# 기계학습된 소프트웨어에 대한 데이터 과학의 역할

KPMG Korea

이광춘 상무

기계(Machine)라 하면 기계장치를 떠올릴 수 있지만, 영어로 머신(machine)은 인공지능을 탑재한 컴퓨터를 의미한다. 자동화 수준을 기계의 도움 없이 모든 결정과 행동을 사람이 취하는 수준부터 인간을 배제하고 기계가 모든 의사결정을 내리고 자율적으로 운전, 판결, 세금계산 등 완전한 자동화 수준으로 구분한다 [1].  이러한 자동화 수준에 차이가 생기는 근본적인 이유는 사람과 기계가 서로 잘하는 영역이 다르기 때문이다. 단어를 찾거나 글자 수를 세는 단순하고 반복적인 작업은 컴퓨터에게는 쉬운 작업이다. 반면, 논문이나 책을 읽고 그 속에 숨겨진 맥락을 파악하는 것은 현재 기술로도 한계가 있다. 인간은 지루하고 반복된 문제를 해결하는 데 적합하지 않고, 컴퓨터도 추상적이고 일반화하는 작업에 적합하지 않았다.

생산성과 임금격차, 보울리의 법칙, 노동인력 참여율을 통해 확인되는 공통된 사항은 1980년 이후 일자리에 구조적인 변동이 생겼다는 점이다[2].  과거 일자리와 관련하여 국가내에 정규직과 비정규직 프레임 혹은 외국인 노동자로 대표되는 국외노동자와 국내 일자리 프레임에 추가하여 사람과 기계 프레임이 추가되었다. 예를 들어, 국가의 근간을 이루는 세무업무를 살펴보자. 과거 숫자를 다룰 수 있는 소수의 사람만이 숫자의 계산을 암산에서 벗어나 주판의 도움으로 생산성을 주판을 사용하지 못한 사람과 비교하여 수십배에서 수천배의 정확도와 함께 빠른 계산을 달성하게 되었다. 이러한 주판은 중간에 기계장치 계산기(찰스 배비지)도 있었지만, 일제 전자계산기로 자리를 내어주지만 사칙연산만 이해하면 기존 주판과 비교하여 어마어마한 생산성을 향상과 정확도를 높인 것은 분명하다. 이후, 개인용 컴퓨터의 보급으로 비지칼크와 로터스 1-2-3가 그 가능성을 열었다면 마이크로소프트 엑셀 스프레드시트 프로그램이 세무사 업무의 생산성을 또한 엄청나게 올린 것도 사실이다. 아마도 여기 까지가 세무사가 기계와 경쟁을 하지 않고 기계가 세무사의 생산성 향상에 도움을 준 것으로 볼 수 있다. 이제부터 문제가 되는 것은 PC 매거진, “The Best Tax Software for 2019”에 소개된 세금관련 프로그램이 $39 달러에 불과하다는 점이다. 1년 세무업무가 개인의 경우 4만원에 불과한데 세무사가 이런 자동화된 기계와 경쟁에서 승리할 수 있는가? 결과는 명확하다. 이제 기계와 세무사간의 일자리 경쟁이 시작된 것이다.

그동안 사람이 직접 설계한 소프트웨어를 연구실에서 꺼내 실생활에 적용함으로써 큰 변화를 이끌어 냈다면, 이제는 그동안 빅데이터 기술을 통해 축적한 데이터를 활용하여 기계로 하여금 소프트웨어를 작성하게 함으로써 학문적으로 뿐만 아니라 실생활도 커다란 변화를 만들고 있으며, 그 중심에는 요즘 자주 언급되는 기계학습(Machine Learning)과 딥러닝(Deep Learning)이 자리하고 있다. 하지만, 기계로 하여금 소프트웨어를 작성하게 하고, 기계가 작성한 소프트웨어를 검증하고, 실생활에 적용시키는 것 또한 모두 사람이 담당해야하는 작업으로 여기에 주목을 받고 있는 것이 데이터 과학이다.

석유화학 산업이 석유를 자원으로 여러 단계의 공정을 거쳐 플라스틱, 섬유, 고무와 같은 상품으로 재탄생시켜 경제적 가치를 창출하듯이, ’데이터 과학(Data Science)’은 데이터 자원을 여러 단계의 과정을 거치면서 기계가 기존에 없던 가치를 창출하는 것으로 대비시켜 볼 수 있다. [5]

기존 통계학이 데이터에서 유용한 가치를 사람이 직접 찾아냈다면, 데이터 과학은 데이터를 원자재로 사용한다는 측면에서는 통계학과 동일하지만, 원자재 데이터를 깔끔한 데이터(tidy data) 형식으로 만들고, 데이터 문법(Grammar of Data)과 그래프 문법 (Grammar of Graphics)을 양대 축으로 삼아 데이터에 내재된 복잡성을 최소화하는 작업을 수행한 뒤에 모형 문법(Tidymodels)에 따라 기계가 데이터로부터 알고리즘을 추출하여 소프트웨어를 제작하게 된다. 마지막으로, 문서 문법(R마크다운)으로 기계가 데이터로 만든 소프트웨어 결과물과 과정을 사람과 커뮤니케이션한다.

미국 CMU 자넷 윙(Jeannette W. Wing) 교수는 2006년 수학적 사고(mathematical thinking)’와 ‘통계적 사고(statistical thinking)’에 이어“읽기, 쓰기, 셈하기와 더불어 컴퓨팅적 사고(Computational Thinking)이 누구나 낮춰야 하는 기본역량”이라고 강조했고, 한국교육에서 추상화와 자동화를 두 축으로 하는 코딩교육이 기계와의 경쟁에서 일자리 문제를 해결할 구세주로 제시되고 있다. 하지만, 사람은 매일 휴대폰, 자동차, 인터넷, SNS, IoT, 스마트-X를 통해 데이터를 만들어내고 기존 방식을 답습해 소프트웨어를 작성하는 것과는 달리 기계는 인간이 생성시킨 데이터를 원자재로 기계학습과 클라우드 인프라를 통해 사람보다 더 뛰어난 소프트웨어를 양산하고 있다. 사실 이렇게 만들어진 기계학습 소프트웨어는 그동안 인식되어 온 양질의 일자리를 빠르게 기계가 대체하고 있고 디지털 불평등도 심화시키고 있다. 단기적으로 데이터 과학은 기계가 만들어낸 소프트웨어 불평등을 완화시킬 수 있고, 더 나아가 기계가 만든 소프트웨어와 사람이 만든 소프트웨어를 아우르는 통합된 소프트웨어를 만들어 냄으로써 새로운 세상을 열어가는데 기여할 것이다.

참고문헌

1. Cummings MM, Man versus machine or man+machine, IEEE Intelligent Systems, 2014;29:62–9.
2. Mishel L., Gould E., Bivens J., Wage stagnation in nine charts. Economic Policy Institute, 2015;6:2–13.
3. Kaplan J., Humans need not apply: A guide to wealth and work in the age of artificial intelligence, Yale University Press 2015.
4. Wing JM., Computational thinking, Communications of the ACM, 2006;49:33.
5. 김기범, 이효정, 박도휘, 기업 운영 혁신을 위한 데이터 과학: 기업의 활용 방안, 삼정KPMG 경제연구원, 2020;제121호.