

빅데이터와 교통

Big Data and Transportation



윤일수



한 음

서론

2009년말 당시 고등학생이었던 유주완군은 수도권 버스 도착정보를 알려주는 '서울버스'라는 앱을 개발하여 선풍적인 인기를 얻었다. 하지만 서울·경기·인천 교통정보 담당자들은 시스템 개발 주체인 경기도와 상의 없이 버스 도착 정보를 무단으로 이용하였고, 더불어 위치정보 사용에 따른 법률적 문제를 이유로 서비스를 중단을 요구하였다.

이에 앱 개발업체들은 공공기관 홈페이지 자료를 사유 재산으로 인정하고 비영리 목적의 앱을 운영하는 개발자들에게까지 공공데이터를 쓰지 못하게 하는 것은 지나친 규제라고 주장하였다.

하지만 2013년 공공데이터를 개방·공유하는 개방형 행정개혁 프로젝트 정부 3.0 비전선포식을 시작으로 공공데이터의 사용과 관련한 갈등은 다소 해소되었으며, 정부는 창조경제 기반조성을 위한 공공데이터를 개방하고 민간 활용을 활성화하기 위한 3대 전략 및 10대 중점추진과제를 추진하

게 되었다.

현재 우리는 민간기관의 포털이나 SNS를 통해 생성·관리되는 수많은 비정형데이터 뿐만 아니라 정부나 공공기관에서 관리하고 있는 정형화된 수많은 공공데이터를 손쉽게 취득·활용할 수 있다.

디지털 환경이 하루가 다르게 발전하고 있는 현 시점에서 생산되는 디지털화된 데이터의 양은 가히 천문학적 수준이며, 데이터의 생성주기는 짧고 범위는 광범위하다. 또한 형태는 텍스트, 이미지, 영상 등 다양한 종류로 생산되며 이와 같이 방대한 양의 디지털화된 정보를 통상 빅데이터라고 부른다.

교통부문에서 빅데이터는 대부분 정형화된 데이터로 한국 도로공사의 OASIS, 도로교통공단의 TAAS등에서 수집 가공되고 있다. 하지만 정형화된 데이터만을 사용하는 것이 아니라 비정형화된 데이터와 혼합으로 분석의 활용도가 높아지고 있다. 활용사례로는 '서울시 심야버스 노선 수립'과 '전국 교통혼잡지도' 등을 예로 들 수 있다. 서울시 심야버스 노선 수립시 버스가 다니지 않는 00

윤일수 : 아주대학교 교통시스템공학과, ilsooyun@ajou.ac.kr, Phone: 031-219-3610, Fax: 031-215-7604

한 음 : 아주대학교 교통시스템공학과, hano3106@ajou.ac.kr, Phone: 031-219-2543, Fax: 031-215-7604

시~05시까지의 통화량 데이터를 이용하여 유동 인구 밀집도 분석을 통해 요일별·노선별 유동인구를 파악하고, 이 패턴을 바탕으로 심야버스의 노선을 최적화 한 사례이다. 그리고 전국 교통혼잡지도는 내비게이션 이용차량의 이동궤적 자료를 이용하여 도로, 교차로, 행정구역별 혼잡강도를 GIS기반 표출시스템으로 구현한 빅데이터 이용사례이다.

이처럼 빅데이터의 사용 목적은 실시간으로 생성되는 수많은 디지털 정보들 중 불필요한 정보를 걸러내고 이용한 정보를 추출하여 새로운 정보의 가치를 창출하는데 있으며, 따라서 이와 같은 빅데이터를 관리하고 분석할 수 있는 연구자의 역할은 점점 그 중요성을 더해가고 있다.

본 논고에서는 IT업계에서 정의하고 있는 빅데이터의 특징에 대해서 알아보고, 이에 대한 분석방법에 주안점을 두었으며, 이를 교통 분야에서 활용할 수 있는 방안에 대해서 기술하였다.

빅데이터의 특징 및 분석기법 소개

기존 빅데이터의 개념이 단순히 데이터의 양이 많은 것을 의미했다면 최근에는 더욱 방대해져서 일반적으로 사용하는 방법이나 도구로 수집, 저장, 검색, 분석, 시각화 등을 하기 어려운 정형 또는 비정형 데이터 집합을 포함한다.

빅데이터란 우리가 알지 못하는 새로운 것을 의미하는 것이 아니며, 우리가 지금까지 데이터라고 인식하지 못했던 것들까지 모두 데이터화 시키는 것이다. 기존 SNS로 대표되는 소셜 미디어의 성장과 최근 스마트폰으로 대변되는 모바일 장치의 확산이 결합되어 일상 속에서 다양한 종류의 대규모 데이터가 급속히 생성, 유통, 축적되고 있다. 또한 RFID와 같이 정보를 감지하는 센서 장비의 이용 확대와 이러한 정보를 수집하는 클라우드 컴퓨팅 기술의 확산은 물류의 이동 및 재고의 변화뿐만이 아닌 개별 소비자들의 개인정보 및 소비행태와 같은 모든 일상에 대한 디지털 기록을 가능케 하고 있다

1. 빅데이터의 특징

일반적으로 빅데이터는 TB(테라바이트)단위의 데이터 크기로 정의되며 데이터 수집 및 분석에 장기적인 시간을 요하므로 데이터 크기가 증가하는 특징을 보인다. 그러나 단순한 데이터 크기의 증가를 넘어서서 빅데이터는 크게 데이터 크기(Volume), 데이터 속도(Velocity), 그리고 데이터 다양성(Variety) 등 세 가지 요소의 복합적인 변화를 그 특징으로 한다.

1) 데이터의 크기(Volume)

빅데이터라는 단어에서 나오는 자연스러운 반응은 데이터의 크기가 크다는 것이다. 하지만 단순 저장되는 물리적 데이터의 크기뿐만이 아닌 이를 분석 및 처리하는데 필요한 네트워크 데이터의 증가는 가장 기본적인 특징이다.

2) 데이터의 속도(Velocity)

빅데이터는 데이터의 실시간 처리 및 장기적 접근을 요구한다. 데이터 생산 및 유통, 수집 및 분석 속도의 증가와 이에 대한 실시간 처리 및 장기간에 걸쳐 데이터를 수집, 분석하는 장기적 접근이 빅데이터의 특성이다. 예시로 전국 도로에 설치된 차량정체 감지센서와 노면상태센서 등은 빅데이터의 속도의 특성을 가지고 있다고 할 수 있다.

3) 데이터의 다양성(Variety)

최근 빅데이터를 이용한 데이터 분석은 고정된 시스템에 저장되어 있지 않고, XML, HTML 등 반정형 데이터를 이용하는 특징을 보인다. 또한 사진, 오디오, 비디오 형식의 소셜 미디어 데이터나 데이터베이스 로그파일(database log) 같이 비정형데이터도 처리할 수 있는 능력을 요구한다. 즉, 빅데이터의 성장이란 단순히 데이터의 크기가 증가하는 것을 넘어서서 다양한 형태의 데이터 처리가 필요하다는 것을 의미한다.

2. 빅데이터의 3대 필요 요소

1) 자원

주어진 빅데이터를 처리·관리하는 측면과 함께 활용할 수 있는 외부 빅데이터 자원을 발견하고 확보하는 전략이 필요하다. 또한 데이터의 품질은 데이터 활용 결과에 중대한 영향을 미치므로 데이터 관리체계 및 데이터의 신뢰성 확보가 매우 중요하다.

2) 기술

데이터를 축적·처리·분석하기 위한 대규모 데이터를 분산 처리하는 빅데이터 플랫폼, 빅데이터 분석기술 및 데이터 분석기법에 대한 개발이 필요하다.

3) 인력

빅데이터를 제대로 활용하기 위해 심층적인 데이터 처리 능력과 분석능력을 동시에 겸비한 전문가가 필요하다.

3. 빅데이터 분석기법

데이터 소스로서 사용할 수 있는 센서 데이터, 기업 데이터, 소셜 데이터, 공공 데이터 등 수없이 많은 데이터들이 존재한다. 이를 수집하고 보관하는 비용이 발생하며, 데이터를 활용하여 새로운 가치로 환산하여야 한다.

빅데이터 분석 기술로는 ‘의미를 발견하는 데이터 처리 및 분석 기술’, 즉, 의미분석 기술과 데이터 마이닝 기술 관련 기법들이 필요하다.

1) 데이터 마이닝(Data Mining)

대용량의 데이터, 데이터베이스 등에서 감춰진 지식, 기대하지 못했던 경향, 새로운 규칙 등의 유용한 정보를 발견하는 과정으로 정보의 연관성을 파악함으로써 가치 있는 정보를 만드는 기법이다.

2) 텍스트 마이닝(Text Mining)

자연어로 구성된 비정형 텍스트 데이터에서 패턴 또는 관계를 추출하여 가치있는 정보를 찾아내는 기법이다.

3) 웹 마이닝(Web Mining)

인터넷상에서 수집된 정보를 데이터 마이닝 기법으로 분석 통합하는 기법이다.

4) 소셜 마이닝(Social Mining)

소셜 미디어에 올라오는 글과 사용자를 분석해 소비자의 흐름이나 패턴 등을 분석하고, 사회의 흐름과 트렌드, 여론변화 추이를 읽어내는 기법이다.

5) 현실 마이닝(Reality Mining)

사람들의 행동 패턴을 예측하기 위해 사회적 행동과 관련된 정보를 기기(휴대폰, GPS 등)를 통해 얻고 분석하는 기법이다.

교통에서의 빅데이터

기존에는 축적된 데이터를 바탕으로 행동 패턴을 도출해 문제점을 발견하는 수준이었다. 하지만 이제는 스마트 기기나 설비에 장착된 다양한 센서에서 수집된 데이터나 웹 검색 키워드 등으로부터 문제점을 발견하고 해결책을 찾는 것이 가능하다.

오늘날 대부분의 도시에 ITS가 구축되어 있으며, 이러한 도시들로부터 수많은 교통데이터가 발생한다. 사람과 자동차, 대형 이벤트들이 어울려 복잡하고 거대한 빅데이터를 만들어 내고 있다. 이러한 빅데이터를 분석하고 활용할 수 있다면 도시의 복잡한 움직임을 한눈에 파악할 수 있는 수단이 될 수 있다.

활용 사례로서 센서 데이터인 GPS 데이터를 분석하여 지능형 교통정보를 제공하는 경우를 들 수 있다. 일본 노무라연구소에서 구축한 UTIS를 활

용해 내비게이션 오류를 최소화한 교통정보 서비스는 2011년 일본 대지진 사고 때 도로 상의 교통 정체로 인한 피해를 최소화하는 데 기여한 바 있다. UTIS는 계약을 맺은 택시들과 개인 데이터 제공에 합의한 사용자의 GPS 데이터로부터 자동차의 주행 스피드를 계산해 교통정체 등 도로 교통 정보를 생성하여 사용자 스마트폰으로 실시간 정보를 송신한다.

빅데이터의 분석을 위한 오픈 API를 자동차에 적용하면 자동차가 운전자의 행동을 효율적으로 학습하고 미래의 행동을 예측한다. 통신 기업들도 클라우드 기반에서 신속, 정확하고 손쉬운 길안내 서비스를 제공하기 위하여 올레navi, Tmap, U+ Navi 등을 제공하고 있다. 이들 앱은 모두 클라우드와 접목되어야 하며, 다양한 개인화 서비스 제공을 위해 센서 데이터 분석 기술들이 확장 적용될 것이다.

서울시에서는 심야버스 노선을 확정하는데 빅데이터를 활용하였다. 일반 시내버스 노선 결정에는 교통카드 정보를 활용하였으나, 심야버스 노선에는 교통카드 정보와 더불어 다산 콜센터에 문의된 내용과 KT와 협조하여 심야에 전화를 이용하는 사람들의 발신자 위치와 번호 소유자의 주소지를 연결하여 이동경로를 예측하였다. 이 데이터를 바탕으로 서울시는 심야버스 노선을 확정하였다.

데이터마이닝 프로그램

전 세계적으로 빅데이터를 분석하는 프로그램들이 개발되고 있다. 각 기업별로는 자체적인 마이닝 프로그램을 통한 분석을 실시하고 있으며 일반적으로 많이 사용되는 2가지의 프로그램에 대하여 소개하고자 한다.

1) SPSS Modeler 15

SPSS Modeler는 SPSS Inc.에서 개발된 Data Mining 전문 패키지이다. 현재 전 세계적으로 가장 널리 보급, 활용되는 마이닝 S/W로 데

이터를 불러오는 것부터 데이터 핸들링, 그래프, 모델링을 거쳐 최종적으로 결과를 출력하는 기능을 지원한다. Modeler의 가장 큰 장점은 통계분석 프로그램인 SPSS와의 연계성에 기인한다.

2) NetMiner 4.0

넷마이너는 순수 국내 기술로 개발된 네트워크 분석 소프트웨어로, 데이터 변환, 네트워크 분석, 통계분석, 네트워크 시각화 기능 등을 유연하게 통합하여 제공한다. 특히 분석뿐만 아니라 시각화 기능까지 연계하여 분석자에게 제공한다.

결론

세계 주요국들은 정책 결정 과정의 투명성 제고를 위해 열린 정부 정책을 추진중이며, 이의 첫 움직임이 바로 공공정보(PSI) 개방 및 활용이다. 우리나라 역시 정부 3.0의 정책을 통해 공공정보를 개방하였으며, 지속적으로 확대 시행 예정이다. 공공데이터가 잘 활용하게 되면 실생활과 기업의 신사업 아이디어에 도움을 줄 수 있는 형태로 얼마든지 가공될 수 있다.

그러나 빅데이터는 기존의 기술로는 처리하기 힘들 정도로 규모가 방대하고 형식이 다양하여 이의 활용에는 전문화된 IT기술을 필요로 한다. 교통에서 빅데이터의 광범위한 활용을 위해서는 기존의 정형화된 데이터 분석이 아닌 비정형 데이터를 분석하여 시각화 할 수 있는 기법들을 살펴보고 새로운 서비스를 생산해낼 수 있어야 한다.

참고문헌

- 시로타 마코토 (2013), 빅데이터의 충격, 한빛 미디어.
 정지선 (2012), 성공적인 빅데이터 활용을 위한 3대 요소, 한국정보화진흥원 연구보고서.
 홍진영 (2013), 빅데이터가 만드는 비즈니스 미래지형, 한스미디어.