

# 빅데이터와 교통부문 활용사례

빅데이터(Big Data)란 기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합<sup>1)</sup> 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술<sup>2)</sup>을 의미한다. 각 분야에서 수집된 대량의 자료를 통해 맞춤형 정보를 생성할 수 있다는 기대와 함께 개인정보 보호에 대한 우려 등 관심이 집중되어 있다.

2016년 30호 DB Trend에서는 교통부문에서 빅데이터 활용사례를 소개하고자 한다.

자료: 1) James Manyika & Michael Chui, << Big data: The next frontier for innovation, competition, and productivity >>, McKinsey Global Institute, (2011년 5월), P.1

2) John Gantz & David Reinsel, << Extracting Value from Chaos >>, IDC IVIEW June, (2011년), p.6

\* 출처: 위키백과

## 빅데이터 현황

- 데이터의 규모만큼이나 활용분야 및 사례가 다양하며 제조업 및 유통업, 의학, 마케팅 등 분야에서 영역별로 고유의 특성을 나타내는 분석을 수행하고 있으며 공통적으로 데이터를 분석하고 패턴을 통해 의미있는 결과를 도출하고 있다.
- 현재 교통부문에서 주로 내비게이션 교통정보, 교통카드 데이터, 차량운행기록(DTG), 누적 속도정보, TCS 등 인프라 데이터, 통신사 기지국 트래픽 DB 등의 자료를 활용하고 있으며 외부 DB 연계 및 기관 간 협력을 통해 그 기반을 다지고 있다.
- 교통부문 빅데이터 활용사례는 한국지능형교통체계협회에서 개최한 「교통 빅데이터 분석 및 활용」 교육자료에 기초하여, 빅데이터 형태별로 활용사례를 구분, 정리하여 재구성하였다.
- KTDB 빅데이터 활용사례는 실적자료가 부재한 승용차 부문에서 내비게이션 자료를 활용하여 승용차 O/D 신뢰도를 제고하고자 한국교통연구원 국가교통DB사업단에서 수행한 「여객교통수요 신뢰도 개선방안 연구」와 「교통혼잡지도 DB구축」 사업 내용을 요약하였다.

## 교통부문 빅데이터 활용사례

### ① 내비게이션 교통정보 제공

- 택시, 배송차량, 버스 등에서 수집한 소통 정보를 가공하여 교통정보 제공(교통상황, 길안내, 도착예정시간 등)

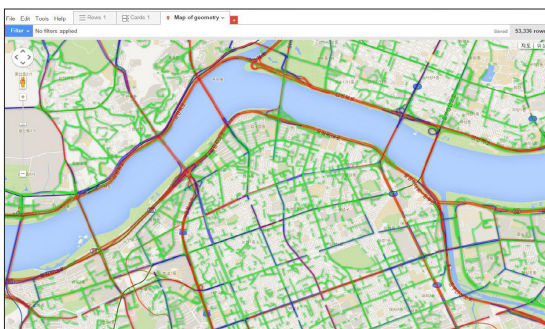


출처: (주)빅스타, 빅데이터 활용 및 플랫폼 분석 기술 이해, 2016년

실시간  
교통정보

### ② 내비게이션 교통정보와 빅데이터 결합

- 전국 실시간 교통량 정보 생성
- 내비게이션 사용 교통량과 통신사 기지국 정보 활용을 통한 상권분석, 입지분석
- 교통링크와 기상관측소 정보 매칭을 통한 교통정보 제공
  - 실시간 폭우·폭설 지역에 대해 교통정보 반영
- 내비게이션 소통정보와 유고정보 연계제공

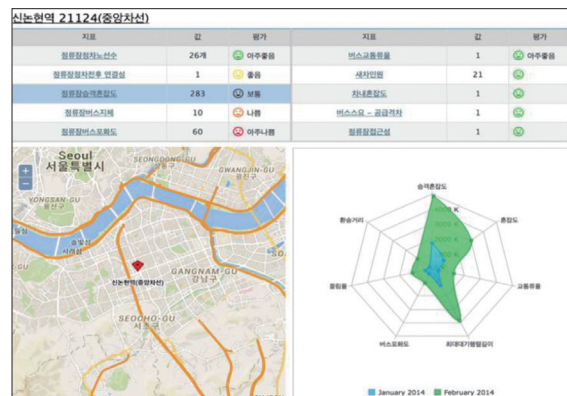


출처: SK 플레닛, Tmap 교통정보 현황 및 Big Data 활용방안, 2016년

실시간  
교통정보

### ③ 교통카드 데이터 분석

- 교통카드 사용정보 및 노선정보, 승하차 정보를 이용하여 출/퇴근 시간대별 이용현황 및 고도화 과정을 통해 정류장별 승객 혼잡도 등 분석

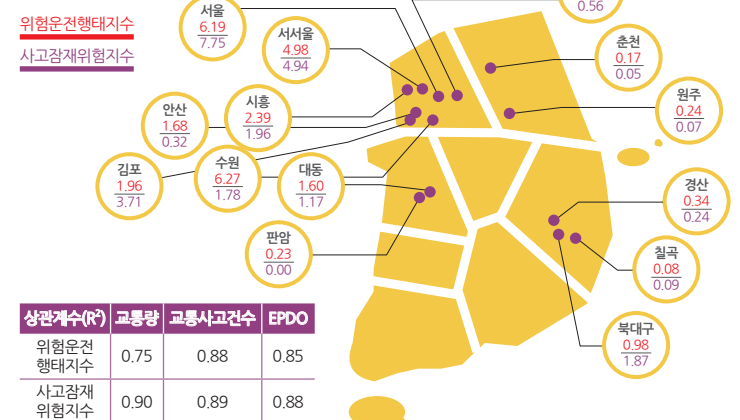


출처: 한국지능형교통체계협회, Splunk를 활용한 Big Data의 이해 및 활용, 2016년

정류장별  
승객 혼잡도

### ④ 차량운행기록(DTG) 데이터 기반 연구

- 위험운전 행태지수 및 잠재 위험지수 개발



출처: 한국도로공사 도로교통연구원, 교통분야 산업화 추진전략 및 계획, 2016년

\* 위험운전 행태지수: 운전자의 위험도를 정규화하여 영업소 간 위험도 비교할 수 있는 지수

\* 잠재 위험지수: 이용 교통량(고객) 중 운전자의 위험행동에 따라 사고 잠재 위험도 및 이를 비교할 수 있는 정규화된 지수

\* EPDO(Equivalent Property Damage Only): 대물피해환산점

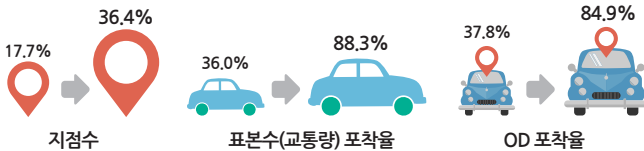


## KTDB 빅데이터 활용사례

### ① 첨단자료를 활용한 O/D 추정방안 연구

- KTDB 수요분석 신뢰성 향상을 위한 지점으로는 위치가 부적합하므로 교통수요 검증에 위한 교통량 조사지점의 재선정을 위해 내비게이션 자료를 활용하고, 다양한 지점 선정기준 (링크 별 표본수, 통과 기종점쌍 수 등)을 고려하여 조사지점을 추가 선정

교통량 조사지점 추가에 따른 효과



<조사지점 추가에 따른 포착을 변화>

구분	과정도	추가지점수	표본수	OD쌍	지점수
	추가 전	1,156	36.0%	37.8%	17.7%
1차	1번 기준	25	36.6%	38.3%	18.1%
	2번 기준	161	50.3%	53.0%	20.6%
	3번 기준	42	51.9%	53.8%	21.2%
	4번 기준	155	63.1%	60.4%	23.6%
2차	2번 기준	150	68.8%	68.1%	25.9%
	4번 기준	146	75.3%	72.0%	28.1%
3차	2번 기준	141	78.9%	77.0%	30.3%
	4번 기준	137	83.2%	79.6%	32.4%
4차	2번 기준	133	85.5%	83.1%	34.4%
	4번 기준	129	88.3%	84.9%	36.4%
최종 추가지점	-	1,219	-	-	-

- 1번 기준: 모든 기종점쌍이 최소 1개 이상의 조사지점을 갖도록 코든라인에서 추가지점 선정  
 2번 기준: 미 관측 지점 중 코든지점 당 기종점 쌍이 많은 순서에 따라 추가지점을 선정  
 3번 기준: 미 관측 지점 중 특정 기종점쌍의 교통량이 기준 비율보다 높은 코든지점 추가  
 4번 기준: 미 관측 지점 중 코든지점 당 기종점 통행량이 많은 순서에 따라 추가지점을 선정

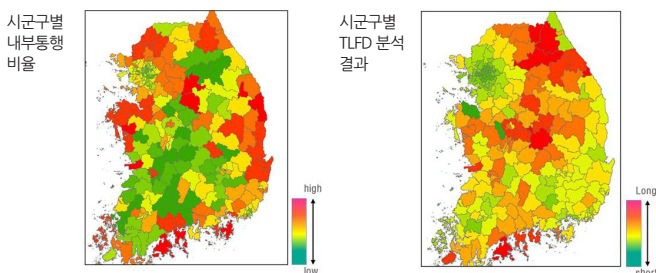
### ② 내비게이션 자료를 활용한 내부통행량 산출방안 연구

- 교통량 정산에 사용되는 관측교통량 중 지역내 통행비율 산정을 위해 내비게이션 자료를 활용하여 내부통행량 산출방안 연구

#### 내부통행량 분석

- 전국 각 VDF 등급별 내 통행량 비율 분석
- 도로의 규모와 통행거리의 상관관계 파악
- 전 등급에서 도시부가 지방부보다 큰 내 통행비율이 낮게 나타남

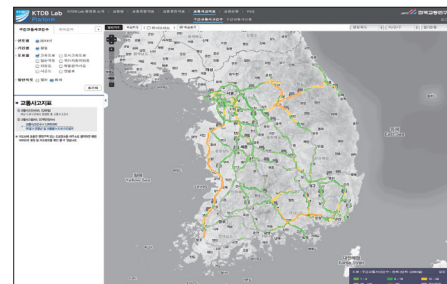
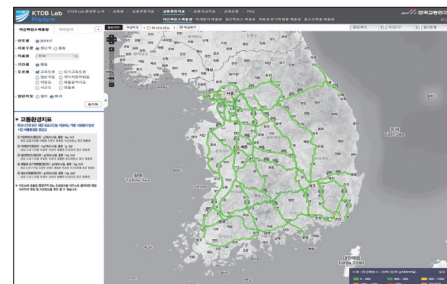
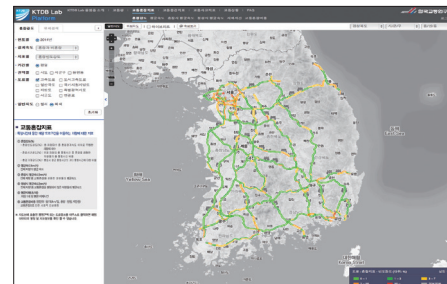
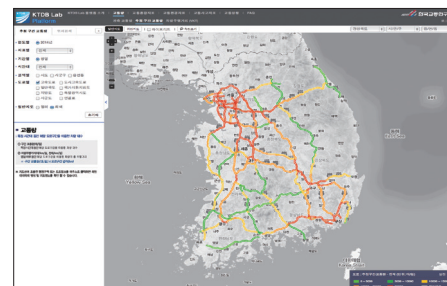
<시군구별 내부통행비율 및 중간통행거리(TLFD) 분석결과>



주: TLFD(Trip Length Frequency Distribution) 중간통행거리

### ③ 교통혼잡지도 DB구축

- 내비게이션 자료를 통해 추정된 전국 범위의 링크교통량을 기초로 하여 네트워크 성능지표의 신규 개발과 기존 교통소통지표에 대한 개선을 수행하여 교통량/교통소통지표의 개선 및 교통사고지표, 환경지표 등을 개발



### ④ 그밖에 분석결과

- 교통카드 자료 분석사례: 대중교통 환승 및 이용실태 등 분석
- 트위터 분석 사례: 교통 관련 트윗에 대한 내용적 분류를 통해 일반인들의 관심 및 경향 파악

출처: 한국교통연구원, 교통 부문에서의 빅데이터 현황 및 활용, 2012년