



자유주의 정보 19-174

본 내용은 아래 기사 및 칼럼 내용을 요약 번역한 내용임

Per Bylund,

No, "Big Data" Can't Predict the Future

12 July, 2015

빅데이터는 미래를 예측할 수 없다

우리는 '공짜의 시대'에 살고 있다고 말할 수 있다. 온라인으로 제공되는 서비스들은 대부분 무료로 제공되기 때문이다. 물론 그게 정말로 '공짜'인 것은 아니다. 소비자가 사용료를 직접 지급하지 않는 것이다. 오늘날의 인터넷 회사들은 소비자에게 서비스를 무료로 제공하면서 수익을 창출한다. 우선 많은 소비자를 끌어 모은 뒤 광고를 해서 돈을 버는 방법이 있고, 최근에는 소비자로부터 수집한 '빅데이터'를 이용해 돈을 버는 방법이 부상하고 있다.

빅데이터가 이전 시대의 데이터와 구별되는 가장 큰 특징은, 방대한 양의 자료들을 수집하고 분석하는 데 필요한 한계비용이 종종 0에 도달할 정도로 매우 낮아졌다는 것이다. 이에 따라 빅데이터가 곧 비즈니스의 미래를 선도할 것으로 많은 사람이 예상한다.

다만 여기서 문제가 하나 발생하는데, 사람들이 사회적 예측에서 대수의 법칙(law of large numbers)과 빅데이터에 지나치게 의존하게 된다는 것이다. 물론 통계는 수학적으로 평균을 형성할 수 있다. 하지만 그것이 곧 현실세계를 정확하게 반영하고 있다 할 수 있는가? 자연과학의 영역에서는 확실하게 그렇다. 통제된 환경에서 반복되는 실험을 통해, 우리는 분명 문제점을 발견하고 잘못된 설명을 제거해낼 수 있다. 적어도 우리가 실험요인들을 통제하는데 충분한 능력이 있다고 가정한다면 말이다.

그렇다면 사회과학의 영역에서는 어떤가? 하이에크는 우리 시대를 '과학주의의 시대'라고 명했다. 사회과학까지 자연과학의 방법론을 따르고 있기 때문이다. 우리는 빅데이터가 생산, 물류, 판매 등을 완전히 변화시키리라 전망한다. 빅데이터를 통해서 공급업체가 고객들의 요구를 더 잘 공략하고 그들이 무엇을 원하는지 '예측'할 수 있다고 생각하기 때문이다. 예컨대 아마존닷컴은 당신의 구매 이력과 유사한 구매 이력을 가진 다른 사람들의 검색결과를 토대로 당신에게 추천물품을 제안한다.

인간행동의 예측 불가능성

빅데이터에 기초해 광고하는 것은 매우 효과적이다. 당신이 관심을 거의 표하지 않은 것들을 광고대상에서 배제한다면, 당신이 더욱 선호하는 물품들을 광고하게 될 확률이 늘어날 것이다. 빅데이터를 통해 그들은 당신에게 물건을 팔기 위해서 어떤 것을 광고해야 하는지 보다 정확하게 식별하게 된다.

그러나 유사성은 예측과 같지 않다. 통계를 통해 발생확률이 제로에 가까운 사건들을 배제하는 것은 예측이 아니다. 예측은 "한가지 혹은 두 가지의 매우 가능성 높은 결과를 제외한 '모든 결과'를 정확하고 완전하게 배제하고, 그 결과가 사실로 밝혀지는 것"이다. 이 기준을 충족하지 못한다면 그저 게임이나 추측에 불과한 것이다. 물론 그것은 상당히 정밀한 추측이지만, 근본적으로 예측이라 말할 순 없다.

빅데이터는 어디서 실패하는가

마이크로소프트는 매우 진보된 알고리즘과 많은 양의 수집된 데이터에 기초해 스포츠를 비롯한 여러 사안의 결과를 예측하고자 하는 예측 엔진을 만들었다. 놀랍게도, 이 것은 매우 잘 작동했고, 피파 월드컵의 결과를 완벽하게 예측하는 데 성공했다. 이 사례는 빅데이터가 미래를 예측할 수 있다는 증거로 사용될 수 있는가?

그렇지 않다. 마이크로소프트는 오스트리아학파, 더 정확히는 인간행동학의 교훈을 학습하고 있을 뿐이다. 오스트리아학파는 인간행동에 상수가 없으며, 따라서 사회현상에 대한 예측이 불가능하다는 것을 파악한 유일한 학적입장이다. 하이에크는 예측은 언제나 제한적이며, 정확한 예측은 불가능하다고 말했다. 예를 들어, 우리는 수요-공급의 원칙과 같은 경제법칙에 의존해 특정 상황에서 가격이 상승하거나 하락할 것이라 예측할 수 있지만, 그 가격이 정확히 얼마가 될지는 결코 말할 수 없다. 스포츠나 다른 경쟁에서도 이는 정확히 적용된다. 더 좋은 기록을 가진 팀이 항상 이기는 것은 아니다.

사회과학은 다르다

사회과학의 주제는 복잡한 사회현상이다. 사회현상은 사회적 상호작용으로서 인간행동의 복합적 결과이다. 스포츠도 마찬가지이다. 양 팀 선수들의 의욕과 집중력, 사생활과 사고방식, 그리고 팀 내 기강 등의 요소가 경기결과에 영향을 미친다. 더 강한 팀이 더 약한 팀을 과소평가하며 대충 경기에 임하다 패배할 수도 있으며, 더 약한 팀이 반드시 이겨

야 한다는 부담을 느끼지 않고 공격적인 전술을 펼치다 승리할 수도 있다. 마이크로소프트의 예측엔진 역시 이러한 요소와 고군분투하고 있다. 최근 이 예측엔진은 미국 미식축구 리그의 14개의 경기 중 단 4번의 결과만 맞히기도 했다.

예측엔진을 통한 추측이 단순한 직감이나 동전 던지기보단 정확하므로, 만약 당신이 내기한다면 이것을 참조하는 것은 현명하다. 하지만 빅데이터를 통한 '추측'이 실제로 결과를 '예측'해낸다고 말하긴 어렵다. 매번 결과를 맞히는 것은 행운이 필요하다. 통계가 우리를 완벽한 예측의 경지로 이끌 수는 없다.

번역: 김경훈

출처: <https://mises.org/library/no-big-data-can%E2%80%99t-predict-future>