

<http://dx.doi.org/10.17703/JCCT.2023.9.3.655>

JCCT 2023-5-75

## 빅데이터 분석은 사회과학 연구에서 방법론적 혁신인가?

### Is Big Data Analysis to Be a Methodological Innovation? : The cases of social science

이상기\*

SangKhee Lee\*

**요약** 사회과학 분야에서 빅데이터 연구는 기존의 사회과학 연구방법을 보완하는 역할을 하고 있다. 사회과학자들이 선호하는 서베이 및 실험 방법이 주로 회상 기억에 의존하여 다소 부정확하다면 빅데이터는 실시간 기록이라 보다 정확하다. 기존의 사회과학 연구가 시간과 비용 등의 이유로 연구대상을 전수 조사하기보다 표집에 의한 표본 조사를 주로 하는 것과 달리 빅데이터 연구는 전수에 가까운 데이터를 분석한다. 그렇지만 시간의 흐름에 따라 사회 분위기가 변할 수 있고, 연구대상도 동일하지 않아 연구의 반복 및 재현은 둘 다 쉽지 않다. 무엇보다 기존의 사회과학 연구가 '이론-방법-데이터'의 삼각구조가 튼튼한 데 반해 빅데이터를 활용한 분석은 이론의 빈약함을 보이고 있어 심각한 문제다. 과학적 설명논리로서의 이론이 없으면 연구결과를 얻고서도 제대로 해석하지 못하거나 온전히 활용 할 수 없기 때문이다. 그러므로 빅데이터 연구가 진정한 방법론적 혁신이 되기 위해서는 새로운 이론(블랙박스)을 창출하기 위한 연구자들의 노력과 함께 빅 씽킹(big thinking)이 필요함을 제안했다.

**주요어** : 빅데이터, 방법론적 혁신, 연구의 삼각구조, 블랙박스, 빅 씽킹, 융복합적 사고, 근시안/원시안

**Abstract** Big data research plays a role of supplementing existing social science research methods. If the survey and experimental methods are somewhat inaccurate because they mainly rely on recall memories, big data are more accurate because they are real-time records. Social science research so far, which mainly conducts sample research for reasons such as time and cost, but big data research analyzes almost total data. However, it is not easy to repeat and reproduce social research because the social atmosphere can change and the subjects of research are not the same. While social science research has a strong triangular structure of 'theory-method-data', big data analysis shows a weak theory, which is a serious problem. Because, without the theory as a scientific explanation logic, even if the research results are obtained, they cannot be properly interpreted or fully utilized. Therefore, in order for big data research to become a methodological innovation, I proposed big thinking along with researchers' efforts to create new theories(black boxes).

**Key words** : bigdata, methodological innovation, triangular structure of research, black box, big thinking, convergent thinking myopia/hyperopia

\*정회원, 부경대학교 미디어커뮤니케이션학부(언론정보전공) 교수 Received: March 5, 2023 / Revised: March 14, 2023

(BK21 “지역사회 혁신성장을 위한 데이터과학 응용인재 양성사업단장”) Accepted: May 8, 2023

접수일: 2023년 3월 5일, 수정 완료일: 2023년 3월 14일

\*Corresponding Author: lsngk@pknu.ac.kr

게재확정일: 2023년 5월 8일

Dept. of Communication Studies, Pukyong National University, Korea

## I. 서론

‘빅데이터’라는 용어가 처음 등장한 1997년 이후 이를 이용한 연구와 실제 활용이 봇물 터지듯 이루어지고 있다[1]. 일반적으로 빅데이터는 인공지능을 학습(machine/deep learning)시키기 위한 원자료(raw data)로 여겨졌지만, 빅테크 기업들(아마존, 구글, 페이스북 등)이 고객 기반 데이터로 다양한 서비스와 마케팅 방법을 개발하고 있다는 사실이 알려지면서 학계에서도 주목하기 시작했다[2][3]. 특히, 2008년은 빅데이터 연구의 분기점으로 일컬어진다[4]. 당시 구글은 하루 5억 건이 넘는 검색기록을 통해 독감추세(GFT: Google Flu Trends) 서비스를 공개했는데, 이는 공식자료보다 일주일에서 열흘 정도 앞서 독감 유행상황을 예측한 것이었고 실제 적중률도 95%를 넘었다. 이에 뉴욕타임즈와 CNN 등 주요 언론이 관련 뉴스를 보도했고, 네이처(Nature)는 빅데이터 특집호를 냈다. 아울러 크리스 앤더슨(C. Anderson)이 와이어드(Wired)지에 “숫자가 스스로 자기 얘기를 할 수 있다. 상관관계가 인과관계를 대체할 것이고, 일관성을 갖춘 모형이나 통합된 이론 등의 체계적 설명 없이도 과학은 발전해나갈 수 있다.”는 도발적 주장(이론의 종언: The End of Theory)을 던진 해이기도 했다[5].

현재 빅데이터와 관련한 연구는 인공지능 분야뿐만 아니라, 인류학, 커뮤니케이션 및 미디어 연구, 컴퓨터 연산 사회과학, 범죄학, 행동경제학, 개발경제학, 정치과학, 인지심리학, 문화 및 사회심리, 발달심리, 과학의 과학(science of science), 사회학, 환경 및 기후변화, 지속가능(식량), 문화의 진화, 유전학, 인지 신경과학, 사회 및 정동 신경과학, 신경학, 정신의학, 역학 등 인간(행위)과 관련한 전 분야로 확산하고 있는 중이다[6]. 국내에서도 2010년 전후로 빅데이터 관련 연구가 출현하기 시작했고[7], 2013년에는 사단법인 ‘한국빅데이터학회’가 출범했다. 초창기 공학 분야(전자, 정보통신, 컴퓨터 공학)와 사회과학 분야(법학, 일반사회과학, 경영학, 신문방송학)에서 많은 연구가 진행되었지만, 최근에는 거의 모든 학문 분야에서 빅데이터 관련 연구가 수행되고 있다.

그럼에도 불구하고, 크리스 앤더슨이 제기한 도발적 문제 제기(“통합된 이론의 체계적 설명 없이도 과학은 발전할 수 있다. 빅데이터 정보는 원리상 어떤 이론/이

론가도 필요로 하지 않는다.”)에 대해 명쾌한 해답 혹은 해결을 위한 단초를 제공하는 연구들은 찾아보기 힘들다. 물론 프랑스 철학자인 스티글러(B. Stiegler[8])와 같이 “이론적 지식을 상실하고 있는 시대”를 우려하는 목소리도 없진 않다. 이 와중에 챗GPT라는 생성형 AI(generative AI)가 등장했다. 챗GPT는 언어와 문자로 표현할 수 있는 것이라면 논문, 보고서, 문학, 코딩 등 분야를 가리지 않고 수준 높은 결과물을 제시한다[9]. 그렇지만 어떤 질문에도 답을 하게끔 설계되었기에 일부 부적절한 답(인간의 편향 반영)을 내놓는 경우도 있고, 있지도 않은 학술논문이 발간된 것(환각: hallucination)처럼 응답하는 경우도 있어, 특히 학술 논문은 반드시 확인·검증이 필요하다. 이러한 상황에서 본 논문은 빅데이터 분석이 사회과학 ‘방법론(methodology)’으로 정립되기 위한 ‘이론’ 창출 방안을 제안하고, 빅데이터 분석 방법이 사회과학 연구에 끼친 긍정적인 영향과 더불어 한계를 지적하고, 향후 사회과학 연구에서 빅데이터의 적절한 활용 방안을 제시하고자 한다.

## II. 연구의 삼각구조

### 1. 이론-방법-데이터

제대로 된 연구는 ‘이론-방법-데이터’가 잘 어우러져야 한다. 과학의 엄밀성을 추구하는 ‘과학의 과학(science of science)’ 분야에서는 삼각측량(triangulation of research)을 권유할 정도다. 즉, 연구 문제를 해결하기 위해 이론, 방법, 데이터셋 각각을 셋 이상 사용할 것을 권장하는 방법이다. 이를 통해 연구결과와 타당성과 신뢰성을 높이고, 연구 편향을 완화시킬 수 있다고 본다[10].

그런데 사회과학연구에서 이론적 사고는 충분하지 않다(<그림1.>의 오른쪽에서 실선이 아닌 점선으로 표시한 이유). 먼저, 추상화가 부족한 이유는 실증주의(경험주의)의 영향 때문이라고 할 수 있다. 즉, 데이터로 입증되지 않으면 어떤 이론도 거부하는 경향이 강하다. 그러나 자연과학의 세계에서는 먼저 이론이 수립된 후 몇 십 년이 지나 관측(입증)된 사례가 허다하다(아인슈타인의 ‘상대성 이론’이나 ‘블랙홀’의 존재 등). 다음으로 개념화가 부족한 경우는 특히, 빅데이터 분석에서 자주

발견되는데, 분석 알고리즘이 공개되지 않아서 혹은 컴퓨터가 제시하는 언어나 개념을 인간이 이해할 수 없는 경우가 있기 때문이다. 그렇다고 인공지능에게 설명을 요구하는 XAI(eXplainable Artificial Intelligence) 역시 인간의 이론적 사고에 역행하는 것이다. 실증주의 이전의 사회과학자들은 ‘이상 사회’에 대한 비전도 서슴없이 내놓았다. 지금 우리에게 필요한 것은 이러한 ‘사회과학적 상상력’이다. 다만 특정 이론을 종교적 신념 차원으로 격상시키는 오류는 절대 피해야 할 것이며, 이론과 현실(방법-데이터)의 간극을 메우는 노력을 끊임없이 경주해야 할 것이다.

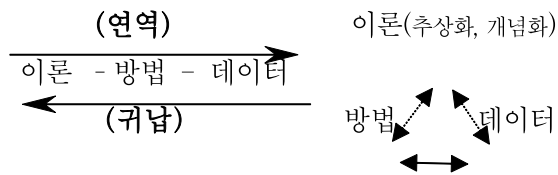


그림 1. 연구의 삼각구조  
 Figure 1. Triangular Structure of Research

연구에서 이론이 필요한 이유는, 설령 데이터가 이론에 맞지 않더라도 왜 그런지에 대한 이해(돌발사건, 편향, 통제되지 않은 변수 등)가 가능하다는 데 있다. 이제까지의 사회과학 연구는 연역적 사고에 기초한 ‘양적 연구(주로 이론/가설을 검증하기 위한 목적의 통계적 방법)’와 귀납적 사고에 기초한 ‘질적 연구(참여관찰/인터뷰 등을 통한 이론의 발견)’가 무수히 수행되었지만 과학적 체계를 흔들기는커녕 더욱 풍성하게 했다. 즉, 이론에 기반 한, 혹은 이론을 (재)정립하기 위한 기존의 연구들은 하나의 이론에서 무수히 다양한 연구사례가 나오더라도, 반대로 다양한 연구사례를 통해 이론을 재해석하거나 새롭게 창출하더라도 이론(과학적 설명 논리)을 수정, 보완, 축적하는 데 기여할 수 있었다(<그림 1>의 왼쪽 상황을 묘사). 그러나 빅데이터를 이용한 연구(주로 귀납적 방식)는 데이터에서 출발하여 방법(python, R)의 정교화 수준에 머물고 있다. 다시 말해 이론(추상화, 개념화)의 단계까지 나아가지 못하고 있다. 기껏해야 다양한 패턴(pattern) 속에서 공통점을 찾기 위한 노력이 경주되고 있는 상태다(<그림 1>의 오른쪽 상황을 묘사).

이런 점에서 “이론을 통해 해석하기 전에는 어떤 데이터도 결코 알지 못한다”는 도이치(D. Deutsch)의 주

장은 설득력이 있다[11]. 예컨대 아래의 <그림 2>는 전자 망원경을 통해 우주 전파를 광대역 3개 주파수 채널로 동시에 수신한 관측 결과 그래프다. 전자 망원경으로 관측하면 우리가 일반적으로 보는 하늘과 별의 모습이 전혀 아니다. 그렇지만 천문학자들은 우주와 별에 대한 이론이 있기 때문에 이 자료를 디지털로 변환하고, 다른 신호와 조합하고, 컴퓨터로 분석할 수 있다. 그 결과 사진이나 영상이 만들어지기도 한다. 일반인들의 이해를 돕기 위해 ‘가짜 색’을 입혀서 더욱 그럴 듯하게 표현할 수도 있다. 우리가 제임스 웹 망원경을 통해 바라보는 모습도 그러하다(<그림 3>).

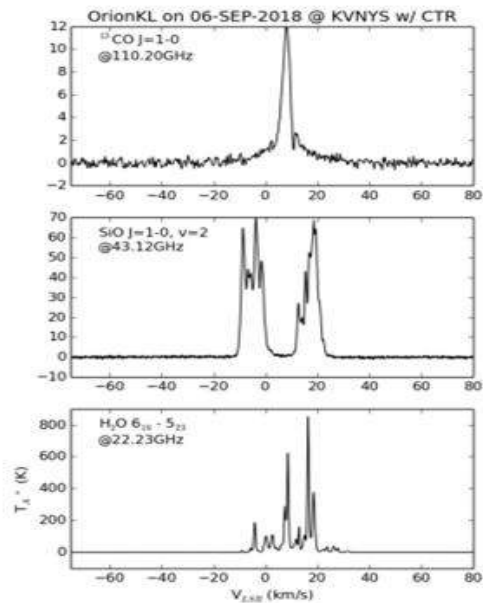


그림 2. 오리온별 관측 결과 (출처: 한국천문연구원)  
 Figure 2. Observation results of Orion (Source: Korea Astronomy & Space Science Institute)

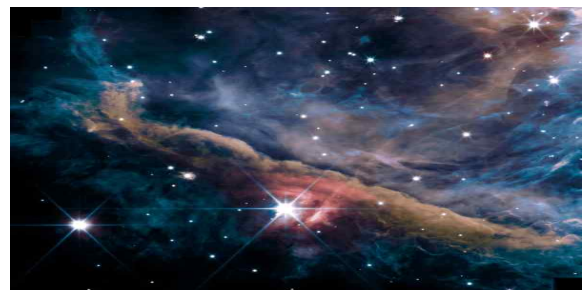


그림 3. 제임스 웹 우주망원경이 촬영한 오리온 성운 모습 (출처: NASA)  
 Figure 3. The Orion Nebula as seen by the James Webb Space Telescope (Source: NASA)

마찬가지로 전자현미경이 관찰한 미시적인 세계도 이와 유사한 경로를 따라 진행된다(<그림4>).

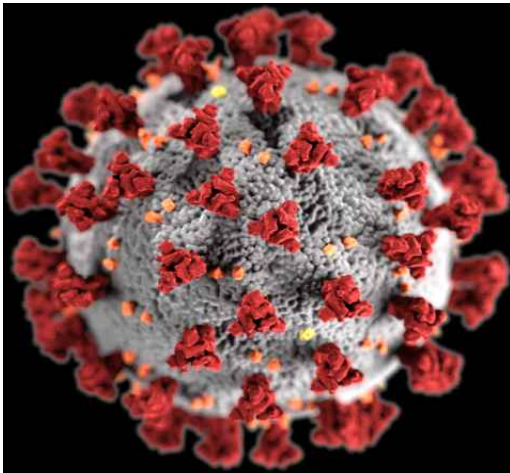


그림 4. 코로나19 바이러스 (출처: ScienceTimes, CDC)  
Figure 4. COVID-19 (Source: ScienceTimes, CDC)

그런데 빅데이터 분석을 통한 사회에 대한 설명(보다 엄밀하게는 데이터에 대한 시각적 표현)은 아직 <그림2>와 같은 상황에서 벗어나지 못했다. 물질적 실체를 가진 자연세계와 달리 사회연구는 사랑, 정의, 공정과 같이 인간행동 및 사회현상을 설명하되 실제(reality)를 규정하기 힘든 요소도 있기 때문이다. 또한 도표, 그래프, 다이어그램, 시각적 데이터 마이닝(visual data mining), 네트워크 시각화 등 다양한 표현 수단이 제시되고 있지만 하나의 통일된 그림을 제한할 수 있는 지도 의문이다. 차칫 단일 모델은 ‘지구촌의 균질화’[12]를 초래함으로써 다양성의 가치를 훼손할 수 있기 때문이다. 더 나아가 예외적 데이터(outlier)를 처리하는 문제도 고려해야 한다. 예컨대 네슬레 소유인 미국의 이유식 회사 거버(Gerber)는 20세기 중반에 서아프리카에서 이유식을 판매하면서 미국과 유럽처럼 ‘웃는 아기 사진’이 붙은 통을 사용했다. 하지만 아프리카의 일부 문화권에서 통에 붙은 사진은 음식의 내용물이라고 여겼다. 즉, 이유식이 ‘아기를 위한’ 식품이 아니라 ‘아기로 만든’ 식품이라고 생각했다는 것이다[13]. (빅)데이터 분석만으로는 아프리카의 이유식 판매량이 정체된 이유를 낮은 경제수준, 음식 선호의 차이 등으로 오관할 가능성이 크다. 이런 측면에서 빅데이터 분석 및 자료에 대한 해석은 다양한 이론과 더불어 연구자들의 협력이 필요하다.

## 2. 사회이론의 블랙박스

필자가 빅데이터 분석에 관심을 갖게 된 계기는 2016년 미국 ‘분석 사회학회(Analytical Sociology)’에서 올린 트윗에 대한 리트윗이 해시태그(#)를 통해 상당한 반향을 불러일으킨 사례를 접하면서다(<그림5>).



그림 5. 미국 ‘분석 사회학회’의 J. Elster(1985) 트윗  
Figure 5. Tweet by Analytical Sociology (J. Elster, 1985)

트윗의 내용과 같이 J. Elster[14]는 미시(개인의 행위)와 거시(사회구조)를 아우를 수 있는 이론 체계야말로 사회과학의 블랙박스(black box)를 여는 것이라고 주장했다. 예컨대 M. Weber의 <프로테스탄트 윤리와 자본주의 정신>의 설명구조는 ‘프로테스탄트 윤리’라는 거시적 변인이 ‘자본주의 제도’라는 거시적 변인에 영향을 미쳤다는 것이다. 그렇지만 이러한 ‘방법론적 전체주의(methodological holism)’는 유럽 각 나라에는 설명력이 있지만, 기독교 국가가 아닌 나라에서 자본주의 제도를 받아들인 현상은 설명할 수 없다(한국, 일본, 중국 등의 동아시아 국가와 중동의 이슬람 국가 등). 따라서 미시적 수준(방법론적 개인주의; methodological individualism)에서 개인 행위자들의 프로테스탄트 윤리 수용에 상응하는 윤리와 동기를 살펴보아야 하며, 또 개인 행위자들의 집합적 행위의 산물로서 자본주의 제도가 도입된 배경과 과정을 살펴봐야 온전한 설명이 될 수 있다[15].

이러한 ‘분석 사회학’ 혹은 ‘합리적 선택 사회학’은 엘스터가 재직했던 시카고 대학에서 1980-90년대를 풍미했다. 시카고 대학에서 석·박사 학위를 받은 김용학은 국내에 <사회구조와 행위(1992)>, <사회 연결망 이론(2004)>, <사회 연결망 분석(2007)> 등을 잇달아 내놓으면서 1990년대와 2000년대의 한국 사회(과)학자들을 매료시켰다. 그렇지만 당시의 방법론적 도구로서는 분석 방법과 데이터의 한계가 있었다. 이런 와중에 빅데이터 분석 방법이 등장했으나, 많은 사회(과)학자들

은 이제야 비로소 블랙박스를 열 수 있지 않을까라는 기대감이 생겼을 수 있다. 2016년의 작은 소동은 그 일면을 보여준 것이다.

그러나 아직까지, 빅데이터를 활용한 사회연구 분야에서 통합된 과학 이론은 정립되지 못했다. 그럼에도 불구하고 많은 사회과학자들이 빅데이터에 주목하고 있으며, 그 활용 가능성을 다각도로 모색하고 있는 것도 사실이다. 서울대학교에 빅데이터 센터가 설립된 지 몇 개월도 되지 않아 150명 이상의 교수가 자발적으로 참여 의사를 밝혔다고 한다[16].

또한 제러드 다이아몬드(J. Diamond), 유발 하라리(Y. Harari), 조지프 헨리크(J. Henrich), 스티븐 핑커(S. Pinker), 제러미 리프킨(J. Rifkin), 마이클 셔머(M. Shermer)와 같은 학자들이 다시금 거대 이론(grand theory)을 시도하고 있다는 점도 주목할 만하다. 즉, 장기간의 인간 역사와 사회문화적 현상을 방대한 데이터와 자료를 통해 설명하고자 하기 때문이다. 따지고 보면 칼 마르크스(K. Marx), 막스 베버(M. Weber), 에밀 뒤르켐(E. Durkheim) 등 사회학의 세 비조(鼻祖)들 역시 거대 이론가였다. 유럽에서 수학하고 미국 사회학을 창시한 탈코트 파슨스(T. Parsons)도 그러했다. 하지만 그의 제자이자 미국 사회학회장을 역임했던 로버트 머튼(R. Merton)에 의해 사회학적 연구는 거대하고 포괄적인 사회현상에 대한 설명보다, 비교적 제한된 중범위 이론(middle-range theories)에 근거한 설명이 적절하다는 의견이 대체로 받아들여졌고, 여전히 사회과학의 주류 패러다임으로 자리 잡고 있다. 상징적으로 미국 및 세계 여러 나라에서 사회과학 연구방법론 교재로 삼고 있는 얼 배비(E. Babbie)의 <사회조사방법론>의 최신판[17]에서 빅데이터 연구방법은 하나의 장(chapter)도 아닌 하위 절(section)로 취급되고 있다. 즉, 내용분석, 비교역사연구, 기존 통계자료 분석 등과 같은 수준이라는 것이다. 그 제목마저도 ‘비개입적 온라인 연구’이고, 온라인 연구의 한 수단으로서 빅데이터 연구방법을 소개하고 있다.

그러므로 빅데이터 분석 및 이와 관련한 방법론이 혁신적으로 패러다임의 전환을 이루기 위해서는 반드시 ‘통일과학(unified science)’의 체계를 확립해야 할 뿐만 아니라, 그 체계 속에서 다양한 연구결과물들이 축적되어야 할 것이다[18]. 아쉽게도 현재까지의 학계 동향을 살펴보면 이러한 시도가 미미한 상태다. 더욱이

빅데이터 관련 연구는 데이터의 방대함(volume)보다 관련 기술의 발달 속도(velocity)를 쫓아가기가 점점 힘들어 보인다. 스코트 갤러웨이[19]가 이러한 현상을 잘 설명했는데, 원 제목(*Post Corona*)보다 번역서의 제목(<거대한 가속>)이 훨씬 더 적절해 보인다. 본 논문에서 통일된 체계를 갖춘 빅데이터 방법론을 제시하기란 지면의 한계가 있어, 연구방향을 제시(4절)하는 것에서만 의미를 찾고자 한다. 아울러 빅데이터 연구방법론(혹은 관련 사회이론)을 정립하기 위한 연구들이 보다 활성화되기를 기대한다.

### III. 빅데이터 분석 방법의 기여와 한계

많은 사회과학 연구자들이 빅데이터에 관심을 기울이는 이유 중의 하나는 기존의 사회과학 연구방법 역시 완전하지 않다는 것을 인식하고 있기 때문일 것이다[20]. 우선, 사회과학자들이 선호하는 서베이 및 실험 방법은 주로 회상 기억에 의존한다. 즉, 구매한 입장권(영화, 뮤지컬, 오페라, 박물관, 미술관 등), 방문한 장소, 지출한 내역, 보낸 시간 및 횟수(관람 요일 및 시간, 참여 횟수) 등은 꼼꼼하게 기록하지 않는 한 어렵짐작으로 응답하는 경우가 대부분이다. 둘째, 시간과 비용 등의 이유로 연구대상을 전수 조사(census)하기보다 표집에 의한 표본 조사를 하기 때문에 여기에서 오류(표집 오차 및 표준오차)가 발생할 수밖에 없다. 셋째, 시간의 흐름(특히, 기술의 급격한 발달)에 따라 사회 분위기가 변하고, 연구대상도 동일하지 않아 연구의 반복(replication) 및 재현(reproducibility)이 쉽지 않다. 넷째, 연구자가 직접 데이터를 수집하기보다 훈련받은 연구보조원이나 다른 전문 기관을 통해 얻는 경우가 많다. 다섯째, 응답자들이 ‘허위’ 응답을 하는 경우도 종종 있다. 아동과 청소년을 대상으로 한 조사에서 부모의 직업을 정확하게 모르는 경우가 있고, 성과를 재촉하는 환경에서 ‘과잉 실적/과소 불량’을 보고할 수도 있다.

이에 비해 빅데이터는 전수 내지 전수에 가까운(volume) 데이터를 얻을 수 있고, 실시간 기록(validity) 및 빠른(velocity) 분석이 가능하며, 정형 데이터뿐만 아니라 비정형(unstructured)/반정형(semi-structured) 데이터 등 다양한(variety) 종류의 데이터를 처리할 수 있다. 이러한 빅데이터의 속성은 기존의 사회과학 데이터

를 보완하는 측면이 분명 있다(기존 사회과학 방법론의 셋째부터 다섯째까지의 문제는 공통의 문제로 여전히 남아 있지만).

그렇지만 앞서 언급한 바와 같이 대부분의 빅데이터 분석은 귀납적으로 접근하기에 분석 결과가 연구하고자 하는 개념 혹은 논제와 정확히 일치하지 않는 경우가 생길 수 있다. 또한 변수들 사이의 인과관계보다 상관관계에 머물러 우연에 의한 관계의 가능성도 내포한다. 예를 들어 비겐(T. Vigen)에 의하면 미국에서 ‘과학-우주-기술’ 등에 대한 재정지출 변수와 자살 변수 사이에 상당한 정도의 상관관계가 발견되었지만 이 변수들 간에 관련성을 찾기란 쉽지 않다[21]. 주가지수(S&P 500)와 방글라데시의 베타 생산량 사이에 높은 상관관계를 찾은 경우도 있다[22].

한편, 빅테크 기업들은 수억 명에 달하는 이용자 관련 빅데이터를 갖고 있지만, 영업기밀 등의 이유로 데이터를 완전히 공개하지 않고 있다[23]. 더욱이 2016년 미국 대통령 선거에서 케임브리지 애널리티카(Cambridge Analytica)가 수백만 명의 페이스북 이용자 정보를 부적절하게 이용한 사실이 드러났고, 연이어 분노·중요 등의 혐오 발언과 허위 정보를 방치함으로써 10대들의 정신건강에 해를 미쳤다는 내부 고발이 잇달았다. 이 때문에 페이스북이 메타로 이름을 바꾼 것은 주지의 사실이다.

빅데이터 연구가 인터넷이나 스마트폰 이용자들이 늘어남으로써 가능해진 만큼 조지프 헨릭(J. Henrich)이 서구 중심의 사회과학 연구가 WEIRD 편향에 빠질 수 있음을 경고한 점도 살펴볼 필요가 있다[24]. 즉, 서구의(western), 교육수준이 높은(educated), 산업화된(industrialized), 부유한(rich), 민주적인(democratic) 시민들(특히, 실험 연구의 참여자는 대부분 연구자 소속 대학생이나 대학원생들인 경우가 많다)을 대상으로 한 연구 결과를 중앙아시아, 아프리카, 중남미의 시민들에게 그대로 적용하기엔 맞지 않는 경우가 많다는 것이다. 리처드 니스벳(R. Nisbett,[25])이나 한스 로슬링(H. Rosling,[26]) 등도 이러한 문제점을 이미 지적한 바 있다. 특히, 빅데이터 관련 연구가 페이스북, 트위터, 인스타그램, 틱톡 등과 같이 전 세계인이 사용하는 소셜 미디어인 경우에는 이 점을 특별히 주의할 필요가 있을 것이다. 아울러 세계적으로 소셜 미디어를 사용하지 않는 인구가 사용하는 인구보다 더 많은 사실도 염두에

두어야 한다.

#### IV. 결론: 빅 씽킹(Big Thinking)의 필요성

서론에서 언급한 크리스 앤더슨의 도발적 주장을, 솔직히 말해 두어 해 전에 처음 접했다. BK21 사업단 주최의 빅데이터 관련 특강에서 초청 강사로부터였다. “이론과 이론가가 필요 없다니? 도대체 누가 이런 말도 안 되는 소리를 했나요?” 존재론적 위기감을 느끼고 초청 강사(교수)에게 절박하게 질문했지만 “위낙 유명한 말인데, 누가 말했는지는 잘 모르겠다”는 답변을 들었다. 이 논문을 시작하게 된 발단이였다. 그런데 해당 발언의 당사자가 크리스 앤더슨이라는 사실을 알고 더 놀랐다. 2008년 이후에 나온 그의 책도 이미 두 권(<프리>, <메이커스>)이나 읽었지만 발언의 예후를 전혀 못 느꼈기 때문이다. 본 논문을 준비하면서 크리스 앤더슨의 도발에 무신경한(사회과학계의 분위기에 더더욱 놀랐다.

학계에서 빅데이터 관련 분석 및 연구결과 출판에만 급급하고, 방법론 및 이론 체계 구성에 소홀한 이유는 학계의 평가 구조와도 무관하지 않다고 본다. 즉, 대학교수의 업적 평가가 연간 단위로 이뤄지다보니 우리 학계에서 생산되는 대부분의 연구는 1-2년 단위의 단기 과제에 집중되어 있다. 또한 사회과학 연구는 앞서 지적한 바와 같이 중범위 이론적 사고가 지배하고 있어 대부분이 집단이나 조직 차원의 연구에 머물러 있다. 개념의 정교화 및 조작적 정의에 충실하려면 연구의 범위를 넓히기보다 좁히는 게 더 적절하다. 특히, 학생들에게는 연구 범위가 넓어지면 넓어질수록 분석이 힘들고 연구해야 할 분량과 고려할 대상들도 늘어나니까, 최대한 좁혀서 생각하고 연구하도록 가르쳐왔다. 또한 연구결과를 연구 범위 내에서만 해석하는 엄격함(구성의 오류fallacy of composition에 빠지지 않도록)이야말로 연구자의 기본자세임을 강조해왔다.

그런데 빅데이터 연구에서도 이러한 연구 태도가 여전히 올바른 것인지 반문하게 된다. ‘새 술은 새 부대에’라고 한 것처럼 방법이 바뀌면 인식론과 방법론(방법에 대한 이론 및 철학) 역시 바뀌어야 하지 않을까? 이러한 측면에서 최소한 다음과 같은 연구 자세 및 연구 방향을 제시하고자 한다. 먼저, 빅데이터 연구는 데이터의 양보다 데이터가 생산되는 속도가 더 감당하기 힘들다

고 했다. 속도가 빨라지면 빨라질수록 우리는 근시안(myopia)에서 벗어나 보다 멀리, 넓은 시야를 가져야 한다. 레비트(T. Levitt)는 일찍이 마케팅이 실패하는 이유가 시장을 보지 않고 상품만 보는 근시안 때문이라고 했다. 또한 소비자의 근원적 욕구(needs)를 보지 못하고 당장 원하는 것(wants)에만 급급한 것도 마케팅 근시안이라고 했다[27]. 빅데이터 연구는 데이터가 방대하고 빠른 속도로 생산되는 만큼 근시안을 벗어나 원시안(hyperopia)을 가져야 할 것이다. 무엇보다 자료의 수집기간만큼은 구할 수 있는 한 최대한 확장시켜야 할 것이다. 금융시장에서 20년의 추세 분석도 부적절한 것으로 드러났기 때문이다. 즉, 2008년의 세계 경제위기는 그 전 20년 동안의 데이터가 호황, 성장 추세를 보였기에 추락할 것은 전혀 생각지도 않은 상태에서, 과생상품을 무분별하게 남발한 데서 비롯되었다. 낙관적인 전망이 지배하는 환경 속에서 비관론자는 외톨이였다[28].

둘째, 빅데이터 연구는 학제간 연구 혹은 융복합 연구가 반드시 필요하다고 본다. 또 홀로 연구하다가 분석 상황 자체가 변하면 그간의 노력이 수포로 돌아갈 수도 있다. 그러므로 적절한 협업을 통해 빠른 시일에 연구결과를 확보하는 것이 좋다. 그런데 교육적 관점에서는 전공 분야(field)에 대한 정체성을 확립한 이후에 영역(domain)을 확장하는 것이 좋은지, 처음부터 다양한 영역을 접하면서 전공 분야를 선택하는 것이 좋은지는 불분명한 상태다. 다중 언어(융복합) 습득이 좋은 건 당연하지만 모국어(전공지식)를 완전히 익힌 후에 외국어(다른 전공)를 익히는 게 좋은지, 모국어가 미숙한 어린 나이 때부터 다중 언어 환경에 노출되는 것이 좋은지 불분명한 것과 유사하다.

셋째, 빅데이터 연구자들 스스로 이론을 정립하기 위해 별도의 노력을 기울여야 한다. 최소한 빅데이터 분석을 통해 연구논문을 작성하고자 한다면 분석과 관련한 기존 연구에 대한 깊이 있는 탐구가 필요하다. 그런데 대학원생들의 연구 행태를 지켜보면 방법을 익히는 데 많은 시간을 소비하지만, 정작 이론에 대한 탐구는 소홀하고 지엽적인 관련 논문 인용에만 그치는 경우가 많다. 방대한 자료를 다루는 만큼 다양한 시각과 관점이 필요함에도 그에 대한 학습이 미진한 것이다. 기계는 방대하고 엄밀한 학습(machine/deep learning)을 하면서 점점 더 스마트 해지는 데, 정작 인간은 시간과 지능의 한계로 얕은(shallow) 학습[29]에만 머문다면 인

공지능과 인간지능의 관계는 역전될 수밖에 없을 것이다. 연구 분야에 한정해서라도 보다 집중적으로 이론의 폭과 깊이를 더할 필요가 있으며, 이러한 목적에서라면 챗GPT와 같은 인공지능을 활용하는 것도 무방할 것이다.

이와 더불어 국내의 빅데이터를 활용한 사회과학연구도 국제 비교연구로 확대해 나가기를 바란다. 국내의 연구결과가 ‘좁은 사회’ 혹은 ‘닫힌 사회’의 연구결과로 그쳐서는 안 될 것이기 때문이다. 또한, 대학교수의 연구업적 평가 및 연구 지원이 1년 단위에서 다년 단위로 확대될 필요가 있다. 연구결과물에 대한 수량 평가 외에 연구과정에 대한 질적 평가를 병행하면 될 일이다.

빅데이터 분석은 결국 ‘과거와 현재’의 방대한 자료를 수집, 분석하고 진단(diagnosis)하는 데 유용할 뿐이다. 그러나 이론은 보이지 않는 것을 보게 하고, ‘미래’를 예측할 뿐만 아니라 해결책을 찾게끔(prescription) 도울 수 있다. 빅데이터가 난무하는 시대에 우리에게 더욱 절실히 필요한 것은 이러한 이론을 찾기 위한 빅씽킹(big thinking: 원시안, 융복합적 사고, 사회이론 창출을 위한 블랙박스 열기 등)이 아닐까 한다.

## References

- [1] Big-data studies of human behaviour need a common language. *Nature*, Vol. 595, pp.149-150. July 2021.
- [2] Seung-Hyeog Moon, “A Study on Securing Global Big Data Competitiveness based on its Environment Analysis,” *The Journal of the Convergence on Culture Technology*, Vol. 5, No. 2, pp.361-366, 2019. DOI:10.17703/jcct. 2019.5.2.361
- [3] Jong-Youl Hong, “Cultural Big Data Platform and Digital Management: Focused on Cultural Contents Industry,” *International Journal of Advanced Culture Technology*, Vol. 10, No. 3, pp.287-294, 2022.
- [4] Shin-Kap Han, “Doing Social Sciences in the Age of Big Data: Rethinking Analytical Strategy in the Changing Data Environment,” *Korean Journal of Sociology*, Vol. 49, No. 2, pp.161-192, 2015.
- [5] C. Anderson, “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete,” *Wired*, June 2008. <https://www.wired.com/2008/06/pb-theory/>

- [6] The future of human behaviour research. *Nature Human Behaviour*, Vol. 6, pp.15-24, January 2022.
- [7] S.T. Kim, "Possibilities and limitations of using big data in medial research," Korean Association for Broadcasting & Telecommunication Studies, Spring Conference Proceedings, pp.119-131, 2017.
- [8] B. Stiegler, *La société automatique 1: L'avenir du travail*, Fayard, 2015.
- [9] J.H. Jung, "The emergence of ChatGPT and challenges in the field of artificial intelligence," *Issues and Perspectives*, NARS, No. 2067, March 2023.
- [10] P. Bhandari, "Triangulation in Research: Guide, Types, Examples," *Scribbr*. March 2023. <https://www.scribbr.com/methodology/triangulation/>
- [11] D. Deutsch, *The Beginning of Infinity: Explanations That Transform the World*, UK: Penguin Books, 2012.
- [12] D. Howes, ed., *Cross-Cultural Consumption: Global Markets, Local Realities*, UK: Routledge, 1996.
- [13] G. Tett, *Anthro-Vision*, Across Publishing Group, Inc., 2021.
- [14] J. Elster, *Making Sense of Marx*, Cambridge: Cambridge University Press, 1985.
- [15] J. Coleman, "Social Theory, Social Research, and a Theory of Action," *American Journal of Sociology*, Vol. 91, No. 6, 1309-1335, 1986.
- [16] Jae-Hyun Lee, "Big Data and Social Sciences: Epistemological and Methodological Issues," *Communication Theories*, Vol. 9, No. 3, pp.127-165, 2013.
- [17] E. Babbie, *The Practice of Social Research* (15<sup>th</sup> Ed.), Cengage Learning, 2020.
- [18] T. Kuhn, *The Structure of Scientific Revolution* (2<sup>nd</sup> Ed.), IL: University of Chicago Press, 1970.
- [19] S. Galloway, *Post Corona: From Crisis to Opportunity*, Portfolio, 2020.
- [20] H. Becker, *Evidence*, IL: University of Chicago Press, 2017.
- [21] Tyler Vigen, "Spurious Correlation," 2014. <http://www.tyklervigen.com>.
- [22] D. J. Leinweber, "Stupid data miner tricks: overfitting the S&P 500," *The Journal of Investing*, Vol. 16, No. 1, pp.15-22, 2007.
- [23] M. J. Salganik, *Bit by Bit: Social Research in the Digital Age*, Princeton University Press, 2018.
- [24] J. Henrich, *The WEIRD People in the World*, Picador Paper, 2021.
- [25] R. Nisbett, *The Geography of Thought: How Asians and Westerners Think Differently... and Why*, Free Press, 2004.
- [26] H. Rosling, *Factfulness: Ten Reasons We're Wrong About the World—and Why Things Are Better Than You Think*, Flatiron Books, 2018.
- [27] T. Levitt, "Marketing Myopia," *Harvard Business Review*, Vol. 38, pp.45-56, 1960.
- [28] N. Taleb, *The Black Swan: The impact of the highly improbable* (2<sup>nd</sup> Ed.), London: Penguin, 2010.
- [29] N. Carr, *The Shallows: How the Internet Is Changing the Way We Think, Read and Remember* (Re-issue edition), Atlantic Books, 2020.

※ 이 논문은 부경대학교 자율창의학술연구비 (2021년)에 의하여 연구되었음