

# 데이터와 인간, 그리고 질문

자유전공학부  
최예지

카카오톡이나 페이스북 담벼락에서 일어나는 대화를 제대로 분석할 수 있는 알고리즘만 개발되면 인간관계에 관한 심리학이나 사회학 이론은 무용지물이 될까? '롱테일 법칙'이라는 용어를 고안한 Chris Anderson은 2008년에 「이론의 종언」이라는 글을 와이어드 지에 게재했다.<sup>1)</sup> 그 글의 요지는 간단하다. 이제 사회과학을 포함한 모든 과학자가야 할 일은 거대한 데이터를 관찰하고 패턴을 찾아내고 상관관계를 찾는 것이다. 굳이 모델을 개발하고, 인과관계를 발견하려 애쓸 것 없다. 이론보다 데이터가 중요하고, 상관관계만 찾으면 충분하다. 기존 과학적 원칙이 무너진 것이다. Frank Pasquale의 『블랙박스 사회』에 따르면 이는 단순한 가설이 아닌 듯 하다. 우리는 이미 빅 데이터로 사회현상과 개인을 분석하는 시대로 들어섰으며 활용 분야는 무궁무진하다. 빅 데이터에 좌우되는 현실은 기존의 사회과학이 새로운 국면에 들어서도록 요구한다.

Frank Pasquale이 언급한 분야 외에도 빅 데이터의 활용은 감정의 영역까지 잠식할 것이다. 인간은 소리로 표현되는 언어뿐만 아니라 상대방의 말투나 표정, 제스처를 통해 순간적으로 드러나는 감정을 서로 주고받는다. 충분한 기술이 발전된다는 전제 하에 빅 데이터 분석 알고리즘 역시 이러한 감정 통찰을 수행할 수 있다. 상대방의 분석된 감정 데이터를 통해 우리는 타인의 감정에 더 잘 공감할 수 있게 될 수도 있다. 영화 Her의 인공 지능 OS까지는 아니더라도 웨어러블 디바이스를 통해 sns 기록과 신체적 변화를 데이터로 변환하여 감정을 파악할 수 있을 것이다.

'감성 컴퓨팅(Affective Computing)'은 바로 이와 관련된 연구를 다루는 학문분야다.<sup>2)</sup> 이 개념은 1995년 미국 MIT (메사추세츠공과대학교)의 로잘린드 피카드 박사가 처음으로 제안했다. 감성 컴퓨팅은 기계가 인간을 이해하고 서로 대화할 수 있는 방법을 연구한다. 아펙티바(Affectiva; MIT 과학자들이 창업한 스타트업)와 피카드 박사 연구진이 함께 개발한 모바일 앱은 이미 표정을 하나의 정보로 처리하여 감정을 파악하는 알고리즘을 실현하고 있다. 이 앱은 눈썹을 치켜뜨는 모습, 얼굴을 찌푸리는 모습을 인식한다. 이걸 통해 긍정적인 표정을 짓는지, 또는 부정적인 표정을 짓는지 인식할 수 있다.

데이터 마이닝을 통한 감정 파악은 새로운 소비 시장의 개척 및 타인의 감정 이해를 바탕으로 한 사회적 갈등 해소 효과를 창출할 수 있다. 겉으로 드러나지 않는 감정을 파악한 뒤 유사한 감정을 느끼고 있는 다른 이들과 연결해주는 서비스를 만들 수도 있다. 또한 감정 분석에 그치는 것이 아니라 해당 감정에 나타날 행동들 역시 빅 데이터 기술로 수집할 수 있다. 이는 제조업, 마케팅과 같은 경제적 영역뿐만 아니라 잠재적 갈등 조정, 조직 내 개인의 극단적 행동 방지와 같은 사회심리학적 분야에서 다시 활용될 수 있다.

문제는 파스칼이 지적했듯이 데이터 분석 알고리즘은 특정 세력의 선입견을 배울 수 있는 가능성은 열려있으나 일반 시민은 이를 확인할 여부가 없다는 것이다. 감정 데이터 해석은 그 과정이 베일에 가려진 블랙박스이다. 따라서 데이터 추출 과정에서 어떠한 이익집단이 정치적, 경제적 압력을 가해 특정 개인이나 타집단의 감정 분석을 조작 및 왜곡해도 시민은 알 리가 없다. 감정의 왜곡은 사회적 불공정성을 약할 수 있다. 특정 인종을 혐오하는 집단이 있다고 하자. 인간의 존엄성 아래 그들은 사회적으로 규탄받아야 마땅하다. 그러나 그들이 경제적으로 감정 빅 데이터 분석 연구소를 세우고 자신들이 혐오하다는 인종의 감정이 정신분열증 환자와 비슷한 증상의 감정을 갖고 있으며 그러한 감정들이 어떠한 사회적 문제를 야기하는지를 끊임없이 언론에 내보낼 수 있다. 이는 일반 대중이 무의식적으로 특정 인종에

1) Chris Anderson, 「The End of Theory: The Data Deluge Makes the Scientific Method Obsolete」, [http://archive.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory), Wired Magazine, 06.23.08

2) 허완, 「인간의 감정을 이해하는 컴퓨터가 세상을 바꾼다」, [http://www.huffingtonpost.kr/2015/05/26/story\\_n\\_7440352.html](http://www.huffingtonpost.kr/2015/05/26/story_n_7440352.html), 허핑턴포스트코리아, 2015년 05월 28일

대한 혐오를 합리화할 수 있는 위험한 상황으로 이어질 수 있다.

감정 데이터 분석에 의해 부정적 진단을 받은 개인 및 집단은 암묵적 낙인이 찍혀 사회적, 경제적 기회를 박탈당할 수도 있다. 채용 과정에서 지원자의 sns 페이스북 타임라인에 자살 관련, 우울증 관련 글들이 많이 올라왔다면 채용하지 않을 수 있다. 또한 감성 컴퓨팅 결과에 오류가 충분히 생길 수 있음에도 불구하고 데이터 분석 결과 정신병에 걸릴 확률이 높다 나오면 채용 기회 뿐만이 아닌 타인과의 인간 관계를 맺는 기회도 잃을 수 있다. 아무리 보안벽을 설치해도 대기업들이 막대한 자본력을 투자하여 개개인의 감정 파악 분석 기기에 침투하여 정보를 빼낼 수 있다. 막대한 자본력으로 감정 분석 결과를 토대로 하여 사회 구성원들을 분류하는 것이다. 이러한 현상은 감정 분석을 바탕으로 한 암묵적 계급의 형성으로 이해할 수 있다. 이 과정에서 특정 세력의 의도로 아무런 이상이 없는데도 비정상적 감정을 소유하고 있다 판단받은 이들이 최하위 계급으로 분류될 수 있다는 것이 큰 문제이다.

감정 분석을 빌미로 개인의 사생활을 검열, 감시하는 가능성도 무시할 수 없다. 빅데이터를 모으는 주체의 감시와 검열에 대한 대중의 불안은 지난 해 카카오톡 유저의 텔레그램으로의 ‘사이버 망명’ 사건에서 드러났다. 심지어 마크 주커버그, 에릭 슈미트를 비롯한 빅 데이터 소스의 큰 원천지 중 하나인 sns의 창업자들도 프라이버시 침해에서 벗어날 수 없었다고 한다. 우리의 감정이 데이터화된다면 이 역시 유출될 수 있다. 다른 이에게만큼은 절대 들키지 않고 싶은 감정이 내가 모르는 사람이 알게 된다면 그보다 더 큰 사생활 침해는 없을 것이다. 더 큰 문제는 소득이 높은 이들은 강력한 보안 시스템을 사들일 여건이 되지만 소득 하위층은 그렇지 못하다는 것이다. 극단적인 시나리오를 예상하면, 부자만이 자신의 감정을 숨길 수 있는 시대가 올 수도 있다.

해결책으로는 감정 데이터에 시한부를 설정하고 데이터가 ‘살아 있을’ 동안 당사자들의 승인 아래 활용되는 시스템의 구축을 들 수 있다. 즉 우선 데이터에 유효기간을 설정하고, 유효기간이 지난 데이터를 원천 봉쇄하는 데이터 플랫폼 밖의 블랙홀을 만드는 것이다. 이 블랙홀에는 데이터 이외에 사회적으로 문제를 일으킨 데이터 추출 기술 또한 봉쇄될 수 있다. 이러한 기술적 해결책 외에 막대한 자본력과 기술력이 집적된 대형 기업들을 중심으로 데이터 경로를 공개하도록 하는 법적 제도를 마련해야 한다.

데이터와 기술의 수동적 소비자가 되지 않기 위해 우리는 계속해서 질문을 던져야한다. 어려운 질문들. 데이터의 숫자를 계산하는 단계에서 나아가 그것들을 이해하기 위해. 사실은 오용되기 용이하고 인간은 왜곡된 사실에 쉽게 의미를 부여한다. 데이터는 스스로 의미를 부여하지 않는다. 인간이 하는 것이다. 이에 사업가, 고객, 환자, 시민으로서 우리는 비판적 사고능력에 집중하여 상당한 시간을 소비해야 할 책임이 있다. 그렇지 않으면 거대 용량의 데이터를 빛의 속도로 가공할 수 있게 됨에 따라 빅 데이터는 지금보다 더욱 빠르게 나쁜 결정이 내려지도록 할 수도 있다.

올바른 질문은 데이터의 규모가 아닌 데이터를 생각하는 방법에 주의를 기울이는 것으로부터 시작되어야 한다. 세계는 인간이 얼마나 많은 양의 데이터가 있는지에 대해 끊임없이 쏟아지는 이야기들로 둘러 쌓여있다. 그러나 빅 데이터 해석은 규모가 전부가 아니다. 축적된 데이터들 사이에는 데이터가 움직이는 속도가 있고 데이터들 종류의 다양성도 문제가 된다. 이미지, 문자, 영상, 음성등이 그것들이죠. 이러한 이질적인 종류의 데이터들을 통합시키는 것은 그것들이 사람들에게 의해 창조되었고 그것들에는 맥락이 필수적으로 필요하다.

이를 실천하기 위해 보다 더 많은 시간을 인류학이나 사회학, 그리고 사회과학 수사학, 철학, 도덕 등에 투자하여 데이터 해석에 비판적으로 질문할 수 있는 능력을 길러야 할 것이다. 컴퓨터 자동 번역이 단 하나의 경우만을 가정하고 맥락을 고려하지 않듯이, 데이터 스스로 자신의 앞 뒤 맥락을 분석할 수는 없다. 그러나 인문, 사회학적 사고는 빅 데이터에 관한 매우 중요한 앞뒤 맥락을 제시하여 정보 수용자들이 비판적으로 생각하는 사람들이 되도록 도와준다. 데이터 수용 과정에서 인문사회학적으로 쟁점이 될 수 있는 부분을 발견할 수 있다면, 더 이상 데이터 규모와 기술은 문제가 되지 않는다. 대신 수용자들은 비판적 시각과 논쟁을 통해 기계적 데이터 해석 속의 확증 편향이나 잘못된 연관성을 발견할 수 있다.

## 참고 문헌

허완, 「인간의 감정을 이해하는 컴퓨터가 세상을 바꾼다」,

[http://www.huffingtonpost.kr/2015/05/26/story\\_n\\_7440352.html](http://www.huffingtonpost.kr/2015/05/26/story_n_7440352.html), 허핑턴포스트코리아, 2015년 05월 28일

Chris Anderson, 「The End of Theory: The Data Deluge Makes the Scientific Method Obsolete」,

[http://archive.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory), Wired Magazine, 06.23.08

Frank Pasquale 「The Black Box Society: The Secret Algorithms That Control Money and Information」,

Harvard University Press: 1 edition, 02. 05. 15

Frank Pasquale,

「7 Ways Data Currently Being Collected About You Could Hurt Your Career or Personal Life」

[http://www.huffingtonpost.com/frank-pasquale/data-collected-hurt-career-personal\\_b\\_6110682.html](http://www.huffingtonpost.com/frank-pasquale/data-collected-hurt-career-personal_b_6110682.html)

, The World Post, 01. 06. 15