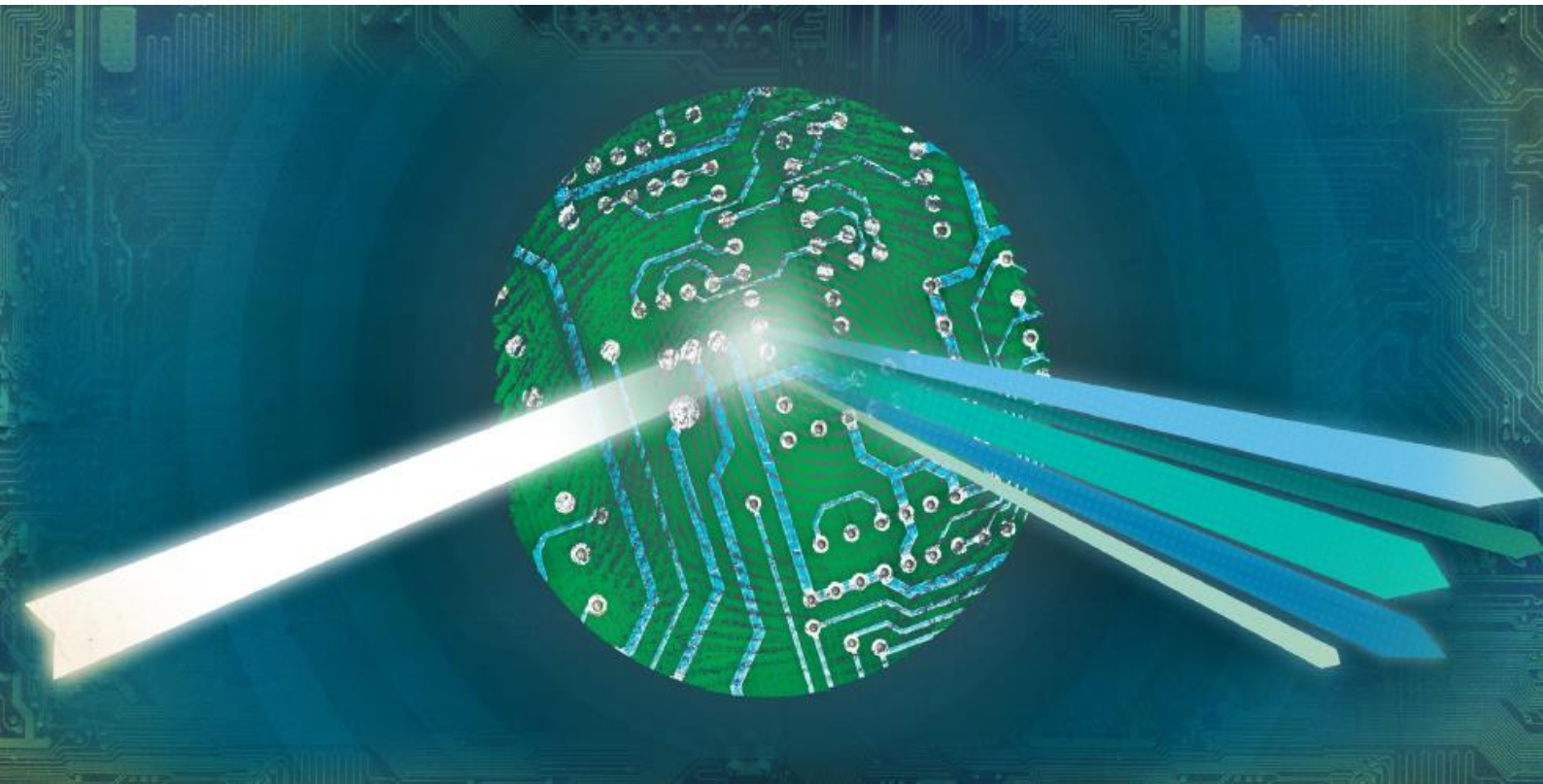


Deloitte.

머신러닝과 발전의 다섯 가지 방향성

전략가를 위한 시그널



By: David Schatsky and Rameeta Chauhan

February 2018

서론

비록 거의 모든 산업이 머신러닝-입력 받은 데이터를 통해 자동으로 패턴과 이상을 발견하고 예측을 하는 인공지능 기술-의 활용 방안을 찾고 있지만, 대부분의 기업들은 이를 아직 활용하지 않는다. 하지만 5가지 발전 방향성이 머신러닝의 배포를 더 쉽고, 빠르며, 저렴하게 만들고 결과적으로 이 기술을 주류화하고 있다. 진입장벽이 무너짐에 따라, 모든 기업들이 이 혁명적 기술의 활용을 탐색하고 있다.

Signals

- 기술 공급사들은 전달 학습이라 불리는 기법을 이용해 필요한 훈련 데이터의 규모를 몇 자릿수 단위로 줄일 수 있다고 주장
- 특화된 칩들이 획기적으로 머신러닝 모델의 훈련을 가속화. 마이크로소프트는 음성 인식 시스템의 훈련 시간을 80% 줄임
- MIT의 연구진은 정확한 예측 그리고 그 예측에 대한 근거를 제공하는 신경망의 훈련 방법을 시연
- 주요 기술 공급사들은 강력한 머신러닝 모델을 모바일 기기에 집어넣기 위한 방법을 찾는 중
- 데이터 과학자 업무 시간의 최대 80%를 차지하는 과업들의 자동화를 목적으로 한 새로운 도구들

머신러닝 사용이 장애물에 직면하다

머신러닝은 오늘날 가용한 가장 강력하고 다양하게 활용할 수 있는 정보 기술 중 하나다. 그러나 대부분의 기업들은 아직 이를 활용하지 않는다. 17개국에 걸쳐 중소기업 및 대기업의 3,100여명의 임원들을 대상으로 수행한 최근의 한 설문조사에 따르면, 10% 미만만이 머신러닝에 투자하고 있다.

몇 가지 요인들이 머신러닝의 도입을 제약하고 있다. 자격이 있는 실무자가 부족하고, 머신러닝 작업을 위한 도구와 프레임워크는 아직 성숙하지 않았고 진화 중이다. 일부 머신러닝 모델 개발에 필요한 대규모 데이터를 구하는 일은 어렵고, 시간이 많이 걸리며, 비용이 많이 들 수 있다.

그리고 블랙박스 문제(black-box problem)가 있다. 머신러닝 모델이 가치 있는 정보를 산출하는 듯해도, 많은 경영진들이 이를 실무에 적용하기를 꺼리는 것처럼 보인다. 그 이유 중 하나는 머신러닝의 내부 작업은 이해가 어렵기 때문으로, 스스로 이해 못하고 명확하게 설명할 수 없는 논리에 기반해 사업을 운영한다는 생각을 일부 사람들이 불편해한다는 점이다. 또 다른 이유로 규제에 의한 제약이 있는데, 규제당국은 기업이 의사결정 근거를 제공하고 보호 계층을 차별하지 않는다는 사실을 입증하도록 요구한다. 이런 상황에서는 결과물이 아무리 정확하거나 유용하다 해도 블랙박스 모델을 채택하기 어렵다.

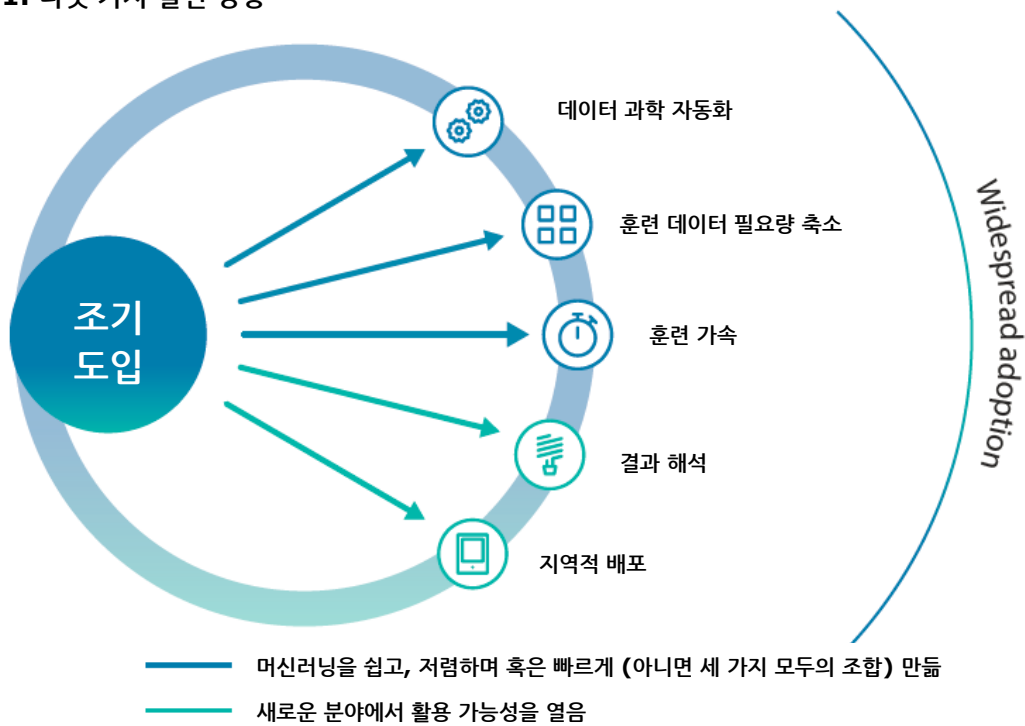
다섯 가지 부문의 발전이 머신러닝 채택에 도움

딜로이트는 기업에서의 머신러닝 도입을 크게 늘려줄 5가지 발전 방향성을 파악했다. 이들 발전 중 세 가지-자동화, 데이터 절감, 학습 가속-는 머신러닝을 더 쉽고, 저렴하며/혹은 빠르게 만든다. 나머지 두 가지인 모델의 해석가능성과 지역적 머신러닝은 새로운 영역에서의 활용 가능성을 열어준다. 다섯 가지 발전 방향성을 활용 범위가 넓은 순으로 정렬해 소개한다.

1. 데이터 과학의 자동화: 머신러닝 솔루션의 개발에는 데이터 과학에 대한 전문 지식이 필요한데, 이 분야 전문가에 대한 수요는 많으나 공급은 부족하다. 데이터 과학은 예술, 과학 그리고 지루하고 고된 디지털 작업의 종합물이다. 실상 데이터 과학자가 소요하는 시간의 80% 정도는 완전 혹은 부분 자동화가 가능하다. 이런 작업에는 데이터의 사전 가공과 정규화, 빠진 값을 채워 넣기, 열 데이터가 날짜인지 숫자인지 해석하기 등의 데이터 정리, 데이터에 대한 가설을 세우기 위해 데이터의 전반적인 특성을 이해할 목적의 시험적인 데이터 분석, 모델이 예측하고자 하는 것과 가장 관련성 높은 데이터 내의 변수를 정하기 위한 특성 파악 및 선택, 가장 정확한 결과를 산출하는 알고리즘을 선택하기 위해 잠재력 있는 수천 개의 알고리즘을 시험하는 알고리즘 선택과 평가 과정이 포함된다.

이들 과업을 자동화하면 데이터 과학자의 생산성과 효과성을 높일 수 있다. 예를 들어, 고객의 투숙객과 대여자로서의 평생가치 모델을 구축하기 위해, 에어비앤비의 데이터 과학자들은 다수의 알고리즘과 설계 접근법을 시험해주는 자동화 플랫폼을 사용했다. 이를 통해 정확성을 5% 이상 높여 의미 있는 영향을 미치도록 알고리즘을 변경할 수 있었다.

그림 1. 다섯 가지 발전 방향



출처: 딜로이트 애널리시스

기존 기업들 혹은 스타트업들이 일부 제공하며, 점점 증가하고 있는 데이터 과학 자동화 도구 및 기술은 머신러닝의 개념 증명에 필요한 시간을 몇 달에서 며칠로 줄일 수 있다. 그리고 데이터 과학자의 생산성을 높여, 데이터 과학 자동화 기술을 채택한 기업들은 심각한 인재부족 상황에서도 머신러닝 활동을 크게 늘릴 것이다.

2. 훈련 데이터 필요량 축소: 머신러닝 모델의 훈련에는 수백만 건의 데이터 요소가 필요할 수 있다. 이는 중요한 문제인데, 데이터를 입수하고 정리하는 일은 시간과 비용이 많이 든다. MRI 화상이 필요한 의료 진단 프로젝트를 예를 들어보자. 방사선 전문의를 고용해 1시간에 6건씩 1,000개의 이미지를 검토하고 정리하려면 3만 달러 이상이 들 수 있다. 사생활과 비밀 보호 우려도 데이터를 구하고 작업하는 일을 어렵게 만든다.

필요한 훈련 데이터를 줄이는 몇 가지 기법이 등장하고 있는데, 한가지 방식은 실제 데이터의 특성을 흉내내기 위한 알고리즘 방식으로 생성된 합성 데이터의 사용이 수반된다. 이는 놀랄 만큼 잘 작동한다. 딜로이트 LLP의 한 팀은 과거에 필요했던 데이터의 1/5만 필요하고, 나머지 80%는 합성 데이터를 사용해 정확한 모델을 만들 수 있게 해주는 도구를 시험해 보았다.

합성 데이터는 외부 인재들의 활용 또한 용이하게 해준다. 다수의 조직들은 보유한 데이터 집합을 온라인에 게시해서 제3자 혹은 크라우드소싱을 이용해 머신러닝 모델을 개발해 왔다. 그러나 만약 데이터 집합에 독점적 성격이 있다면 이런 방식은 불가능하다. MIT의 연구진은 이 문제를 합성 데이터를 이용해 우회하는 방식을 제시했다. 그들은 실제 데이터를 기반으로 합성 데이터를 만들어 외부의 데이터 과학 커뮤니티와 공유했고, 커뮤니티의 연구자들은 이 합성 데이터를 이용하는 머신러닝 모델을 만들었다. 15가지 테스트 중 11건에서, 합성 데이터를 통해 만들어진 이 모델은 실제 데이터로 훈련된 모델과 대등한 성과를 기록했다.

훈련 데이터 필요량을 줄일 수 있는 또 다른 기법은 전송 학습(transfer learning)이다. 한 가지 데이터 집합으로 사전 훈련된 머신러닝 모델은 언어 번역이나 이미지 인식과 같은 유사한 영역에서 새로운 데이터를 통한 학습의 지름길 역할을 한다. 일부 기술 공급사는 전달 학습을 이용하는 그들의 머신러닝 도구가 고객이 필요로 하는 데이터의 규모를 몇 자릿수 줄여준다고 주장한다.

3. 훈련 가속: 대규모 데이터와 복잡한 알고리즘이 관계되기 때문에 머신러닝 모델을 훈련시키는 연산 과정에는 긴 시간-몇 시간, 며칠 심지어 몇 주-이 걸릴 수 있다. 그러나 이제 반도체 기업과 컴퓨터 제조사들 - 기존 기업과 스타트업 -은 GPU(Graphics Processing Units), FPGA(Field-Programmable Gate Arrays), ASIC(Application Specific Integrated Circuits) 과 같은 전문화된 프로세서를 만들어 계산과 칩 내부에서의 데이터 전송을 가속함으로써 머신러닝 모델의 훈련시간을 줄이고 있다.

이들 전용 프로세서들은 기업들의 머신러닝 훈련과 실행을 가속하고, 관련 비용을 절감하는데 도움을 준다. 예를 들어, 마이크로소프트의 연구진은 GPU를 이용해 사람만큼 음성 인식이 가능한 시스템을 1년 만에 완성했다. 만약 CPU만을 사용했다면, 5년이 걸렸을 것이라고 한다. 구글은 자체 개발한 AI 칩인 TPU(Tensor Processing Unit)가 높은 성능을 제공해 추가로 십여 곳의 데이터 센터를 구축해야 하는 비용 부담을 피할 수 있었다고 말한다.

이런 특화된 AI 칩의 얼리 어답터에는 주요 기술 제공사들과 데이터 과학 및 머신러닝 분야의 연구소들이 포함되어 있다. 칩의 채택은 유통, 금융 서비스, 통신 분야로도 확대 중이다. IBM, 마이크로소프트, 구글, 아마존을 포함한 모든 주요 클라우드 서비스 공급자들이 GPU 클라우드 컴퓨팅을 제공하는 상황에서 가속화된 훈련은 모든 조직의 데이터 과학자들에게 가용해질 것인데, 이는 그들의 생산성을 높이고 기업들이 추진하는 활용방안의 숫자를 크게 늘릴 것이다.

4. 결과 해석: 머신러닝 모델은 심각한 약점이 있는데, 많은 모델들이 블랙박스란 점이다. 즉 왜 그런 결과가 나오는지 명확히 설명하기가 불가능하다. 이는 모델들을 많은 활용 분야에 부적합하거나 받아들여지기 어렵게 만든다. 의사나 기업 리더는 신뢰할만한 설명이 없는 의료 진단 혹은 투자 의사결정 결과를 받아들이지 않을 것이다. 일부 경우에는 규제가 그러한 설명을 의무화한다. 예를 들어, 미국의 은행업은 연방준비은행의 지도 사항 SR 11-7를 준수하는데, 이는 모델의 행동에 대한 설명을 요구한다.

그러나 특정한 머신러닝 모델의 블랙박스 내부를 조명해 이들을 보다 해석 가능하고 정확하게 만들어 주는 기법들이 등장하고 있다. 예를 들어, 한 MIT 연구진은 보다 정확한 예측 결과와 그러한 예측의 근거를 제공하는 신경망의 훈련 방식을 시연했다. 이들 기법 중 일부는 이미 상업적인 데이터 과학 상품에 포함되고 있다.

해석 가능한 머신러닝 모델의 구축이 가능해 짐에 따라, 금융 서비스, 생명과학, 헬스케어와 같이 고도로 규제된 산업의 기업들도 머신러닝을 사용할 매력적인 기회를 발견할 것이다. 신용 평점, 추천 엔진, 고객 이탈 관리, 부정 적발, 질병 분석 및 치료 등이 잠재적 활용 분야다.

5. 지역적 배포: 머신 러닝의 채택은 효율성과 결과물을 개선할 수 있는 곳에 기술을 배포할 수 있는 능력과 함께 증가할 것이다. 소프트웨어와 하드웨어 모두의 발전이 모바일 기기와 스마트 센서에서의 기술 사용을 점점 더 실행 가능하게 만들고 있다. 소프트웨어 분야에서 애플, 페이스북, 구글, MS와 같은 기술 공급사들이 상대적으로 적은 메모리를 요구하지만 모바일 기기에서 이미지 인식과 언어 번역과 같은 과업을 처리할 수 있는 가벼운 머신러닝 모델을 만들고 있다.

하드웨어 측에서는, 인텔, 엔비디아, 퀄컴 등과 같은 반도체 업체만이 아닌 구글과 MS 같은 업체들도 모바일 기기에 머신러닝 기능을 부여하기 위해 자체적으로 전력 효율적인 AI 칩을 개발 중이다.

머신러닝 플랫폼으로서의 모바일 기기의 부상은 기술의 활용 잠재력을 높이고 기업들의 스마트 홈과 도시, 자율주행차, 웨어러블 기기, 산업용 IoT와 같은 활용 방안 개발을 유도할 것이다.

머신러닝 주류화를 준비하라

종합적으로, 머신러닝 발전의 다섯 가지 방향성은 일부 기업들의 머신러닝 투자를 방해하는 마찰을 줄이는데 도움을 준다. 그리고 이미 기술을 사용하고 있는 이들의 활용 강화를 도울 수 있다. 이들 발전은 또한 여러 산업에 걸쳐 새로운 활용을 가능케 하고 인재, 인프라, 모델 훈련에 필요한 데이터와 같은 한정된 자원으로 인한 제약을 극복하게 해줄 것이다.

기업들은 과중한 업무에 시달리는 데이터 과학자들의 부담을 덜어주기 위해 일부 업무의 자동화 기회를 탐색해야 한다. 또한 훈련용 데이터 획득의 어려움을 완화해주는 데이터 합성과 전달 학습과 같은 떠오르는 기법에 관심을 기울여야 한다. 그리고 클라우드 서비스 공급자가 제공하는 머신러닝에 최적화된 컴퓨팅 자원에 대해서도 배워야 한다. 기업 자체적인 데이터 센터에서 작업을 실행한다면, 전문화된 하드웨어를 추가하는 것에 대해서도 조사해볼 수 있을 것이다.

비록 머신러닝의 해석 능력은 아직 초기단계에 불과하지만, 높은 가치의 활용방안을 구상하는 기업들은 해석 능력을 개선하기 위한 첨단 기법을 탐색하고자 할 수도 있다. 마지막으로, 모바일 또는 기기 기반의 머신러닝 활용 방안을 고려하는 기업들은 기기에서의 머신러닝 배포가 구현 가능해질 때 바로 준비할 수 있도록 차세대 칩 제조사들이 발표하는 성능 벤치마크를 계속해서 지켜봐야 한다.

머신러닝은 이미 많은 활용분야에 있어 가치 있는 기술임이 드러나고 있다. 5가지 방향성을 따른 발전은 주류 도입에 방해가 되는 일부 장애물의 극복을 도와줄 것이다.



Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee (“DTTL”), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as “Deloitte Global”) does not provide services to clients. Please see www.deloitte.com/kr/about for a more detailed description of DTTL and its member firms.

Deloitte provides audit, consulting, financial advisory, risk management, tax and related services to public and private clients spanning multiple industries. With a globally connected network of member firms in more than 150 countries and territories, Deloitte brings world-class capabilities and high-quality service to clients, delivering the insights they need to address their most complex business challenges. Deloitte’s more than 220,000 professionals are committed to making an impact that matters..

This communication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or their related entities (collectively, the “Deloitte network”) is, by means of this communication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser. No entity in the Deloitte network shall be responsible for any loss whatsoever sustained by any person who relies on this communication.