SPRi Issue Report
2016. 1. 13. (2015-017호)

기계학습의 발전 동향, 산업화 사례 및 활성화 정책 방향

- 딥러닝 기술을 중심으로

한동대학교 김인중 교수
(ijkim@handong.edu)

SPRi 소프트웨어정책연구소

○ 본 보고서는 미래창조과학부 및 정보통신기술진흥센터의 SW공학경쟁력강화 사업의 일환으로 수행한 결과이며, 미래창조과학부의 공식의견과 다를 수 있습니다.

○ 본 보고서의 내용은 연구진의 개인 견해이며, 본 보고서와 관련한 의문사항 또는 수정・보완할 필요가 있는 경우에는 아래 연락처로 연락해 주시기 바랍니다.

- 저자 : 한동대학교 김인중 교수(ijkim@handong.edu)
- 편집 : 소프트웨어정책연구소 김정민 연구원(jungmink26@spri.kr)

《 Executive Summary 》

□ 인공지능 기계학습의 산업적 중요성

- SW는 이제 국가 전체의 경쟁력 확보를 위한 필수 요소가 되었다. 해외 선진 기업들은 고도화된 SW기술을 기반으로 전통적인 IT 분야 뿐 아니라 자동차, 의료, 경제, 교육, 문화 등 거의 모든 분야에서 혁신을 주도하고 있다.
- 인공지능 기계학습의 발달은 지적 활동의 자동화에 대한 가능성을 열고 있다는 점에서 그 파급 효과가 매우 크고 광범위할 것으로 전망된다.
- 최근 딥러닝을 중심으로 급격히 발전한 기계학습 기술은 실용화를 위한 요구 수준과 실제 인공지능 기술 간의 격차를 크게 좁히며 다양한 지능형 시스템의 출현을 예고하고 있다.
- 인공지능 기계학습은 많은 양의 데이터가 발생하는 빅데이터나 사물인터넷 시대에 필수적인 핵심 SW기술이다.
- 구글, 테슬라를 비롯한 선진 SW기업들은 기계학습 기술을 스마트카, 핀테크, 스마트 헬스케어 등 고부가가치 융합분야에 적용하여 새로운 가치를 창출하고 있다.

□ 기계학습 및 딥러닝 기술 소개

- 현재 기계학습 기술은 딥러닝을 중심으로 매우 급격히 발전하고 있다.
- 딥러닝은 데이터로부터 고수준의 정보를 학습하는 기술로 주로 깊은 신경망에 기반한다.
- 딥러닝의 핵심 방법론으로는 사전학습(pre-training) 알고리즘, 컨볼루션 네트워크(CNN), 순환신경망(RNN) 등이 있다.
- 딥러닝은 다양한 분야에 적용되어 기존 방법을 압도하는 탁월한 성능을 보이며, 인공지능 시스템의 실용화에 대한 기대를 높이고 있다.

□ 기계학습 기술의 최근 발전 동향 및 산업화 사례

○ 기계학습 분야에서 최근 이루어진 주요 발전 내용은 새로운 딥러닝 모델 및 계층의 개발, 새로운 학습 알고리즘 및 학습 결과의 이해 방법, 딥러닝과 기존 방법론을 결합한 새로운 응용의 발굴 등이 있다.

○ 구글, 마이크로소프트, 페이스북, 지멘스, 캐스피다, 크리테오, 아마존, NVIDIA 등의 선진 기업들은 기계학습 기술을 인터넷 서비스, 생산공정, 우편자동화, 의료, 보안, 광고, 배송, 지능형자동차 등 다양한 분야에 적용하여 수익을 창출하고 있을 뿐 아니라 새로운 시장을 개척하면서 주도권을 확보하고 있다.

□ 해외에서의 기계학습 기술 발전 및 산업 활성화 성공 요인

○ 기계학습의 학문적/산업적 성공의 배경에는 학계의 뛰어난 리더들, 풍부한 공개 데이터, 지속성 있는 기술 컨테스트에 의한 성능 중심의 기술경쟁, 오픈소스, 연구성과 공유를 위한 빠른 출판 미디어, 연구 성과의 언론 보도에 의한 대중적 이슈화 등이 있다.

○ 반면, 우리나라에서는 전문 인력의 양적/질적 부족, 학습/평가 데이터의 부족, 성능보다 SCI 논문 중심의 경직된 평가제도 등이 기술 발전과 산업화의 장애물이 되고 있다.

○ 기계학습 기술 발전 및 산업 활성화를 위해서는 이 같은 장애물들을 극복하고 해외에서 성공적이었던 제도 및 문화를 도입하여 좋은 기반을 구축하는 것이 필요하다.

□ 기계학습 기술 발전 및 산업 활성화를 위한 정책적 방안

○ 기계학습 기술에 의한 변화에 효과적으로 대응하지 못할 경우 산업 전반에 있어서 세계적 추세를 따라가지 못하고 경쟁력을 잃어버릴 가능성이 있다. 따라서, 기계학습 기술 발전 및 산업화에 대한 정부 차원의 큰 그림과 체계적 추진 및 지원이 요구된다.

○ 깊이 있는 기계학습 전문가를 육성하기 위해서는 오픈소스를 활용한 응용연구 뿐 아니라 방법론 중심의 심도 있는 연구에 대한 지원이 필요하다.

- 기술 컨테스트를 지속적으로 개최하여 현실적 문제들을 제시하고, 참가팀들의 기술을 성능 중심으로 평가한 후, 그 결과를 공개한다면 해외의 사례와 같이 학계 및 산업계의 투명성을 높이고, 논문보다 성능 중심의 실질적 연구를 장려하는 효과를 거둘 수 있을 것이다.

- 기계학습에 필수적인 학습/평가 데이터와 컴퓨팅 인프라를 구축하고 기술 컨테스트와 연동하여 제공한다면 정부의 지원이 실질적인 핵심 기술의 개선으로 이어지도록 유도할 수 있을 것이다.

《 목 차 》

1. 서 론 ····· 1
2. 인공지능, 기계학습, 딥러닝의 발전과 그 영향력 ····· 3
3. 기계학습 기술의 소개 및 성공사례 ····· 10
(1) 개요 ····· 10
(2) 깊은 신경망 ····· 11
(3) 사전학습(pretraining)에 의한 딥러닝 ····· 13
(4) 컨볼루션 네트워크 ····· 16
(5) 순환 신경망 (RNN, recurrent neural networks) ····· 18
(6) 딥러닝의 성공 사례 ····· 20
4. 딥러닝 기술의 최근 발전 동향 ····· 22
(1) 계층 및 네트워크 구조의 발전 ····· 22
(2) 학습 알고리즘의 발전 ····· 24
(3) 시각화에 의한 학습 결과의 이해 ····· 25
(4) CNN과 기존 방법을 결합한 인식 시스템 ····· 26
5. 기계학습 기술의 산업화 사례 ····· 28
(1) 구글: 기계학습 제국 ····· 28
(2) 마이크로소프트의 RankNet, Cortana, Azure ····· 31
(3) 페이스북의 사진 공유 서비스 Moments ····· 32
(4) 지멘스의 기계학습 응용 사례 ····· 32
(5) 오바마 선거캠프의 유권자 맞춤 선거 전략 ····· 34
(6) 캐스피다(Caspida)의 기계학습 기반 지능형 보안 시스템 ····· 34
(7) 크리테오(Criteo)의 기계학습 기반 개인 맞춤형 퍼포먼스 광고 ····· 35
(8) 아마존(Amazon): 구매추천 및 예측, AWS 기계학습 클라우드 ····· 35
(9) NVIDIA: 기계학습을 위한 컴퓨팅 인프라 ····· 35
6. 기계학습 관련 산업의 활성화를 위한 제안 ····· 37
(1) 해외 기계학습 기술의 발전 및 산업화 성공 비결 ····· 37
(2) 우리나라의 기계학습 발전의 장애물 ····· 40
(3) 기계학습 연구 및 상용화 활성화를 위한 제안 ····· 41

1. 서 론

몇 년 전까지만 해도 인공지능 및 기계학습은 전문가들이 사용하는 용어였다. 그러나, 이제는 누구나 한 번쯤은 들어봤을 정도로 친숙한 용어가 되었다. IBM의 인공지능 '왓슨(Watson)'이 제퍼디(Jeopardy) 퀴즈쇼에서 사람과 경쟁해 우승한 것과 구글이 딥러닝(deep learning) 알고리즘을 이용해 대량의 영상으로부터 고양이의 얼굴을 스스로 학습해 낸 사례 등은 일반인들에게도 이미 널리 알려졌다. 또한, 인공지능을 소재로 한 다수의 영화들이 상영되어 사람들의 관심을 끌기도 했다. 세계적으로 인공지능이 주목 받는 이유는 인공지능 기술이 단순한 화제 거리에 그치지 않고 여러 분야에서 상당한 성과를 보이며 사람들의 일상생활과 산업에 변화를 일으키고 있을 뿐 아니라 앞으로 그 영향력이 더욱 커질 것으로 예상되기 때문이다. 구글의 CEO 에릭 슈미츠가 "머신러닝 기술의 출현은 앞으로 5~10년 안에 많은 산업에 영향을 끼칠 것이라 확신합니다." 라고 말 한 것과[3], 마이크로소프트의 창업자 빌 게이츠가 "다음 30년 동안은 역사상 가장 많은 진보가 이뤄질 것입니다" 라고 말한 것은[4] 기계학습을 중심으로 한 SW 기술 발전이 광범위한 분야에서 큰 변화를 가져올 가능성을 시사한다.

[그림 1] 대중에게 널리 알려진 인공지능의 성과

(a) Jeopardy 쇼에서 우승한 Watson [1]

(b) Google의 기계학습 알고리즘이 스스로 학습한 고양이 영상[2]

이와 같이 인공지능과 기계학습이 기대를 모으고 있으나 현실적으로 이러한 기술이 어떻게 발전하고 있으며, 어떠한 변화를 일으키고, 또 일으킬 것인지에 대해 깊이 이해하는 사람은 많지 않다. 인공지능과 기계학습에 대한 시각은 막

연하고 비현실적인 장밋빛 기대를 가진 사람부터 지나치게 회의적인 사람까지 다양하다. 인공지능 기술의 발전이 빠르긴 하지만, 다양한 분야에 활용하기 위해서는 아직도 기술이 좀 더 성숙되어야 한다. 이는 대중의 기대보다 더 많은 시간을 요구할지도 모른다. 그러나, 한편으로 인공지능과 기계학습 기술은 다양한 분야에서 이미 혁신을 일으키기 시작했고, 그러한 혁신들은 또 다른 혁신을 견인하며 연쇄작용을 일으키고 있다. 또한, 현재 주목 받고 있는 빅데이터, 사물인터넷, 스마트카(smart car), 핀테크(fin-tech), 헬스케어(health-care) 등 고부가가치 응용분야가 인공지능 및 기계학습 기술을 요구하기 때문에 궁극적으로 인공지능이 가져올 변화는 매우 광범위하고 강력할 것으로 예상된다. 따라서, 인공지능 및 기계학습 기술이 가져올 변화에 슬기롭게 대처하지 못한 기업과 국가는 거대한 세계적 트랜드를 따라가지 못해 경쟁력에 큰 타격을 입을 수 있다.

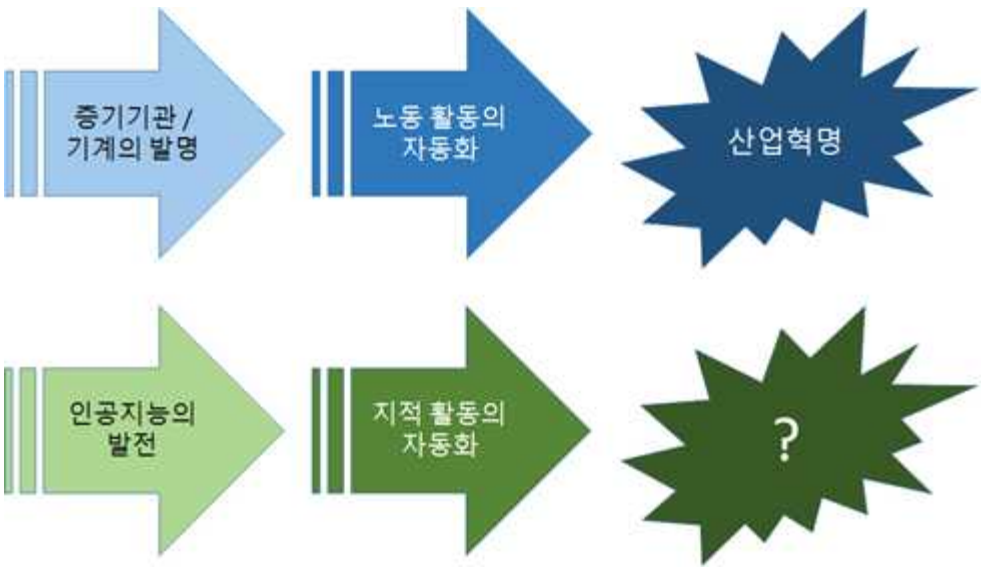
본 보고서는 인공지능 및 기계학습 분야에서 이루어지고 있는 기술의 발전과 이들이 실제 사람들의 생활과 산업을 어떠한 영향을 미치고 있는 있는지를 설명함으로써 인공지능 기계학습 기술이 가져올 변화에 효과적으로 적응하기 위한 방향을 수립하는 것을 돕기 위해 작성되었다. 해외에서 이루어진 인공지능 기계학습 기술의 발전 및 산업화의 내용과 성공 요인들을 분석하고, 우리나라의 기계학습 기술발전을 유도하고 산업화를 활성화하기 위한 방안을 제안한다.

본 보고서의 구성은 다음과 같다. 2장에서는 인공지능, 기계학습, 딥러닝의 개념과 이러한 기술들이 발전할 경우 생활과 산업에 가져올 임팩트에 대해 설명한다. 3장에서는 딥러닝을 중심으로 기계학습의 주요 기술들을 소개하고, 이들의 성공사례들을 소개한다. 4장에서는 최근 기계학습 분야의 기술 발전이 어떤 추세와 방향으로 진행되고 있는지 설명한다. 내용의 특성상 3장과 4장에는 기술적인 내용이 많이 포함되어 있다. 5장에서는 급속도로 발전하고 있는 기계학습을 실제로 산업에 적용한 성공 사례들을 소개한다. 마지막 6장에서는 해외에서 기계학습의 기술발전과 산업화가 성공적으로 수행된 비결 및 원동력을 기술하고, 우리나라의 학계 및 산업계가 기계학습 분야에서 경쟁력을 갖추기 위한 정책적 방안을 제시한다.

2. 인공지능, 기계학습, 딥러닝의 발전과 그 영향력

인공지능은 사람의 지능을 컴퓨터로 구현하기 위한 기술이다. 컴퓨터로 구현된 지능이 큰 의미를 갖는 이유는 지금까지 기계가 할 수 없던 여러 가지 작업들을 자동화 할 수 있는 가능성을 열고 있기 때문이다. 18세기 증기기관을 비롯해 다양한 기계들이 발명되면서 사람이나 짐승이 하던 많은 일들이 자동화되었다. 그 결과, 생산이 비약적으로 증가하였으며, 사람들의 생활과 산업, 더 나아가 인류 문명에 막대한 영향을 끼쳤다. 지금까지 자동화의 대상은 정해진 틀에 따라 수행할 수 있는 단순한 작업들로 국한되었다. 수행 과정을 미리 정의할 수 없는 복잡한 작업이나, 변화하는 환경에서의 적응이 필요한 작업 등은 지능을 요구하기 때문에 지금까지의 기술로는 자동화가 어려웠다. 그런데, 인공지능이 일정 수준 이상 성숙할 경우 복잡도가 비교적 낮은 작업들은 자동화가 가능해질 것이며, 그 후, 기술이 발전할수록 자동화 가능한 작업들은 급속도로 증가할 것이다. 노동 활동의 자동화가 산업혁명으로 이어졌던 사실에 비추어 볼 때 지적 활동의 자동화가 이루어진다면 그 영향이 매우 크고 광범위할 것이다.

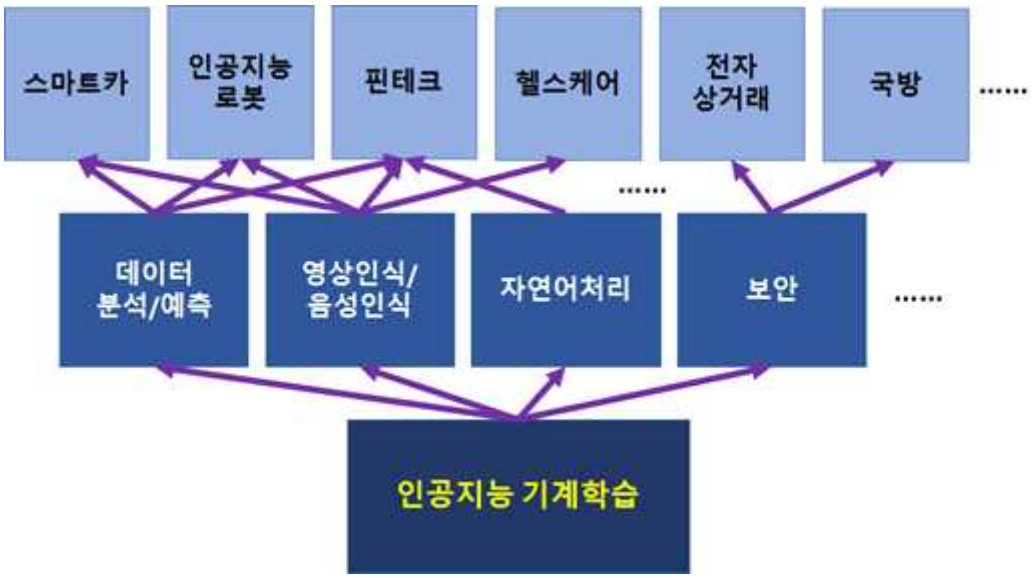
[그림 2] 인공지능이 가져올 지적 활동의 자동화



다양한 기술 중에서도 특히 인공지능의 영향력이 큰 이유는 다양한 분야에 적용될 수 있는 범용성 높은 기술이기 때문이다. 증기기관의 물리적 동력이 다양한 분야에 적용되어 광범위한 변화를 이끌었던 것처럼 인공지능이 제공할 지

적 동력은 다양한 분야에 파괴적 혁신을 일으킬 수 있는 잠재력을 가지고 있다. 최근 수년 째 주목 받고 있는 빅데이터, 사물인터넷, 클라우드 역시 인공지능 기술의 직접적인 응용 분야에 포함된다. 예를 들어, 최근 인공지능 기계학습 분야에서 각광받고 있는 딥러닝(deep learning)은 데이터 분석 및 예측, 영상인식, 음성인식, 자연어처리, 보안 등 다양한 관련 분야의 기술 수준을 크게 향상시켰다. 이러한 기술들에 힘입어 자율주행이 가능한 스마트카, 인공지능로봇, 핀테크, 헬스케어, 전자상거래, 국방 등 광범위한 분야에서 과거에는 구현할 수 없었던 제품과 서비스들이 실용화되고 있으며, 더 나아가 과거에는 상상하기 어려웠던 개념들이 제안되고 있다. 이와 같이, 인공지능의 발전은 다양한 응용기술의 수준을 크게 향상시키고, 그러한 응용기술의 조합에 의해 새로운 상품과 서비스를 창출하여 여러 분야에서 파괴적 혁신을 일으킬 것이다.

[그림 3] 인공지능 기계학습 기술의 범용성

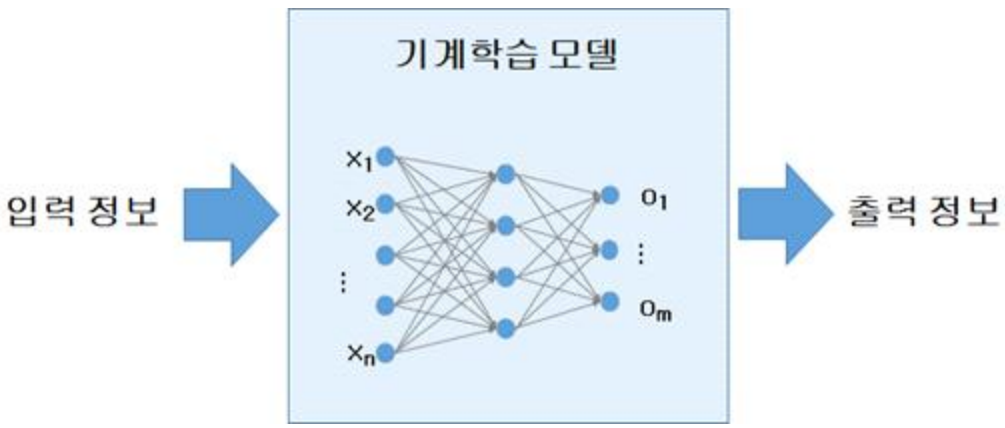


그러나, 현실적으로 인공지능 기술은 최근까지 주목할 만한 변화를 일으키지 못했다. 가장 중요한 이유는 인공지능 기술이 낮은 수준에 머물러 있으면서 현실적 응용에 요구되는 수준에 미치지 못했기 때문이었다. 그로 인해 인공지능은 몇몇 분야에서만 응용되었으며 그 영향력도 제한적이었다. 그런데, 최근 인공지능 기술이 놀라운 속도로 발전함에 따라 지금까지는 기술의 미성숙으로 인해 현실적으로 적용되지 못했던 분야에 성공적으로 적용되기 시작하면서 기대를 높이고 있다.

전통적으로 인공지능의 방법론은 크게 두 가지로 분류된다. 첫째는 지식에 기반한 방법이고, 두 번째는 데이터에 기반한 방법이다. 지식기반 시스템은 사람이 가진 지식을 컴퓨터로 표현하고 이를 활용해 현상을 분석하거나 문제를 해결한다. 반면, 데이터에 기반한 시스템에서는 지식을 직접 제공하기 보다는 지식과 정보가 포함된 데이터를 제공하고 컴퓨터가 필요한 정보를 스스로 학습하게 한다. 이러한 데이터 기반 인공지능에서 가장 중요한 것은 데이터로부터 작업에 필요한 정보를 효과적으로 추출하고 추상화 할 수 있는 학습 모델과 학습 알고리즘이다. 이러한 기술을 연구하는 학문분야를 기계학습이라고 한다. 기계학습은 학습데이터가 많을수록 높은 성능을 얻을 수 있다는 장점이 있는데, 최근에는 인터넷과 빅데이터 기술의 발전에 의해 대용량의 학습데이터를 수집하는 것이 과거에 비해 용이해졌다. 이에 힘입어 기계학습에 기반한 시스템들이 다양한 분야에서 우수한 성과를 나타내었다. 뿐만 아니라, 빅데이터 기술 및 컴퓨팅 인프라가 발전할수록 그 위력이 점점 더 강력해질 것으로 기대되고 있다.

인공지능 기계학습은 오랜 역사 동안 여러 번의 흥망성쇠를 겪어 왔다. 한 예로, 1990년대 후반부터 딥러닝이 각광받기 시작한 2000년대 중반까지는 기계학습의 발전이 상당히 둔화되었다. 기계학습은 보통 수학적인 모델을 기반으로 이루어진다. 기계학습 모델은 특정 정보나 데이터를 수치로 입력 받아 그로부터 도출한 결론을 수치로 출력한다. 예를 들어, 영상 인식을 위한 모델은 영상을 입력 받아 그 영상의 의미를 수치화 된 값으로 출력하고, 의료 진단을 위한 모델은 병의 증상을 수치 형태로 입력 받아 진단 결과를 수치화된 값으로 출력한다.

[그림 4] 기계학습 모델

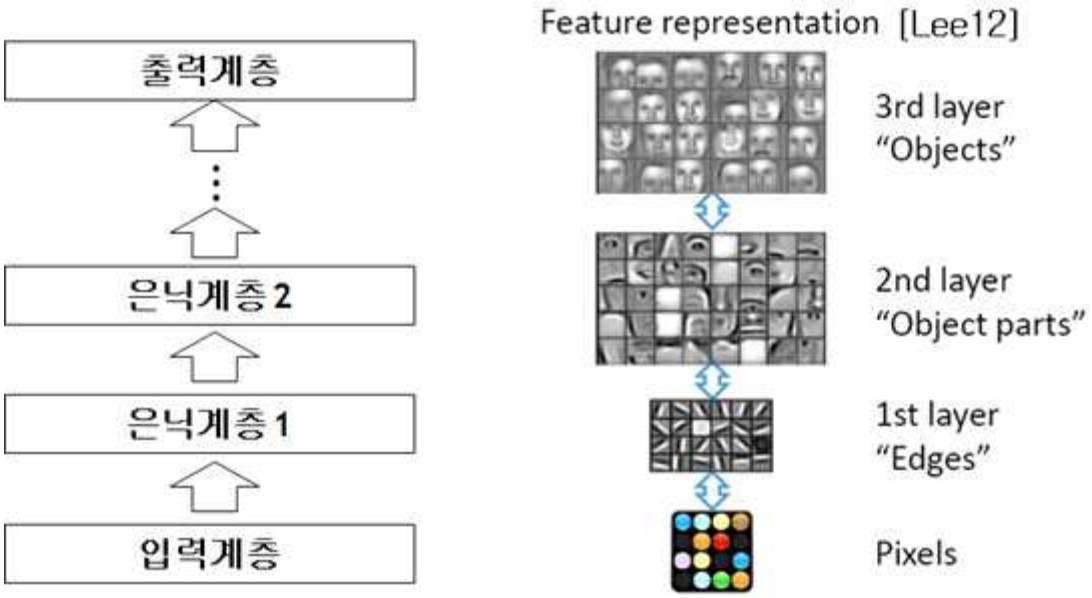


기계학습 모델의 동작은 입력정보와 출력정보 간의 매핑 관계에 의해 정의된다. 학습 과정은 학습데이터에 대하여 원하는 값을 출력하도록 모델을 최적화하는 과정이다. 특정 데이터에 대하여 잘 동작하도록 모델을 학습시키면 실제 상황에서 그 데이터와 유사한 새로운 데이터가 입력되었을 때 학습된 것과 유사한 결과를 출력할 것이라고 기대할 수 있다.

기계학습 모델 및 학습 알고리즘의 성능은 입력정보와 출력정보간의 매핑을 학습데이터로부터 얼마나 효과적으로 학습할 수 있는가에 의해 평가된다. 그런데, 수 년 전까지의 기계학습 모델들은 단순한 기능은 잘 학습하지만, 입출력 관계가 복잡한 기능은 잘 학습하지 못했었다. 학습하려는 정보가 복잡하면 복잡할수록 정교한 모델이 요구되었는데, 기존의 모델로는 복잡한 기능을 학습하기에 역부족이었기 때문이었다. 따라서, 간단한 문제에는 좋은 성능을 보이지만, 복잡한 문제에 대해서는 성능이 급격하게 저하되는 것이 기존 기계학습 모델의 한계였다.

2000년대 중 후반부터 이러한 기술적 한계를 극복할 수 있는 모델 및 학습 방법론이 개발되었는데, 그것이 바로 딥러닝(deep learning)이다. 복잡한 기능을 학습하기 위해서는 데이터로부터 고수준특징(high-level feature)을 학습할 수 있는 능력이 요구된다. 예를 들면, 사람의 얼굴을 인식할 때 기존 모델들은 화소, 에지(edge), 명도 변화의 방향, 좁은 영역의 텍스처(texture) 등의 저수준특징(low-level feature)을 이용할 경우 높은 성능을 얻기 어려울 뿐 아니라, 환경 변화에 잘 적응하지 못한다. 그러나, 눈, 코, 입의 형태 및 위치와 같은 고수준 특징을 이용할 경우 훨씬 정확하고 안정적인 성능을 얻을 수 있다. 이와 같이 고수준 특징을 학습하기 위한 기술이 바로 딥러닝이다. 딥러닝 기술은 많은 수의 계층으로 구성된 깊은 신경망(deep neural networks) 모델을 중심으로 발전하고 있다. 신경망의 각 계층들은 하위계층으로부터 정보를 입력 받아 추상화함으로써 좀 더 고수준 정보로 변환해 상위계층으로 전달한다. 따라서, 신경망의 계층이 많을수록 더 높은 수준의 특징을 효과적으로 추출할 수 있다.

[그림 5] 깊은 신경망과 고수준 특징



그러나, 많은 계층을 보유한 깊은 신경망 모델을 잘 학습하는 것은 기술적으로 쉽지 않다. 깊은 신경망 모델을 실제로 기존 알고리즘을 이용해 학습할 경우 학습이 전혀 이루어지지 않거나, 학습이 이루어지더라도 다른 기계학습 모델들보다 더 우수한 성능을 이끌어내지는 못했었다. 그러나, 2006년 Hinton이 깊은 신경망을 효과적으로 학습할 수 있는 알고리즘을 개발하고, 컨볼루션 네트워크(CNN, convolutional neural networks)의 학습 알고리즘이 영상인식에서 탁월한 성능을 보이면서 딥러닝의 잠재력이 현실화되었다.

딥러닝 모델 및 학습 알고리즘의 발견을 통해 기술적 장애물을 극복할 수 있게 되자, 지금까지 어렵다고 여겨져 왔던 복잡한 문제들에 대한 성능이 크게 개선되었다. 특히, 음성인식, 영상인식(얼굴인식, 물체인식, 텍스트인식 등), 자연어처리 등의 분야에서는 기존 방법들과 확실히 차별화되는 높은 성능을 보였다. 이러한 성공에 힘입어 매우 많은 연구자들이 딥러닝을 더욱 개선하거나 새로운 문제에 적용하려는 연구를 수행하고 있으며, 이는 더욱 빠른 기술의 발전으로 이어지고 있다. 이에 대한 자세한 내용은 3장과 4장에서 설명한다.

딥러닝이 급속도로 발전하는 두 번째 이유는 제품과 서비스에 실제로 적용되면서 산업계에 혁신을 일으키며 수익을 창출하고 있기 때문이다. 딥러닝으로 인

해 기계학습 시스템의 성능이 상용화 가능한 수준에 이르자 여러 기업들이 이를 활용해 새로운 제품 및 서비스들을 제공하였다. 특히 구글, 페이스북 등의 선진 글로벌 IT 기업들은 자사가 보유한 우수한 기계학습 기술을 활용해 혁신적인 서비스들을 제공하였는데, 이들은 사용자들로부터 좋은 반응을 얻고 있다. 딥러닝이 세계적인 열풍을 일으키자 컴퓨팅 H/W 전문 기업이나 클라우드 서비스 기업들도 이를 지원하기 위한 제품들을 사업화 하고 있다. 이에 대한 구체적 사례들은 5장에서 기술한다. 이와 같이 딥러닝에 의해 발전된 기계학습은 수익을 창출하기 시작하였고, 반대로 기계학습 기술을 확보하지 못한 기업들은 경쟁력 저하로 인해 위기를 맞을 수 있는 상황이 되었다. 그로 인해 많은 기업들이 기계학습에 투자를 늘리게 되었고, 이는 다시 기계학습 기술의 발전으로 이어지고 있다. 즉, 기술이 수익을 창출하고 수익은 투자를 유도하며, 투자는 기술을 더욱 발전시키는 선순환 구조가 만들어졌다.

현재에도 기계학습이 빠르게 발전하면서 활발히 실용화되고 있지만, 미래에는 응용 분야가 더욱 넓어지면서 그 발전 속도와 영향력이 더욱 증가할 것으로 예상된다. 그 이유는 먼저 관련 분야의 발전을 들 수 있다. 기계학습의 발전에 빅데이터 기술에 의해 수집된 학습데이터들이 큰 역할을 했음은 이미 기술하였다. 그런데, 급속히 증가하는 데이터의 볼륨은 학습데이터만을 제공하는 것이 아니라 기계학습 기술의 새로운 응용 분야들을 창출한다. 처리해야 할 데이터가 적다면 사람이 수작업으로 처리할 수 있으나, 데이터의 양이 많아질수록 자동화의 필요성이 높아진다. 특히, 대용량의 데이터로부터 고급 정보를 추출하는 분석과정에 기계학습을 활용할 경우 사람이 할 수 있는 것보다 더욱 강력하고 빠른 분석을 수행할 수 있다.

사물인터넷 분야 역시 기계학습과 밀접한 관계를 갖는다. 다양한 사물에 계산 능력과 센서가 부착되고 상호 연결된다면 많은 양의 데이터가 실시간으로 생성/교환된다. 이러한 데이터로부터 유용한 정보를 추출하고 활용하기 위해서는 센서의 데이터를 해석해서 의미 있는 정보를 찾아내고 판단에 반영해야 하는데, 이를 위해서는 기계학습 기술이 필수적이다. 특히, 사물인터넷 중 대표적 분야인 스마트카(smart car)에는 카메라를 비롯한 다양한 센서들이 부착된다. 자동차에

지능을 구현하기 위해서는 이러한 센서들이 제공하는 데이터로부터 주행상황을 파악해야 한다. 이를 위해 카메라영상으로부터 차선, 교통신호, 보행자 등을 인식하해야 하는데, 이 과정에 기계학습 기술이 사용된다. 이미 스마트카의 핵심 기술인 자율주행이나 자율주차, 운전자보조 시스템(ADAS: advanced driver assistance system)에는 기계학습에 기반한 컴퓨터 비전 기술이 활용되고 있다. 스마트카의 보급이 증가할 경우 기계학습에 대한 수요 역시 증가할 것으로 예상된다.

헬스케어, 핀테크 분야에서도 유사한 수요가 발생하고 있다. 예를 들어, 건강보험공단의 대용량 데이터로부터 정보를 추출하여 건강의 이상 징후를 발견한다거나, 다수 기업의 지표로부터 부도가능 기업을 찾아내는 것, 주식투자에서 다수의 종목을 관찰하고 평가하는 것 등 다량의 데이터를 자동으로 분석/이해함으로써 가치를 창출할 수 있는 분야는 모두 기계학습의 응용분야라고 할 수 있다.

컴퓨팅 기술의 발전도 기계학습의 발전을 가속하고 응용 분야를 넓히는데 매우 중요한 역할을 하고 있다. 과거에는 컴퓨팅 속도의 한계로 인해 대용량 신경망을 학습하는 것이 어려웠다. 구글 등 많은 서버를 보유한 기업들만 다수의 컴퓨터를 연결하여 대규모 학습을 실시할 수 있었다. 그러나, 수년 전부터 컴퓨터 그래픽을 위해 개발된 GPU들을 수치 연산에 사용하면서 기존 CPU 중심 컴퓨팅보다 수십 배 빠른 계산 속도를 얻을 수 있게 되었다. 즉, GPU에 의해 고성능 컴퓨팅의 대중화가 이루어진 것이다. 바이두(Baidu)는 GPU 대신 FPGA를 이용해 딥러닝에 요구되는 계산을 수행하고 있다. 이러한 컴퓨팅 기술이 발전하면 할수록 기계학습을 응용할 수 있는 기회는 더욱 많아질 것이다. 예를 들어 GPU 기반 컴퓨팅 기술이 더욱 발전하여 모바일 단말기에서 GPU 기반 인식을 수행할 수 있게 된다면 현재 클라우드 컴퓨팅에 의존하고 있는 딥러닝 기술을 모바일 단말기에서 직접 사용할 수 있게 된다. 이 경우 딥러닝은 지금보다 훨씬 다양한 응용분야에 활용될 수 있다.

3. 기계학습 기술의 소개 및 성공사례

(1) 개요

본 장에서는 최근 수년 동안 기계학습 분야에서 급속한 발전을 이끌어온 딥러닝 연구의 내용과 성공사례에 대해 기술한다. 기계학습 분야에서는 오랫동안 데이터로부터 특정 업무를 수행하기 위한 정보를 학습하려는 연구들이 수행되어 왔다. 그 결과, 다양한 학습 모델과 알고리즘들이 개발되었다. 대표적인 기계학습 방법론으로는 신경망(neural networks), 가우시안 혼합 모델(Gaussian mixture model), CRF(conditional random field), 은닉 마르코프 모델(hidden Markov model), SVM(support vector machine), 볼츠만 머신(Boltzmann machine) 등이 있다. 각 방법론들은 각각의 특성과 장단점을 가지고 있어서 문제의 특성에 따라 다양한 방법론들이 사용되어왔다.

그런데, 2000년대 후반부터 딥러닝 기술이 발전하며 다양한 벤치마크 컨테스트에서 다른 방법들을 압도하는 높은 성능을 보이면서 그 영역을 점점 넓혀가고 있다. 딥러닝은 데이터로부터 고수준특징(high-level feature)을 학습하기 위한 기술이다. 주로 깊은 신경망(deep neural networks)을 이용해 구현되는데, 고수준특징을 학습하기 때문에 복잡하고 변화가 많은 응용 분야에서 탁월한 성능을 보인다.

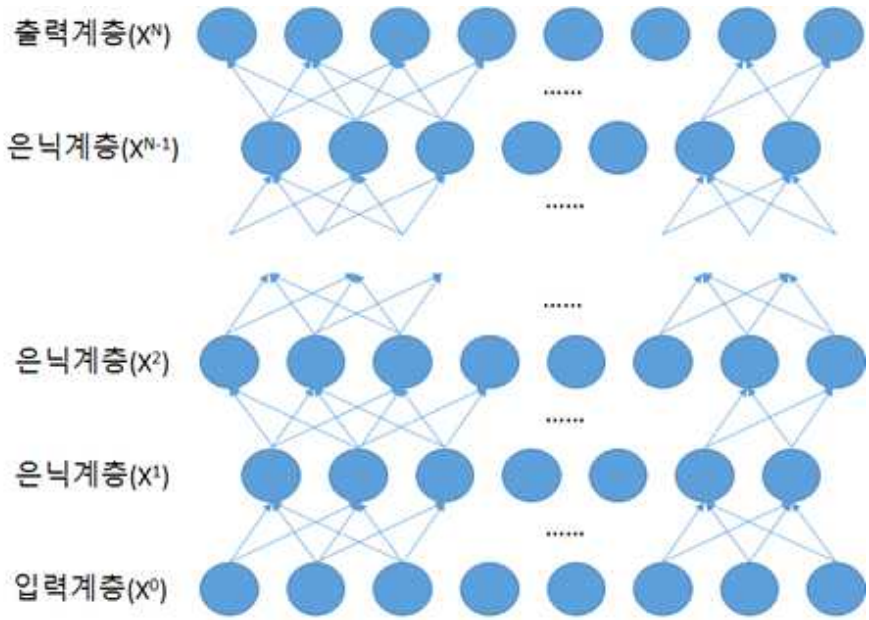
딥러닝이 중요한 또 하나의 이유는 강력한 성능뿐 아니라, 새로운 구조와 알고리즘, 심지어 다른 방법론을 위해 개발된 방법론들까지도 받아들여 적용할 수 있는 유연성을 갖추고 있기 때문이다. 예를 들어, 성능 개선에 도움이 되는 새로운 아이디어가 있을 때 그 아이디어를 신경망 계층의 형태로 구현할 수 있다면 딥러닝의 틀 속에서 통합하는 것이 매우 용이하다. 실제로 딥러닝 모델의 일종인 컨볼루션 네트워크의 경우 다양한 종류의 계층들이 결합되어 각각의 역할을 수행함으로써 영상인식 분야에서 탁월한 성능을 보이고 있다. 또한, 과거에 그 효과가 검증된 다양한 기법들을 딥러닝에 적용하여 성능을 개선한 사례도 매우 많다.

이러한 유연성은 딥러닝이 향후에도 더욱 발전할 수 있는 가능성을 열어준다는 점에서 딥러닝의 미래를 밝게 한다. 딥러닝이 기계학습 분야에서 미친 영향력은 매우 커서 이제는 딥러닝을 제외하고는 기계학습을 논하기 어려울 만큼 중요한 위치를 차지하고 있다. 따라서, 본 보고서도 딥러닝을 중심으로 기계학습의 발전 동향을 서술한다.

(2) 깊은 신경망

깊은 신경망은 다수의 계층으로 구성되며, 각 계층은 여러 개의 노드들을 포함한다. 각 노드는 사람의 뇌세포의 동작을 모방하여 만든 계산 유닛이다. 각 계층의 노드들은 하위 계층 노드들의 출력 값들을 결합함으로써 추상화한다. 예를 들어, 널리 사용되는 퍼셉트론(perceptron)의 각 노드는 하위 계층 노드의 출력값의 노드 간 연결 가중치(connection weight)에 대한 가중 합(weighted sum)을 계산한다.

[그림 6] 깊은 신경망



여기에서 각 계층이 수행하는 동작은 그 계층의 노드와 하위 계층의 노드간의 연결 가중치에 의해 결정되는데 가중치가 적절히 학습되었을 경우 각 계층은 하위 계층이 출력한 낮은 수준(low-level)의 특징으로부터 좀 더 높은 수준(high-level)의 특징을 추출한다. 따라서, 최하위 계층에서는 가장 낮은 수준의 특징을 추출하며, 상위 계층으로 갈수록 점점 더 높은 수준의 특징으로 결합된다. 최상단 계층에서는 가장 높은 수준의 특징을 추출하거나 출력 결과를 결정한다. 이러한 원리로 인해 신경망을 구성하는 계층의 수가 증가하면 할수록 더 높은 수준의 특징을 추출할 수 있게 된다. 그러므로, 깊은 신경망은 높은 수준 특징을 필요로 하는 복잡한 작업을 수행하는데 있어서 탁월한 성능을 갖는다.

깊은 신경망이 갖는 또 다른 장점은 특징 추출과 인식(classification)를 하나의 신경망에서 수행할 수 있다는 점이다. 기존 방법들은 먼저 입력 패턴으로부터 특징 추출 알고리즘을 통해 특징 벡터를 추출한 후, 특징 벡터를 인식기에 입력하여 인식 결과를 얻는다. 그러나, 깊은 신경망에서는 각 계층이 특징 추출, 또는 좀 더 높은 수준의 특징으로의 추상화를 수행하기 때문에 깊은 신경망의 중하단 계층들은 특징 추출 및 추상화를, 최상단 계층은 인식을 수행하는 것으로 해석할 수 있다. 따라서, 깊은 신경망은 별도의 특징 추출 알고리즘 없이 입력 패턴을 직접 인식할 수 있다. 이와 같이 특징 추출과 인식이 통합되어 수행되는 구조는 추가적인 장점을 갖는다. 기존의 특징 추출 방법들이 고정된 알고리즘으로 구성되어 학습이 불가능한데 반해, 깊은 신경망의 특징 추출 계층들은 데이터로부터 학습된다는 점이다. 뿐만 아니라, 특징 추출 및 추상화를 수행하는 중하위 계층의 학습과 인식을 수행하는 상단계층이 동일한 신경망 내에서 통합적으로 학습되기 때문에 더 좋은 성능을 얻을 수 있다.

또한, 깊은 신경망은 많은 수의 가중치, 즉 학습 파라미터를 포함할 수 있다. 학습 파라미터가 많아지면 오버피팅(over-fitting)이 발생하여 학습 데이터에 대해서는 높은 성능이 측정되지만 학습되지 않은 패턴에 대한 성능이 저하된다. 반면, 학습 파라미터가 많아지면 학습 수용력(capacity)은 증가하기 때문에, 매우 많은 수의 학습 데이터가 사용 가능한 경우는 그로부터 풍부한 정보를 학습할 수 있게 된다. 이러한 관점에서 신경망은 계층 및 노드의 수를 변화시킴으로써 학습 파라미터의 수를 조절할 수 있다는 장점을 갖는다. 특히, 딥러닝 기술의 발전으로 인해 계층 수의 제한이 사실상 사라짐에 따라, 많은 수의 학습 데이터로부터 많은 정보를 학습하기에 충분한 규모의 신경망을 사용할 수 있게 되었는데, 이러한 신경망은 풍부한 학습 데이터로부터 많은 정보를 학습하여 높은 성능을 제공할 수 있다.

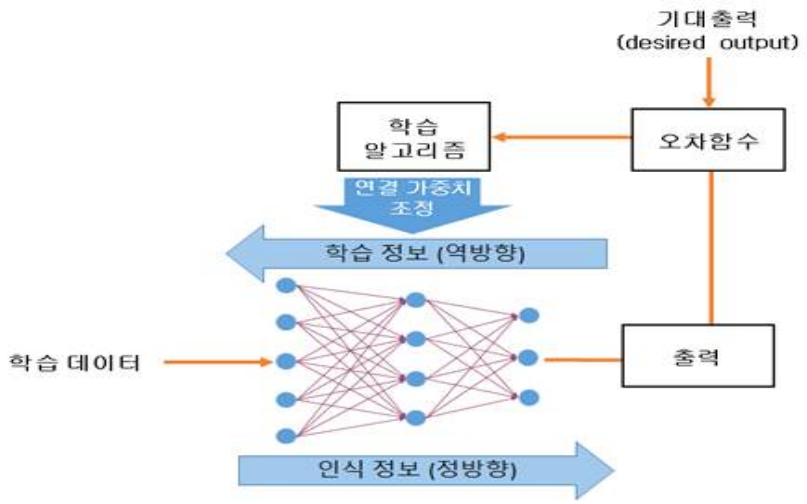
이와 같은 깊은 신경망의 장점에도 불구하고, 최근까지는 깊은 신경망을 실제적 응용분야에 적용하는데 현실적인 어려움이 있었다. 가장 중요한 이유는 기존의 알고리즘으로는 깊은 신경망을 학습하기 어렵다는 것이다. 뿐만 아니라, 깊은 신경망을 학습할 수 있는 충분한 학습 데이터를 수집하기 어려웠던 것과 컴퓨터

의 성능이 딥러닝에 필요한 계산량을 감당하기에는 충분하지 않았던 것 역시 현실적인 어려움이었다. 그러나, 최근에는 이와 같은 기술적, 현실적 장벽들이 기술의 발전으로 인해 극복되면서 다양한 응용분야에 깊은 신경망을 성공적으로 적용할 수 있게 되었다. 먼저, 빅데이터 기술의 발전로 인해 대규모 학습 데이터를 수집할 수 있게 되었다. 이는 깊은 신경망이 많은 수의 학습데이터를 요구한다는 단점을 극복하게 함과 동시에 학습 수용력이 크다는 강점을 극대화할 수 있는 계기가 되었다. 또한, 2000년대 중반부터 실용화 된 GPU 기반 병렬처리 기술은 딥러닝을 위해 요구되는 계산 능력을 제공하였다[9]. 무엇보다도 기존 학습 알고리즘의 한계를 극복할 수 있는 새로운 알고리즘들이 개발되었다.

(3) 사전 학습(pre-training)에 의한 딥러닝

신경망 학습에 가장 널리 사용되는 알고리즘은 오류 역전사 알고리즘 (error back-propagation algorithm)이다[10]. 오류 역전사 알고리즘은 경사도 기반(gradient-based) 학습 방법의 일종이다. 먼저, 학습 데이터를 입력하였을 때 신경망이 생성하는 출력 값과 입력 데이터에 대해 기대되는 출력 값(desired output)의 차이로부터 오차 함수 E를 정의한다. 신경망의 출력 값은 입력 데이터와 가중치 벡터에 의해 결정된다. 입력 데이터와 기대 출력 값이 고정될 경우 오차 E는 가중치 벡터 W의 함수가 되어 E(W)로 표기할 수 있다. 오류 역전사 알고리즘은 먼저 E(W)를 미분하여 오차를 증가시키는 경사도를 계산한 후, 그 반대 방향으로 가중치 벡터를 이동시킴으로써 학습 데이터에 대한 오류를 점차적으로 감소시키는 알고리즘이다.

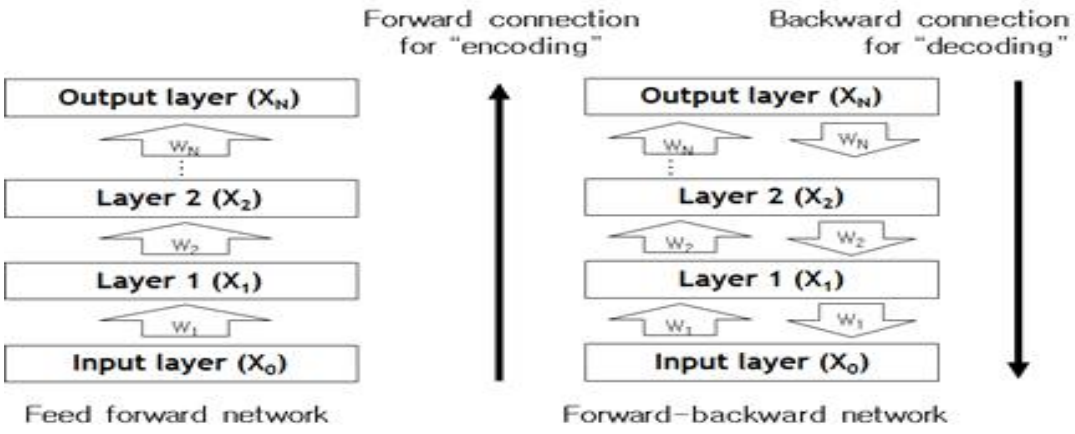
[그림 7] 신경망 학습 알고리즘



이러한 알고리즘에서 하위 계층 노드의 오차 신호는 상위 계층 노드들의 오차 신호를 가중합의 형태로 결합하여 계산하는데, 이 과정에서 서로 다른 방향을 갖는 오차 신호들이 섞이기 때문에 하위 계층으로 내려갈수록 오차 신호가 점점 희미해진다. 따라서, 계층이 많은 깊은 신경망에 오류 역전파 알고리즘을 적용할 경우, 상위 계층은 학습이 잘 이루어지지만 하위 계층에서는 학습이 잘 이루어지지 않는다. 이러한 현상을 방향성 소실 문제(diminishing gradient problem)이라고 한다. 이와 같이 오차 신호가 약해지는 문제는 가중치 값이 학습되지 않았고 노드 간 연결이 많을 때 특히 심각해진다. 각 계층의 노드들이 모두 연결된(fully-connected) 신경망을 랜덤 가중치로부터 학습한다면 위의 두 조건에 모두 해당되므로 학습이 잘 이루어지지 않는다.

수 년 전부터 이러한 문제점을 극복할 수 있는 학습 알고리즘들이 개발되었는데, 이들은 딥러닝 연구를 활성화하는데 결정적인 공헌을 하였다[6][7]. 딥러닝 알고리즘은 크게 두 가지로 분류할 수 있다. 첫 번째 방법은 먼저 자율학습(unsupervised learning) 알고리즘에 의해 가중치 값을 사전 학습(pre-training)한 후 가중치 값이 성숙하면 기존의 학습 알고리즘에 의해 미세 조정(fine-tuning)하는 것이다. 가중치의 초기 값을 랜덤 지정하였을 때에는 기존 학습 알고리즘의 문제점이 심각하게 발생하지만, 사전 학습 알고리즘을 통해 어느 정도 학습된 가중치를 추가적으로 학습할 때에는 그와 같은 문제가 심각하게 발생하지 않는다. 두 번째 방법은 각 계층의 노드들 간 연결의 수를 제한함으로써 오차 신호가 덜 섞이도록 하는 것이다. 전자에 해당하는 방법들로는 DBN(deep belief networks)[6], stacked auto-encoder[11] 등이 있으며 후자의 방법으로는 컨볼루션 네트워크 (CNN, convolution neural networks)이 대표적이다[7][12][13].

[그림 8] 순방향 신경망과 양방향 신경망



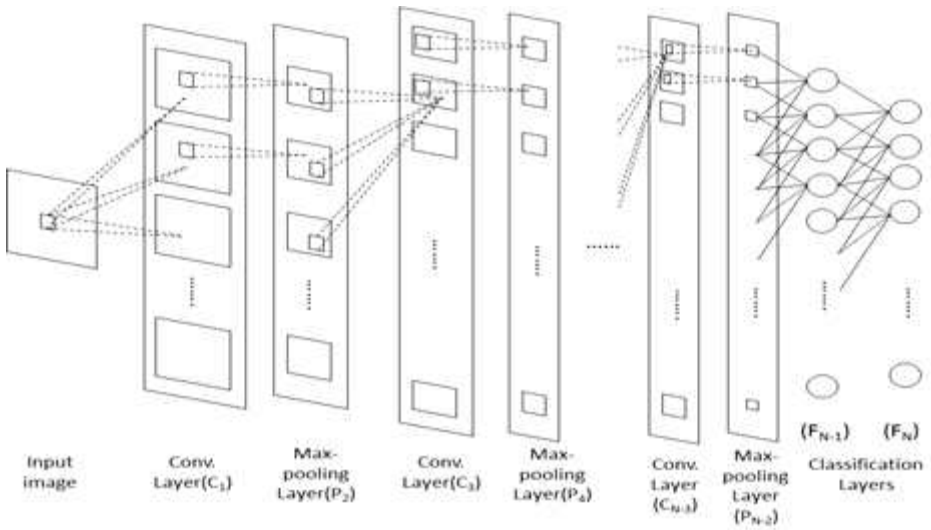
사전 학습을 이용한 딥러닝 알고리즘들은 순방향 연결과 역방향 연결을 모두 가진 양방향 신경망 모델에 기반 한다. 순방향 신경망이 [그림 8]의 좌측과 같이 하위 계층에서 상위 계층으로의 순방향 연결만 갖는데 반해 양방향 신경망은 [그림 8]의 우측과 같이 순방향 연결뿐 아니라, 상위 계층에서 하위 계층으로 연결된 역방향 연결도 갖는다. 이 때, 순방향 연결은 하위 특징에서 상위 특징을 추출하는 인코딩을 수행하며, 역방향 연결은 상위 특징으로부터 하위 특징을 생성하는 디코딩을 수행한다. 따라서, 순방향과 역방향 연결을 모두 가진 양방향 신경망은 하위 계층에서 상위 계층으로 전파(propagate)한 후 다시 상위 계층에서 하위 계층으로 역 전파함으로써 하위 계층의 특징, 또는 입력 패턴을 복원할 수 있다. 이와 같은 신경망으로는 DBN(deep belief network)에 사용된 RBM(restricted Boltzmann machine)이나 auto-encoder 등이 있다.

양방향 신경망을 이용한 사전학습은 다음과 같이 수행한다. 학습 데이터를 입력 계층에 입력한 다음 첫 번째 은닉 계층(layer 1)으로 인코딩한 후 다시 입력 계층으로 디코딩한다. 이렇게 재생성(reproduce)된 입력 계층의 값이 원래 입력된 학습 데이터와 최대한 일치되도록 첫 번째 계층의 가중치를 학습한다. 첫 번째 계층의 학습이 끝나면 학습된 첫 번째 계층의 가중치를 이용해 두 번째 가중치를 학습한다. 학습데이터를 입력한 후 layer 1의 가중치를 이용해 layer 1의 출력값을 계산한다. 이렇게 생성한 layer 1의 출력값을 학습 데이터로 사용하여 동일한 방법으로 layer 2를 학습한다. 이러한 방법으로 최하위 계층에서 시작해 점차 상위 계층으로 학습을 진행한다. 이 알고리즘은 신경망에 입력 데이터가 주어졌을 때 순방향 및 역방향 전파를 통해 동일한 데이터를 복원하도록 신경망의 가중치를 학습한다. 그 결과 학습된 신경망은 입력 데이터들의 정보를 최대한 보관해야 하기 때문에 입력 데이터로부터 중요한 특징을 추출하는 동작을 수행한다. 이와 같은 사전 학습 과정이 끝나면, 기존의 관리 학습(supervised learning) 알고리즘을 이용해 목표로 하는 동작을 수행하도록 추가로 최적화 한다.

(4) 컨볼루션 네트워크

딥러닝 연구의 또 다른 중심인 컨볼루션 네트워크(CNN, convolutional neural networks)은 사람이나 동물의 시각 처리 과정을 모방하기 위해 개발된 신경망이다. CNN의 주요 개념 및 네트워크 구조는 1980년에 제안된 Neocognitron에서 시작되었다[14]. 그 후, 1990년대 후반 LeCun이 경사도 기반 학습 알고리즘을 CNN 성공적으로 적용함으로써 다양한 영상 인식 분야에 널리 적용되었다[7].

[그림 9] Convolutional Neural Network의 구조



기본적인 CNN의 구조는 [그림 9]와 같다. 기본적인 CNN은 세 종류의 계층으로 이루어진다. CNN의 중하위 계층은 convolution 계층과 max-pooling 계층으로 구성된다. 이들은 특징을 추출하고 추상화하면서 점차로 높은 수준의 특징을 추출한다. CNN의 최상위 계층들은 fully-connected 계층으로 구성된다. 이들은 convolution 계층과 max-pooling 계층이 추출한 높은 수준의 특징으로부터 최종 결과를 계산한다. Convolution 계층과 max-pooling 계층은 복수의 특징맵(feature map)으로 구성된다. 각 특징맵은 다수의 노드들이 2차원적으로 배열된 구조를 갖는다.

Convolution 계층의 특징맵은 복수의 입력 특징맵과 연결된다. 특징맵의 노드들은 연결된 하위 계층 특징맵 중 특정한 위치의 윈도우 내의 노드들과 연결된다. 이 때 동일한 특징맵의 노드들은 동일한 가중치를 공유하는데, 이는 2차원 영상 처리 및 특징 추출에 많이 사용되는 컨볼루션과 동일한 효과를 갖는다. 여

기에서 가중치는 컨볼루션 마스크의 역할을 수행하는데, 기존 컨볼루션 알고리즘이 고정된 마스크를 이용한 반면 CNN에서는 컨볼루션 마스크가 신경망 가중치로 구성되기 때문에 데이터로부터 자동으로 학습된다.

Max-pooling 계층은 하위 계층에 위치한 convolution 계층과 동일한 수의 특징맵을 가지며 각 특징맵은 입력 특징맵과 1:1로 연결된다. Max-pooling 계층의 각 노드들 역시 연결된 입력 특징맵 중 특정좌표에 위치한 윈도우내의 입력 노드들과 연결된다. Max-pooling 노드들은 자신과 연결된 입력 노드들의 값 중 최대값을 선택해서 자신의 값으로 가져온다. Max-pooling의 동작은 매우 단순하나, 영상 인식에 있어서 매우 중요한 두 가지 역할을 담당한다. 첫째는 각 특징이 추출되는 위치상의 변이를 흡수하는 것이다. Max-pooling node들은 윈도우 내의 특징값 중 최대값을 취하지만, 그 좌표는 무시한다. 그 결과 특징의 위치가 윈도우 내에서 변할 경우 출력값에 아무 영향을 받지 않기 때문에 특징값은 보존하되 작은 위치 변이는 흡수하는 역할을 한다. 또한 max-pooling 계층은 입력 특징맵을 그보다 작은 크기의 특징맵으로 매핑하는데, CNN에 여러 개의 max-pooling 계층이 있다면 이들을 지나는 동안 특징맵의 크기가 빠르게 감소하여 적은 수의 계층으로도 큰 영상을 인식할 수 있게 한다.

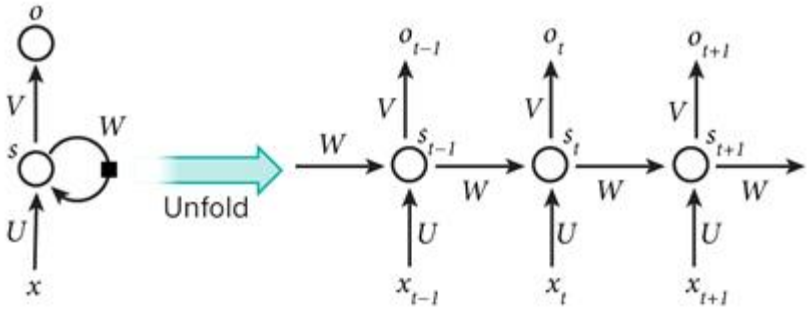
CNN의 최상단에는 fully-connected 계층이 위치하는데, 이들은 하위 계층에서 추출한 높은 수준의 특징으로부터 최종 인식 결과를 결정한다. Fully-connected 계층의 각 노드들은 하위 계층의 모든 노드들과 연결된다.

CNN은 기존 오류 역전파 알고리즘과 동일한 원리의 경사도 기반 방법에 의해 학습될 수 있다. CNN의 중하위 계층을 이루는 convolution 계층과 max-pooling 계층의 노드들은 매우 제한된 수의 입력 노드들과 연결된다. 따라서, 계층간 연결이 많은 신경망에 비해 오차 신호가 섞어서 상쇄되는 문제가 훨씬 적게 발생한다. 또한 CNN은 특징 추출에 매우 효과적인 convolution 계층과 위치 변이 흡수에 효과적인 max-pooling 계층을 통해 높은 수준의 특징을 추출하기 때문에 다양한 변이에도 잘 적응한다. 그 결과, CNN은 다양한 영상 인식 분야에 적용되어 매우 높은 성능을 보이고 있다.

(5) 순환 신경망 (RNN, recurrent neural networks)

일반적인 신경망 모델들은 여러 개의 입력 패턴이 있을 경우 이들을 서로 무관한 독립적인 정보로 처리한다. 그런데, 여러 개의 신호가 연속적으로 들어오고 앞뒤의 신호가 서로 상관관계를 가질 경우 입력 패턴들간의 관계를 모델링 할 필요가 있다. 이러한 예로는 음성신호, 자연어 문장, 전자펜을 사용한 필기, 생체신호, 경제 지표 등이 있다. 이러한 문제들을 효과적으로 처리하기 위해서 순환 신경망(RNN)이 개발되었다. 순환 신경망의 가장 큰 특징은 [그림 10]의 좌측 그림과 같이 한 노드에서 동일 노드로의 순환연결(W)이 있는 점이다. 즉, 다음 시간 노드 s의 값은 입력 x 뿐 아니라 s의 현재 상태와 입력 등 두 가지 정보에 의해서 결정된다. 이러한 순환적 동작을 시간에 따라 펼쳐 보면 [그림 10]의 우측 그림과 같이 표현할 수 있다. 동일 노드로부터 연결된 순환 연결은 과거 시점의 정보를 전달해 주고 입력 노드로부터의 연결은 현재 시간의 새로운 입력을 전달해 준다. 과거와 현재로부터 전달된 두 가지 정보를 결합하여 새로운 상태를 결정한다. 순환 연결은 이와 같이 과거의 정보를 반영하기 때문에 메모리의 역할을 담당한다고 볼 수 있다.

[그림 10] 순환 신경망(RNN)의 구조(좌)와 동작(우) [22]

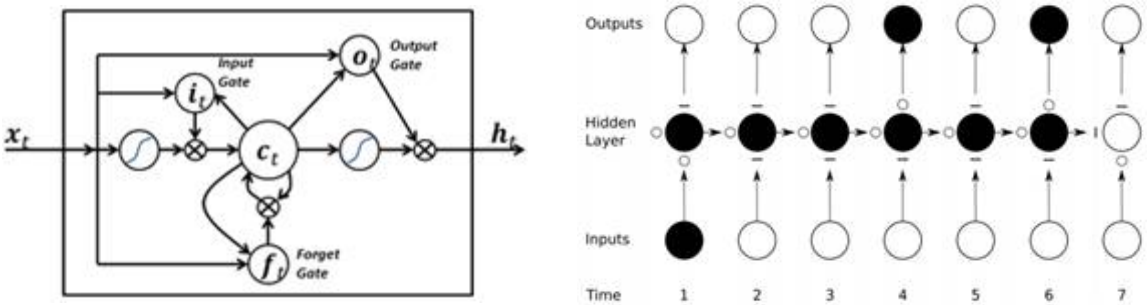


그런데, 순환 연결은 시간의 흐름에 따라 반복적으로 동작하기 때문에 순환 신경망을 긴 시간 동안 동작시킬 경우 깊은 신경망과 유사하게 동작한다. 예를 들어, [그림 10] 좌측의 순환 신경망은 구조적으로 매우 단순하지만, 우측 그림에서 수평 연결의 길이는 동작한 시간에 비례한다. 즉, 오랜 시간 동작하는 순환 신경망의 동작은 수평 방향의 계층이 많은 깊은 신경망과 유사하다. 따라서, 순환 신경망은 딥러닝과 밀접한 관계를 가지며, 딥러닝 모델의 일종으로 분류되는 경우도 많다.

그런데, 순환 신경망이 과거의 정보를 반영하긴 하지만, 현재로부터 시간적으로 멀어질수록 정보의 강도가 약화되기 때문에 시간 거리가 먼 정보간의 상관관계는 잘 학습하지 못한다. 예를 들어, 자연어 처리 등에서는 상호 중요한 관계를 갖는 정보가 시간적으로 많이 떨어져 있는 경우가 자주 발생하기 때문에 일반적인 RNN은 이러한 상관 관계를 학습하는데 효과적이지 못했다.

이러한 문제를 해결한 새로운 순환 신경망 모델이 제안되었는데, 최근 가장 우수한 성능을 보이고 있는 모델은 LSTM(long short-term memory) 네트워크이다[23]. LSTM은 과거 정보를 매우 오랫동안 보관할 수 있도록 설계되었다. LSTM의 기본 노드의 구조는 [그림 11] (a)와 같다. 기존 RNN에서는 입출력 연결 및 과거 정보를 반영하는 순환 연결이 매우 단순하였다. 그러나, LSTM은 그보다 복잡한 구조를 갖는다. 각 노드는 입력, 출력, 그리고, 현재 노드의 상태 정보를 제어하는 세 종류의 게이트(입력, 출력, 망각 게이트)를 갖는다. 입력 게이트와 출력 게이트는 각각 입력과 출력 신호를 제어하고, 망각 게이트(forget gate)는 현재 노드에 저장된 정보를 유지할 것인지 버릴 것인지 여부를 결정한다. 이 세 종류의 게이트에 의해 각 시간에 입력을 받아들일지, 현재 저장된 정보를 보관할지, 값을 출력할지 여부를 결정한다. 이와 같이, 과거의 정보를 보관할지 여부를 별도의 구조를 이용해 판단하기 때문에 의미가 있는 정보들이 오래되더라도 소실되지 않고 보존될 수 있다.

[그림 11] LSTM의 구조 및 동작 [24]



(a) LSTM의 노드 구조 (b) LSTM 의 동작

LSTM은 최근 음성 인식, 필기 인식, 그리고, 기계번역을 비롯한 자연어 처리에서 탁월한 성능을 보이며 기계학습의 핵심 방법론으로 자리 잡았다. 뿐만 아니라, 로봇 컨트롤, 동작인식, 시간적 예측 등에도 널리 활용된다.

(6) 딥러닝의 성공 사례

딥러닝은 다양한 패턴 인식 문제에 적용되어 매우 높은 성능을 나타내었다. 필기숫자인식에 있어서는 가장 널리 사용되는 MNIST 데이터에 대하여 0.23%의 오류율을 보였다[15]. 수 년 전까지 성능이 좋다고 평가되었던 SVM의 최저 오류율이 0.56%였던 것과 비교하면 딥러닝을 통해 성취한 오류율은 매우 낮은 수준이다[15].

딥러닝은 중국어 인식에도 매우 높은 성능을 보이고 있다. 최근까지 중국어 인식에는 mQDF(modified Quadratic Discriminant Function)이 가장 널리 사용되었다. 그러나, 2011년도 ICDAR(International Conference on Document Analysis and Recognition) 에서 개최된 컨테스트에서 CASIA(Chinese Academy of Science, Institute of Automation) 데이터에 대해 CNN을 사용한 시스템이 가장 높은 성능을 보였다[16]. 이 같은 트랜드는 2013년에도 이어졌는데, CNN은 최고 94.77%의 인식률을 보여 기존 방법으로 얻은 가장 높은 인식률 92.64%를 크게 상회하였다[17]. 최근에는 CNN에 문맥정보(domain-specific knowledge)를 결합하여 96.87%까지 성취한 연구결과가 발표되었다[25].

또한 가장 큰 규모의 물체 인식 컨테스트인 ILSVRC(ImageNet Large Scale Visual Recognition Challenge) 2013부터는 1위 뿐 아니라 대부분의 상위 랭커들이 CNN에 기반을 둔 방법을 사용하였다[18]. 특히 ILSVRC는 매년 새로운 기술이 발표되며 그 성능을 높이고 있기 때문에 딥러닝 기술의 발전을 가장 잘 볼 수 있는 기술 컨테스트이다.

얼굴인식 분야에서도 CNN은 기존의 최고 기록을 갱신하였다. 얼굴인식에서 널리 사용되는 LFW (Labeled Faces in the Wild) 데이터에 대하여 기존 최고 성능이 96.33%였으나, 2014년 CNN을 적용한 결과 97.35%로 더 높은 성능을 보였다[19]. 그 후 CNN에 기반한 얼굴인식 기술은 계속 발전되어 최근에는 Google에서 개발한 FaceNet이 무려 99.63%의 높은 성능을 기록하였다[27].

필기한글 인식에서도 CNN은 높은 성능을 나타내었다. 가장 널리 사용되는 필기한글 데이터는 SERI95a와 PE92이다. 2000년대 초반까지 구조적 방법이 가장 높은 성능을 보였고 2011년도에 이르러서는 통계적 방법도 그와 유사한 성능을 나타내었다. 기존 방법에 의한 최고 인식률은 SERI95a에 대하여 93.71%, PE92에 대하여

87.7%에 머물렀었다. 그러나, CNN을 필기한글에 적용하였을 때 SERI95a에 대하여 95.96%, PE92에 대하여 92.92%의 성능을 나타내었다[21]. 이는 기존 최고 성능과 비교하여 35.71%, 42.44%의 상대적 오류 감소율에 해당한다. 최근에는 더욱 발전하여 SERI95a에 대하여 97.67%, PE92에 대하여 96.34%까지 성능이 향상하였다[28].

Deep learning은 음성인식 분야에서도 높은 성능을 제공하였다. 기존에 널리 사용되는 방법은 HMM(hidden Markov Model)과 GMM(Gaussian mixture model)을 결합한 continuous HMM이었다. 그러나, GMM대신 딥러닝 모델인 DBN(deep belief network)을 HMM과 결합함으로써 오류율을 20%까지 낮추었다 [20]. 또한, 2013년에는 LSTM이 TIMIT(Texas Instruments and Massachusetts Institute of Technology) 데이터에 대하여 17.7%의 오류율을 기록하면서 LSTM이 음성 인식에서 가장 우수한 방법이 되었다. 또한, LSTM은 2009년 ICDAR의 필기인식 컨테스트에서 우승을 차지하면서 온라인 필기 인식에도 우수한 성능을 보이고 있다.

이 외에도 딥러닝을 새로운 분야에 적용하려는 시도는 지속적으로 이루어지고 있으며 계속 새로운 성공 사례들이 만들어지고 있다.

4. 딥러닝 기술의 최근 발전 동향

(1) 계층 및 네트워크 구조의 발전

CNN은 2차원 영상으로부터 고수준 특징을 추출하는데 효과적이지만, 더욱 우수한 계층 및 네트워크의 구조를 개발하기 위한 연구가 활발하게 이루어졌다. 컨볼루션 연산은 2차원 특징을 학습하고 추출할 수 있으나, 선형 연산에 의존하기 때문에 입력 벡터에 대한 선형 조합으로 구성된 특징만을 학습할 수 있다. 또한, CNN은 고정된 크기의 영상을 입력 받기 때문에 인식대상 물체의 크기가 어느 정도 일정해야 한다는 제약이 있다. 이러한 한계를 극복하고 더욱 효과적으로 특징을 추출하기 위해 max-out 네트워크[29], NIN(network-in-network)[31], 공간적 피라미드 통합(SPP, spatial pyramid pooling) 계층[32], 인셉션(Inception) 계층[33] 등이 개발되었다. Max-out 계층은 복수의 내부 은닉 노드들을 가지며, 각 은닉 노드는 입력 벡터에 대해 선형 연산을 수행한다[29]. Max-out 계층은 여러 개의 은닉 노드가 계산한 선형 특징의 값 중에서 가장 큰 값을 값을 출력한다. 이와 같이 복수의 선형 특징 중 최대값을 선택하는 연산은 선형 컨볼루션으로는 표현할 수 없는 다양한 비선형 함수를 근사할 수 있다.

NIN은 max-out 계층의 아이디어를 좀 더 일반화 한 MLPconv 계층으로 구성된다[31]. 기존의 컨볼루션 계층에서는 각 특징맵이 하나의 선형 컨볼루션 연산자를 갖는데 반해, MLPconv 계층의 특징맵은 복수의 선형 컨볼루션 연산자와 내부 MLP를 갖는다. 각 입력 영역으로부터 컨볼루션 연산자들이 추출한 특징들은 내부 MLP에 입력되어 한 번 더 조합된다. MLPconv 계층은 이와 같이 복수의 내부 계층을 포함하기 때문에 단일 계층만으로도 구분력이 강한 비선형 특징을 추출할 수 있다. 다수의 MLPconv 계층들로 구성된 NIN은 특징 추출 성능이 우수하기 때문에 최상위 계층에서 복잡한 완전 연결 계층대신 간단한 전역 평균 통합(global average pooling)만을 사용해도 잘 인식할 수 있다. 또한, 간단한 인식 계층을 사용할 경우 오버피팅(overfitting) 문제가 적게 발생하는 추가적인 장점도 얻을 수 있다.

SPP 계층은 CNN이 인식대상 물체의 크기가 균일해야 한다는 제약을 극복하기 위해 개발되었다[32]. 일반적인 CNN은 특징 통합을 위해 최대값 통합 계층을 이용한다. 최대값 통합 계층은 입력 특징맵을 균일한 크기의 영역으로 분할하고 각 영역의 최대값을 선택하기 때문에 인식대상 물체의 크기 변이를 고려하지 않는다. 그러나, SPP 계층은 입력 특징맵을 다중 스케일 피라미드로 분할함으로써 입력 특징을 다양한 스케일로 통합한다. 예를 들어, 입력 특징맵들을 각각 3x3, 2x2, 1x1등의 다른 스케일로 분할한 후 각 영역의 최대값들을 다음 계층으로 전달하면 입력 영상 중 인식 대상 물체의 크기 변화에 대한 적응력을 개선할 수 있다. 이러한 SPP 계층은 특히 CNN을 물체의 검출에 사용할 때 효과적이다.

인셉션 계층은 신경망의 연결 수를 최소화 하면서 성능이 우수한 CNN을 구성하기 위해 개발되었다[33]. 인셉션 계층은 1x1, 3x3, 5x5 등 세 가지 컨볼루션 연산과 3x3 최대값 통합 연산으로 구성하여 다양한 스케일의 특징을 추출한다. 또한, 3x3, 5x5 컨볼루션 및 최대값 통합에는 1x1 컨볼루션을 추가로 적용함으로써 특징벡터의 차원을 축소하였다. 이와 같은 구조는 다양한 스케일의 정보를 학습할 수 있으면서도, 노드간 연결을 최소화 함으로써 비교적 적은 파라미터로 많은 계층을 사용할 수 있게 한다.

네트워크 구조의 측면에서 CNN은 더 많은 계층과 특징맵들을 포함하며 규모가 커지는 방향으로 발전하였다. LeCun이 90년대에 개발한 LeNet은 7개의 계층으로 구성되었다[7]. 2012년 Cireşan은 10개의 계층을 사용하여 중국어를 인식하였고[34], Krizhevsky는 11개의 계층을 사용하여 ILSVRC2012에서 우승하였다[35]. CNN의 계층 수가 증가할수록 더 복잡한 패턴을 효과적으로 학습할 수 있다. 그러나, 학습해야 할 파라미터의 수가 증가하여 계산량과 메모리가 더 많이 필요하게 되고, 더 많은 학습 데이터를 요구하는 등 현실적 문제가 발생한다.

최근에는 더 많은 계층을 사용하면서도 이러한 문제점을 최소화 하기 위한 방법들이 제안되었다. GoogLeNet은 무려 22개의 계층을 포함하여 매우 강력한 성능을 갖는다[33]. 그럼에도 불구하고, 적은 연결로도 우수한 표현력을 갖는 인셉션 계층으로 구성되었기 때문에 네트워크가 갖는 파라미터는 비교적 적다.

ILSVRC2014에서 우승한 VGG 팀은 가중치가 있는 계층만 11~19개, 최대값 통합 계층을 포함해 15~24개의 계층으로 구성된 깊은 CNN을 사용하였다[36]. 특히, 일반적인 CNN들이 5x5~7x7 크기의 컨볼루션 연산자를 사용하고 컨볼루션 계층과 최대값 통합 계층을 교대로 배치한 것에 반해, VGG 팀은 3x3연산자를 사용하는 컨볼루션 계층을 2~4개씩 중첩하였다. 여러 개의 작은 컨볼루션 연산자를 중첩할 경우 하나의 큰 컨볼루션 연산자와 유사한 영역을 처리할 수 있으면서도 학습 파라미터의 수는 적은 수로 유지할 수 있기 때문에 학습 효율과 성능이 개선된다.

(2) 학습 알고리즘의 발전

LeCun이 기울기 기반 학습 알고리즘을 LeNet에 성공적으로 적용한 이후에도 다양한 학습 방법이 개발되었다. 딥러닝에 사용되는 깊은 신경망들은 많은 수의 파라미터를 포함하기 때문에 오버피팅 문제가 자주 발생한다. 이를 방지할 수 있는 근본적인 방법은 학습 데이터를 충분히 수집하는 것이지만 현실적으로 쉽지 않기 때문에 정형화 기법(regularization techniques)을 적용함으로써 보완한다. 가장 기본적으로 사용되는 정형화 기법은 신경망의 학습과정에 목적 함수의 최적화 이외에도 파라미터들이 극단적인 값을 갖지 않도록 추가적인 제약을 가하는 것이다 [37].

2012년 Hinton이 개발한 dropout 알고리즘도 오버피팅을 방지하는데 효과적이다[38]. 학습 단계에서 신경망의 은닉 노드들을 특정 확률로 비활성화함으로써 노드들의 공동적응(co-adaptation)을 방지하여 각 노드들이 독립적인 특징을 학습하도록 유도한다. 이론적으로 dropout은 노드들을 공유하는 다수의 신경망들의 출력을 기하 평균으로 통합한 것과 유사한 효과가 있다. Dropout은 많은 실험에서 그 효과가 검증되어 널리 사용되고 있다.

사람이나 동물에게 새로운 지식을 가르칠 때 먼저 쉬운 예제들을 보여줌으로써 중요한 개념을 익히게 한 후, 점점 어려운 예제들을 제시하여 복잡한 개념을 익히게 할 경우 교육효과가 개선된다. Bengio는 이러한 사실을 신경망의 학습에 적용한 커리큘럼 학습(curriculum learning) 방법을 제안하였다[39]. 신경망을 학습할 때 학습 데이터를 임의의 순서로 제공하는 것이 아니라, 의도적으로 결정된 순서에 따라 제공한다. 잘 설계된 커리큘럼에 따라 신경망을 학습할 경우 학습 속도가 증가할 뿐 아니라 학습 결과도 개선된다.

최근 Ioffe는 깊은 신경망의 학습이 느린 이유 중의 하나가 학습 도중 발생하는 내부 공변인 이동(internal covariate shift)임을 지적하였으며 이에 대한 해결책으로 배치 정규화(batch normalization) 방법을 제시하였다[40]. 신경망의 상위 계층은 하위 계층이 출력한 특징 벡터에 대하여 학습된다. 그런데, 학습이 진행되는 동안 하위 계층의 파라미터가 바뀌기 때문에 출력 특징의 분포가 변하고, 이는 상위 계층의 학습을 방해한다. 이러한 문제를 최소화하기 위해서는 학습률(learning rate)을 낮추어야 하는데, 이 경우 학습 속도가 저하된다. Ioffe가 제시한 새로운 해결책은 학습 과정 중 학습 데이터의 배치별로 하위 계층의 출력 특징의 분포가 일정해지도록 정규화 하는 것이다. 이러한 배치 정규화를 적용하면 내부 공변인 이동이 상쇄되기 때문에 학습 속도가 개선될 뿐 아니라, 학습 후 성능도 더욱 개선된다.

(3) 시각화에 의한 학습 결과의 이해

CNN의 영상 인식 성능은 매우 우수하나, CNN이 데이터로부터 학습한 내용이 무엇인지 확인하는 것은 간단하지 않다. 최하단 컨볼루션 계층의 경우 각 컨볼루션 연산자의 가중치를 시각화 함으로써 어떠한 특징을 학습했는지 확인하는 것이 가능하다. 그러나, 그 외의 계층이 학습한 정보를 시각화 하는 것은 쉽지 않다. CNN이 전향적(feed-forward) 네트워크이기 때문에 입력영상, 또는 저수준 특징으로부터 고수준 특징으로의 상향식 추상화는 잘 수행하는 반면, 반대 방향의 하향식 연산은 제공하지 않기 때문이다. 따라서, CNN이 학습한 내용을 시각화하고 이해하기 위해서는 추상화를 위한 상향식 연산 이외에 고수준 특징으로부터 저수준 특징, 또는 영상을 생성할 수 있는 하향식 연산이 요구된다.

CNN에 하향식 연산을 구현할 수 있는 방법으로는 컨볼루션 DBN(CDBN, convolutional deep belief networks)[41], 디컨볼루션 네트워크(deconvolutional networks)[42][43] 등이 있다. CDBN은 컨볼루션 RBM(CRBM, convolutional restricted Boltzmann machine)과 확률적 최대값 통합 계층(probabilistic max-pooling layers)으로 구성된다[41]. CRBM은 컨볼루션 계층과 유사한 구조를 갖는다. 그러나, 각 노드의 동작은 RBM(restricted Boltzmann machine)과 동일하여 추상화를 위한 상향식 연산뿐 아니라 패턴생성을 위한 하향식 연산을 제공한

다. 디컨볼루션 네트워크는 영상의 분석과 합성을 위한 특징을 추출하기 위해 개발되었다[42]. 디컨볼루션 네트워크는 복수의 디컨볼루션 계층(deconvolution layers)으로 구성된다. 디컨볼루션 연산은 컨볼루션 연산의 역함수에 해당한다. 디컨볼루션 계층과 역통합 계층(unpooling layers)을 조합하면 컨볼루션 계층과 최대값 통합 계층으로 구성된 CNN의 역작용에 해당하는 하향식 연산을 구현할 수 있다[43]. 입력 영상에 대해 CNN으로 특징을 추출한 후 은닉 계층의 특정 특징 맵으로부터 하향식 연산을 수행하면 그 특징 맵이 입력 영상의 어느 부분에 반응했는지를 시각화 할 수 있다.

이와 같은 하향식 연산을 이용해 고수준 특징 맵을 시각화 한 결과 사람이 직관적으로 중요하다고 느끼는 부분이 생성되었다. 이러한 결과들은 CNN이 학습데이터로부터 실질적으로 중요한 특징을 잘 학습할 수 있는가에 대한 질문에 긍정적으로 대답하였다.

(4) CNN과 기존 방법을 결합한 인식 시스템

CNN을 이용해 분할된 물체의 인식뿐 아니라 장면 인식, 다중 물체 인식 등 좀 더 확장된 문제를 해결하기 위한 연구도 수행되었다. 이러한 연구들은 CNN과 전통적인 방법론을 창의적으로 결합하여 CNN만으로는 수행하기 어려운 기능들을 효과적으로 구현하였다. Farabet은 CNN을 이용해 장면 영상 전체를 분석하였다[44]. 다중 스케일 CNN을 이용해 영상의 각 화소가 어떤 물체의 일부인지를 대략적으로 판별한다. 그러한 화소들을 슈퍼픽셀(super-pixel)단위로 군집화하고, 슈퍼픽셀들을 연결하여 세그먼트 그래프를 구성한다. 이렇게 구성된 세그먼트 그래프에 대하여 다층 컷(multi-level cut) 알고리즘을 이용해 CNN의 판별 결과와 가장 잘 조화되는 분할 조합을 선택한다.

R-CNN(regions with CNN feature)은 영상으로부터 물체의 검출과 인식을 함께 수행한다[45]. 먼저 기존의 컴퓨터비전 기술을 이용해 물체가 존재할 것으로 추정되는 후보 영역들을 추출한 후 각 후보 영역에 대해 CNN을 활용해 실제 물체인지를 판별하고 인식을 수행한다. R-CNN은 전통적인 컴퓨터비전 기법과 CNN을 결합하여 우수한 성능을 성취한 대표적인 사례로 알려져 있다.

CNN을 이용해 순환적 신경망(RNN)을 구성한 연구도 활발히 진행되었다. Ba는 CNN기반 RNN을 이용해 사람이 물체에서 물체로 시선을 이동하며 순차적으로 인식하는 과정을 모방하였다[46]. RNN으로 구현된 glimpse 네트워크를 이용해 영상 중 주목해야 할 영역을 결정하고 CNN을 이용해 선택된 영역을 분석한다. CNN의 분석 결과는 내부 표현에 반영되고, 다시 glimpse 네트워크에 입력되어 다음에 주목할 영역을 추출하는 과정에 반영된다. Sermanet는 Ba의 아이디어를 계승 발전시켜 물체의 세부 분류(fine-grained categorization)에 활용하였다[47]. 물체를 세부 분류하기 위해서는 유사도가 높은 아종(sub-categories)들 간 차이점에 주목해야 한다. 이를 위해 glimpse 네트워크 이용해 입력 영상을 세부 분류하기 위해 주목해야 할 영역을 결정하였다.

5. 기계학습 기술의 산업화 사례

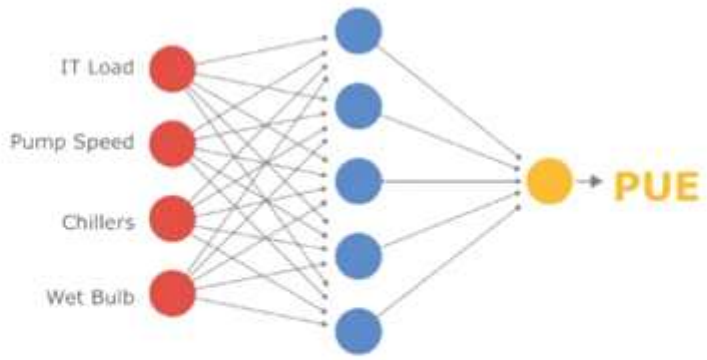
본 장에서는 3, 4장에서 설명한 기계학습 기술들이 실제 산업에 어떻게 적용되고 있는지를 설명한다.

(1) 구글: 기계학습 제국

세계적으로 기계학습 분야에서 가장 앞서가는 기업은 역시 구글이다. 구글 CEO 레리 페이지가 2014년 3월 TED 강연에서 기계학습을 “오랫동안 보아온 것 중에 가장 흥미로운 기술“이라고 말했을 정도로 기계학습 연구에 많은 투자를 하고 있다[48]. 구글은 2013년 딥러닝의 대가인 Hinton 교수를 영입하고, 기계학습 전문 기업 딥마인드(DeepMind)와 네스트(Nest)를 거액에 인수하는 등 우수한 연구진을 확보하였다. 또한, 막대한 양의 데이터와 컴퓨팅 인프라를 보유하고 있어서 기계학습 연구에서 가장 이상적인 환경을 갖추고 있다. 그 결과, GoogLeNet[33], FaceNet[27], Batch Normalization[40] 등의 우수한 연구성과를 발표하면서 기계학습 관련 연구에서 앞서나가고 있다. 또한, 그 결과를 적극적으로 활용하여 혁신적인 서비스를 개발할 뿐 아니라 회사의 운영에도 널리 활용하고 있다. 더 나아가 자사의 딥러닝 엔진을 공개하고 구글포토(Google Photo)를 이용해 데이터를 수집하는 등 기계학습 기술의 주도권을 확보해 나가고 있다.

먼저, 구글은 검색에 기계학습 알고리즘을 적극적으로 활용하고 있다. 구글 검색 엔진의 근간이라고 할 수 있는 페이지 랭크 알고리즘 랭크 브레인(Rank Brain)은 기계학습에 기반을 둔 텍스트마이닝(text mining) 기술이다[49]. 특히, 기계학습 기술을 사용한 음성 인식 기술과 영상인식 기술을 이용해 멀티미디어 콘텐츠의 검색 서비스도 이용한다. 또한, 구글은 대규모 데이터센터를 운영하는데 있어서 성능 및 에너지 관점에서 최적화하기 위해 기계학습을 활용하고 있다. 데이터센터의 서버 및 장비들의 사용 시간 및 에너지 사용량을 측정한 후 이러한 정보에 대해 기계학습 기술을 이용하여 서버 및 기타 장비를 관리하고, 냉각 시스템을 운영한다. 그 결과 PUE(power usage effectiveness) 예측에 99.6%의 높은 정확도를 얻을 수 있었고, 에너지 절약을 포함한 데이터센터의 운영을 크게 효율화 하였다[50].

[그림 12] Google 데이터센터 PUE 예측 신경망 [50]



A simplified version of what the models do: take a bunch of data, find the hidden interactions, then provide recommendations that optimize for energy efficiency.

구글은 사용자의 데이터로부터 정보를 활용하여 서비스를 제공하고 있다. 구글의 개인비서 서비스 '구글나우(Google Now)'는 자연어 인터페이스를 갖추고 있으며, 사용자의 검색 활동, 기기정보, 방문한 장소 등 분석하여 사용자가 원하는 정보를 능동적으로 제공한다[51]. 구글나우에 사용되는 음성인식, 자연어처리 및 사용자 데이터로부터 제공할 정보를 추정하는 기능 등은 기계학습 기술을 이용해 구현된다.

얼마 전 공개한 스마트 리플라이(Smart Reply) 서비스 역시 기계학습 기술로 만들어졌다. 이메일의 내용을 분석해 상황에 맞는 답글의 후보를 사용자에게 제시한다[52]. 따라서, 사용자는 단말기에서 타이핑에 소요되는 시간을 낭비하지 않고 빠르게 응답할 수 있다. 이러한 지능적 서비스에 요구되는 자연어 처리 기술은 앞 절에서 설명한 LSTM을 이용해 구현되었다.

2015년 5월 발표된 사진 저장 서비스 구글포토(Google Photo)는 기계학습을 이용한 서비스이면서 동시에 기계학습 성능을 개선하기 위한 데이터 수집을 위해 만들어졌다[53]. 모바일 단말기에서 촬영한 사진을 구글 클라우드에 백업한 후 자동으로 시간, 위치, 주제 별로 분류해 주고, 다른 사람들과 공유도 가능하게 한다. 또한, 사진을 스토리로 구성하거나, 여러 장의 사진을 연결하여 파노라마 사진을 만들어 주는 등 여러 가지 지능적인 서비스를 제공하고 있다. 그런데, 구글은 구글포토 서비스에 저장된 데이터를 딥러닝 엔진의 학습에 사용한다. 서

비스 시작 6개월 만에 1억 명이 넘는 사용자를 확보하였으며, 무려 50억장의 사진을 수집하였다. 이와 같이 방대한 데이터를 이용해 구글의 딥러닝 엔진을 학습할 경우 앞으로 구글의 기계학습 기술은 더욱 발전할 것으로 예상된다.

구글 번역은 기계학습을 이용한 다국어 번역 서비스이다[54]. 2015년 12월 기준으로 90개의 언어를 지원하고 있으며, 매일 2억 명 이상의 번역을 제공하고 있다. 구글 번역은 문법적 규칙 등 전통적인 자연어 처리 알고리즘이 아닌 기계학습을 이용한 통계적 기계번역 기술로 개발되었다. 구글 번역기는 한 언어(L1)에서 다른 언어(L2)로 번역할 때 먼저 L1을 영어로 번역한 후 그 결과를 L2로 번역한다. 이 과정에서 코퍼스 기반 기계번역(CBMT, corpus-based machine translation)이라는 기술을 사용한다. 두 언어 간에 병렬 코퍼스를 기반으로 비교분석하여 번역 결과물을 생성한다. 즉, 컴퓨터에게 언어의 문법을 가르치는 것이 아니라, 다량의 문서들로부터 문법을 스스로 배우도록 하는 것이다. 따라서, 데이터의 양이 많아질수록 번역 성능도 개선될 것으로 예상된다.

구글은 2010년 자율주행자동차 개발 계획을 발표한 이후 지속적으로 연구에 투자해 왔다[55]. 2014년 12월에는 자율주행차의 시제품을 공개하였다. 구글의 자율주행차에는 3D 카메라를 비롯한 여러 가지 센서들이 장착되어 있어서 도로 상황을 실시간으로 파악한다. 3D 카메라는 30m 거리까지의 물체를 탐지할 수 있으며, 레이저 센서는 360도 주변의 물체를 감지한다. 지난 3월까지 3년간 130만 Km의 무사고 기록을 달성하였으며 가까운 미래에 무인차를 상용화 할 계획을 가지고 있다. 특히, 구글은 복잡하고 어려운 판단, 예를 들어 인간의 윤리적 판단이 요구되는 상황에서의 판단까지도 기계학습을 통해 익힐 수 있을 것으로 예측한다.

[그림 13] 구글 자율주행차

(a) 자율주행차 초기버전 (b) 자율주행차 시제품

구글은 자사의 기계학습 엔진인 텐서플로우(TensorFlow)를 오픈소스로 공개하였다[56]. 구글의 제품 중 50개 이상이 텐서플로우를 사용한다고 있다고 알려져 있다. 그 이유는 다음과 같이 분석된다. 세계적으로 기계학습의 적용 분야는 급증하고 있으나, 기계학습 전문가가 부족하여 구하기 어려운 상황이다. 따라서, 텐서플로우를 공개함으로써 많은 전문가들이 자신들의 기계학습 엔진을 사용하도록 만들고, 더 나아가 개발에 참여하도록 유도함으로써 기계학습 전문가들의 영입과 네트워킹을 강화할 수 있다. 두 번째는 텐서플로우를 공개함으로써 많은 사람들이 사용하도록 유도하여 협업의 발판을 마련하고 기계학습 분야에 구글의 영향력을 강화할 뿐 아니라, 더 나아가 기계학습 분야의 주도권을 갖기 위한 전략으로 추정된다. 에릭 슈미츠는 심지어 상당수의 경쟁자들도 구글의 기술을 활용할 것으로 예측하였다. 또한, 구글은 텐서플로우를 공개하더라도 자신들의 기술적 경쟁력에는 문제가 없다고 판단한 것으로 추정한다. 기계학습에는 알고리즘과 소프트웨어뿐 아니라 데이터와 컴퓨팅 인프라가 요구된다. 구글은 우수한 기계학습 기술을 보유하고 있을 뿐 아니라 데이터 및 인프라 측면에서 우월적인 위치에 있기 때문에 핵심 엔진을 공개하더라도 경쟁력에는 문제가 없을 것이다.

(2) 마이크로소프트의 RankNet, Cortana, Azure

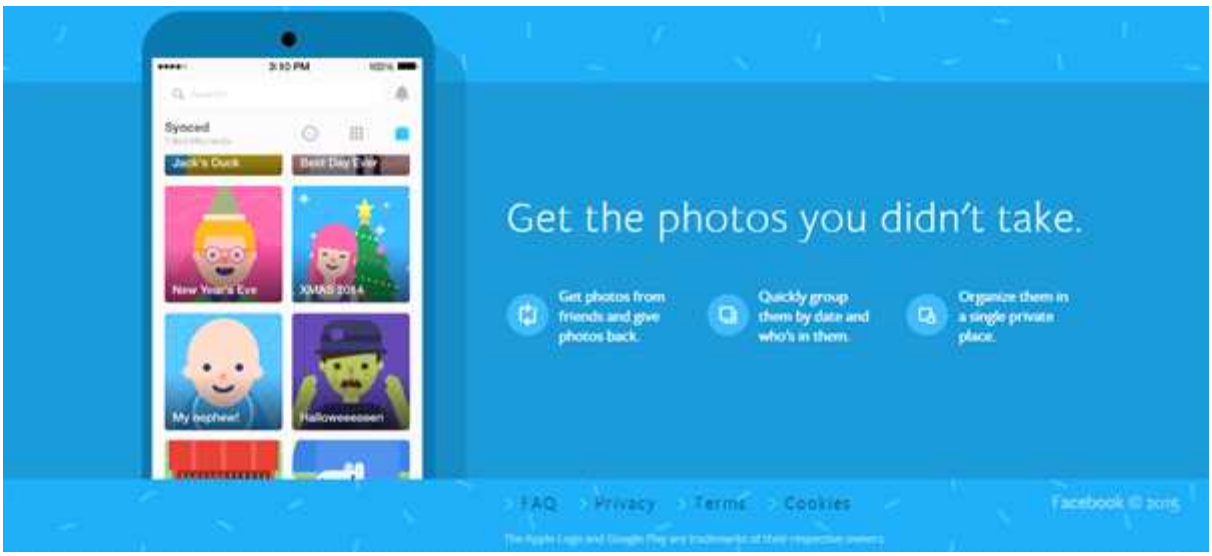
마이크로소프트는 기계학습 기술을 이용해 검색 서비스 Bing과 인공지능 비서 코타나(Cotana)를 개발하였다. Bing에 사용된 RankNet은 기계학습을 기반으로 웹페이지의 순위를 결정한다[57]. 코타나는 애플의 시리(SIRI)와 유사한 지능형 개인비서 서비스이다[58]. 음성인식을 이용해 대상을 입력 받아 검색엔진을 통해 사용자가 원하는 결과를 보여준다. 또한, 메시지나 메일, 일정 등을 확인하고 관리할 수 있도록 돕는다. 구글나우처럼 사용자에게 유용한 정보를 자동으로 추천해 주기도 한다.

마이크로소프트는 자사의 기계학습기술을 서비스로 만들어 애저 기계학습(Azure Machine Learning)이라는 이름으로 출시했다. 애저 서비스는 기계학습 알고리즘과 컴퓨팅 파워를 제공하여 쉽게 기계학습 시스템을 구성할 수 있도록 돕는다. 승강기 기업 티센크루프는 애저 기계학습을 통해 74%의 고장을 예측하였으며, 카네기 멜론 대학은 애저를 이용해 유동인구 데이터를 활용함으로써 에너지소비를 약 30% 절감하였다[58].

(3) 페이스북의 사진 공유 서비스 Moments

페이스북은 인공지능연구소를 설립하고 두 명의 유명한 전문가 러쿤(Yann LeCun) 교수 및 퍼거스(Rob Fergus) 교수를 영입하였다. 페이스북은 2013년 Deep Face라는 딥러닝 기반 얼굴인식 엔진을 개발하였다. 당시 LWF 얼굴 영상 데이터에 대하여 97.35%의 높은 성능을 발표한 바 있다[19]. 이를 기반으로 페이스북은 2015년 6월 얼굴인식기술을 탑재한 사진공유 서비스 Moments를 공개하였다[59]. Moments는 사진을 스캔해서 사진 내 특정 인물의 앨범을 자동으로 생성한다. 흥미로운 것은 사용자가 아니라 다른 사람이 촬영한 사진들까지도 모을 수 있는 점이다.

[그림 14] Facebook Moments

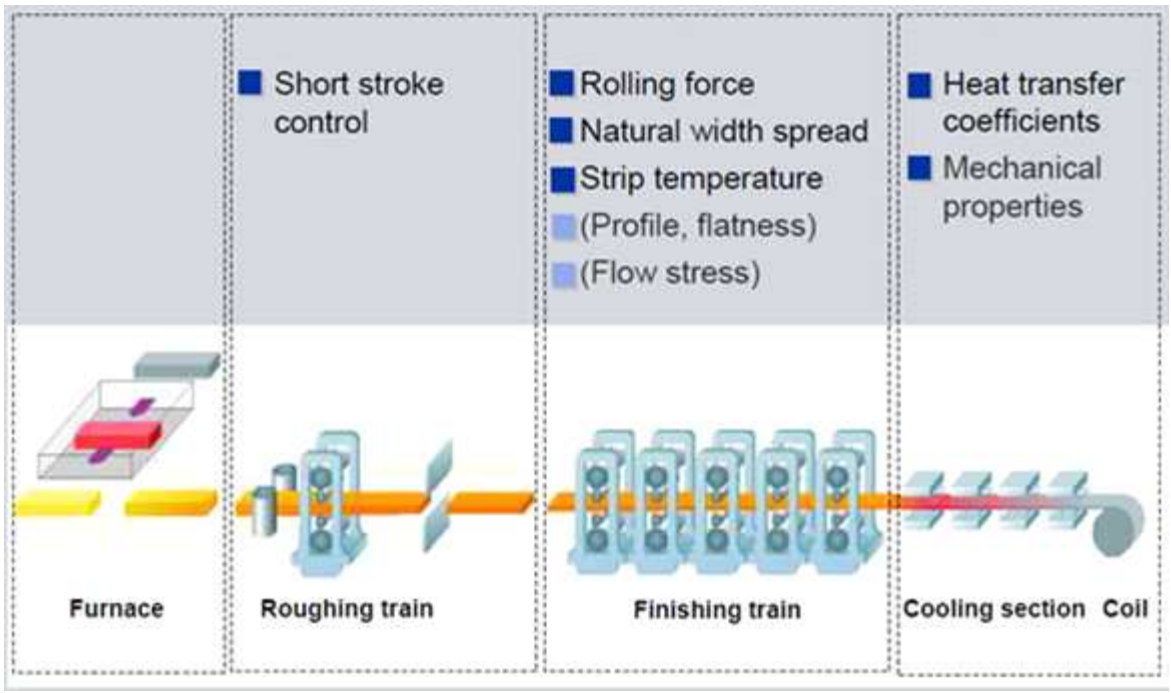


(4) 지멘스의 기계학습 응용 사례

지멘스(Simens)는 오래 전부터 기계학습을 활용해 온 대표적인 기업이다. 기계학습을 성공적으로 실용화한 경험을 바탕으로 적용 분야를 점점 넓히고 있다[59].

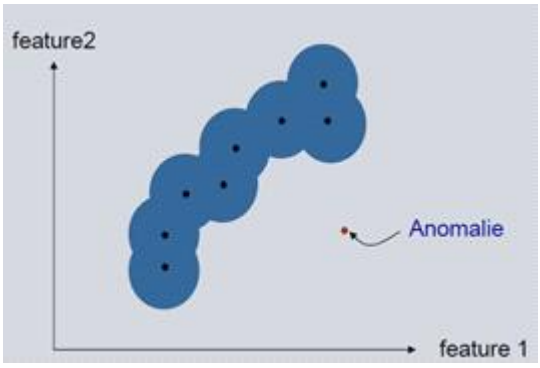
먼저 지멘스는 철강 생산과정 중 압연에 기계학습을 적용하였다. 철강의 품질은 압연 과정 중 로울러의 힘, 온도 등 다양한 요소의 영향을 받는다. 이러한 상관관계를 기계학습을 통해 학습함으로써 효과적으로 압연 공정을 제어하였다. 기존의 방법은 물리적 공식만을 이용하였는데, 기계학습을 적용하자 오차가 크게 감소하였을 뿐 아니라 환경의 변화에도 잘 적응하였다.

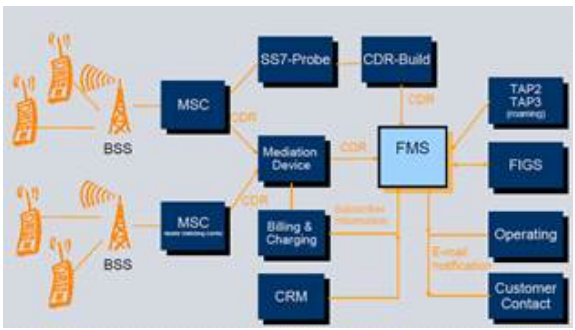
[그림 15] 지멘스의 철강 생산과정



또한, 지멘스는 시스템의 동작을 관찰하다가 비정상적으로 보이는 이상 현상(anomaly)을 검출하는데 기계학습 알고리즘을 사용하였다. 정상적인 상황에서의 동작을 확률 분포, 또는 클러스터의 형태로 학습한 후 학습된 패턴과 다른 현상이 발견되면 이상 상황으로 검출한다. 이러한 이상 현상 검출 기술은 사고상황 관리, 금융 사기 검출, 원격 기계 진단 등에 폭넓게 응용되었다.

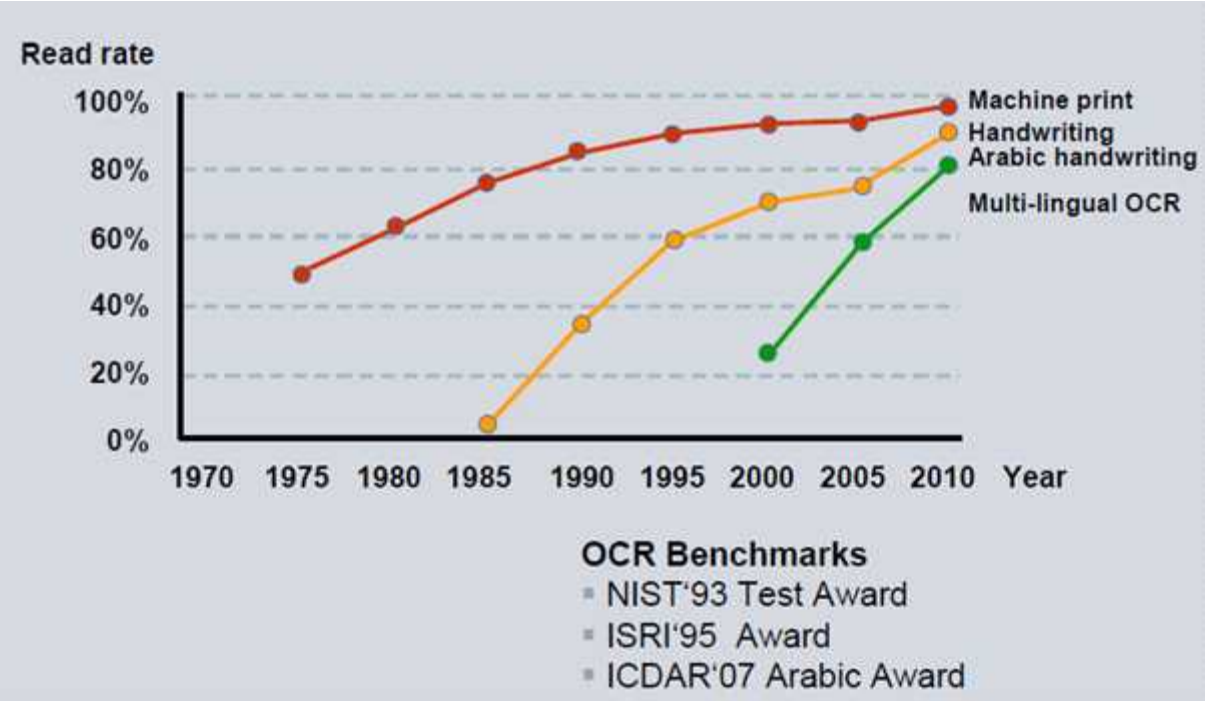
[그림 16] 이상 현상 검출 방법 및 이를 이용한 금융사기 검출 시스템





주소인식을 통한 우편자동화 및 의료영상분석은 지멘스가 오래 전부터 기계학습 기술을 적용해 온 분야이다. 오랜 기간 동안 다양한 기계학습 방법론을 적용시키며 지속적으로 성능이 개선되고 있다. [그림 17]는 기계학습 기술의 발전에 따른 주소 인식 성능의 개선을 보여준다.

[그림 17] 주소 인식 성능의 발전 (Simens)



(5) 오바마 선거캠프의 유권자 맞춤 선거 전략

미국의 오바마 대통령의 재선캠프에서는 빅데이터와 기계학습을 효과적으로 이용한 사례로 손꼽힌다. 근사적으로 접근하던 모든 데이터를 수치화 하고 그 값을 바탕으로 효율적으로 자원을 분배하였다. 또한, 유권자들을 투표성향, 성별 인종 등으로 나누어 유권자 명부를 작성하고, 이를 바탕으로 유권자들의 성향에 다라 다른 전략을 사용하였다. 이와 같은 기법은 선거 자금 모금에도 사용되었다. 이와 같이 기계학습을 활용한 덕분에 오바마 대통령은 선거를 매우 효율적으로 치를 수 있었다.

(6) 캐스피다(Caspida)의 기계학습 기반 지능형 보안 시스템

캐스피다는 기계학습을 이용하여 세계최고 수준의 보안 탐지 기술을 보유하였다[62]. 보안사건 중 다수가 자격증명을 악용한 공격으로 파악되고 있는데, 공격자가 유효한 자격 증명을 사용할 경우 기존 보안 기법으로는 탐지하기가 매우 어렵다. 이와 같이 보이지 않는 위협을 탐지할 수 있는 가장 좋은 방법은 기계학습이다. 기계학습을 이용하면 고객들의 행동 분석을 통해 보안 위협을 사전에 감지하고 대응할 수 있다. 캐스피다는 기계학습을 이용해 지능형 위협과 악의적 내부자를 탐지하여 보안 업계에서 혁신을 주도한 업체로 평가받고 있다. 이와 같은 기술력을 인정받아 2015년 7월 캐스피다는 스플렁크에 인수되었다[62].

(7) 크리테오(Criteo)의 기계학습 기반 개인 맞춤형 퍼포먼스 광고

크리테오는 맞춤형 광고 디스플레이 기업이다[63]. 기계학습 기술을 이용하여 대량 소비자 데이터를 분석해 개인 맞춤형 퍼포먼스 광고를 집행한다. 그 결과 사용자가 관심을 가질 만 한 광고를 디스플레이할 수 있고, 그 결과 구매 전환 비율이 높다. 뿐만 아니라 데이터가 많아지면 성능이 개선되는 기계학습의 특성상 광고가 집행될수록 예측력과 추천 정확도가 높아진다.

이와 같은 기술을 바탕으로 측정 가능한 ROI를 제공하는 것이 가능하다. 기존의 광고 엔진이 클릭률을 기준으로 광고비를 결정하지만 클릭이 실제 구매로 이어지지 않는 경우가 많기 때문에 광고주 입장에서는 불확실성이 많다. 이에 반해, 크리테오는 기계학습을 이용해 30%에 이르는 높은 구매 전환율을 기반으로 광고비를 제시할 수 있어 광고주들이 기대 수익을 예측하기 용이하다.

(8) 아마존(Amazon): 구매추천 및 예측, AWS 기계학습 클라우드

아마존은 기계학습을 이용해 고객의 구매 패턴을 분석해 관심 있을 만한 상품을 추천한다. 구매 추천 서비스는 매우 효과적이어서 아마존 매출의 1/3 이상이 구매 추천 서비스에 의해 발생한다. 더 나아가 2014년에는 고객이 주문하기도 전에 배송을 시작하는 기술을 실용화 하였으며, 관련 특허도 확보하였다[64]. 기계학습을 이용해 고객이 구매할 것으로 예상되는 물품을 미리 포장해서 고객과 가까운 물류창고에 옮겨 놓음으로써 배달 시간과 물류비용을 절감하였다.

이와 같이 기계학습 기술의 효과를 활용한 경험을 바탕으로 아마존 웹서비스(AWS, Amazon Web Services)에서는 클라우드 서비스에 기계학습 기능을 추가해 제공하고 있다[65]. 아마존이 보유한 기계학습 알고리즘을 클라우드 사용자가 활용하여 부정거래탐지, 요구예측, 콘텐츠 개인화, 사용자행동 예측, 소셜미디어 확인, 텍스트 분석 등 인공지능을 요구하는 애플리케이션을 구축할 수 있다.

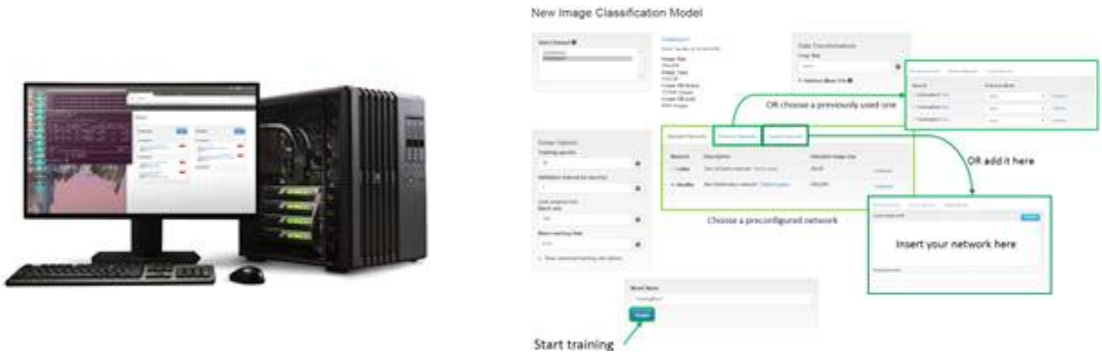
(9) NVIDIA: 기계학습을 위한 컴퓨팅 인프라

그래픽 장비 전문 기업 NVIDIA는 딥러닝 열풍으로 큰 기회를 맞이한 H/W 제조사 중 하나이다. 딥러닝은 많은 계산량을 요구하기 때문에 GPU를 이용한 대규모 병렬 처리가 필수적이다. NVIDIA는 CUDA라는 툴킷을 이용해 자사의 GPU를 수치 계산에 사용하기 편리하도록 상품화 하였는데, 딥러닝은 CUDA의 성능

을 활용하기에 매우 적합하다. 그로 인해 NVIDIA에서는 고성능 VGA의 판매를 크게 늘릴 수 있었다. 더 나아가 NVIDIA는 이러한 흐름에 적극적으로 대응하였다. CUDA 툴킷에 딥러닝 기술을 추가하여 cuDNN이라는 라이브러리를 보급하였는데, 딥러닝 연구자들로부터 좋은 반응을 얻었다.

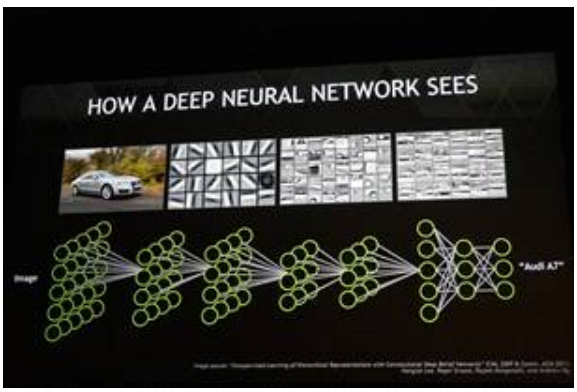
또한, 자사의 고성능 VGA를 탑재한 기계학습용 컴퓨터 DIGITS DevBox를 구성해 시판하였다[66]. DIGITs DevBox에는 4대의 Titan X GPU와 i7-5930K CPU, 512G M.2 SSD 등 고성능 부품과 NVIDIA의 S/W, 그리고 유명 오픈소스 딥러닝 S/W들을 탑재하고 있다.

[그림 18] NVIDIA DIGITS DevBox



또한, NVIDIA는 스마트카에 기계학습이 중요하게 사용된다는 점을 빠르게 인지하고 이에 빠르게 대응하고 있다. NVIDIA는 아우디 등 자동차 업체들과 협력하에 스마트카의 무인 주행을 위한 컴퓨터 Driver-PX를 개발하였다[67]. 동시에 12개의 카메라 영상을 인식해 주변 상황을 파악하고 주행할 수 있는데, 여기에 사용된 테그라 X1칩은 자율주행에 필요한 딥러닝 알고리즘을 가동하는데 필요한 GPU 기반 컴퓨팅을 지원하기 때문이다.

[그림 19] NVIDIA의 스마트카를 위한 딥러닝 GPU 플랫폼 Drive-PX



6. 기계학습 관련 산업의 활성화를 위한 제안

지금까지 해외에서 이루어지고 있는 기계학습 기술의 급속한 발전과 성공적인 실용화 사례들을 기술하였다. 기계학습 기술은 일반인들이 크게 인식하지 못하는 사이 이미 많은 변화를 일으키고 있다. 기계학습의 향후 발전 잠재력과 아직 미개척 상태인 수많은 응용 분야, 그리고, 기계학습에 투자되고 있는 막대한 인력과 자원을 고려할 때 기계학습의 중요성은 당분간 크게 증가할 것으로 예상된다. 본 장에서는 이처럼 임팩트가 큰 기계학습 기술과 관련해서 어떻게 경쟁력을 확보할 것인가에 대해 기술한다.

먼저, 해외에서 이루어진 기계학습 연구 및 산업화의 성공 요인을 서술함으로써 기계학습 연구 및 산업화를 위해서는 어떠한 환경이 필요한지 살펴본다. 그리고, 현재 우리나라의 학계와 산업계에서 기계학습 기술을 습득/발전시키고 산업화 하는데 어떠한 장애물들이 있는지 정리한다. 마지막으로, 이러한 장애물을 극복하고 기계학습 분야의 경쟁력을 확보하기 위해 필요한 정책들을 제안한다.

(1) 해외 기계학습 기술의 발전 및 산업화 성공 비결

기계학습의 발전을 견인하고 있는 딥러닝 신화의 뒤에는 몇 명의 매우 뛰어난 연구자들이 있었다. 토론토 대학의 제프리 힌튼(Geoffrey Hinton), 뉴욕 대학의 얀 러쿤(Yann LeCun), 몬트리올 대학의 조수아 벤지오(Joshua Bengio), 그리고 스텐포드 대학의 앤드류 응(Andrew Ng) 등이 그들이다. 이들은 RBM, DBN, CNN, Stacked Auto-encoder 등 딥러닝의 핵심 방법론들을 발표하며 기계학습 분야의 기술적 장애물을 돌파해 냈다.

딥러닝 연구가 활성화된 후에도 많은 연구 결과를 발표하며 딥러닝 연구의 리더로서의 역할을 수행하였다. 오랫동안 기계학습 연구를 직접 수행하면서 다양한 경험과 고찰을 통해 지니게 된 이들의 통찰력은 기계학습의 발전에 매우 중요한 가이드가 되고 있다. 특히 이들이 이끄는 그룹에서는 새롭고 효과적인 아이디어들이 많이 발표했을 뿐 아니라 다양한 벤치마크에서 탁월한 성능을 보임으로써 전 세계 연구자들에게 기술적인 성공 모델을 제시하였다.

딥러닝의 리더들은 기술적인 공헌 외에도 개방적인 연구 문화를 만드는데 큰 역할을 하였다. 자신들이 개발한 딥러닝 방법론들을 빠르게 공개하였고, 직접 구현한

딥러닝 엔진까지도 공개하면서 다른 사람들이 비교적 쉽게 진입할 수 있도록 도왔다. 또한, 딥러닝 연구자들을 위해 홈페이지, 페이스북, Youtube 등을 통해 자신들의 연구 결과를 포함한 딥러닝 분야의 기술 발전 상황을 잘 정리해서 배포하였다.

해외 딥러닝 연구의 중요한 요소는 풍부한 공개 데이터이다. 학습 및 평가를 위한 데이터는 딥러닝 연구에 필수적이다. 해외에서는 빅데이터가 활성화되기 훨씬 이전부터 NIST(미국 국립표준기술연구소) 등의 정부기관이 연구를 위한 데이터를 구축해 보급하였다. 또한, 스텐포드, 버클리 등 우수 대학에서는 자신들이 수집한 데이터를 연구자들에게 공개하였다. 특히, 스텐포드, 프린스톤, A9, 구글이 후원하는 ImageNet 프로젝트는 대용량의 영상데이터를 제공함으로써 딥러닝 연구에 필요한 데이터를 풍부하게 제공하였다[68]. 따라서, 딥러닝 연구에 입문하는 사람들이라도 이러한 데이터를 활용해 자신들의 아이디어를 구현해 볼 수 있다.

여러 학술단체 및 커뮤니티에서는 기술 수요가 많은 문제들의 표준적인 데이터를 기준으로 성능 컨테스트를 개최하고 있다. 이러한 컨테스트는 많은 연구자들이 참가하여 기술의 경연장이 되면서 기술 교류를 활성화 하고 있다. 이러한 성능 컨테스트로는 ILSVRC(ImageNet Large Scale Visual Recognition Challenge), ICDAR의 문서인식기술 contest 등이 대표적이다. 특히, 많은 컨테스트에서 상위권을 차지한 팀들은 자신들의 알고리즘이나 소스 코드를 공개하였는데, 그로 인해 새로운 기술들이 매우 빠르게 보급되었다. 또한, 기술력이 뛰어난 기업들은 컨테스트에서 상위에 입상할 경우 단시간에 기술력을 인정받을 수 있기 때문에 지명도를 높일 수 있는 기회가 되기도 한다.

딥러닝 모델과 알고리즘은 수학적으로 복잡할 뿐 아니라, GPU 컴퓨팅이 필수적이기 때문에 일반적인 CPU 기반 컴퓨팅보다 구현하기 어렵다. 따라서, 딥러닝에 입문하는 연구자들은 상당한 시간과 노력을 투자해야 하는 현실적 장벽에 부딪치게 된다. 그런데, 선진 연구팀들이 딥러닝 엔진들을 오픈소스로 공개하면서 때문에 이러한 현실적 장벽을 크게 완화하였다. 이러한 오픈소스 엔진들은 뛰어난 전문가들에 의해 개발되었고, 많은 개발자들이 참여하면서 우수한 성능과 확장성을 가지고 있어서 딥러닝에 입문하는 연구자들에게 매우 큰 도움이 된다. 이러한 오픈소스로는 Caffe[69], Theano[70], Torch[71] 등이 있으며, 최근 구글이

공개한 텐서플로우(TensorFlow)[56]도 많은 관심을 받고 있다.

새로운 딥러닝 기술이 빠르게 공유되는 데에는 전통적인 학술지보다 훨씬 빠른 고속 출판 미디어들이 중요한 역할을 하였다. 기존 학술지들이 투고에서 개제까지 수개월에서 수년이 소요되는데 반해 새로운 미디어들은 온라인으로 매우 짧은 시간 내에 게재가 가능하다. Simons재단의 후원을 받아 코넬 대학이 운영하는 e-Print 아카이브[72]는 이러한 고속 미디어의 대표적인 예이다. 다수의 우수한 딥러닝 연구 결과가 이를 통해 처음 발표되었으며 나중에 정식 학술지에 게재되었다. 또한, CVPR(computer vision and pattern recognition conference)을 비롯한 학술대회에도 많은 우수 논문이 발표되었는데, 학술대회는 학술지에 비해 게재속도가 훨씬 빠르고 그 파급효과도 매우 크다. 특히 해외에서는 유명 학술대회에 논문을 발표할 경우 SCI 학술지 이상으로 인정받기 때문에 많은 논문들이 학술대회를 통해 발표되고 있으며 수준도 매우 높다. 딥러닝 기술이 이처럼 빠르게 발전한 배경에는 이와 같이 연구 성과를 빠르게 공유할 수 있는 출판 시스템의 역할이 컸다.

딥러닝이 여러 분야에서 우수한 성능을 보이자, 구글, 아마존, 페이스북을 비롯한 다양한 기업들이 관심을 가지고 참여하게 되었다. 이들은 학계의 연구자들과 긴밀한 협력을 통해 우수한 연구 성과를 발표하기도 하고, 학회나 프로젝트를 후원하면서 딥러닝 연구의 동력을 제공하였다. 무엇보다도 선진 IT기업들은 딥러닝 기술을 실제 제품과 서비스에 응용하면서 상업적인 성공 사례들을 다량으로 만들었으며, 이는 전 세계 산업계에서 딥러닝을 주목하게 되는 계기가 되었다.

딥러닝 연구를 이끌어온 연구자들과 기업들은 상호 협력을 통해 새롭고 흥미로운 연구를 수행해 그 결과들을 공개하였다. 이런 결과들은 언론에 크게 보도되면서 많은 사람들에게 화제가 되었고, 딥러닝에 대한 대중들의 관심을 유발하는데 큰 역할을 하였다. 1000대의 컴퓨터와 16,000개의 코어를 이용해 대용량 자율학습 알고리즘을 통해 고양이 얼굴을 학습해낸 성과는 산업계의 구글과 대학의 앤드류 응(Andrew Ng) 교수 연구팀이 협력해서 이루어낸 성과이다[2]. 또한, 구글의 자율주행차가 네바다주 운전면허를 취득한 일이나 자율주행차의 사고원인이 대부분 상대방 차의 부주의였다는 분석 결과 등이 언론에 회자되면서 대중들의 관심을 끌만한 이슈들을 많이 제공하였다. 더 나아가 인공지능에 대한 영화들이 만들어지고

몇 가지 철학적인 질문까지 유발하면서 인공지능과 딥러닝에 대한 관심은 더욱 높아졌다.

(2) 우리나라의 기계학습 발전의 장애물

해외에서 기술적, 상업적 성공 사례들이 많이 나타난 것과는 대조적으로 우리나라에서는 아직 딥러닝관련 연구와 산업화가 활성화되지 못하고 있다. 그 이유는 다음과 같이 정리할 수 있다.

가장 큰 원인은 딥러닝에 정통한 전문가의 부족이다. 2000년대까지 인공지능 기계학습 기술의 실용화 수준과 거리가 많았다. 그로 인해 기업, 정부, 대중의 관심을 별로 받지 못했으며 연구 지원도 적었다. 그러자, 기계학습 분야의 많은 연구자들이 타 분야로 이동하였다. 2013년 이후 딥러닝의 출현에 의해 기계학습 분야가 부흥하면서 갑자기 기계학습 전문가들의 수요가 급증하였다. 그러나, 그 동안 기계학습 분야에서 꾸준히 연구하던 전문가들은 매우 소수에 불과했기 때문에 현재는 산업적 수요에 비해 딥러닝 전문가의 수가 크게 부족한 상황이다.

또한, 국내에는 딥러닝 연구를 위해 필요한 데이터셋이 부족하다. 해외에서는 다양한 데이터셋이 구축/공개되어 연구에 사용되고 있으나, 우리나라에 적용하기 위해 필요한 데이터와는 성질이 다른 경우가 많다. 예를 들어, 얼굴인식 및 문서인식 기술을 우리나라에서의 실용화하려면 한국인의 얼굴영상데이터, 필기한글데이터 등 우리나라에서 수집된 데이터가 필요한데, 이러한 데이터들이 제대로 준비되어있지 못하다. 예를 들어, 필기한글 영상데이터의 경우 90년대에 만들어진 PE92나 SERI95a 이후 다양성과 수량이 개선된 데이터가 구축되지 못했다. 그 결과, 실제 우리나라에서 사용될 만한 응용분야에 딥러닝을 적용하려는 연구자들은 학습데이터 확보의 어려움에 직면하고 있다.

해외의 연구자와 기업들이 실질적인 성능 개선을 목표로 연구하는데 비해 우리나라에서는 논문을 목표로 연구하는 분위기가 팽배해 있다. 이는 SCI 논문만이 우수한 연구 실적으로 인정받도록 제도화 되어 있고, 정작 연구 결과물의 성능은 제대로 인정받을 만한 제도나 사회적 분위기 형성되어있지 않기 때문이다. 몇몇 언론사에서 실시하는 대학 평가가 SCI 논문 중심으로 되어 있는 것은 이러한 문제를 더욱 심화시키고 있다. 이와 같이, 성능보다 논문을 우선시하는 연구 문화로 인해 논문 발표는 증가하였음에도 실질적인 성능의 개선이나 새로운 응용분야의 발굴로 이어지지 못하고 있다.

또한, 해외에서는 성능 컨테스트에서 상위권에 진입하거나 학술대회에서 우수한 결과를 발표할 경우 높게 평가 받을 수 있도록 제도적, 사회적 환경이 형성되어 있으나 우리나라는 그렇지 못하다. 그로 인해 SCI를 중심으로 연구 결과를 발표하게 되는데, 연구에서 게재까지 많은 시간이 소요되기 때문에 딥러닝의 빠른 발전 속도를 따라가기 어려움이 있다.

(3) 기계학습 연구 및 상용화 활성화를 위한 제안

기계학습은 복잡하고 생소하여 일반인들이 이해하기는 쉽지 않은 기술이다. 그러나, 응용 범위가 광범위하고 사회적, 경제적으로 영향력이 크기 때문에 효과적으로 대응하지 못할 경우 산업 전반에 있어서 세계적 추세를 따라가지 못하고 경쟁력을 잃어버릴 가능성이 있다. 따라서, 기계학습이 빠르게 발전하면서 응용 범위를 넓혀가고 있는 이 시점에 정책적으로 기술발전 및 산업화에 대한 큰 그림을 그리고 체계적으로 추진해야 한다.

예를 들어, 핀테크, 헬스케어와 같은 융합 분야에는 기계학습을 잘 적용할 경우 파괴적 혁신을 일으킬 수 있는 기회가 많다. 그러나, 기계학습 전문가들과 협력할 수 있는 여건이 갖춰져 있지 않다. 예를 들어, 금융 및 의료 분야에는 SW를 잘 이해할 수 있는 융합 역량을 갖춘 전문가가 절대 부족하다. 또한, 스마트카에 응용을 위해서는 기존의 자동차 업계 및 H/W 업계와의 긴밀한 협력이 필요하고, 관련 법규의 적절한 보완 등 제도적 지원이 필요하다. 이러한 장애물들을 극복하기 위해서는 해당 융합 분야에 대한 큰 그림과 일관성 있는 추진이 필요하다.

현재 기계학습의 연구와 산업화에 가장 큰 걸림돌은 기계학습 전문가의 부족이다. 이를 극복하기 위해서는 인내심을 가지고 체계적인 고급인재 양성 정책이 필요하다. 기계학습 기술은 수학적으로 복잡하여 고급 전문가 육성에 많은 시간이 소요된다. 따라서, 국내의 여러 대학과 기업에서는 원천기술에 접근하기 보다는 오픈소스를 활용한 응용 연구 등 피상적인 연구에 그치고 있다. 오픈소스를 이용할 경우 짧은 시간 내에 응용 시스템을 구성할 수 있다는 장점이 있지만, 딥러닝 기술의 핵심에 접근하는 데에는 한계가 있다. 또한, 해외 그룹이 개발한 오픈소스에 지나치게 의존할 경우, 자체경쟁력을 갖출 기회를 놓치게 될 위험이 크다. 따라서, 정부에서는 딥러닝과 관련해서는 실용적 응용을 위한 오픈소스 기반 사업과 고급 핵심 인재양성을 위한 딥러닝 핵심 알고리즘 개발 사업으로 구분해 지원하는 것이 바람직하다. 특히, 고급인재양성을 위해서는 단기간 동안 대규모 연구단 단위로 거액의 연구비를 지급하는 것보다는 중소규모의 연구비를

다수의 소규모 그룹에 장기간 지원하는 것이 효과적이다. 또한, 금융, 의료 등 비IT분야에 SW과정을 확대함으로써 추후 기계학습 및 SW전문가들과 협력하여 해당 분야에 혁신을 주도할 수 있는 융합인력을 육성하는 것도 필요하다.

논문보다 성능 중심의 경쟁을 유도하기 위해서는 해외 사례와 같이 성능 중심의 기술 컨테스트를 개최하는 것이 효과적이다. 우리나라에서의 산업 활용을 위한 실질적 문제들을 제시하고, 이에 대한 표준 데이터를 구축한 후 그 데이터에 대하여 컨테스트를 지속적으로 개최한다면 해외 사례와 같이 실질적인 연구를 장려하고 촉진하는 효과를 얻을 수 있을 것이다. 또한, 연구 그룹들의 역량을 객관적으로 평가할 수 있기 때문에 학계 및 산업계의 투명성을 높이는 효과도 얻을 수 있다. 특히, 우수한 딥러닝 기술을 보유한 벤처기업에서는 자신의 기술력을 증명하여 M&A의 기회를 확대할 수 있게 될 것이며 이러한 성공사례가 발생할 경우 딥러닝 전문가들에게 보유기술의 산업화에 대한 동기를 부여할 것이다.

그런데, 컨테스트의 문제는 가까운 장래에 실용화할 수 있는 현실적 문제와 원천기술의 발전을 유도할 수 있는 고난도 문제로 구분하는 것이 바람직하다. 전자의 경우 오픈소스 등 가용 자원을 모두 활용하여 우수한 성능을 얻음으로써 가까운 장래에 산업화를 촉진할 수 있다. 후자의 경우, 기존 오픈소스의 피상적 응용만으로는 접근하기 어려운 수준의 문제를 제시함으로써 원천기술의 발전을 유도하며, 핵심 기술에 접근 가능한 깊이 있는 전문가의 육성을 촉진할 수 있다.

앞에서 기술한 바와 같이 기계학습 연구 및 실용화에는 데이터와 컴퓨팅 인프라가 필수적이다. 따라서, 정부의 주도하에 데이터와 컴퓨팅 인프라가 구축, 지원된다면 기계학습 연구에 큰 힘이 될 것이다. 그런데, 이 같은 지원이 기계학습의 연구와 산업 활성화에 기여하게 하기 위해서는 앞에서 언급한 기술 컨테스트와 연동하는 것이 바람직하다. 즉, 컨테스트를 통해 기술적, 산업적으로 중요한 문제를 제시하고 데이터를 수집하여 평가하며, 컨테스트 참가자를 대상으로 컴퓨팅 인프라를 지원한다면, 데이터 및 컴퓨팅 지원이 실질적인 핵심 기술의 성능 개선으로 이어지도록 유도하는데 도움이 될 것이다.

[참고문헌]

[1] https://en.wikipedia.org/wiki/Watson_(computer)

[2] https://googleblog.blogspot.kr/2012/06/using-large-scale-brain-simulations-for.html

[3] 에릭 슈미트, "머신러닝이 우리 삶 바꿀 것", http://www.bloter.net/archives/242228

[4] 빌 게이츠에게 물었다! Q&A 10가지, http://www.huffingtonpost.kr/2015/01/29/story_n_6568786.html

[5] Y. Bengio, Learning deep architectures for AI, Foundations and Trends in Machine Learning, vol. 2, iss. 1, pp. 1-127, 2009.

[6] G. E. Hinton, S. Osindero, Y. Teh, A fast learning algorithm for deep belief nets. Neural Computation vol. 18, no. 7, pp. 1527-1554, 2006.

[7] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[8] http://en.wikipedia.org/wiki/Multilayer_perceptron

[9] https://developer.nvidia.com/cuda-zone

[10] http://en.wikipedia.org/wiki/Backpropagation

[11] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, Extracting and composing robust features with denoising autoencoders, in Proceedings of the Twenty-fifth International Conference on Machine Learning (ICML'08), (W. W. Cohen, A. McCallum, and S. T. Roweis, eds.), pp. 1096-1103, ACM, 2008.

[12] D. C. Cireşan, U. Meier, J. Schmidhuber. Multi-column Deep Neural Networks for Image Classification. IEEE Conf. on Computer Vision and Pattern Recognition 2012.

[13] D. C. Cireşan and J. Schmidhuber. Multi-column Deep Neural Networks for Offline Handwritten Chinese Character Classification, IDSIA Technical Report No. IDSIA-05-13, 2013.

[14] K. Fukushima, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, vol. 36, no. 4, pp. 193-202, 1980.

[15] http://yann.lecun.com/exdb/mnist

[16] Cheng-Lin Liu, Fei Yin, Qiu-Feng Wang, Da-Han Wang, ICDAR 2011 Chinese Handwriting Recognition Competition, 2011. (http://www.nlpr.ia.ac.cn/events/HRcompetition/Report.html)

[17] Fei Yin, Qiu-Feng Wang, Xu-Yao Zhang, Cheng-Lin Liu, ICDAR 2013 Chinese Handwriting Recognition Competition, 2011. (http://www.nlpr.ia.ac.cn/events/CHRcompetition2013/competition/Report.html)

[18] http://www.image-net.org/challenges/LSVRC/2014/results

[19] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf, DeepFace: Closing the Gap to Human-Level Performance in Face Verification, CVPR2013.

[20] Hinton, Geoffrey, et al. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups." Signal Processing Magazine, IEEE 29.6 (2012): 82-97.

[21] In-Jung Kim, Xiaohui Xie, Handwritten Hangul recognition using deep convolutional neural networks, International Journal on Document Analysis and Recognition (IJDAR), vol. 18, issue 1, pp. 1-13, Sep. 2014.

[22] http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

[23] S. hochreiter and J. Schmidhuber, LONG SHORT-TERM MEMORY, Neural Computation 9(8):1735{1780, 1997

[24] Klaus Greff, Rupesh Kumar Srivastava, Jan Koutník, Bas R. Steunebrink, Jürgen Schmidhuber (2015). "LSTM: A Search Space Odyssey". arXiv:1503.04069.

[25] W. Yang, et. al, Improved Deep Convolutional Neural Network For Online Handwritten Chinese Character Recognition using Domain-Specific Knowledge, http://arxiv.org/ftp/arxiv/papers/1505/1505.07675.pdf

[26] http://www.image-net.org/challenges/LSVRC/2014/results

[27] F. Schroff, D. Kalenichenko and J. Philbin, FaceNet: A Unified Embedding for Face Recognition and Clustering, http://arxiv.org/abs/1503.03832

[28] I.-J. Kim, C. Choi, and S.-H. Lee, Improving discrimination ability of convolutional neural networks by hybrid learning, International Journal of Document Analysis and Recognition, online published (http://link.springer.com/article/10.1007/s10032-015-0256-9?wt_mc=internal.event.1.SEM.ArticleAuthorOnlineFirst), 2015

[29] Graves, Alex; Mohamed, Abdel-rahman; Hinton, Geoffrey (2013). "Speech Recognition with Deep Recurrent Neural Networks". Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on: 6645-6649.

[30] I. Goodfellow, et al. “Maxout networks“, arXiv preprint arXiv: 1302.4389, 2013.
[31] M. Lin, Q. Chen, and S. Yan, “Network in network” , Proc. ICLR, 2014.
[32] K. He, et al. “Spatial pyramid pooling in deep convolutional networks for v isual recognition“, Computer Vision-ECCV 2014. Springer International Publi shing, 346-361, 2014.
[33] C. Szegedy, et al. “Going deeper with convolutions.“ arXiv preprint arXiv:1 409.4842, 2014.
[34] D. Cireşan and J. Schmidhuber, “Multi-column deep neural networks for of fline handwritten Chinese character classification“, arXiv preprint arXiv:13 09.0261, 2013.
[35] A. Krizhevsky, I. Sutskever, and G. Hinton. “Imagenet classification with de ep convolutional neural networks“, Advances in neural information processi ng systems. 2012.
[36] K. Simonyan, and A. Zisserman, “Very deep convolutional networks for lar ge-scale image recognition“, arXiv preprint arXiv:1409.1556, 2014.
[37] https://en.wikipedia.org/wiki/Regularization_(mathematics)
[38] G. Hinton, et al. “Improving neural networks by preventing co-adaptation of feature detectors“, arXiv preprint arXiv:1207.0580, 2012.
[39] Y. Bengio, et al. “Curriculum learning.“ Proc. 26th annual international con ference on machine learning. ACM, 2009.
[40] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network traini ng by reducing internal covariate shift“, arXiv preprint arXiv:1502.03167, 2015.
[41] H. Lee, et al. “Convolutional deep belief networks for scalable unsupervise d learning of hierarchical representations.“ Proc. 26th Annual International Conference on Machine Learning. ACM, 2009.
[42] M. Zeiler, et al. “Deconvolutional networks“, Computer Vision and Pattern Recognition (CVPR), IEEE Conference on. IEEE, 2010.
[43] M. Zeiler and R. Fergus. “Visualizing and understanding convolutional netw orks“, Computer Vision-ECCV 2014, pp. 818-833, 2014.
[44] C. Farabet, “Towards Real-Time Image Understanding with Convolutional Network” , Diss. Université Paris-Est, 2013.
[45] R. Girshick, et al. “Rich feature hierarchies for accurate object detection and semantic segmentation“, Computer Vision and Pattern Recognition (CV PR), IEEE Conference on. IEEE, 2014.
[46] J. Ba, V. Mnih, and K. Kavukcuoglu. “Multiple object recognition with visua l attention“, arXiv preprint arXiv:1412.7755, 2014.
[47] P. Sermanet, A. Frome, and E. Real. “Attention for Fine-Grained Categoriz ation“, arXiv preprint arXiv:1412.7054, 2014.
[48] Larry Page TED, https://www.youtube.com/watch?v=6HQKkZzadlg
[49] RankBrain: How Google Is Using Artificial Intelligence To Rank Web Pages

http://marketingland.com/rankbrain-how-googles-using-artificial-intelligence-to-rank-web-pages-148794
[50] Machine Learning Drives Google Green Datacenter Ambitions, http://www.nextplatform.com/2015/06/08/machine-learning-drives-google-green-datacenter-ambitions
[51] Google Now, https://www.google.com/landing/now
[52] Smart Reply, http://www.wired.co.uk/news/archive/2015-11/03/google-smart-reply-machine-learning-email
[53] Google Photo, https://photos.google.com
[54] Inside Google Translate (https://www.youtube.com/watch?v=_GdSC1Z1Kzs)
[55] Google Self-driving Car Project, https://www.google.com/selfdrivingcar
[56] TensorFlow, https://www.tensorflow.org
[57] C. Burges, et.al, Learning to Rank using Gradient Descent, http://research.microsoft.com/en-us/um/people/cburges/papers/ICML_ranking.pdf
[58] What is Cortana?, http://windows.microsoft.com/en-us/windows-10/getstarted-what-is-cortana
[59] Facebook Moments, http://www.momentsapp.com/
[60] V. Tresp, “On the Growing Impact of Machine Learning inIndustry” , Simens
[61] Obama Changed The Political Campaign With Big Data, https://datafloq.com/read/big-data-obama-campaign/516
[62] Caspida, http://caspida.com
[63] Criteo, http://www.criteo.com
[64] Amazon Patents Anticipatory Shipping, http://techcrunch.com/2014/01/18/amazon-pre-ships
[65] What is Amazon Machine Learning?, http://docs.aws.amazon.com/machine-learning/latest/dg/what-is-amazon-machine-learning.html
[66] NVIDIA DIGITS DevBox, https://developer.nvidia.com/devbox
[67] NVIDIA Driver-PX, http://www.nvidia.com/object/drive-px.html
[68] ImagetNet homepage, http://www.image-net.org
[69] Caffe homepage, http://caffe.berkeleyvision.org
[70] Theano homepage, http://deeplearning.net/software/theano
[71] Torch homepage, http://torch.ch
[72] e-Print Archive, http://arxiv.org

주 의

1. 이 보고서는 소프트웨어정책연구소에서 수행한 연구보고서입니다.
2. 이 보고서의 내용을 발표할 때에는 반드시 소프트웨어정책연구소에서 수행한 연구결과임을 밝혀야 합니다.