

보건·복지 ISSUE & FOCUS

Korea Institute for Health
and Social Affairs

ISSN 2092-7117
제 230호 (2014-09) 발행일 : 2014. 03. 07

KIHASA 한국보건사회연구원
Korea Institute for Health and Social Affairs

정부 3.0과 빅데이터: 보건복지 분야 사례를 중심으로

빅데이터의 4가지 특성(Volume, Variety, Velocity, Value)과 정부 3.0의 세 가지 추진전략은 유기적인 연관성이 있음

여기서는 정부 3.0의 구현을 위한 핵심적 요소인 빅데이터의 중요성과 보건복지 분야에서의 빅데이터 활용 사례를 제시하고, 빅데이터 활용과 관련하여 제기되는 이슈와 한계를 살펴봄

빅데이터 분석 시 거시적 통찰과, 빅데이터 수집과 관리·활용에 관한 사회적 합의가 뒷받침된다면 개인정보보호의 문제로 인한 한계와 빅데이터를 효과적으로 활용함으로써 새롭게 창출할 수 있는 가치 사이의 상충관계는 어느 정도 해결 가능할 수 있을 것임



오미애
사회통계연구실 부연구위원

1. 정부 3.0과 빅데이터

■ 정부 3.0과 빅데이터 출현 배경

- 정부 3.0이란 공공정보를 적극 개방·공유하고, 부처간 칸막이를 없애 소통·협력함으로써 국민 맞춤형 서비스를 제공하고, 일자리 창출과 창조 경제를 지원하는 새로운 정부운영 패러다임을 의미함
 - 정부 3.0이 제기된 배경으로 IT 환경의 변화로 인한 빅데이터의 등장을 들 수 있음
- 빅데이터란 기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 의미함¹⁾
 - 빅데이터의 주된 출현 배경은 기하급수적으로 증가하는 데이터의 양과 IT의 발달임

■ 빅데이터의 4가지 특성(4V)과 정부 3.0의 추진전략은 유기적인 연관성이 있음

- 정부 3.0의 첫 번째 추진전략인 ‘소통하는 투명한 정부’는 빅데이터 활용 활성화를 위해 공공정보를 적극 공개하고 공공데이터를 개방하며, 이에 따라 활용 가능한 자료의 양(Volume)이 매우 방대해짐

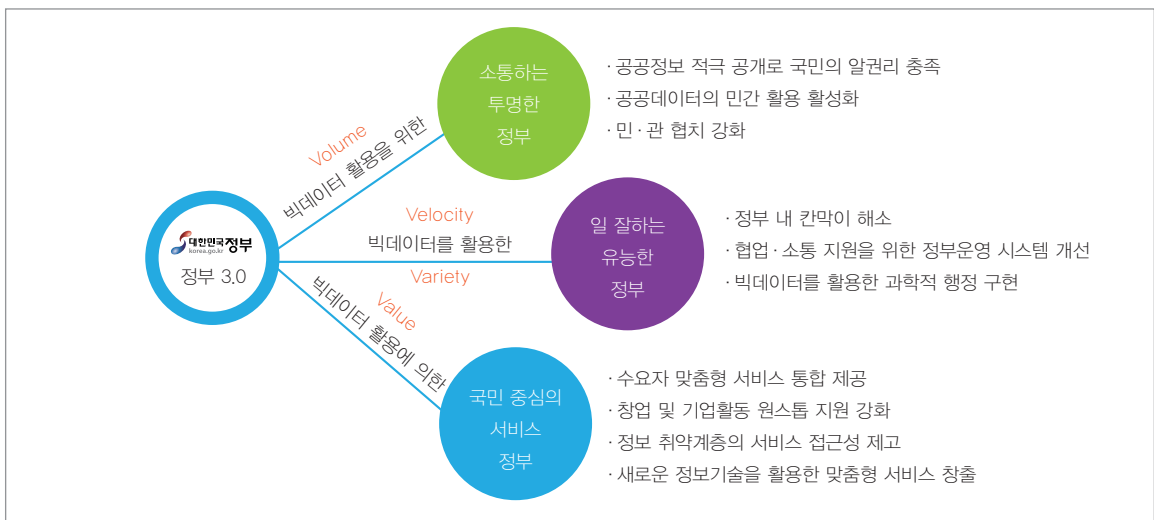
1) 출처: 위키백과.

- 정부 3.0의 두 번째 추진전략인 ‘일 잘하는 유능한 정부’는 빅데이터를 활용한 과학적 행정구현으로, 다양한 정보의 결합(Variety)이 가능하고, 정부운영시스템 개선으로 인하여 자료가 축적되는 속도(Velocity)가 빠름
- 정부 3.0의 세 번째 추진전략인 ‘국민중심의 서비스 정부’는 빅데이터를 활용한 분석결과를 기초로 수요자 맞춤형 서비스를 통합 제공함으로써 빅데이터의 목적인 새로운 가치(Value)를 창출함

■ 이 글에서는 정부 3.0의 성공적인 추진을 위한 핵심적 요소인 빅데이터의 중요성을 살펴보고, 보건복지 분야에서의 빅데이터 활용 사례를 제시함

○ 이와 함께, 빅데이터 활용 과정에서 제기되는 다양한 이슈와 고려되어야 하는 한계점을 짚어봄

[그림 1] 빅데이터 특성과 정부 3.0 추진전략



2. 빅데이터의 중요성

■ 빅데이터의 시대적 흐름

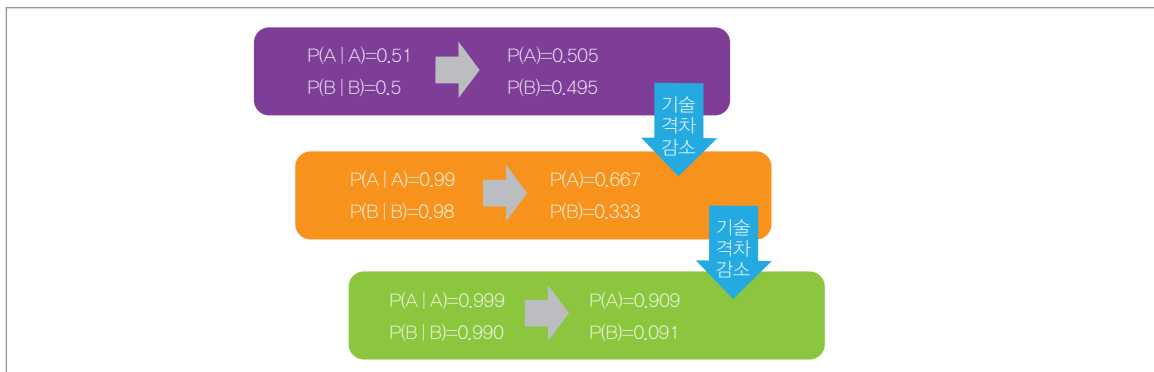
- 1990년대 중반부터 2000년대 중반까지는 ‘Data warehousing era’로, 사용자의 의사결정을 지원하기 위하여 축적한 데이터를 사용자 관점에서 고객정보, 교통정보, 서비스정보 등 주제별로 관리함으로써 주요한 정보를 적절한 시점에 제공할 수 있도록 함
- 2000년대 중반부터 2010년대 중반까지는 ‘Data convergence era’로, 데이터의 통합 관리를 통해 분석하고자 하는 소비자 등 대상자에 관한 입체적인 이해를 가능하게 함
- 2010년대 중반 이후는 ‘Big data era’로, 통합된 고객정보, 모바일, SNS의 결합을 통해 기존의 정형 데이터로는 파악하기 어려운 변화를 감지하고 소비자의 욕구를 즉시 확인하고 활용할 수 있게 됨

■ 치열한 경쟁사회일수록 빅데이터의 중요성은 증가함²⁾

2) 김용대, 조광현(2013). 빅데이터와 통계학, 한국데이터정보과학지, 24(5), pp.959-974.

- 데이터로부터 얼마나 많은 부가가치를 창출할 수 있는가의 관점에서, 자료분석으로부터 새롭게 얻을 수 있는 지식 또는 부가가치의 양과 그 차이는 크지 않음
- 하지만, 국가 간 또는 회사 간 기술격차의 감소로 인해 경쟁이 치열해지는 환경 속에서 이러한 작은 차이가 큰 결과를 가져올 수 있으며, 빅데이터의 활용은 이러한 차이를 만들어낼 수 있음
 - 예를 들어, A상품과 B상품만 존재하는 상황에서 A상품 고객이 A상품을 다시 구매할 확률이 0.51, B상품 고객이 B상품을 다시 구매할 확률이 0.50이라고 할 때, A상품의 시장 점유율은 0.505이며, B상품의 시장 점유율은 0.495임³⁾
 - 두 상품의 기술격차가 감소하여 A상품의 고객이 A상품을 다시 구매할 확률이 0.999이고, B상품의 고객이 B상품을 다시 구매할 확률이 0.990이라고 하면, A상품의 시장 점유율은 0.909이며, B상품의 시장 점유율은 0.091으로 두 상품의 시장 점유율 차이는 무려 90% 정도가 됨
 - 기술격차의 감소에도 불구하고 두 상품 사이에 존재하는 재구매 확률의 미세한 차이는 이와 같은 커다란 결과를 낳을 수 있으며, 이러한 작은 차이는 빅데이터의 분석을 통해 고객의 욕구 변화를 즉시 파악하여 품질 개선이나 홍보에 활용함으로써 창출할 수 있는 가치임

[그림 2] 경쟁사회에서 기술격차 감소로 인한 빅데이터 활용의 중요성



3. 정부 3.0 추진전략별 중앙행정기관의 빅데이터 활용 사례⁴⁾

■ 투명한 정부: ‘병원·약국 정보검색 서비스’

- 정부 3.0 공공정보 공개 차원에서 건강보험은 그 동안 축적된 빅데이터를 융합·활용하여 환자들이 궁금해 하는 병원별·질환별 진료비와 입원일수 등에 대한 정보를 스마트폰 어플리케이션과 건강보험심사평가원 홈페이지를 통하여 언제 어디서나 정보에 접근 가능하도록 제공
- 이러한 서비스를 통해 제공되는 정보는 다음과 같음
 - 병원별로 급성심근경색증, 고혈압·당뇨병 등 질병 치료 및 수술별 평가자료
 - 주요 수술별 진료비 정보 및 평균 입원일수 정보
 - 병원별 주소, 전화번호와 함께 주요 의료장비(CT, MRI) 보유여부, 야간 진료(소아과) 여부 등 정보

3) 시장 점유율은 마코프 체인의 정상분포로 계산한 결과임.

4) 안전행정부(2013), 정부3.0 우수사례집.

■ 유능한 정부: ‘건강검진결과 공유를 통한 운전면허 발급 시 신체검사 생략’

- 2013년 8월부터 운전면허 적성검사 시 국가건강검진(시력·청력) 정보를 활용하여 별도의 신체검사나 자료 제출없이 운전면허 신규취득 및 갱신이 가능해짐
- 연간 약 300억원의 경제적 절감효과(직접 160억, 간접 140억)
 - 직접적으로는 신체검사를 별도로 시행하지 않음에 따라, 신체검사 실시에 따른 검사 비용과 검사인력의 인건비 등을 절감함
 - 간접비용 절감효과는 별도의 신체검사에 소요되는 시간(약 20분)을 기회비용으로 환산하였음
- 이는 안전행정부, 보건복지부 등 5개 부처 및 공공기관이 부처간 칸막이를 제거하고 협업에 의해 빅데이터를 활용함으로써 구현된 통합 행정서비스의 대표적 사례임

〈표 1〉 운전면허증 발급시 건강검진정보 공동이용 협업기관 역할

관계기관	역할
안전행정부	- 행정정보 공동이용 시스템 개발 - 제도개선, 민원사무처리기준표 개정 및 관보 고시
보건복지부	- 국가건강검진 결과 중 경찰청에서 시력·청력 정보를 공동으로 이용할 수 있도록 승인 및 지원
국민건강보험공단	- 보유하고 있는 국가건강검진 결과 중 시력·청력 정보를 행정정보 공동이용 시스템을 통해 제공
경찰청	- 도로교통법 시행규칙 개정
도로교통공단	- 운전면허시험장을 통한 시력·청력 정보 공동이용 실시 - 고객정보 유출 방지 등 교육 실시

자료: 보건복지부 보도자료(2013. 8. 1)

■ 서비스 정부: ‘개인별 평생 맞춤형 건강서비스 제공’

- 국민건강보험공단은 전 국민의 출생에서 사망까지 건강보험자격 및 보험료 자료, 병의원 이용내역과 건강검진 결과, 가입자의 희귀난치성 질환 및 암 등록정보 등 10여년 동안 축적된 1조 3,034억 건의 빅데이터를 보유하고 있음
- 이러한 빅데이터를 활용하여 개인별·인구집단별 다양한 맞춤형 건강관리 서비스(건강상태, 4대 중증질환, 만성질환 등)와 개인별 진료내역, 건강검진, 장기요양서비스 등을 연계한 평생 건강관리 포털서비스를 제공할 예정임
- 검진결과와 진료이력 등 개인별 건강정보를 연계하여 건강위험 요인을 분석하고, 질병별 위험군에 따라 예방 프로그램 제공과 건강상담자를 연결하여 지속적인 개인별 맞춤형 사전예방 건강서비스를 실시할 예정임

4. 정부 3.0 구현을 위한 지방자치단체의 빅데이터 활용 사례: 광주광역시 광산구의 GIS와 공공데이터 연계

- 광주광역시 광산구는 보유하고 있는 각종 공공데이터와 지리정보시스템(GIS; Geographic Information System)을 결합하여 사회현상에 대해 정확한 원인을 분석하고 근본적인 대책을 수립하여 정책에 활용

○ 활용 가능한 공공데이터 실태조사를 통하여 데이터 현황을 파악하고, 조직 내·외부 7개 분야에서 380여종의 공공데이터를 수집 및 확보(2014. 1. 기준)

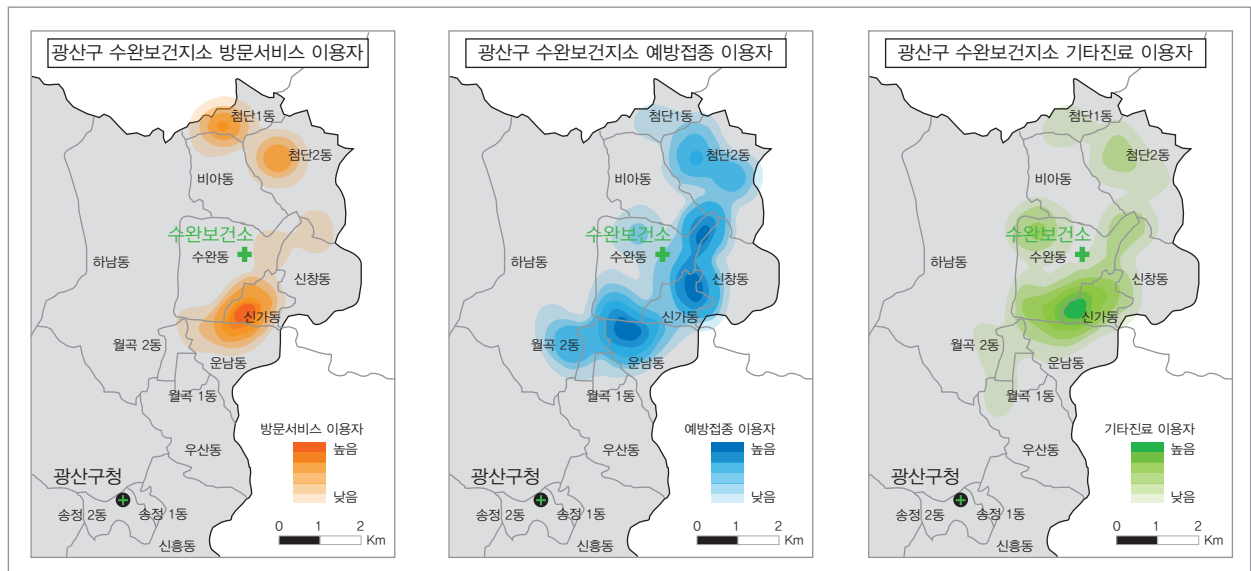
○ GIS를 활용하여 공공데이터를 시각화함으로써 복지, 교육, 안전 등 총 7개 분야의 현황 및 문제점을 분석하였고, 7개 분야, 52개 소주제, 205개의 정책 지도를 제작하여 지역주민들에게 제공함

■ 보건·건강을 그리다 – 보건소 이용자 분포 분석결과

○ 광산구 수완보건지소 지역의 만성, 재활 등의 기타 진료 이용자 평균 연령은 42.9세이고 30대의 경우 보건소 이용이 적음을 알 수 있음

○ 예를 들어, 신가동을 중심으로 이용자 분포를 살펴보면, 예방접종과 기타진료 이용자의 분포는 다른 양상을 보여주고 있으며, 신가동 서쪽 지역 주민의 기타진료 이용 비율이 가장 높음을 알 수 있음

[그림 3] 광산구 수완보건지소 이용자 분포



자료: GIS로 그리는 광산구 행정지도

■ 생애주기별 맞춤형 복지를 그리다 – 영유아 보육시설 입지 분석결과

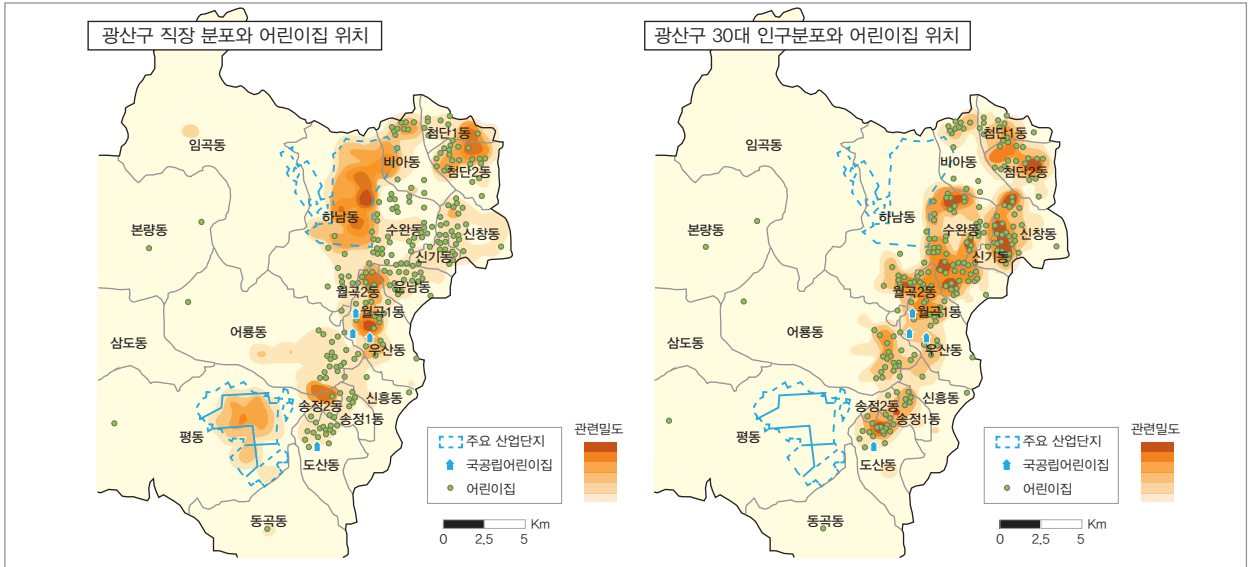
○ 광산구 내에 출근지가 있는 인구의 직장 분포와 영유아 관련 보육시설의 입지를 살펴보면 약 40% 정도의 근로자는 산업단지로 출근하고 있으나, 주요 산업단지 내에는 어린이집이 거의 없음

○ 30대 인구 분포와 영유아 관련 보육시설 입지를 살펴보면, 현재 어린이집은 인구가 밀집한 주거지 중심으로 배치되어 있음을 확인할 수 있음

○ 이러한 분석결과로부터, 향후 추가로 어린이집을 설치할 경우 아이의 부모가 근무하는 산업단지나 월곡, 송정 지역의 직장 밀집지역 인근에 배치하는 것을 고려할 수 있음

○ 현재 국공립어린이집은 전체 448개소 중 4개소로 영유아 보육시설 운영은 거의 전적으로 민간영역에 의존하고 있으므로, 산업단지 주변으로 신규 국공립어린이집의 배치를 고려해 볼 수 있음

[그림 4] 광산구 직장 및 30대 인구분포와 어린이집 위치



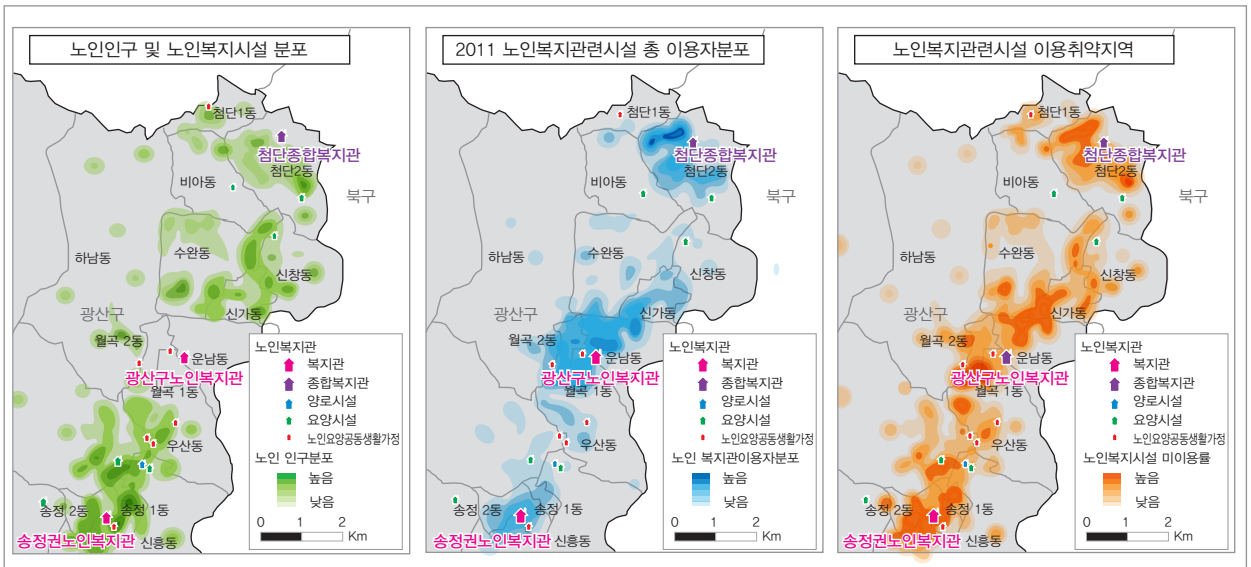
자료: GIS로 그리는 광산구 행정지도

■ 생애주기별 맞춤형 복지를 그리다 – 노인복지시설 취약지역 분석결과

○ 노인복지시설을 이용하지 않는 노인의 분포를 살펴보기 위해, 전체 노인인구의 분포에서 노인 관련 복지시설 이용자에 가중치를 적용하여 제거한 결과, 노인인구의 분포와 노인복지 관련 시설 이용자 분포가 균등하지 않고 편중되어 있음을 확인할 수 있음

○ 노인복지 관련 시설 이용 취약지역을 살펴보면, 진하게 나타난 송정동, 우산동, 월곡1동, 첨단2동은 노인인구에 비해 노인복지시설 이용률이 상대적으로 저조한 지역이므로 신규 노인복지 관련 시설의 설치를 고려해야 하는 것으로 판단됨

[그림 5] 광산구 노인인구 및 노인복지시설 분포와 이용자 분포

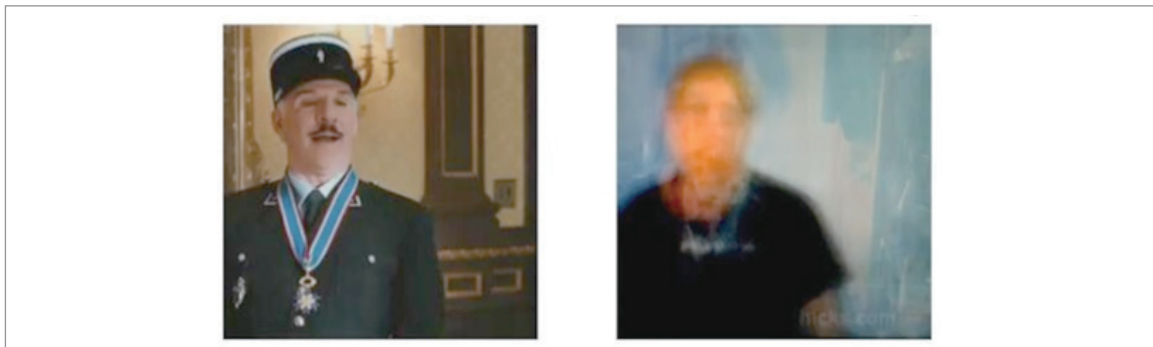


자료: GIS로 그리는 광산구 행정지도

5. 해외 빅데이터 활용 사례: 꿈을 찍는다 그리고 마음을 읽는다⁵⁾

- UC Berkeley의 Gallant Lab(Neuroscience 분야)에서는 움직이는 영상 정보를 처리하는 대뇌 신호를 모형화하여 뇌신호만을 이용해 시각 정보를 재구성
 - 실험자에게 몇 시간동안 영화를 보게 하면서 대뇌에서 일어나는 변화를 fMRI(functional magnetic resonance imaging)를 활용하여 뇌신호를 파악하고, 이러한 뇌신호와 실험자가 본 영상간의 관계를 모형화 함
 - 유튜브(Youtube) 동영상 자료 등을 이용하여 뇌신호와 같은 반응을 보이는 동영상 이미지의 평균을 가지고 영상을 재구성함

[그림 6] 실제 영상정보와 뇌신호로 재분석된 영상정보



자료: Gallant Lab UC Berkeley

- 이 연구는 유튜브와 같은 영상 빅데이터가 있기 때문에 가능하였으며, 대뇌의 영역별 기능을 확인하고 그것이 작동하는 방식을 알려주었다는 점에서 큰 의미가 있음
 - 이러한 기술은 뇌졸중 환자, 혼수상태 환자 등 언어적 의사소통이 힘든 사람들의 머릿속으로 생각한 것을 영상으로 나타낼 수 있어 마음에 어떤 일이 일어나는지에 대한 이해를 가능하게 만들었음
 - 또한, 뇌성마비 등 뇌병변 장애를 가진 사람들이 자신들의 마음을 컴퓨터가 읽을 수 있는 뇌-기계 인터페이스를 위한 기반을 마련하였다는데 의의가 있음

6. 빅데이터 활용의 이슈와 한계

- 과적합(overfitting) 문제⁶⁾
 - 방대한 양의 자료가 축적되어 있는 빅데이터는, 자료를 세분화하여 분석할 수 있게 만들어 세밀한 정책적 판단을 가능하게 함
 - 그러나, 현재 축적된 빅데이터를 활용하여 지나치게 정교한 분석을 수행할 경우, 미래에 대한 예측력은 오히려 떨어질 가능성이 있음

5) Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, 21(19), pp.1641-1646.

6) 김용대, 조광현(2013). 빅데이터와 통계학, *한국데이터정보과학지*, 24(5), pp.959-974.

- 빅데이터를 분석할 경우, 현재의 자료가 미래를 가장 잘 설명할 수 있다는 통계학의 기본가정이 성립되는 지를 살펴보아야 하며, 현재 빅데이터가 생산·축적되는 모집단 자체가 변화될 수 있다는 사실에 유의해야 함⁷⁾

○ 많은 경우 SNS와 같이 빅데이터를 생산하는 모집단은 시간에 따라 변화되며, 이러한 변화를 인지하는 것이 성공적인 빅데이터 분석의 필수적인 전제조건임

■ 차원의 저주(curse of dimensionality)⁸⁾

○ 차원의 저주란, 분석의 정확도를 어느 정도 유지하기 위해서 변수의 수를 증가시킬 때마다 필요한 데이터의 수는 기하급수적으로 증가한다는 것으로, 다양한 정보의 결합으로 자료의 차원이 커지면서 발생할 수 있는 현상임

○ 변수의 개수가 무수히 많아지면, 아무리 빅데이터라고 할지라도 감당하기 어려운 수준일 수 있으므로 이러한 차원의 저주를 피하기 위한 여러 가지 통계적 기법들을 고려해야 함

■ 개인정보보호와 법적 문제

○ 개인정보보호법의 제정 취지는 개인정보 침해로 인한 국민의 피해 구제를 강화하여 사생활의 비밀을 보호하고 개인정보에 대한 국민의 권리와 이익을 보장하는데 있음

○ 빅데이터 활용의 가치를 논의하면서 동시에 개인정보보호와 관련된 문제를 제기하는 것은 데이터를 수집하는 공공기관 및 민간 기업들이 어떤 경로를 통해 어떤 데이터를 수집하는지, 그리고 수집된 데이터를 어떻게 관리하고 이용하는지에 대한 정보제공자의 불안감 때문일 수 있음⁹⁾

- 최근 사회적 이슈가 되고 있는 주요 신용카드사의 대량의 개인정보 유출 사건은 이러한 측면에서 빅데이터 구축 및 활용을 제약하고, 정부 3.0의 구현을 어렵게 만드는 한계로 작용할 수 있음

○ 그럼에도 불구하고, 빅데이터 수집과 관리·활용에 관한 투명성 제고를 바탕으로, 개인정보가 포함된 데이터에 대한 충분한 수준의 익명화(masking)를 거쳐 분석에 활용한다면, 개인정보보호의 문제로 인한 한계와 빅데이터를 효과적으로 활용함으로써 새롭게 창출할 수 있는 가치 사이의 상충관계는 어느 정도 해결 가능할 수 있을 것임

7) 예로, 트위터 자료를 분석하고 이를 주식투자에 이용하여 좋은 수익률을 기록하였던 어느 빅데이터 회사가 2012년 대선 이후로 많은 어려움을 겪었는데 그 이유는 2012년 대선 이후에 진보적 성향의 트위터 이용자들이 대거 트위터를 탈퇴했기 때문으로 사후 조사됨.

8) Hastie, T., Tibshirani, R., Friedman, J.(2001), The elements of statistical learning, New York, Springer, p.22.

9) 서울대학교 법학연구소 · 빅데이터센터 공동학술대회자료. 개인정보보호의 법과 정책, p.33.

집필자 | 오미애 (사회통계연구실 부연구위원) 문의 | 02-380-8267

발행인 | 최병호 발행처 | 한국보건사회연구원

(122-705)서울특별시 은평구 진흥로 235 | TEL 02)380-8000 | FAX 02)352-9129 | <http://www.kihasa.re.kr>

한국보건사회연구원 홈페이지의 발간자료에서 온라인으로도 이용하실 수 있습니다. <http://www.kihasa.re.kr/html/jsp/publication/periodical/focus/list.jsp>