

최동진(조선대학교 컴퓨터공학과 대학원생)

김관구(조선대학교 컴퓨터공학과 교수)



1. 서론

초고속 인터넷망과 무선 인터넷망의 획기적인 발달로 인해 데이터는 더 이상 시간과 장소에 제약받지 않고 신속하게 전달 및 공유되고 있다. 더욱이, 스마트폰과 SNS의 혁명으로 엄청난 양의 데이터가 생산되고 그 종류 또한 다양해지고 있는 상황이다. 빅데이터(Big Data)란 기존의 방식으로 저장/관리/분석하기 어려울 정도로 큰 규모의 자료 [1]를 의미하는 것으로 최근 글로벌 경제지, 컨설팅 그룹 등에서 화두가 되고 있는 주제이다. 빅데이터는 주관적이며 바라보는 시각에 따라 그 활용가치를 다르게 예측하고 있는데 Gartner의 경우 데이터를 원유에 비유하여 미래 경쟁의 우위를 좌우하는 원석으로, McKinsey는 빅데이터의 활용에 따라 특히 의료, 공공행정 등 5대 분야에서 큰 가치를 창출할 것으로 예상하고 있다. 표 1은 각 기관들이 조사한 2012년 IT기술 전망에 관한 것으로 모든 기관은 이구동성으로 빅데이터의 중요성을 예측 전망하고 있다. 이렇듯, 빅데이터는 현재 화두가 되고 있는 빅 이슈로써 국내뿐만 아니라 세계적으로 관심을 받고 이를 활용한 연구를 선도적으로 수행하려는 움직임이 나타나고 있다. 빅데이터에 관한 연구는 크게 두 가지 관점으로 나눌 수 있는데 데이터의 처리속도 및 능력에 관한 관점과 데이터의 마이닝

에 관한 관점이다. 제타바이트(1조 Gigabyte)에 이르는 대용량 데이터를 분산 병렬 처리하여 실시간 데이터 처리를 가능하게 하는 하드웨어적 연구와 산재되어 있는 대규모 데이터 속에서 데이터의 정재를 통해 숨겨진 의미를 추출하고 데이터를 가공하는 소프트웨어적 연구이다. 학계와 기업들은 이 두 목적을 해결하기 위하여 발 빠르게 다양한 연구를 진행하고 있는 상황이다.

본고에서는 빅데이터 시대를 대비하기 위하여 전략적인 방안을 제공하기 위하여 국내외 대표적인 연구 동향과 함께 빅데이터 기술 전략 및 방안을 살펴보고자 한다.

2. 빅데이터(Big Data)

빅데이터는 크게 세 가지 종류로 나눌 수 있는데 그 중 첫 번째는 자연과학 정보이고 두 번째는 기업의 생산라인에서 발생하는 공장기기의 상황정보이고 마지막으로 소셜 미디어(Social Media) 정보를 들 수 있다. 위 세 가지의 데이터는 서로 상이한 성격을 띠는 데이터로써 데이터의 형태에서부터 마이닝, 분석 방법까지 모두 상이하다. 빅데이터의 성격에 걸맞은 데이터 처리가 핵심으로 부각되는 만큼 본 장에서는 빅데이터의 3가지 유형들과 이를 이용한 국내외 연구동

향을 살펴본다.

표 1 2012년 IT 기술 전망

기관	2012년 IT 기술 전망
가트너	<ul style="list-style-type: none"> - 미디어 태블릿 - 모바일 중심 어플리케이션과 인터페이스 - 상황 및 소셜 사용자 경험, 사물 인터넷 - 앱스토어와 마켓플레이스 - 차세대 분석기술, 클라우드 컴퓨팅 - 빅 데이터, 인메모리 컴퓨팅 - 초절전(저전력) 서비스
딜로이트	<ul style="list-style-type: none"> - 비주얼리제이션 - 올모스트 엔터프라이즈 어플리케이션 - 사이버 정보, 혁명가로서의 CIO - ERP의 부활, 역량 클라우드 - 실분석 (Real Analytics), 소셜 컴퓨팅 - 사용자 참여, 어플라이드 모빌리티
EMC	<ul style="list-style-type: none"> - 새로운 킬러 어플리케이션 = 간편함 - 모바일 우선(Mobile First) 전략 - 디지털 비즈니스 모델 - 빅 데이터 전문가 "데이터 과학자" - 기업 IT, 외부 IT 서비스 업체와 경쟁 - 클라우드 기술 - IT보안의 변화, 새로운 보안 기업 등장 - 전문화된 IT 서비스 기업의 각광 - 빅 데이터와 분석학, 새로운 가치와 기회 창출
NIPA	<ul style="list-style-type: none"> - 클라우드, 4G LTE - 정보보호, 차세대 TV - IT융합, 차세대 부품 - 차기정부 IT정책, 윈도8 - 스마트 기기, 빅데이터
삼성 SDS	<ul style="list-style-type: none"> - 소비자 지향적 기술 대중화 - 네트워크 통한 협력적 소비 - 게임 메커니즘 활용 비즈니스 - 웹 기반 사물 연결과 제어 - 소셜 분석, 삶의 질을 향상시키는 IT - 모바일 컴퓨팅 보안, 플랫폼 전쟁

2.1 자연과학에서의 빅데이터

자연과학은 빅데이터의 보고라고 할 정도로 그 종류는 다양하고 그 양 또한 방대하다. 그 중에서도 현재 가장 많이 활용되고 있는 분야는 단연, 의학 분야라고 말할 수 있다. 국내외 각종 기관들은 서둘러 빅데이터 처리를 활용한 의료정보 사업에 뛰어들고 있다. 미 대통령 직속 기관인 과학기술 정책실(Office of Science and Technology Policy)은 2억달러 이상을 투입하여 빅데이터 기술을 개발한다고 발표하였다. 유전자 연구 및 의료, 교육, 지구과학

등 빅데이터 활용 효과가 뛰어난 분야를 우선으로 유래 없는 지원을 시행한다는 것으로 매년 \$3,300억의 가치를 창출 할 것으로 예측하고 있다. 대용량의 의료 데이터를 분석하여 효과적인 진료방법을 파악하고 투약효능을 향상시키고 의료협회 간 데이터의 공유로 치료효과를 극대화시키는 것을 목표로 한다. 미국 온타리오 공과대학의 캐롤린 맥그리거 교수는 안타깝게 꺼져가는 미숙아의 생명을 구하기 위하여 최첨단 의료시스템에서 발생하는 의료데이터를 실시간 분석 및 처리하여 패턴을 발견하고 이 패턴에 해당하는 징후에 따른 감염여부를 판단하는 시스템을 적용하였다[2].

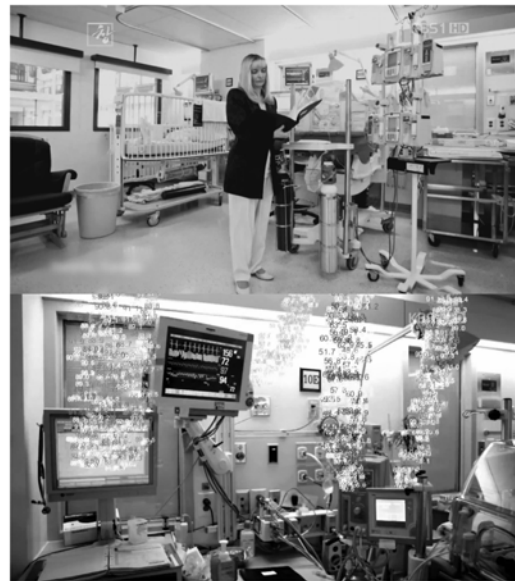


그림 1. 미국 온타리오 공과대학의 빅 데이터 분석 방법을 이용한 미숙아 증상 예측의 예

그동안 수기로 작성한 신생아 상태표에 의존하였던 의료체계를 획기적으로 바꿀 수 있는 계기를 마련한 예이다. 그동안 활용되지 않고 버려지는 데이터(심장박동수, 체온 등)를 재발견하여 신생아 문제상황 예측 시스템을 개발하여 수많은 생명을 구할 수 있게 되었다.

IBM에서 개발한 빅데이터 처리 슈퍼컴퓨터인 Watson을 적용하여 영국 더럼대학(Durham University)의 ICC 및 OCF와 함께 은하계의 우주 연구를 위한 빅데이터 처

리 프로젝트를 진행하고 있다. 페타바이트 이상에 달하는 우주 정보를 대상으로 은하계의 형성과 소멸에 관한 현상을 규명하는 연구이다[3]. 뿐만 아니라, 스웨덴 Uppsala 대학의 천체물리연구소에서는 우주의 날씨를 효과적으로 분석하기 위하여 IBM의 빅데이터 분석 솔루션을 도입하여 초당 21.6 테라바이트를 분석하여 플라즈마 클라우드가 우주 내에서 어떻게 움직이는 지 그 경로를 예측할 수 있는 결과를 도출할 수 있게 되었다[4].

국내 과학분야 빅데이터 시스템을 구축중인 KISTI는 극지연구소와 함께 극지 기후 변동 연구를 개발하기로 하였다. 연각 45억 원에 달하는 대규모 프로젝트로써 고정밀, 대용량 위성사진 500만장 정도를 분석하여 기후예측 모델과 융합하여 세계최고의 빅데이터 분산처리 기술을 개발하는 것을 목표로 하고 있다.



그림 2. 스웨덴 Uppsala 대학에서 연구중인 우주환경과 우주날씨 그리고 지구에 미치는 영향의 예

2.2 기업을 위한 빅데이터

기업의 생산라인의 기기에서 발생하는 데이터는 하루에도 수 십 테라바이트에 이른다. 끊임없이 발생되고 소멸되어 지는 원석과 같은 대규모 데이터는 지금 이 순간에도 가공되지 못하고 버려져만 가고 공장에 문제가 발생하였을 시 원인규명의 단순 목적에만 활용되고 있다. KAIST의 장영재 교수는 반도체 공장에서 생산되는 제품의 시간

이 서로 상이한 것에 의문을 갖고 데이터를 분석한 결과, 공정에서의 문제점을 도출하고 제품 생산시간을 단축시켜 원가절감효과를 달성할 수 있었다고 밝혔다[5].

세계적인 기업 중 하나인 Walmart는 실시간 재고 분석을 위한 시스템을 도입하여 공급자들에게 상품 재고 현황에 대한 실시간 정보를 제공하여 납품업체의 재고 관리 효율성을 극대화하는 효과를 창출하였고 Li&Fung의 실시간 관리 시스템은 해외 의류브랜드에 IT시스템을 접목하여 주문, 생산, 선적 전과정의 실시간 모니터링 기술을 바탕으로 발생할 수 있는 리스크를 최소화하는 효과를 얻을 수 있었다[1]. 뿐만 아니라, 서울대학교의 조성준 교수는 반도체 공정에서 발생하는 센서 데이터를 분석 및 모니터링에 활용하였고, 실내 냉난방의 에너지 관리를 위하여 시스템 냉난방기에서 발생하는 데이터를 수집, 정제, 분석하여 에너지 효율을 높이는 연구를 수행하였다[6].

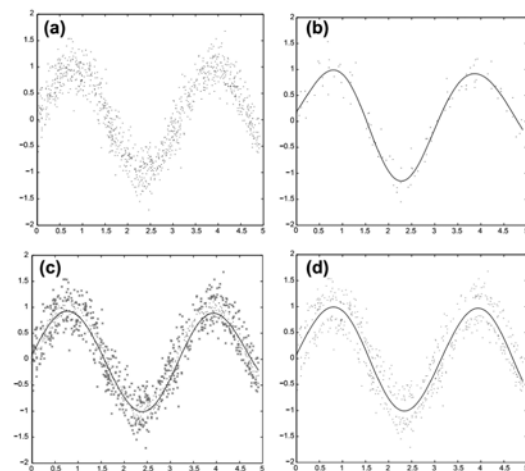


그림 3. 실시간 추출된 데이터를 바탕으로 SVR 알고리즘을 적용한 예

LG CNS의 통합품질 분석은 공정과정에서 센서 등을 통해 수집되는 기계 상태와 온도, 공정 상황 데이터를 분석하여 어떤 부분에 이상이 있는지를 판단하여 분석 처리시간을 97% 단축시켰다. MIT분석자료에 의하면 기업성과가 높은 기업들이 성과가

낮은 기업들보다 데이터분석 활용률은 5.4 배나 높아 직관적인 의사결정보다는 데이터 분석자료에 근거하여 리스크를 최소화하고 성과를 극대화할 수 있다고 보고하였다. 이렇듯, 빅데이터는 기업에 활발히 적용하여 기업의 이윤창출뿐만 아니라 산업발전에 기여할 수 있음은 분명하다.

2.3 소셜 미디어(Social Media)에서의 빅데이터

소셜 미디어는 소셜 네트워크 기반 위에 개인의 생각이나 의견, 경험, 정보 등을 서로 공유하고 타인과의 관계를 생성하는 개방형 온라인 플랫폼을 의미한다[7]. 그 종류로는 블로그(Blog), 소셜 네트워크 서비스(SNS), 위키(Wiki), UCC, 마이크로 블로그(Micro-Blog)등을 들 수 있는데, 다양한 형태의 콘텐츠가 웹 또는 스마트폰을 통해 생산 및 가공 공유되어 지고 그 데이터의 양은 트위터(Twitter)를 기준으로 할 때 하루 수억 만 건에 달하는 글들이 매순간 발생되고 있고 YouTube에서는 매분마다 24시간 분량의 비디오가 업로드 되고 있다. 기업가, 정치가, 소비자, 언론가, 교수 등 다양한 사용자들이 자신을 기준으로 발생한 일들 또는, 사건들을 기록하면서 생각을 공유하고 의견을 전달하는 장소로 활용되고 있다. 그 전파속도는 뉴스와 같은 언론매체를 뛰어넘어 새로운 언론매체로 부상하고 있다. 이와 같이 사용자의 주관적인 생각을 분석하여 현재 이슈가 되는 토픽이나 대중의 관심사등을 분석하여 정치, 사회, 경제에 적용하는 연구가 활발히 수행되고 있다.

구글(Google)은 전세계 인터넷 사용자의 이용패턴 빅데이터 분석을 통하여 사용자의 의도를 예측하고 맞춤형 서비스를 제공하는 검색엔진으로 널리 각광을 받고 있다. 수억 만건에 달하는 웹문서를 수집한 구글은 학술자료에서부터 특허, 논문자료 등과 더불어 사용자가 입력하는 질의어를 분석하여

트렌드를 분석하고 이슈가 되는 기술을 제공하고 있다. 미국의 대형 인터넷 서점인 아마존(amazon)은 사용자의 구매 로그 데이터를 분석하여 향후 구입할 상품을 예측하여 추적하는 시스템을 도입하여 전체 매출의 30% 향상을 이룩하였다.

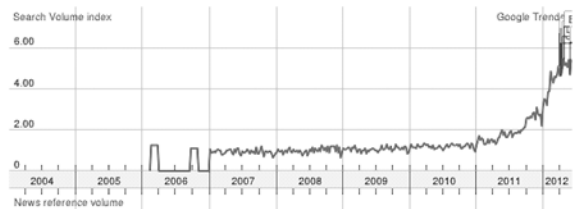


그림 4. 구글 트렌드를 이용한 빅데이터의 트렌드 분석에

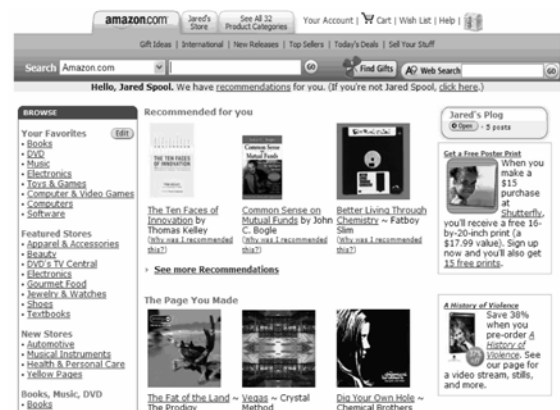


그림 5. 아마존에서 제공하는 추천 시스템의 예

페이스북(Facebook)은 자사의 데이터분석팀에서 페이스북에 올라오는 글을 분석하여 전 세계적으로 페이스북을 통한 데이터의 흐름을 시각적으로 보여주었다. 누가 어디에서 어느 시간에 어떤 데이터에 열광을 하는지를 분석함으로써 소셜 미디어 데이터 분석의 가능성을 증명하였다.

이와 같이 페이스북은 대용량의 데이터를 이용하여 새로운 의미를 부여하고 사용자와 주변인들과의 관계를 규명함으로써 지능형 맞춤 서비스를 제공하기 위한 연구를 수행



그림 6. 전 세계 페이스북 사용자의 연령별 분포도

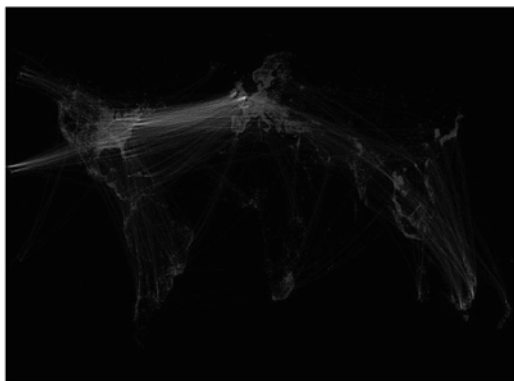


그림 7. 페이스북 사용자의 이동 경로

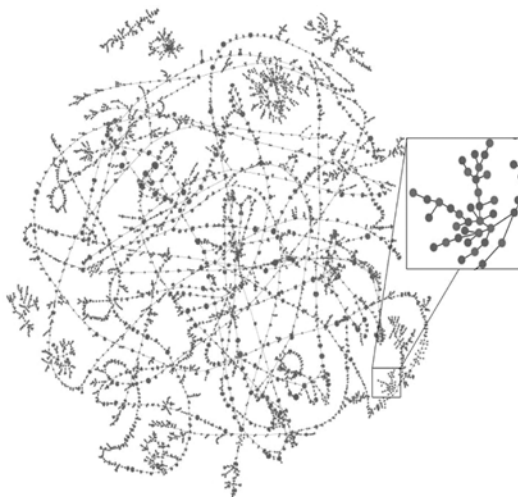


그림 8. 페이스북 사용자의 인맥 관계도 예 (85 million)

하고 있다. 또한, 트위터에서 추출한 집단 정서 상태가 다우존스 지수값과 서로 관련이 있다는 연구결과도 보고되었다[10]. rm 결과로 다우존스 지수값과 비교하였을 때 무려 87.6%에 달하는 정확도를 달성한 것을 볼 수 있는데 이는 소셜 빅데이터가 갖

는 큰 잠재력을 증명한 예로 볼 수 있다.

국내 업체의 경우 트위터와 같은 소셜 미디어를 대상으로 대중의 목소리를 분석하는 연구를 활발히 수행하고 있는데 그 중 대표적 예로 솔트룩스(Saltlux)에서 개발한 트루스토리(True Story)를 들 수 있다. 뉴스, 블로그 트위터 등에서 기록되는 다양한 사용자의 글을 수집 및 분석하여 현재 정치인들을 비교분석하는 재미있는 빅데이터 분석 플랫폼을 개발하였다. 트루스토리는 클라우드 컴퓨팅과 인공지능 기술이 결합된 정형, 비정형 데이터의 분석을 수행함으로써 기업 빅데이터 분석, 통신 및 금융 빅데이터 분석 등 다양한 도메인에 활용되어지고 있다 [8]. 자연어처리, 텍스트마이닝, 기계학습, 시맨틱 기술과 같은 인공지능 기술이 폭넓게 활용된 첨단 기술의 결과물이라고 말할 수 있다.



그림 9. 트루스토리에서 분석되는 소셜 데이터의 예

국내 포털 사이트의 대표적인 기업 중 하나인 다음소프트(DaumSoft)은 자사에서 그

동안 수집한 블로그 데이터를 분석하여 마케팅에 적용시키는 연구를 수행하였다. 여러 사람의 생각과 의견을 분석하여 객관적인 데이터로 가공하기 위하여 웹상에서 대중의 의견을 분석 한 뒤 대중의 생각을 수집하고 분석 판단하였다. 다음소프트의 소셜미디어 분석 시스템은 인터넷에 자연어로 표현되어 있는 사용자의 의견을 수집하여 고품질 자연어처리(Natural Language Processing)기술과 정교한 텍스트마이닝 기술을 통해 분석을 하고 제품 또는 서비스의 모델, 항목, 세부 속성에 대한 긍정, 부정 등의 평가 결과를 도출해내는 지능형 의견 분석 소프트웨어이다[9].



그림 10. 다음소프트에서 실시한 웹 데이터 분석을 통한 사용자 선호도 분석에 관한 연구 예

소비자 및 사회를 이해하고 트렌드를 감지하고 예측하기 위하여 단순히 컴퓨터 알고리즘을 적용한 것이 아닌, 인문학, 사회학, 경제학, 철학, 수학 등 다양한 전문가들의 조언을 통하여 의사결정을 내리고 데이터를 분석하였다.

3. 결론

통신망의 발달과 최근 스마트폰의 급격한 보급과 발전으로 인해 데이터의 생산력은 과거와 비교할 수 없을 만큼 획기적으로 증가하였다. 이에 따라 기존의 방법으로 해결할 수 없을 만큼 방대한 빅데이터는 수많은 광석을 품은 광산으로 간주되어 국내외 학계 및 기업들의 커다란 관심을 받고 있다. 광산에서 광물을 신속히 캐낼 수 있는 방법으로 빅데이터를 해결하려는 관점과 산재되어있는 광물 속에서의 보석을 찾아내려

는 데이터 분석 관점으로 볼 때 본고에서는 후자에 그 중요도를 부여한다. 빅데이터는 학계와 기업, 정부 등 모든 기관에 걸쳐서 향후 획기적인 변화를 가져올 수 있는 광석임은 분명하다. 하지만 빅데이터를 바라볼 때 간과해서는 안되는 점은 빅데이터는 광산이 큰 것이지 매장량이 크다는 것을 보장하지는 않는다. 산재되어 있는 대규모 데이터 속에서 의미를 파악하고 가공하는 작업은 단순히 컴퓨터공학만의 문제가 아닌 자연과학, 인문학, 사회학, 경제학, 철학, 물리학, 수학, 천문학, 지질학 등 모든 분야의 전문가들이 함께 풀어나가야 할 임무인 것이다. 비정형 데이터를 어디에 어떤 목적으로 사용할 것인가를 분명히 하여 산재되어 있는 데이터 중에서 의미 있는 데이터만을 정제한 뒤, 그 속에서 가치를 부여하고 정형화된 형태로 가공하여 공유하는 문제로 초점을 맞추어야만 현재 화두가 되고 있는 빅데이터 처리에 선도적으로 앞서 나갈 수 있을 것이다. 빅데이터의 빠른 처리능력도 중요하지만 그보다 더 강조되어야 할 부분은 빅데이터를 가지고 무엇을 할 것인지 그리고 목적에 걸맞은 데이터 분석 처리를 수행할 때 결국, 빅데이터 활용으로 인한 기대효과를 극대로 창출 할 수 있을 것이다.

참고 문헌

- [1] 이성춘, 임양수, 안민지, “Big Dat, 미래를 여는 비밀 열쇠” KT경제경영연구소
- [2] 시사기획창 빅데이터 세상을 바꾸다, 2012년 1월 31일
- [3] www.ibm.com/storage
- [4] <http://www.space.irfu.se/>
- [5] 장영재, “빅데이터 환경과 과학적 경영”, 2012 IT 21 Global Conference
- [6] 조성준, “Big data 기술: 압축, 선택 및 클러스터링”, 2012 IT 21 Global Conference
- [7] www.wikipedia.org

[8] www.saltlux.com

[9] 송길영, "Trend Analysis via Social Big Data Mining", 2012 IT 21 Global Conference

[10] J. Bollen, H. Mao, X. Zeng, "Twitter mood predicts the stock market", Journal of Computational Science, 2(1), March 2011, Pages 1-8