

빅데이터 선진국을 기대하며

임병화 한국경제연구원 선임연구원

1990년대 말 메이저리그 야구팀인 ‘오클랜드 어슬레틱스’는 만년 최하위에 돈이 없어 그나마 실력 있는 선수들은 다른 구단에 빼앗겼던 중소구단에 불과하였다. 선수에 의존하지 않고 적은 비용으로 성적을 내야만 했던 구단주는 개개인 성적이 우선시되는 기존 선발방식을 바꿔 ‘머니볼’ 이론을 따르기로 했다. 오직 경기 데이터에만 의존하여 선수 자질 중 팀 승리에 기여하는 요소를 분석하여 선수를 선발하는 상식을 깬 발상을 한 것이다. 사생활 문란, 잦은 부상, 최고령 등의 이유로 다른 구단에서 외면 받던 선수들이 팀에 합류하였고 당시 메이저리그 최고 부자 구단인 뉴욕 양키스 팀 연봉에 불과 삼분의 일도 안 되는 저 연봉의 선수들로 팀이 꾸려졌다. 그러나 놀랍게도 이런 데이터 분석에 기초한 운영은 메이저리그 역사상 최초 20연승이라는 대기록과 4년 연속 포스트 진출이라는 경이적인 기록을 세웠으며 이들이 보여준 투자 대비 성과는 140년이 넘는 메이저리그 역사상 최대 이변으로 남아있다. 얼마 전 영화로 제작되기도 한 빌리 빈 구단주의 경영 패러다임 변화는 이후 모든 메이저리그 팀에 영향을 미쳐 수학이나 통계학을 전공한 분석가를 활용하여 선수기용이나 세부 작전에 관여하게 하고 있다. 현미경 야구로 표현되는 일본 프로야구와 국내 프로야구에서도 데이터 분석은 팀 운영의 핵심 요소로 뽑히고 있으며 야구가 데이터의 스포츠라는 것은 야구에 관심이 있는 사람이라면 누구나 인정하는 부분이 되었다.

2011년 2월 미국 TV 방송에 인간 퀴즈 달인과 IBM에서 개발한 슈퍼컴퓨터 왓슨(Watson)과의 퀴즈대결이 펼쳐졌다. 진행자가 문제를 읽어주면 빠른 시간 내에 부저를 눌러 정답을 답변하는 일반적인 퀴즈쇼 방식이었다. 왓슨은 인간이 말하는 자연어를 분석하여 원하는 대답을 말해야 하기 때문에 방대한 자료 축적과 빠른 분석 능력이 요구되는 상황이었다. 하지만 왓슨은 당시 74회 연속 우승을 차지했던 퀴즈 달인을 여유롭게 제치고 승리하는 모습을 보이며 진정한 인공지능의 면모를 보여주었다. 최근에는 월가에 취직하여 투자설명서나 대출 실적 등의 분석을 담당할 것이라고 알려져 있으며 장기적으로는 고객의 돈을 실제 주식투자 등에 자동으로 투입하는 단계까지 기대하고 있다. 뿐만 아니라 뉴욕 소재 메모리얼 슬로안 케

터링 암센터와 업무협약을 맺고 의사들의 암 진단 및 치료방법 선택을 돕는 전문의 실습과정을 밟고 있다. 즉, 금융, 의료 분야에서 생성되는 방대한 자료를 신속·정확하게 분석하여 사용자가 원하는 정보를 보여주고 의사결정에 직접적인 영향을 주고 있는 것이다. 금융, 의료 분야에서의 데이터 분석은 새로운 패러다임 혁신으로 다가오고 있다.

바야흐로 빅데이터(Big Data) 시대이다. 빅데이터란 기존 데이터에 비해 크기가 방대하고 형식이 다양하며 순환속도가 빨라 일반적인 방법으로 수집하거나 분석하기 어려운 디지털 데이터를 뜻한다. 트위터의 짧은 기록들이나 문서, 페이스북의 사진 또는 '좋아요' 기록, 스마트폰 안의 위치정보, 휴대폰의 통화기록이나 전자상거래 현황 등 현대인들의 모든 생활 패턴이나 활동 모습을 파악할 수 있는 막대한 양의 데이터들이 빅데이터에 속하며 최근에는 대규모 데이터의 수집, 분류, 분석 도구, 플랫폼, 분석기법 등을 포괄하는 용어로 의미가 확대되었다. 2012년 세계경제포럼(다보스포럼)에서 국제 개발을 이끌 가장 중요한 기술로 꼽히기도 한 빅데이터 기술은 향후 비즈니스의 모습과 나아가 산업 구조까지 바꿀 정도의 잠재력이 있다고 평가받고 있으며 경제, 사회, 문화, 정치 등 우리 삶의 혁신적 패러다임으로 부각되고 있는 상황이다.

빅데이터의 부상은 디지털 데이터의 급증과 연관성이 높다. 2011년 전 세계에서 생성된 디지털 정보량은 1.8제타바이트¹⁾로 대한민국 모든 사람이 18만 년 동안 쉬지 않고 1분마다 트위터에 3개의 글을 게시하는 양과 달한다. 하루에도 미국의 모든 도서관에 소장된 정보의 8배에 달하는 데이터가 매일 생산되고 있으며 2년마다 2배의 디지털 정보량이 증가하고 있다고 한다. 이들 데이터를 활용해 사람들의 생각이나 행동의 변화를 읽을 수 있다면? 무수히 떠다니는 데이터 속에서 내가 알지 못했던 정보를 알려줄 수 있다면? 빅데이터는 바로 이런 물음에 대한 대답이라고 할 수 있다. 새로운 사업기회를 찾거나 경영 효율성을 높이길 원하는 기업 비즈니스나 효율적인 자원배분과 치안, 안보 등을 목적으로 하는 정부 입장에서 반드시 필요한 기술이라 하겠다.

새로운 패러다임 혁신을 불러올 빅데이터

사실 방대한 양의 자료 분석을 활용해 새로운 가치를 창출한다는 생각은 전혀 새로운 것이 아니다. 그러나 최근 하드웨어의 가격 하락과 클라우드 서비스²⁾ 도입에

1) 1,800,000,000,000,000,000 bytes

2) 구름(cloud)에서 유래된 말로 무형의 자원(하드웨어, 소프트웨어)을 web을 통해 빌려 쓰는 방식을 뜻한다. 인터넷 이용만 가능하면 언제 어디서든지 PC 작업이 가능하기 때문에 휴대성이 뛰어나며 고가의 PC

따른 정보 저장 비용의 감소와 대용량 데이터 분석기술의 발달로 데이터 분석 시간이 짧아지면서 실시간 분석 결과가 기업 경영이나 공공 부문의 의사결정 단계에 절대적 역할을 하고 있다. 영화 속 추격 장면에서나 볼 수 있었던 실시간 정보 전달은 이미 현실이 되었고 소비자의 니즈나 행동을 파악하기 위해 빅데이터 기술을 기업 경영에 활용하고 있는 사례도 증가하고 있다. 가장 대표적인 기업으로 작년 3월 세계 패션의 메카인 뉴욕 5번가의 빌딩을 미국 최대 매매가로 매입하면서 세계 패션계를 경악하게 하였던 패션브랜드 ‘자라(Zara)’를 꼽을 수 있다.³⁾ 패션 산업에서는 급변하는 유행을 따라잡기 위해서는 디자인, 생산, 매장 진열 과정을 최대한 단축시켜야하고 다양한 소비자 욕구를 충족시키기 위해서는 다품종이 필수 요소이다. 실제 자라는 일반 패션 브랜드의 3배가 넘는 다품종을 소량 생산하고 있으며 저가에 상품을 신속하게 공급하고 있다. 그러나 재고처리 문제까지 생각한다면 이러한 경영 방식은 상당한 비용이 요구된다. 과연 자라는 어떻게 그 많은 비용을 감당하고 있는 것일까? 바로 빅데이터에 그 해답이 있다. 전 세계 매장에서 본사로 실시간 유입되는 판매 데이터를 실시간 분석하여 상품 수요 예측, 각 매장별 적정 재고 산출, 그리고 상품별 가격 결정을 하면서 비용을 최소화 하고 있는 것이다. 1년 후의 트렌드를 예측하는 것이 아닌 실시간 의사결정에 필요한 정보를 생산한다는 것이 바로 빅데이터의 새로운 패러다임 혁신인 것이다.

또 다른 혁신은 소비자 니즈를 미리 파악하여 정보를 제공해 준다는 데 있다. 미국의 DVD 렌탈 회사인 넷플릭스나 온라인 서점 아마존은 기존 고객들의 소비 정보를 바탕으로 수많은 제품 가운데 특정 고객이 선호하는 제품과 관련된 정보를 제공하여 고객들의 구매를 촉진하고 개별 고객의 정보과부화를 줄여주고 있다. 아마존의 경우 이러한 추천 시스템에 의한 매출비중이 30%에 달하고 넷플릭스는 무려 80%에 달한다고 한다.

무궁무진한 빅데이터의 활용 분야

직관이나 느낌에 의존하지 않는 의사결정을 가능케 하고 실시간 소비자 욕구 파악이 가능한 빅데이터 기술은 대부분의 산업에서 활용 가능하다. 먼저 IBM의 왓슨 처럼 금융투자, 의료 산업 분야는 이미 활용하고 있고 개인 맞춤형 교육산업에도 활용이 가능하다. 유통사들은 고객들의 이미 주문 정보를 분석하기 시작하였고 카드사들은 카드 회원들의 구매 패턴에 관심을 기울이기 시작하였다. 교통 물류 기업들은 상품의 이동 경로와 여행객들의 이동 경로를 파악하여 새로운 사업기회를 찾

나 소프트웨어를 구매할 필요가 없어 저렴한 특징이 있다. 구글의 크롬북, 애플의 아이클라우드(icloud), KT의 유클라우드(Ucloud) 등이 대표적인 클라우드 서비스이다.

3) “빅데이터의 시대-경영을 과학으로”, 디지예코 포커스 2012.

고 있으며 자동차를 비롯한 제조업 기업들은 제품에 전자단말기를 설치하여 소비자의 불만사항을 파악하거나 제품에 이상이 발생하였을 시에 신속하게 문제점을 발견, 해결하고 있다.⁴⁾

하지만 무엇보다 빅데이터 활용이 기대되는 분야는 공공부문에 있다. 기후변화, 해양오염, 방사능 유출 등 광범위한 지역에 생성되는 대용량 정보를 실시간으로 처리하여 선제적으로 대응하거나 고객행동 분석을 통해 자금세탁 방지와 금융사기 예방을 위한 솔루션을 제공할 수 있다. 또한 테러 네트워크 정보를 수집하여 국가 안보 위협 요인을 사전에 감지, 대응할 수 있으며 상품의 시장가격과 교통정보의 실시간 분석으로 물가안정과 교통 정책 수립에 각각 도움을 줄 수 있다. 실제 미국 FBI는 1시간 안에 범인 DNA 분석을 위한 주정부 데이터 연계 및 빅데이터 실시간 분석 솔루션을 확보하고 있으며 빅데이터 분석을 통해 범죄자의 동선을 미리 파악하여 검거하는 성과를 거두기도 하였다. 영국과 호주는 공공부문의 정보를 오픈하여 일반인들의 참여를 장려하고 경제적 사회적 가치 증대를 꾀하고 있으며 싱가포르의 테러 및 전염병 등 미래의 불확실성을 대비하여 빅데이터 기반 위험 관리 계획을 추진하는 등 세계적으로 공공부문에서의 빅데이터 활용이 증가하고 있다.

그렇다면 우리나라에서의 빅데이터 현주소가 어떻게 될까? 스마트폰 보급률이 높아 엄청난 양의 데이터를 생산하고 있는 것과는 대조적으로 국내 빅데이터 시장은 거의 신생아 수준이라 할 수 있다. 일부 포털사이트에서 데이터 처리 기술 향상으로 업무 생산성이 높아진 정도이고 SK텔레콤의 티맵(Tmap) 네비게이션 정도가 거의 유일하다. 공공부문에서는 지난해 재난 전조감지, 구제역 예방, 맞춤형 복지 실현과 물가관리, 과학기술·의료 선진화 등의 분야에 빅데이터 활용방안을 제시하긴 하였지만 데이터 수집단계도 시작하지 못한 상황에서 빅데이터 상용화까지는 상당한 시간이 필요해 보인다.

글로벌 IT 기업 출현을 주도하게 될 빅데이터

현재 구글, 아마존, 페이스북, 애플이 빅데이터 선도기업으로 꼽히며 작금의 빅데이터 열풍을 주도하고 있다. 최근 구글이 내놓은 ‘빅 쿼리’와 같이 자체적인 기술 개발은 물론 기업 경영이나 공공부문에서 데이터 수집, 처리, 분석, 시각화 과정의 플랫폼 제공과 동시에 활용방안을 제안하고 있다.⁵⁾ 따라서 이제는 기업들이 관심만

4) 월마트, 리엔펑, P&G, 베스트바이, 포드, 히타치 등의 기업이 빅데이터 기술을 활용한 경영 기법을 사용하고 있다. “빅데이터: 산업 지각변동의 진원”, 삼성경제연구소 2012.5.2

5) 이들 인터넷 서비스 업체들은 공공부문에도 두각을 보이고 있는데 구글은 2008년 글로벌 금융 위기 도래와 경기 저점의 탈출 시기를 검색에 분석을 통해 미국정부보다 6개월 앞서 예측 했으며 현재 독감 유행 정보도 미국 정부보다 1주일 이상 앞서 하고 있다.

가지면 얼마든지 빅데이터 기술을 경영에 활용할 수 있는 시대가 왔고 빅데이터 활용여부에 따라서 시장의 판도가 바뀔 수도 있는 상황이 되었다. 최근 기업공개를 앞둔 페이스북의 가치가 지나치게 높다는 평가가 있지만 페이스북이 가지고 있는 개인정보 및 빅데이터 기술의 경제적 효과와 성장 잠재력을 고려하면 오히려 저평가 되어 있다고 볼 수도 있는 것이다.

실제 최근 영국에서 나온 보고서에 의하면 2012년부터 5년간 빅데이터로 영국 산업에 미칠 경제적 효과가 2160억 파운드(약 395조원)에 달할 것이고 5만 8000개의 일자리 창출을 가져올 것으로 전망하고 있다. 또한 맥킨지의 보고서에 따르면, 미국 의료 산업에서 빅데이터가 활용될 경우 비용효과가 연간 3조 달러에 달하는 등 유럽의 공공행정, 소매업, 제조업 등 분야에서 상당한 경제적 효과를 예상하고 있다. 그리고 금융, 공공기관, 소매, 여행, 교육 등 다양한 분야에서 빅데이터 취급 능력이 새로운 영역을 개척해 나갈 것으로 전망하고 있다. 이렇듯 빅데이터 시대에 뒤처지는 것은 앞으로 국가 경쟁력에 큰 위협으로 다가올 것이다.

이런 점에서 국내 인터넷 기반은 선진국 수준임에도 불구하고 아직까지 세계적인 IT 기업이 출현하지 못하고 있는 점은 상당히 아쉬운 대목이다. IT분야의 낮은 연봉과 열악한 근로 환경이 원인이라 할 수 있겠지만 무엇보다 광고 이외의 마땅한 수익구조가 없었던 점이 가장 크게 작용했었다. 그러나 이제는 빅데이터 기술에 대한 기업과 정부의 수요 증대에 따라 지속적인 수익확보가 가능할 것이며 우리나라 환경에 특화된 플랫폼 및 솔루션 제공은 글로벌 기업들과 경쟁력을 갖추는 지름길이 될 것이다. 즉 빅데이터 산업이 글로벌 IT 기업 출현을 주도할 가능성이 크다는 것이다.

빅데이터 선진국을 기대하며

글로벌 IT 기업 출현이나 빅데이터 상용화를 앞당기기 위해서는 넘어야 할 산이 많다. 사실 스마트폰 보급률이 높고 소셜 네트워킹을 비롯한 인터넷 사용도 세계적 수준인 우리나라는 기술적인 면에서 크게 뒤쳐져 있는 상황은 아니다. 특히 수도권에서는 인구밀도가 높고 인터넷 기반 시설이 잘 마련되어 있어 빅데이터 활용 가치가 높으며 오히려 세계시장을 선도할 잠재력이 충분하다고 볼 수 있다. 그러나 문제는 정보의 생성, 수집, 활용에 있어 규제수준이 높아 기술개발 단계까지 이르지 못하고 있다는 점이다. 앞으로의 빅데이터 시대는 실시간으로 생산되는 개인정보와 위치정보 활용에 그 핵심이 있는데 국내의 경우 개인정보보호법과 위치정보보호법에 의해 관련 사업은 규제에 묶여 성장이 멈춘 상태이다.

클라우드 서비스 관련 제도 정비도 중요한 과제이다. 빅데이터의 성공열쇠는 바로 클라우드 서비스에 있다고 해도 과언이 아니다. 클라우드 서비스를 통해 개인과 조직의 암호화된 방대한 데이터가 한곳으로 축적되고 실시간 데이터 분석이 이뤄지기 때문이다. 그러나 국경이 없는 인터넷 세상에 나라마다 규제차익이 발생하고 있어 규제 수준이 낮은 국가 주도로 클라우드 시장이 형성되고 있다. 국내에는 단순히 소비자들에게 스토리지 서비스를 제공하는 정도에 불과하지만 앞으로는 컴퓨터를 비롯하여 자동차, TV, 스마트폰 등의 다양한 단말기로부터 수집된 방대한 자료를 실시간 분석하여 소비자나 기업, 또는 정부가 원하는 정보를 제공해 주는 수준까지 확장될 것으로 예상된다. 따라서 우리나라도 해외와의 규제차익을 파악하고 그로 인해 피해가 발생하지 않도록 빅데이터를 위한 클라우드 서비스 관련 인프라 개선에 집중해야 한다.

마지막으로 무엇보다 중요한 것은 개인정보나 위치정보 수집과 활용에 부정적인 인식의 개선이다. 근래의 포털사이트와 금융기관의 개인정보 누출 사건, 그리고 아이폰의 위치정보 수집에 따른 소송이 이어지면서 개인정보 및 위치정보 수집은 더욱 민감한 상황이 되었다. 물론 사생활 보호와 보안, 그리고 ‘빅브라더’에 대한 우려는 매우 중요한 부분이다. 다만 앞으로 빅데이터의 중요성이 더욱 높아질 것이고 데이터의 암호화 기술이나 저장 기술이 발전할 것을 감안하여 득과 실을 따져볼 필요가 있다는 것이다. 미국의 경우 개인정보나 위치정보 수집에 대한 규제가 없는 반면, 엄격한 사후 관리 시스템을 도입하고 있어 IT 업체들이 정보 수집 및 저장 단계를 넘어 기술개발에 전념하는 것이 가능한 점은 우리에게 시사하는 바가 크다. 결국 ‘머니볼’이나 ‘왓슨’에서 보여준 패러다임의 변화도 결국 우리 스스로 원해야 가능한 것이라는 점을 명심하고 빅데이터의 중요성에 대한 인식과 공감대를 형성하는 것이 바로 빅데이터 선진국으로 가는 첫걸음이라고 하겠다.