필요한 역을 선정하거나, 특정 지역 간 통행이 주로 어떤 환승역을 거쳐서 이루어지는지 등도 파악해 볼 수 있을 것으로 기대된다.



[그림 4-5] 지하철 노선 간 환승 통행 비율 추정 예시

4) 포털지도 길 찾기 안내정보 기반 대중교통 거점 보행 접근 거리 추정

승용차를 줄이고 대중교통의 이용 촉진을 위해서는 버스정류장이나 지하철역과 같은 대중교통 거점까지의 접근이 편리해야 한다. 대중교통 거점까지의 접근 편리성은 일반적으로 이용 가능한 접근수단, 접근 거리, 접근시간 등의 지표를 이용하여 평가할 수 있다.

하지만 이러한 지표는 현장을 방문하여 직접 조사하지 않으면 정확하게 측정할 방법이 많지 않다. 현장을 방문하더라도 특정 지점 간 조사가 가능할 뿐, 대중교통 거점 주변의 수많은 임의의 지점에서 거점까지의 평균 지표 값은 얻기 힘들다.

교통카드 기록으로는 카드를 태그하는 버스정류장 또는 지하철역 간 통행을 세부적으로 분석할 수 있으나, 가정에서 출발하여 지하철역에 도달하거나, 버스정류장에 내려서 학교로 가는 것과 같은 통행의 말단에 대한 정보는 확인할 수 없다.

5년 주기로 시행되는 전국단위 가구통행실태조사에서는 대중교통 거점까지의 접근수단과 통행시간을 묻는 문항이 포함되어 일부 조사가 이루어지고 있으나, 표본의 수가 전체 모집단 대비 너무 적어 분석에 어려움이 있다. 교통카드 기록과 가구통행실태조사의 대중교통 접근 통행의 조사결과를 연계하여 분석하면 대중교통 말단에서의 통행 행태를 어느 정도 유추해 볼 수 있겠으나, 대중교통 접근 편리성을 지역적으로 정교하게 파악하는 데는 한계가 있다.

하지만 대부분 포털 사이트에서 제공하는 지도 서비스의 길 찾기 기능을 활용하면 임의의 지점에서 대중교통 거점까지의 접근수단, 접근시간 등 접근 통행에 대한 세부적인 정보를 얻을 수 있다. 지도 서비스의 길 찾기에서는 이용자가 원하는 출발지와 도착지의 정보를 입력하면 [그림 4-6]과 같이 이용 가능한 수단별로 최적의 경로를 안내해 주며 수단별 이동시간, 이동거리, 환승 횟수 등도 함께 보여준다. 특히, 대중교통 이용 안내 경로에는 가장 가까운 대중교통 거점이 어디이고, 여기까지의 예상 도보 이동시간이 얼마나 되는지, 목적지 인근의 대중교통 거점에서 최종 목적지까지의 도보 이동시간이 얼마나 되는지 등에 대한 정보도 함께 제공한다.



[그림 4-6] 포털지도 길 찾기 안내 시 제공되는 도보 시간 관련 정보 예시

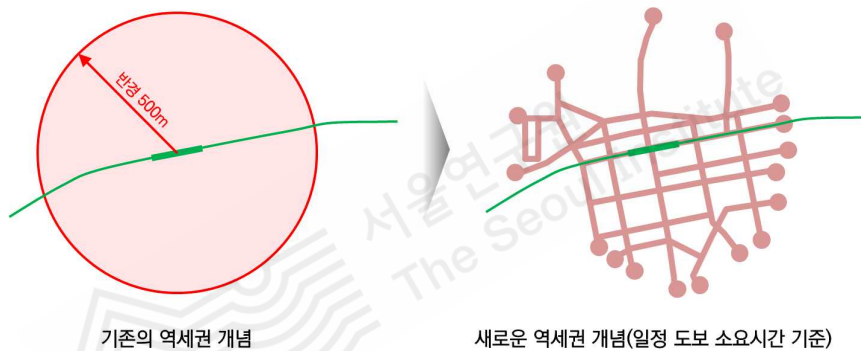
포털지도 서비스에서 제공되는 도보 이동거리 및 이동시간은 실제 보행 네트워크 기반으로 산출되므로 비교적 정확한 값이라고 말할 수 있다. 또한 포털 회원과 비회원에 상관없이 무료로 매일 많은 사람이 포털지도 서비스의 길 찾기 기능을 사용하고 있어, 그동안 포털운영업체에 쌓인 검색기록의 양은 방대할 것으로 생각된다. 이러한 길 찾기 검색기록 빅데이터를 분석하면 지역별·대중교통 거점별로 접근 통행에 대한 평균 접근 거리 등을 산출할 수 있으며, 이 외에도 대중교통 소외지역 등 다양한 정보를 얻을 수 있을 것으로 판단된다. 이렇게 빅데이터 분석을 통해 얻은 정보는 향후 대중교통 접근 개선, 소외지역 해소 등 관련 정책수립 및 시행에 유용하게 활용할 수 있을 것이다.



[그림 4-7] 다양한 기준에 따른 도보 접근 거리 산출 예시

또한, 포털 서비스 길 찾기 검색기록은 [그림 4-8]과 같이 도보 이동시간을 기준으로 새로운 개념의 역세권을 정의하는 데 사용할 수도 있을 것으로 보인다. 통상적으로 지하철 역세권은 역의 중심이나 출구를 기준으로 직선반경 500m 지역 등 거리를 기준으로 정의하고 있지만, 통행자로서는 역으로부터 일정 반경 내에 있다 하더라도 실제 도보 이동시간이 지나치게 오래 걸린다면, 역세권에 거주하고 있다는 느낌을 받기 어려울 수도 있다.

그러므로 역세권을 단순 직선거리로 정의하지 않고 실제 도보 이동시간을 기준으로 정의한다면 대중교통 거점까지의 보행환경 개선 등 더욱 현실적인 정책을 마련하여 시행할 수 있고, 더불어 부동산 정보 파악이나 상권 분석 등에서 더 잘 활용될 수 있을 것으로 판단된다.



[그림 4-8] 새로운 역세권 개념 정의 예시

5) 교통약자의 심야 안전 개선을 위한 지점 선정

교통약자에는 주로 어린이, 임산부, 노약자, 장애인 등이 포함되며, 신체적 기능이 저하되어 있거나 상황 판단력 등의 인지 능력이 상대적으로 낮아 통행에 불편을 느끼거나 사고 노출 위험도가 높은 집단도 포함된다. 우리나라에서는 2005년에 「교통약자의 이동편의증진법」이 제정되어 이듬해부터 시행되었으며, 보도 위 점자블록 설치, 촉지안내도 설치, 저상버스 도입, 장애인 콜택시 운영 등 다양한 정책들이 도입되어 시행되고 있다.

이러한 정책 시행에도 불구하고 교통약자들이 체감하고 있는 통행 불편은 여전히 존재하며, 특히 야간에는 주간보다 시야가 제한적이기 때문에 이들의 통행 불편은 더욱 커지고,

심지어는 범죄에까지 노출되어 안전이 위협을 받는다. 이에 서울시에서는 심야시간대 시민들의 안전한 귀가를 지원하기 위해 다양한 정책들을 마련하여 시행하고 있다.

여성 안심 귀가 도우미 서비스는 전화나 애플리케이션을 이용하여 도착 30분 전까지 신청하면 버스정류장이나 지하철역으로부터 자택까지 안심 귀가 도우미가 동행함으로써 귀갓길에 발생할지도 모르는 범죄에 사전 대응하는 정책이다.

안심귀가 마을버스 서비스는 늦은 밤 마을버스를 이용하는 시민이 정류장이 설치되어 있지 않더라도 원하는 장소에서 하차할 수 있는 서비스이다. 향후 시민의 요구사항을 세부적으로 파악하고 정교한 이용수요 예측을 통해 재시도해 볼 수 있는 서비스이다. NFC(근거리무선통신)를 이용해 탑승한 택시 정보를 미리 보호자에게 전송할 수 있도록 한 택시 안심귀가서비스는 이용률이 저조하여 종료되기는 했지만, 교통약자의 안전을 위한 서비스 중 하나로 시민들에게 제공되었다.



〈안심귀가 마을버스〉



〈여성 안심 귀가 도우미〉

출처: 구글 이미지 검색(<http://image.google.com>)

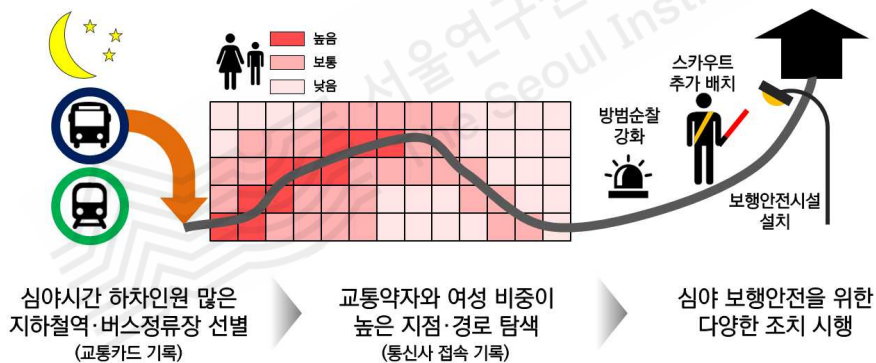
〈그림 4-9〉 심야시간대 안전한 귀가를 위한 정책 예시

여성 안심 귀가 도우미 서비스는 많은 인력이 필수적으로 투입되어야 하는데, 한정된 재원으로 모든 지하철역이나 동네에서 이와 같은 밀착형 서비스를 제공하는 것은 분명히 한계가 있다. 따라서 이러한 서비스가 꼭 필요한 지역을 선별하여 인력을 효과적으로 투입하는 것이 필수적이다. 현재는 혼자 사는 여성의 밀도가 높은 원룸 밀집 지역 등을 중심으로 도우미가 배치되어 운영되고 있는데, 향후 서비스를 확대 시행할 때에는 심야시간대에 여

성 등 교통약자의 통행이 많이 이루어지면서 안전이 취약한 지역을 우선 선별하여 귀가 도우미를 배치하거나 야간 안전시설물 추가 설치 등의 정책을 시행해야 할 필요가 있다.

이러한 지역 및 교통약자 이동 경로 선별에 있어 교통카드 기록과 통신사 기지국 접속 기록 등 빅데이터를 충분히 활용할 수 있을 것으로 판단된다.

우선 교통카드 기록을 이용하여 심야시간대에 하차 인원수가 많은 지하철역과 버스정류장을 선별한다. 그 다음으로 통신사 기지국 접속 기록을 활용해 분석된 통행 인구를 이용하여 가로별 유동인구를 산출하고, 성별 및 연령대별 통행 비율을 구하여 교통약자로 분류될 수 있는 집단과 여성 비중이 높은 지역 및 이동 경로를 탐색한다. 이 결과를 토대로 심야 보행 안전이 취약한 지역에 안심 귀가 지원 스카우트를 추가 배치하고, 교통약자의 이동 경로상에 가로등이나 CCTV 설치 확대, 자율방범대와 경찰의 순찰 강화 등 맞춤형 대응 방안을 마련·시행할 필요가 있다.



[그림 4-10] 심야 보행 안전 지원 과정 예시

6) 위험물 수송차량 통행빈도에 따른 위험요소 식별

서울시에는 많은 주유소와 LPG 충전소가 있으며, 도시가스가 공급되지 않는 지역에서는 석유와 가정용 LPG를 난방 연료로 사용하고 있다. 이를 수송하기 위해서는 탱크로리와 같은 위험물 수송차량이 불가피하게 서울 시내를 운행하게 되며, 이들 차량에 예기치 않은 사고가 발생하게 되면 대형 인적·물적 피해가 발생할 수 있다.

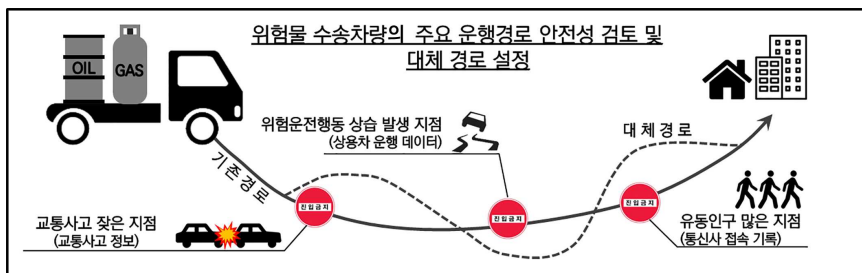
이러한 사고의 심각성 때문에 현재도 위험물 및 독극물 수송차량은 상수도 보호구역이나

어린이보호구역과 같은 일부 지역에서 통행을 제한하고 있다. 또, 특정 시간대에는 도심으로의 진입이나 교량과 터널의 통행을 제한함으로써 위험물 수송차량의 사고 발생 시 그 피해를 최소화함과 동시에 시민의 안전을 지키고자 하는 정책을 시행하고 있다.

하지만 특정 지역과 시간대에 대해 위험물 수송차량의 운행을 제한하는 것만으로는 대형 사고의 위험성을 줄이고 시민의 안전을 담보하는 데 여전히 부족하다. 특히 연료 수송차량처럼 전면적인 통행제한이 어려운 생활 밀착형 위험물 수송차량 운행의 상시 모니터링을 강화하고 이 차량들이 주로 운행하는 지역이나 경로를 파악하여 사고의 위험요소를 사전 제거하는 등 사고 발생 시 피해를 최소화하는 방안을 마련하여 시행할 필요가 있다.

이를 위해서는 우선 서울 시내에서 위험물 수송차량과 연관된 교통사고가 자주 발생하는 지점과 이들 차량이 주로 이용하는 경로, 운행 중 급가속·급정거·급회전 등 위험운전행동이 자주 일어나는 지점을 파악할 필요가 있다. 이러한 지점은 교통사고 자료와 상용차 운행 데이터, 자동차 등록원부 등 관련 빅데이터 분석을 통해 충분히 파악 가능할 것으로 판단되며, 파악된 지점에 대해서는 현장 조사 등을 통해 원인을 세부적으로 살펴 사고의 위험을 미리 제거하여야 한다.

또한, 상용차 운행 데이터로부터 위험물 수송차량이 자주 운행하는 지역이나 도로 주변에 유동인구가 많은 경우에는 이 지역을 우회할 수 있는 경로를 적극적으로 안내해 주거나 유동인구가 많은 시간대에 통행을 제한할 수도 있을 것이다. 유동인구가 많은 지역에서의 위험물 수송차량 교통사고는 자칫 대형 인명피해로 이어질 가능성이 크기 때문이다. 이때 위험물 수송차량의 운행경로 주변의 시간대별 유동인구 파악은 통신사 기지국 접속 기록을 활용해 가능할 것이다.



[그림 4-11] 위험물 수송차량 통행 관리 예시

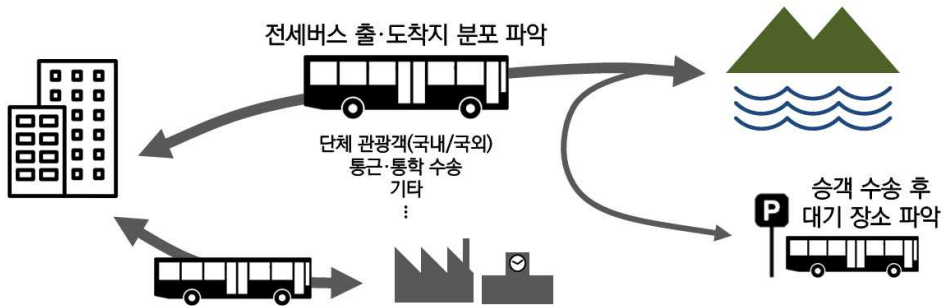
7) 전세버스 통행 O/D 파악 및 통행패턴 분석

고정된 시간표와 출발지·도착지·경유지에 따라 운행되는 정기 노선버스와 달리, 전세버스는 이용을 원하는 고객과의 계약을 통해 출발지·경유지·목적지와 통행 시간대가 결정되어 승객을 운송하게 된다. 전세버스는 주로 국내 여행객이나 국제공항을 통해 입국하여 다시 출국할 때까지의 외국인 관광객을 상대로 운송서비스를 제공한다. 또한, 결혼식장례식 등 경조사 참석을 위한 단체 승객을 운송하거나 공공기관·기업·대학 등과의 계약을 통해 통근자나 학생들을 정기적으로 운송하기도 한다. 이처럼 전세버스는 여행객과 같이 일시적인 통행 수요뿐만 아니라 통근·통학생과 같이 정기적인 통행 수요에 대응하는 교통수단으로써의 역할을 하고 있다.

하지만 통행의 적지 않은 부분을 담당하고 있는 전세버스의 출·도착지와 이들이 주로 이용하는 이동 경로, 주·정차지, 이용자의 통행 특성 등에 대해서는 지금까지 세부적으로 알려진 것이 많지 않다. 이러한 정보는 전세버스의 주정차 및 운행 관리, 관광객 편의 증진, 관광상품 개발을 위한 수요분석 등 관련 정책에 활용될 수 있다.

5년마다 시행되는 가구통행실태조사에서 전세버스와 외국인 관광객의 통행실태에 대한 조사가 일부 이루어지고 있지만, 표본 수가 적어 세부적인 파악에는 한계가 있다. 이러한 한계는 위험물 수송차량과 같이 상용차 운행 데이터와 자동차 등록원부의 빅데이터 연계 분석을 통해 다소 극복될 수 있을 것으로 판단된다.

아울러 전세버스들의 출·도착지 분포와 지역 특성으로부터 전세버스를 이용하는 통행자의 특성도 분석해 볼 수 있다. 대부분 전세버스가 앞에서 기술한 목적을 위하여 활용되는 만큼 지역 간 통행 비중이 상당히 높을 것으로 생각되며, 차량의 위치정보로부터 해당 지역이 어떤 특성을 가지는지 공간정보와 연계하여 추정해 볼 수 있을 것이다. 예를 들어, 서울 시내에서 경기도 외곽의 산업단지로 통행하는 전세버스라면 이용자들은 통근자일 가능성이 크고, 도착지가 국립공원이나 항구, 유원지 등과 가깝다면 이용자들은 관광객일 가능성이 클 것이다. 이처럼 전세버스 이용자의 통행 특성을 파악함으로써, 대중교통이나 승용차를 이용하지 않는 통근·통학 통행이나 관광객의 통행 행태를 유추해볼 수 있다.

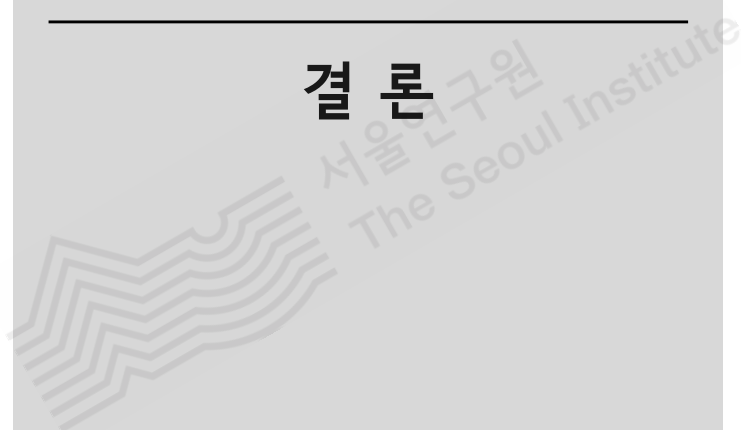


[그림 4-12] 전세버스 통행 O/D 파악 및 행태 분석 예시

또한 상용차 운행 데이터를 분석하면, 전세버스 이용자들이 통행 목적을 달성할 때까지 대기하거나 다음 운행까지 휴식을 취할 때 전세버스가 주로 어디에 주차하고 있는지도 파악할 수 있다. 이는 대기 지점에서의 전세버스 주차가 적절한지를 확인하거나 주차장 설치 필요성 여부를 검토하는 데 활용할 수 있으며, 휴게시설 설치 등 전세버스 기사들의 편의 증진과 불법주차 단속 및 계도, 주차요금 정책수립 등에 활용할 수 있을 것으로 기대된다.

05

결 론



05 | 결론

이 연구에서는 교통정책 수립과 연관성이 있는 빅데이터의 현황을 원자료 수준에서 세부적으로 파악하고 그 활용 가능성을 검토하였다. 빅데이터는 통행의 전 과정과 여러 장소에서의 다양한 활동을 통해서 생성되고 있으며, 각 교통수단의 운행 및 교통운영 과정에서도 끊임없이 생성되고 있는 것으로 나타났다.

특히, 이 연구에서는 조사된 빅데이터를 토대로 총 11가지의 교통정책 또는 서비스를 제시하였고, 그 중 7가지에 대해서는 향후 심층적인 추가 연구를 위해 세부적인 활용 구상안을 기술하였다. 이 연구를 통해 얻은 결론은 다음과 같다.

1) 교통정책 수립에 있어 빅데이터 활용 가능성은 매우 커

빅데이터를 교통계획, 교통운영, 교통안전, 대중교통 등 분야별 교통정책 수립에 활용함으로써 정책의 질을 획기적으로 높이고, 기존 정책의 시행 효과를 분석하거나 새로운 정책 또는 서비스를 발굴하는 데 많은 도움이 될 수 있을 것으로 판단된다.

또한, 교통 정책수립에 필요한 각종 조사를 빅데이터로 대체하거나 보완하여 정책수립에 드는 비용을 절감할 수 있고, 빅데이터 활용을 통해 정책 실무자들이 수행하고 있는 업무 중 일부에 대한 부담을 덜어줄 수 있을 것으로 기대된다.

2) 빅데이터 확보 및 체계적 관리를 위한 노력이 계속되어야

이 연구에서 살펴본 바와 같이 교통정책 수립에 활용 가능성이 있는 빅데이터는 공공기관 뿐만 아니라 민간기관에서도 다양한 형태로 수집·가공·보관되고 있으나, 개인정보 보호, 영업비밀, 데이터의 불완전성 등 여러 가지 법적/제도적 문제와 정보에 대한 인식의 차이로 정책수립에 빅데이터를 활용하기가 매우 어려운 것이 현실이다.

서울시도 빅데이터의 중요성을 인식하고 관련 조직 정비, 민간기관과의 제휴 확대, 빅데이터캠퍼스 설립, 중앙정부 법/제도 개선 건의 등 많은 노력을 해오고 있으며, 실제 빅데이

터를 활용한 다양한 정책수립 지원 사례를 제시하고 있다. 이러한 노력에도 불구하고 다양한 이유로 인해 기대만큼 활용되지 못하고 있는 빅데이터의 활용성을 높이기 위한 작업이 필요하고 더 많은 민간 빅데이터에 대한 확보 노력이 계속되어야 한다.

또한, 그동안 개인정보 침해에 대한 우려로 제대로 활용되지 못하고 있는 빅데이터의 활용을 위해 민간공공의 보유기관과 업무 제휴 및 협력을 강화할 필요가 있다. 각 기관에서 수집되고 있는 빅데이터의 특성을 파악한 후, 정책수립 및 시행에 도움이 될 것으로 기대되는 데이터에 대해서는 적극적인 제공 또는 공유 요청이 필요할 것으로 보인다. 아울러, 빅데이터의 활용에 있어 법적/행정적 제약사항들을 완화하기 위해 중앙정부에 대한 건의와 관계기관의 공감대 형성을 통해 관련법과 제도의 개선이 이루어질 수 있도록 노력할 필요가 있다.

이와 함께 서울시에서 자체적으로 보유하고 있는 각종 데이터의 통합과 관리가 체계적으로 이루어질 수 있는 시스템을 마련하여 정책수립 시 정책입안자 및 관련 연구기관에서 쉽게 빅데이터에 접근하여 다방면으로 활용할 수 있도록 하여야 한다. 나아가 빅데이터를 활용한 새로운 가치 창출에 있어 민간의 창의성 및 효율성이 접목될 수 있는 환경도 조성할 필요가 있다.

3) 이 연구의 빅데이터-교통정책 연계 활용 구상안 구체화 필요

이 연구에서는 조사된 빅데이터의 원자료 및 가공자료를 기반으로 서울시의 교통정책과 연계 가능성이 있는 여러 가지 빅데이터 활용 구상안을 제시하였다. 그러나 교통계획, 교통운영, 교통안전, 대중교통 등 다양한 부문에 대하여 제시된 활용 구상안은 실제 빅데이터를 확보하여 구체화한 후, 이를 관련 교통정책 수립에 얼마나 활용할 수 있고 그 효과가 어느 정도 발생할지에 대한 추가 연구를 통한 검증이 필요하다.

이를 위해서는 전체 모집단의 데이터셋이 아니더라도 서울시와 민간-공공의 빅데이터 보유 기관이 서로 협의협력하여 연구용 표본 데이터라도 공유하고, 이를 기반으로 선행 연구를 진행하여 빅데이터 활용의 실효성이 있는 것으로 판단된다면 실제 빅데이터를 관련 정책 수립에 활용할 수 있을 것이다.

4) 미래 빅데이터의 체계적인 수집·관리를 위한 선제 대응 필요

현재 빅데이터의 적극적인 활용을 저해하는 가장 큰 원인은 개인정보 유출에 대한 우려이지만, 이 외에도 데이터 자체의 완결성이 떨어지거나 여러 기관에 산재되어 있어 데이터 간 연계가 되어 있지 않은 것도 큰 이유이다. 이러한 문제는 민간주도로 해결이 어려울 것으로 판단되므로 중앙정부 등 공공기관이 나서 데이터의 수집에서부터 제공에 이르기까지 체계적인 관리가 가능하도록 법/제도 개선 및 시스템 마련이 필요해 보인다.

현재까지 공공 및 민간 부문에서 축적되고 있는 빅데이터의 수집·제공에 있어 드러난 이슈와 문제점을 거슬삼아, 앞으로 다양한 부문에서 생성·축적될 것으로 보이는 교통 관련 빅데이터의 체계적인 수집 및 관리 시스템에 관한 연구를 지속 수행하여 다가오는 빅데이터 시대에 대비할 수 있어야 한다.

미래교통 분야의 빅데이터는 현재 급격히 발전하고 있는 자율주행차량 및 관련 인프라와 사물인터넷으로부터 대량으로 생성될 것으로 보인다. 특히, 자율주행차량은 수많은 센서가 장착되어 주변의 차량이나 보행자, 시설물 등을 감지하고 이에 기초한 주행 전략을 실시간으로 수립하여 차량을 움직인다. 또한, 도로상에 설치된 통신장비와 실시간으로 차량 운행상태, 도로·교통 상태에 관련된 정보를 교환하면서 차량-도로 간 자율협력주행이 이루어질 것으로 생각된다. 이 과정에서 차량운행 및 도로·교통에 대한 상태정보가 대량으로 센터에 수집되고 축적될 것이다.

이러한 대규모 빅데이터의 효과적인 활용을 위해서는 자료수집체계, 가공 및 관리방안, 데이터 간 연계처리 및 통합방안에 관한 사전연구가 활발히 이루어져야 하고, 빅데이터를 기반으로 한 새로운 정책 및 서비스의 발굴을 계속해 나갈 필요가 있다.

이를 위해 서울시에서도 이미 보유하고 있는 교통 관련 빅데이터의 연계 및 통합에 대한 노력을 계속하여야 할 것이다. 뿐만 아니라 미래의 대규모 빅데이터 시대에 적극적으로 대응할 수 있도록 현재 운영 중인 교통정보센터의 위상을 높여, 여러 곳에 분산되어 있는 교통 부문 빅데이터 관련 업무를 총괄하도록 하는 등 그 기능을 크게 강화할 필요가 있다.

참고문헌

- 경찰청, 2016, 「2015년 경찰통계연보」.
- 김경태, 2014, 「SKT 통신인구 자료를 이용한 오디 산정」.
- 김종학·고용석·김준기, 2015, 「모바일 빅데이터의 국토교통 분야 활용 및 시사점」, 국토정책 Brief, 499, 국토연구원.
- 서울특별시, 2015, 「택시운행분석 데이터셋 활용가이드」.
- 서울특별시, 2017a, 「서울시 빅데이터 분석 사례집: 1. 빅데이터 기반 심야버스 노선수립을 위한 분석」.
- 서울특별시, 2017b, 「서울시 빅데이터 분석 사례집: 2. 택시 운행 데이터 분석」.
- 서울특별시, 2017c, 「서울시 빅데이터 분석 사례집: 3. 장애인콜택시 운영 분석」.
- 윤서연·이춘용·김호정·육동형·김상록, 2015, 「지역 간 교통수요예측의 신뢰성 제고를 위한 빅데이터 활용방안 연구」, 국토연구원.
- 이석주·연지운·천승훈, 2013, 「빅데이터를 이용한 교통정책 개발 및 활용성 증대방안」, 한국교통연구원.
- 카카오, 2017, 「택시는 미래 교통의 중심이다」, 한국교통연구원 교통빅데이터포럼 발표자료.
- 한국도로공사, 2016, 「2015 고속도로 교통량 통계」.
- 한국은행, 2016, 「지급수단 이용행태 조사결과 및 시사점(2016년)」.
- EMC, 2014, *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of the Things*, IDC.
- Gao, Song., Yang, Jiue-An., Yan, Bo., Hu, Yingjie., Janowicz, Krzysztof., McKenzie, Grant., 2014, "Detecting origin-destination mobility flows from geotagged Tweets in greater Los Angeles area", Presented at the 8th International Conference on Geographic Information Science (GIScience'14), Sep 24~26, Vienna, Austria.
- McKinsey Global Institute, 2011, *Big data: The next frontier for innovation, competition, and productivity*, McKinsey & Company.
- META Group, 2001, *3D Data Management: Controlling Data Volume, Velocity, and Variety*, META Group Inc.
- Resource Systems Group, 2015, "Utilizing Big Data Methods for a Long-Distance Automobile

Origin-Destination Study”, Presented by Polly Ramsey at the 15th TRB National Transportation Planning Applications Conference, May 17~21, Atlantic City, New Jersey.

Ryan Kinsky, 2014, “AirSage: Population Analytics”, Presented at the 60th Annual Traffic and Safety Conference, May 13~15, Columbia, Missouri.

Transport for London, 2015, “Innovations in London’s transport: Big Data for a better customer experience”, Presented at European Data Forum 2015, Nov 16~17, Luxemburg.

Tuck Yew, Lui, 2014, “Singapore Urban Mobility”, Presented at the International Transport Forum Summit 2014, May 21~23, Leipzig, Germany.

[http://bus.go.kr/\(서울대중교통\)](http://bus.go.kr/(서울대중교통))

[http://data.seoul.go.kr/\(서울_열린데이터_광장\)](http://data.seoul.go.kr/(서울_열린데이터_광장))

[http://golmok.seoul.go.kr/\(서울특별시_우리마을가게_상권분석서비스\)](http://golmok.seoul.go.kr/(서울특별시_우리마을가게_상권분석서비스))

[http://stat.seoul.go.kr/\(서울통계\)](http://stat.seoul.go.kr/(서울통계))

[http://taas.koroad.or.kr/\(도로교통공단_교통사고분석시스템\)](http://taas.koroad.or.kr/(도로교통공단_교통사고분석시스템))

[http://tfl.gov.uk/\(Transport for London\)](http://tfl.gov.uk/(Transport_for_London))

[http://www.airsage.com/\(AirSage_홈페이지\)](http://www.airsage.com/(AirSage_홈페이지))

[http://www.data.go.kr/\(공공데이터_포털\)](http://www.data.go.kr/(공공데이터_포털))

[http://www.seoul.go.kr/\(서울특별시\)](http://www.seoul.go.kr/(서울특별시))

[https://www.mot.gov.sg/\(Singapore Ministry of Transportation\)](https://www.mot.gov.sg/(Singapore_Ministry_of_Transportation))

[https://www.si.re.kr\(서울연구원_인포그래픽스\)](https://www.si.re.kr(서울연구원_인포그래픽스))

[https://www.tmap.co.kr/\(SK텔레콤_티맵_홈페이지\)](https://www.tmap.co.kr/(SK텔레콤_티맵_홈페이지))

Abstract

A Study on Application of Big Data for Transportation Policy Development and Improvement

Gyeong-Sang Yoo · Chang Lee · Gyu-geun Yoon

Big data is defined as dataset that is so voluminous that researchers can not collect, store, manage, or analyze such data with normal database software. Because of the large number of transportation system users, potential benefits of collecting and distributing real-time transportation data, and increasing demand for processing and analyzing new kinds of data from cutting-edge technologies such as autonomous vehicles, it is expected that transportation sector can achieve positive changes by applying big data analysis.

In this study, we investigated the status of big data relative to the transportation and examined their applicability to develop or improve transportation policies and services. Moreover, we suggested new policies and services that can be derived from our study and sorted them into transportation planning, safety, public transportation, and operation. To promote future research, we described the mechanism of each policy and service at the conceptual level.

Applying big data to transportation system will reduce data collection cost and time and enhance social benefits. Although there are several barriers to using big data analysis, professionals should overcome these obstacles by establishing strategic partnerships with organizations retaining data, proposing amendments to current laws to the national government and the National Assembly, systemically managing data presently owned and sharing data with research groups.

Contents

01 The Era of Big Data Is Emerging

- 1_Big Data Has Come to Us
- 2_Transportation Sector Also Cannot Avoid the Paradigm Shift
- 3_The Current State of Big Data Analysis and Its Adaptability to the Transportation Sector Should Be Understood

02 Implications of Big Data for the Transportation Sector

- 1_Foreign Implication Cases
- 2_Domestic Implication Cases
- 3_Expecting Effective Application in Transportation Sector with Large Populations

03 Big Data That Can Be Applied to the Transportation Sector

- 1_Investigating and Surmising Attributes of Big Data
- 2_Each Data Has Various Ranges of Availability

04 Several Proposals for Transportation Policies That Could Possibly Be Developed with Big Data Implication

- 1_Navigating Approaches to Implicating Big Data Analysis in the Practical Affairs of the Seoul Metropolitan Government
- 2_Convergence with Spatial Data May Be Necessary
- 3_Some Examples of Transportation Policies Derived from Big Data

05 Conclusion

서울연 2016-BR-25

빅데이터와
교통정책의
연계 방향

발행인 _ 서왕진

발행일 _ 2017년 5월 15일

발행처 _ 서울연구원

ISBN 979-11-5700-225-2 93530 8,000원

06756 서울특별시 서초구 남부순환로 340길 57

본 출판물의 판권은 서울연구원에 속합니다.