

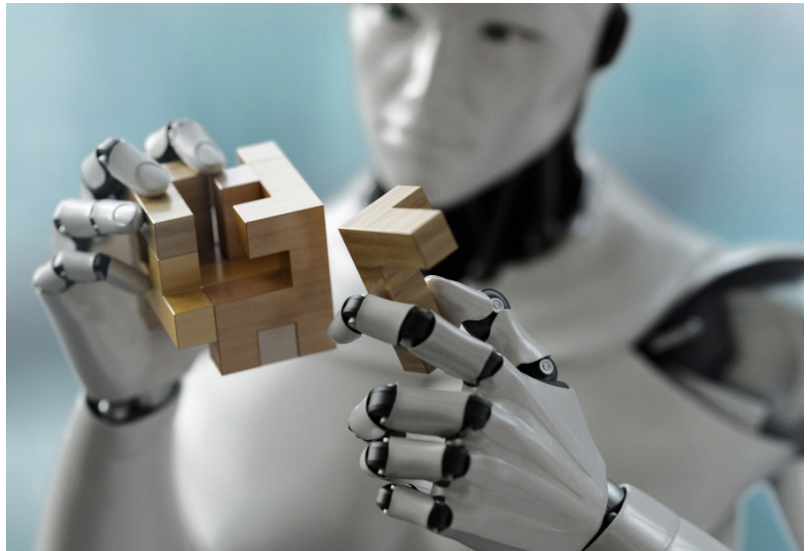
Artificial Intelligence

최근 인공지능 개발 트렌드와 미래의 진화 방향

이승훈

1. 인공지능 혁신의 시작
2. 인공지능의 최근 개발 트렌드
3. 한계와 극복
4. 새로운 시도의 시작
5. 맺음말

2017. 10. 10



요약

인공지능은 최근 5년간 매우 빠른 속도로 진화하고 있다. 과거 이론에 머물거나 제한된 기능만을 수행했던 인공지능은 이제 실제 구현을 통해 그 성능을 증명해내고 다양한 현실 세계의 문제에 하나씩 적용되기 시작했다. 2010년을 전후해 혁신적으로 발전한 알고리즘, 컴퓨팅, 빅데이터 기술이 서로 융복합되며 이런 성과를 이루어 내고 있다. 인공지능 분야의 혁신적인 논문, 이러한 논문들을 실제 구현 가능하게 하는 컴퓨팅 인프라(클라우드 및 GPU), 인공지능을 학습 시킬 수 있는 충분한 데이터가 확보되면서 이론에서 현실로 잘 나오지 못했던 인공지능이 우리의 곁으로 다가오게 된 것이다.

빠르게 발전하고 있는 인공지능은 이제 여러 분야에서 사람의 능력을 넘어서는 수준으로 구현되고 있다. 광범위한 분야에 걸쳐 인간처럼 외부의 정보를 인식하고, 학습하며, 추론하고, 행동하는 인공지능에 대한 연구가 활발히 진행되고 있다. 특히 시각, 청각지능 분야의 발전으로 인해 인공지능은 이제 사람 보다 더 높은 정확도로 사물을 인식할 수 있고, 사람과 비슷한 수준으로 언어를 이해할 수도 있게 되었다. 이러한 인식분야의 발전으로 인공지능은 이제 외부의 수 많은 데이터를 스스로 인식하고 이해해 지식화할 수 있는 '정보'로 받아드릴 수 있게 되었다. 그 동안 축적되어 온 엄청난 빅데이터를 기계가 스스로 학습할 수 있게 되면서 인공지능의 지능이 혁신적으로 발전하고 있는 것이다.

특히 최근 2년 간은 강화학습 및 관계형 추론, 예측 기반의 행동 분야 연구가 활발히 진행되며 인공지능이 인간의 사고 영역에 한걸음 더 다가 섰다. 알파고의 핵심 기술 중 하나인 강화학습(Reinforcement Learning)에 대한 연구가 2016년 이후 빠르게 발전하고 있다. 강화학습 분야의 발달로 인해 인공지능은 이제 목적 달성을 위한 방법을 시행착오를 통해 스스로 깨우치며 알아간다. 수 십만 번 이상의 반복 학습을 통해 터득하게 된 인공지능의 방법은 때로는 사람들이 전혀 생각해 내지 못했던 방식으로 문제를 해결해 내기도 한다. 게임과 같은 가상의 환경을 중심으로 연구되어 온 강화학습은 최근에는 3차원 환경, 현실 세계를 반영한 환경에서 연구가 진행되고 있다. 특히 일부 기업들의 연구소에서는 향후 제품, 서비스 탑재를 목적으로 강화학습 기반의 인공지능을 연구, 개발하기 시작하고 있다.

한 걸음 더 나아가 인공지능은 이제 다양한 정보들을 조합해 자신의 관점으로 새로운 명제를 추론(Inference/Reasoning)하거나 미래를 예측하고 행동하기도 한다. 인간의 고유 영역이라고 여겨져 온 추론/행동 분야의 연구는 2017년을 전후해 빠르게 발전해오고 있다. 특히 알파고를 구현한 딥마인드는 인공지능이 마치 인간처럼 추론하고 행동하는 논문을 잇따라 발표하며 인간처럼 유연한 사고가 가능한 인공지능 구현의 가능성을 보였다. 영상 혹은 텍스트로 주어진 정보를 개별적으로 인식하는 수준을 넘어 다양한 정보 간의 상대적인 관계를 직관적으로 파악해 추론해 내거나, 어떤 행동을 실행할 때 단순히 현재 상황에서 최선을 선택하는 것이 아니라 미래에 일어날 일들을 예측해서 행동하기도 한다. 불가능 할 것 같았던 관계형 추론, 예측 기반의 행동 분야의 인공지능 연구가 그 가능성을 보이면서 향후 인공지능의 발전은 한 단계 더 진화할 것으로 전망된다.

아직 한계는 많다. 막대한 양의 데이터와 컴퓨팅 파워가 필요하다. 알파고의 구현을 위해 서만 3000만개의 착점 정보가 필요했고 약 1200개에 달하는 CPU가 동시에 활용되었다.

이러한 한계 극복을 위한 노력들 또한 현재 진행 중이다. 데이터를 인위적으로 생성해 인공지능의 학습 과정에 활용하거나 현실을 정교하게 반영한 시뮬레이터를 구현해 반복학습이 가능한 환경을 가상으로 만들어 내기도 한다. 혹은 구현된 인공지능을 매우 단순화 시키거나 이미 학습된 지능을 다른 인공지능에 이식하여 새로운 지능 구현에 활용함으로써 학습 과정에 필요한 데이터나 컴퓨팅 파워를 최소화하기도 한다.

기존 인공지능과는 다른 전혀 새로운 방식으로 인공지능을 구현하려는 시도들도 시작되고 있다. 최근 5년간 인공지능이 엄청난 발전을 이루었지만 자율적인 판단과 능동적인 행동에 기반하는 인간의 지능과는 큰 차이가 있는 것이 현실이다. ‘인간처럼 계산(Computing like Human)’하는 지능을 넘어 ‘인간처럼 생각(Thinking like Human)’하는 지능을 구현하기 위한 노력들이 요구되고 있는 것이다. 이러한 노력들 중 하나로 신경과학(Neuroscience), 뇌과학(Brain Science) 분야에서의 인간 뇌에 대한 근본적인 연구를 컴퓨터 과학 분야의 연구에 접목 시켜 전혀 새로운 방식으로 인공지능을 구현하려는 시도도 시작되고 있다.

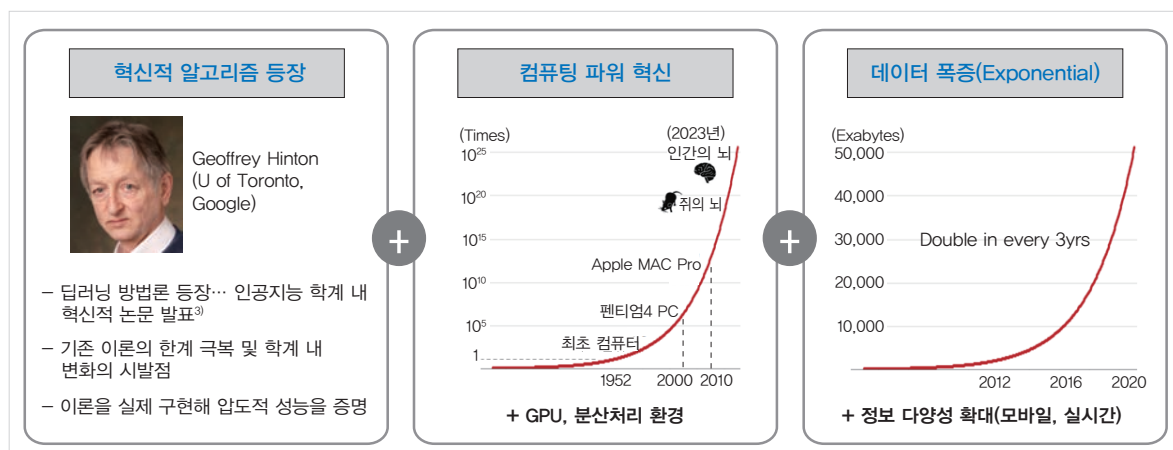
이렇듯 인간의 고유 영역이라고 생각되었던 분야에서 하루가 다르게 인공지능이 구현되고 있으며 그 성능 또한 인간의 수준을 빠르게 따라잡고 있다. DeepMind, OpenAI 등을 중심으로 혁신적인 논문이 연이어 발표되며 새로운 연구분야가 개척되고 다양한 연구 기관, 기업들이 후속 연구를 통해 단지 몇 달 만에 높은 완성도의 인공지능으로 구현해 내고 있는 상황이다. 주요 기업들은 이러한 연구 결과들을 자신들의 제품과 서비스에 빠르게 적용해 상용화하고 있다.

반면, 선도 연구기관 및 주요 기업들과 우리나라의 격차는 더욱 심화되고 있는 상황이다. 국내 기업들은 실리콘밸리의 기업들에 비해 상대적으로 소프트웨어 역량과 축적된 데이터 측면에서 상당히 열위에 있다. 단기적으로는 Tensorflow 등과 같은 오픈소스 기반의 개발 및 참여를 통한 역량 축적이 시급하며 보다 근본적으로는 중장기적인 관점의 양질의 데이터 확보, 경쟁력있는 인공지능의 개발 역량을 높이기 위한 노력이 병행되어야 할 것이다.

1. 인공지능 혁신의 시작

2000년대 까지만 하더라도 인공지능 연구자들은 주로 인간이 만들어 놓은 지식을 기계에게 학습 시키는 방법으로 인공지능을 구현해 왔다. 각 분야의 전문가들이 정교하게 모델링한 규칙들을 기계가 학습하면서 특정 분야의 인공지능이 만들어졌다. 이러한 방법으로 만들어진 인공지능은 일반적인 상황에서는 물론이고 다소 예외적인 상황들에서도 적절히 대응하며 꽤 높은 수준의 성능으로 구현될 수 있었다. 그러나 전문가의 역량과 상당한 시간, 투자가 수반되어야 했고 인간의 언어, 기호학적 표현의 한계, 데이터의 한계, 계산능력의 제약 등으로 적용 가능한 분야가 제한적이었다. 이로 인해 인공지능은 실제 현실에서의 활용보다는 주로 TV속의 ‘쇼’에서 존재해 왔던 것이다.

2012년, 인공지능 학계에서는 혁신적인 연구 결과가 나오게 된다. 인간의 사전 작업 없이도 기계가 데이터를 분석해 이미지 속의 사물을 구별해 내고¹, 고양이가 무엇인지 사전적 정보가 전혀 없이 천만 개의 동영상 스스로 학습해 영상 속에서 고양이를 구분해 낼 수 있게 되었다². ‘딥러닝’이라 불리는 새로운 기법을 활용해 구현된 이들 인공지능은 기존의 방법론에 비해 압도적인 성능을 나타내기 시작했다. 게다가 오랜 시간과 비용이 들었던 인간의 개입 과정도 획기적으로 줄어들었다. 물론 이러한 혁신적인 인공지능 이론³은 2000년대 중반 혹은 그 이전부터 제안되어 왔지만 최근의 IT, 전자 기술의 기하급수적 발전에 힘입어 비로소 실제 구현되기 시작했다. 과거 수 개월이 소요되었던 기계학습 과정은 이제 단지 수 시간, 혹은 몇 분만에 처리가 가능해졌다. 현실 세계를 반영하고 있는 방대한 데이터를 통해 기계는 마치 실제 세상 속에서 인간처럼 정보를 인지하고 학습해 지식으로 발전시켜 나가기 시작했다. 불과 최근 수

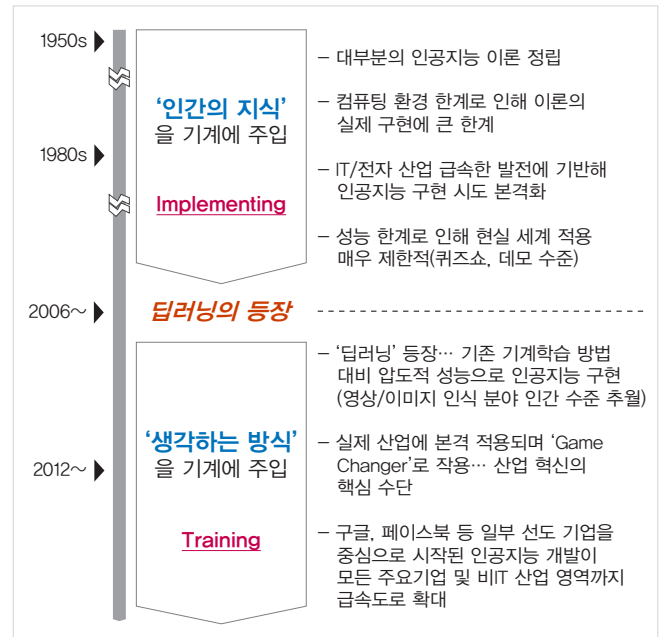


1 인공지능 발전의 핵심 요소

- 1 A. Krizhevsky, et al., ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems, 2012
- 2 Q. Le, et al., Building High-level Features Using Large Scale Unsupervised Learning, ICML 2012
- 3 G. Hinton, et al., A fast learning algorithm for deep belief nets, Neural Computation, 2006

년 사이에 이러한 일들이 일어나고 있다.

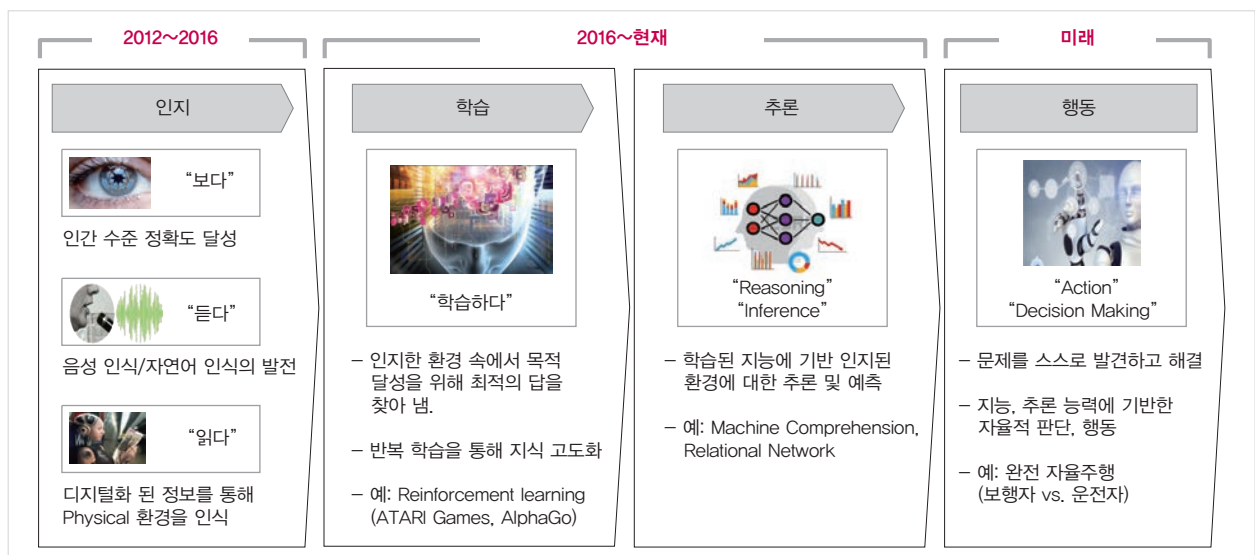
인간과 체스 대결에서 승리하거나, 퀴즈 대결⁴에서 승리하며 TV쇼에서만 존재해 왔던 인공지능은 이제 인간을 대신해서 운전을 하거나(Google: 자율주행 자동차), 월스트리트의 금융 전문가 보다 월등한 수익을 내며 투자를 하기도 하고(Kensho: 연봉 \$30~\$50만에 이르는 퀀트/애널리스트 15명이 4주 걸렸던 분석을 5분 만에 해결), 전문의 보다 더욱 정확한 진단을 내리기까지 한다(IBM Watson Health). 즉 '인공지능'이라는 단어가 처음 사용된 1956년부터 약 60년의 시간에 걸쳐 제대로 된 구현방법을 모색해 오던 인공지능이 이제 그 방법을 깨닫기 시작하면서 엄청난 속도로 발전할 수 있는 토대를 마련하게 된 것이다.



2 인공지능 구현 방법론의 전환

2. 인공지능의 최근 개발 트렌드

딥러닝으로 인한 인공지능의 발전은 인지, 학습, 추론, 행동과 같은 인간 지능 영역의 전 과정에 걸쳐 혁신적인 진화를 만들어 내고 있다. 시각, 청각과 같은 감각기관



3 인공지능 발전 분야

4 IBM DeepBlue(세계 체스 챔피언 카스파로프와 체스 대결 승리, 1997), IBM Watson(제퍼디 퀴즈쇼 인간과 대결 승리, 2011)

에 해당하는 인지지능에서부터 인공지능이 스스로 지능을 발전 시키는 학습, 새로운 상황을 추론하고 행동하는 단계에 이르기까지 다양한 분야의 연구가 동시 다발적으로 빠르게 발전되고 있다. 2012년을 기점으로 본격적으로 발전하고 있는 인지 분야의 지능은 이미 인간 능력 이상의 수준으로 구현되고 있다. 지능 발전의 가장 큰 걸림돌이었던 인지 분야의 해결은 인공지능이 현실 세계를 인간처럼 인식하는 것을 가능하게 하였고 이에 기반한 학습/추론/행동 분야의 연구가 매우 활발하게 진행되고 있다.

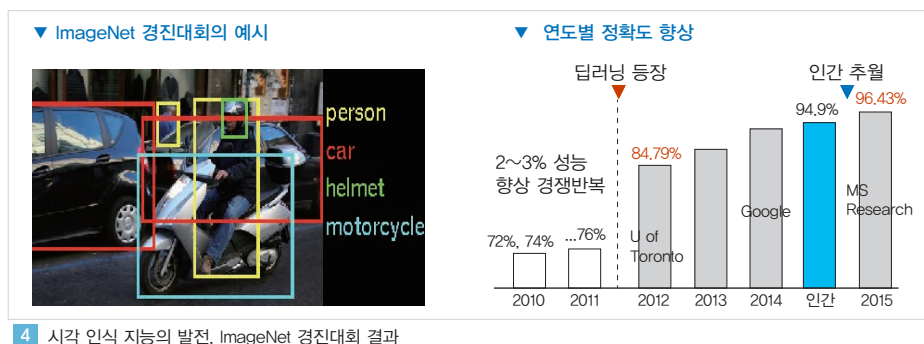
(1) 인지 지능의 발전: '인간처럼 보고/듣는 기계의 등장'

① 인간의 능력을 뛰어 넘는 시각 지능

이미지/영상의 인식과 이해

인간의 시각 지능에 해당하는 이미지 인식분야에서는 이미 인간 수준을 넘어서는 인공지능이 구현되고 있다. 2012년 유튜브 영상 속의 고양이를 스스로 구분해 낸 구글의 인공지능⁵과 이미지 속의 사물 인식 정확도를 혁신적으로 향상 시킨 토론토 대학의 인공지능⁶이 구현 된지 약 3년만에 인간의 수준을 뛰어 넘은 것이다. 매년 다양한 연구기관이 참여해 이미지 내 사물 인식의 정확도를 경쟁하는 ImageNet⁷ 경진대회는 2015년 마이크로소프트가 96.43%의 정확도를 달성하며 인간의 인식률(94.90%)을 추월하였다. (2017년 정확도: 97.85%⁸)

단순히 이미지속의 사물의 종류를 인식하는 것을 뛰어넘어 인공지능은 이제 영상/이미지속의 상황을 이해하게 된다. 사람 얼굴 사진을 보면 남성, 여성 등과 같은 외형적 특성을 인식할 수 있을 뿐 아니라 눈, 코, 입 모양의 상관 관계를 분석해 표정을



5 Q. Le, et al., Building High-level Features Using Large Scale Unsupervised Learning, ICML 2012

6 A. Krizhevsky, et al., ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012

7 스탠포드대에서 주관하는 영상인식 분야 경진대회... 1000가지 종류의 사물로 구성된 100만장의 이미지가 주어져서 각 이미지 속에 존재하는 사물의 종류를 알아 맞추는 경쟁

8 <http://image-net.org/challenges/LSVRC/2017/>



5 이미지/영상 이해: 사람의 표정/감정 인식(좌), 이미지 내 상황을 언어로 표현(우)

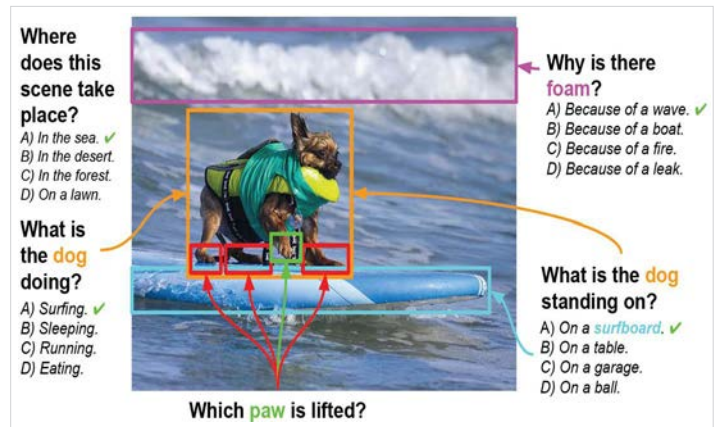
Source: Microsoft(좌), Google(우)

인지하거나 감정을 추측⁹한다. 게다가 2015년 구글이 발표한 논문에서는 이미지 속 상황을 정확히 이해해 인간의 언어로 표현¹⁰하기도 한다.

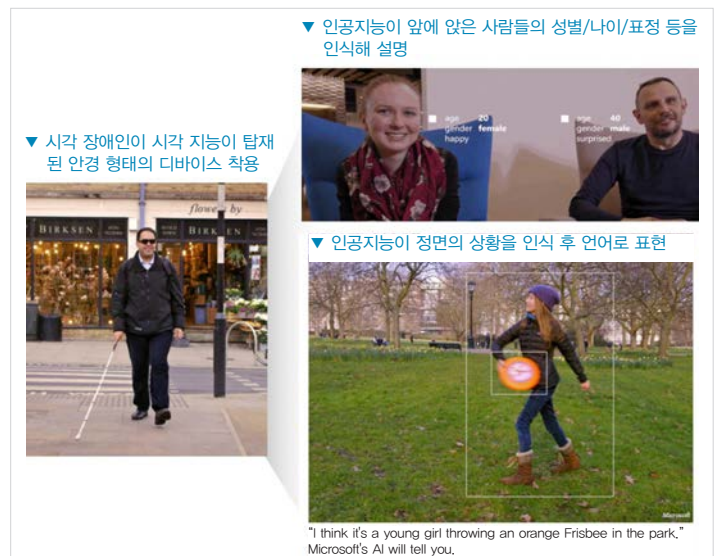
시각 지능을 통해 이미지를 인식하고 이해하게 된 인공지능은 이미지 속 상황에 대한 물음에 대해 정확히 답을 하기도 한다¹¹. <그림 6>과 같이 인공지능은 인간의 질문을 정확히 이해하고 답을 해낸다. ‘Attention’이라고 불리는 방법을 통해 인공지능은 이미지 내 다양한 사물 중 질문의 답에 해당하는 부분에 스스로 집중¹²하며 답을 찾아 낸다.

이처럼 진화하고 있는 시각 지능을 기반으로 한 인공지능은 연구 단계를 넘어서 실제 생활에 적용되며 다양한 혁신을 만들어 갈 것으로 전망된다. Microsoft는 고도화된 시각 지능을 활용한 ‘Seeing AI’라는 시각 장애인용 인공지능을 발표했다. 앞을 볼 수 없는 시각 장애인의 시각 지능을 인공지능이 대신하는 것이다. 시각 장애인에게 눈 앞의 상황을 인간의 언어로 설명해 주거나, 앞에 앉아 있는 상대방의 성별, 나이, 표정 등의 정보를 제공해 준다 <그림 7>. 즉 시각 장애를 갖는 사람들의 시각 인지 기능을 인공지능이 대신 제공해 장애인들의 일상 생활을 혁신적으로 변화 시킬 수 있을 것으로 전망된다.

실제 산업 영역에서는 자율주행(지능형 주행)



6 Visual Q&A: 이미지와 관련된 질의 응답.

Source: Y. Zhu(2016)¹¹

7 SeeingAI: 시각 인지 지능의 활용 예시

Source: Microsoft

9 Microsoft Emotion API

10 O. Vinyals, et al., Show and Tell: A Neural Image Caption Generator, CVPR 2015

11 Y. Zhu, et al., Visual7W: Grounded Question Answering in Image, CVPR 2016

12 Z. Yang, et al., Stacked Attention Networks for Image Question Answering, CVPR 2016

과 관련해 인공지능 기반의 시각 지능이 빠르게 적용되고 있다. 몇년전까지만 해도 구글 및 주요 자동차 제조사들의 자율주행/ADAS 기술은 LiDAR, Radar 등과 같은 특화 센서의 동원, 이들 센서로부터의 정보를 분석하고 소프트웨어로 구현할 수 있는 전문가와 막대한 투자/R&D 역량 확보가 가능한 소수의 기업만이 구현할 수 있었다.

하지만 최근 등장하고 있는 Startup들은 카메라, 초음파 센서 등과 같은 저가의 범용 센서와 딥러닝 기반의 시각 지능으로 지능형/자율 주행 기술을 구현해 내고 있다. Comma.ai, Drive.ai는 인공지능에게 다양한 주행 영상 데이터를 학습시키며 지능형 주행 기술을 구현한다. 마치 사람이 운전을 배워가는 과정과 같이 학습 초기에는 매우 단순한 충돌 방지, 차간 거리 유지, 조향 기능등을 학습 시킨다. 기본적인 주행 기능을 학습 후에는 차량 주행이 많지 않은 도로에서부터 시작해 고속도로까지 주행하며 다양한 환경에서 주행 방법을 익힌다. 교통 표지판, 신호등을 인식하거나 차간 거리를 유지 하는 등 대부분의 기능이 카메라를 통해 입력되는 시각 영상에 기반해 구현되며 일부 기능 구현에 필요한 정보들은 초음파, 레이더 등과 같은 센서를 활용하기도 한다. AutoX라는 또 다른 Startup은 이런 기능들을 다른 센서는 전혀 사용하지 않고 오직 카메라를 통해 입력된 영상만으로 구현한다. 6개의 카메라를 통해 입력되는 자동차 주변 환경 정보를 딥러닝으로 학습한 시각 지능이 마치 사람처럼 인지하고 자동차를 제어한다. AutoX의 창업자이자 프린스턴대 교수인 Xiao 박사는 자율주행 분야의 전문가가 아닌 컴퓨터 비전(Vision) 분야의 전문가로서 해당 기술을 구현해 내고 있다.

이와 같이 딥러닝에 기반한 시각 지능이 실제 산업에 적용되면서 산업 내 소수의 기존 경쟁자들이 갖고 있었던 기술 진입 장벽을 허물며 산업의 핵심 패러다임을 전환시키기 시작하고 있다.

이미지/영상의 합성과 생성

인간 수준 이상의 시각 지능을 갖게 된 인공지능은 이제 시각 정보를 자유롭게 변형하거나 전혀 새로운 이미지를 생성해 내기도 한다. 인간의 인식 수준 이상의 시각 지능에 기반해 만들어 지고 있는 이러한 가상의 이미지는 사람들이 쉽게 구분해 낼 수 없을 정도의 높은 완성도를 보이고 있다. UC버클리대 연구팀은 딥러닝을 활용해 이미지를 다양하게 변형하는 논문¹³을 발표하였다. <그림 8, 좌>와 같이 풍경 사진의 계절적 특성을 이해한 인공지능은 하나의 풍경 사진을 여름, 겨울 사진으로 변경하거나 동물/식물의 특성을 정확히 이해해 동물의 외형을 자유롭게 변형하기도 한다. 또한 모네, 고흐 등과 같은 유명화가의 그림을 반복적으로 학습해 일반 풍경 사진을 특

13 J. Zhu, et al., Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks ICCV 2017

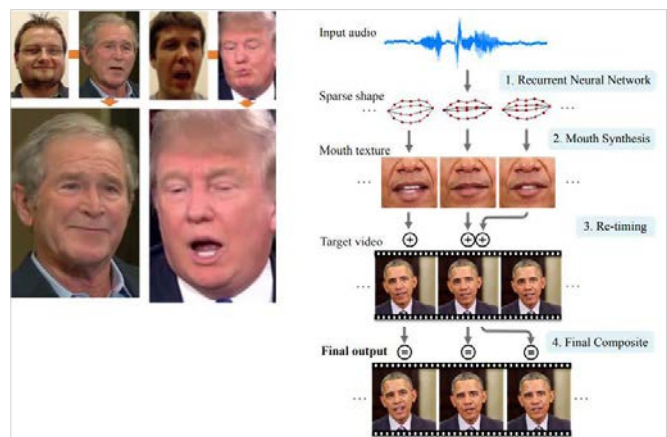


8 시각 지능 이해에 기반해 변형/합성된 이미지

Source: J. Zhu(2017)¹³, L. Gatys(2016)¹⁴

정 화가의 화풍이 접목된 그림으로 변환¹⁴하기도 한다(그림 8, 우). 이는 수많은 학습 과정을 통해 각 이미지의 특성을 정확히 이해하고 지식화해 새로운 이미지의 생성에 적용해 내는 것이다. 인공지능이 만들었다는 것을 사전에 알려주지 않으면 인공지능에 의해 인위적으로 생성된 가상의 이미지라는 것을 인간의 시각 지능으로는 구분해 내기 어려울 정도로 높은 완성도를 보이고 있다.

단순히 정지된 이미지를 합성하는데 그치지 않고 실시간의 동영상상을 합성하기도 한다. 스탠포드대의 연구팀은 유명인의 영상에 전혀 다른 사람의 표정을 합성한다. 실시간으로 바뀌는 표정이 그대로 유명인의 얼굴 표정에 반영¹⁵된다. 워싱턴대에서 발표한 'Synthesizing Obama'라는 논문¹⁶에서는 오바마 대통령의 목소리만을 가지고 입 모양을 생성해 오바마 대통령의 전혀 다른 연설 영상에 합성한다. 즉 단순히 정지된 이미지 정보의 합성 수준을 넘어 실시간의 영상 변형, 합성까지 가능한 것이다.



9 영상 정보의 실시간 합성: 유명인의 영상에 표정 합성(좌). 오바마 대통령의 목소리만으로 입 모양을 생성해 다른 연설 영상에 합성(우)

Source: J. Thies(2016)¹⁵, S. Suwajanakorn(2017)¹⁶

더 나아가 인공지능은 이제 세상에 존재하지 않는 전혀 새로운 사물을 만들어 내기도 한다. GAN(Generative Adversarial Networks)¹⁷이라 불리는 이 방법은 새로운 데이터를 생성하는 인공지능과 생성된 데이터가 진짜인지 혹은 가짜인지를 판별하는 두 인공지능이 서로 경쟁하며 진짜와 같은 가상의 이미지를 만들어 낸다. 구글의 Ian Goodfellow에 의해 제안된 GAN은 2016년 이후 매우 빠르게 성능이 향상되고 있으며(그림 10)과 같이 2017년 4월 발표된 논문¹⁸에 나

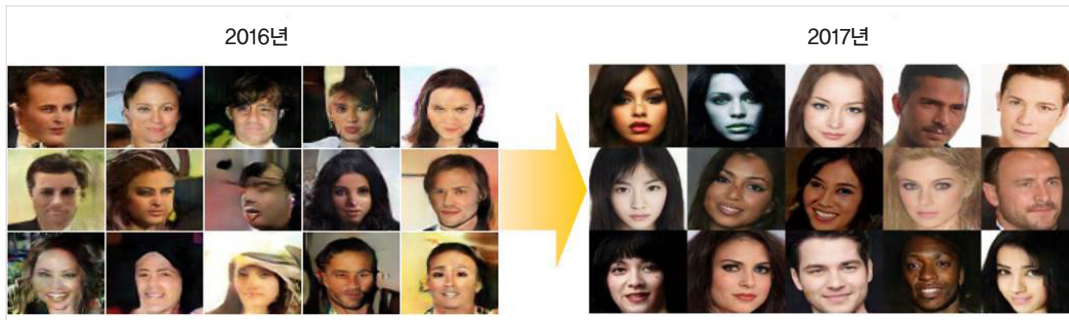
14 L. Gatys, et al., Image Style Transfer Using Convolutional Neural Networks, CVPR 2016

15 J. Thies, et al., Face2Face: Real-time Face Capture and Reenactment of RGB Videos, SIGGRAPH 2016

16 S. Suwajanakorn, et al., Synthesizing Obama: Learning Lip Sync from Audio, SIGGRAPH 2017

17 I. Goodfellow, et al., Generative Adversarial Networks, 2012

18 D. Berhelot, et al., BEGAN: Boundary Equilibrium Generative Adversarial Networks, 2017



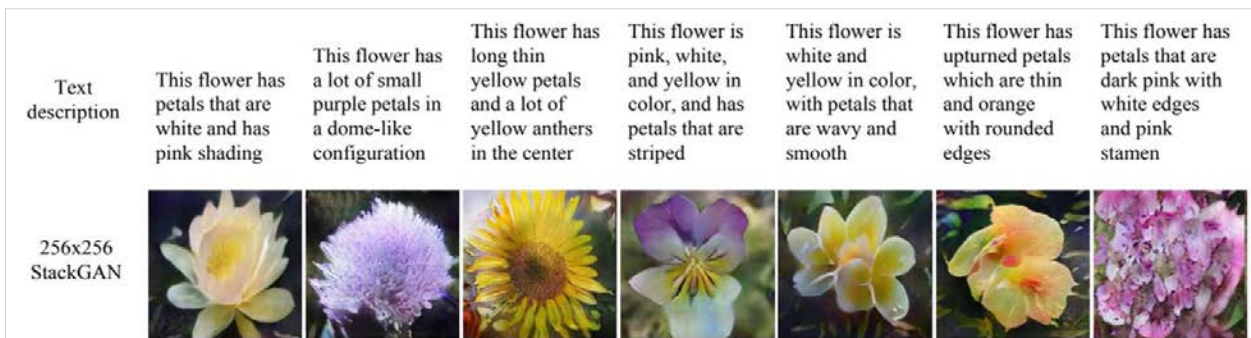
10 GAN 기반의 이미지 생성 기술 발전: 2016년 생성 이미지(좌), 2017년 생성 이미지(우)

Source: D. Berhelot(2017)¹⁸

탄안 이미지들은 가상으로 생성된 이미지라는 것을 구분할 수 없을 정도로 높은 완성도로 구현되고 있다.

2016년 발표된 StackGAN¹⁹은 인간의 언어로 기술된 텍스트를 이해해 특정 사물을 생성하기도 한다. <그림 11>과 같이 특정한 모양, 색깔의 꽃 사진을 생성해 내거나 새로 만들어 내기도 한다. 물론 이러한 사물은 GAN과 마찬가지로 세상에 존재하지 않는 전혀 새로운 가상의 형상들이며 생성된 결과만으로는 인간의 시각 지능으로 판별이 매우 어렵다.

이러한 시각 인지 지능의 산업적 영향력은 매우 클 것으로 전망된다. 단기간에 직접적으로는 엔터테인먼트, VR/AR 과 같은 영상 콘텐츠와 관련된 주요 산업에 핵심 역량으로 작용할 가능성이 높다. 인공지능이 애니메이션의 영상 뿐 아니라 음성까지도 스스로 생성해 내는 것이 가능하며, 이를 통해 유명 배우의 외형을 학습한 인공지능은 다양한 모습으로 배우의 영상을 변형하거나 새롭게 생성하는 것도 가능할 것이다. 단순한 콘텐츠 산업을 넘어 인간의 시각이 관련된 거의 모든 산업에 직/간접적으로 영향을 미칠 수도 있다. 교육, 쇼핑, 교통 등 모든 영역에서 산업의 핵심 요소 기술로 작용해 기존 산업의 경쟁 방식을 혁신 시킬 가능성도 있다.



11 인간의 언어를 이해해 이미지를 생성: 생성하고 싶은 꽃의 형태를 언어로 기술(상), 기술된 언어에 기반해 인공지능이 생성한 꽃 이미지(하)

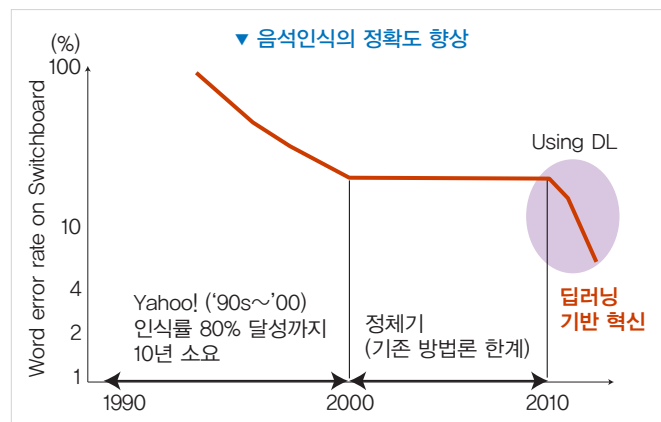
Source: H. Zhang(2016)¹⁹

19 H. Zhang, et al., StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks, 2016

② 인간 수준의 언어 인식/이해 지능의 구현

음성 인식분야의 인공지능 연구는 매우 오래 전부터 다양한 기업들이 연구, 개발을 해오고 있지만 현재까지도 자유로운 대화가 가능한 수준까지는 구현되지 못하고 있다. 1990년대부터 당시 거대 IT 기업이었던 Yahoo 등 주요 기업들은 엄청난 투자를 통해 음성 인식 기능을 구현하려 노력하였으나 그 정확도는 높지 않았다. 약 80% 정도의 정확도를 달성하는데 10년 이상의 시간이 소요되었으나 음성 인식은 그 후 정체를 맞이하였고 현재까지도 인간 수준의 인식률에는 크게 미치지 못하며 널리 상용화 되지 못하고 있는 상황이다.

언어 인식 분야의 지능이 빠르게 발전되지 못했던 것은 기존 사람(전문가) 중심 방법론의 한계 때문이다. 기계가 인간의 언어를 인식하기 위해서는 개별 단어의 의미를 이해하는 것을 시작으로 구문/문장 등 매우 복잡하고 다양한 단어들의 관계들이 정확하고 정교하게 모델링 되어야 한다. 과거에는 이러한 단어 간 관계 정의를 언어학을 전공한 전문가가 중심이 되어 모델링해 왔다. ‘Ontology’라 불리는 이러한 언어 모델은 전문가가 일일이 단어간의 관계를 설정해 놓는 방식으로 구현되며 그렇기 때문에 전문가의 능력, 경험, 투자 비용 등이 언어 인식의 핵심 역량으로 작용해 왔다. 하지만 이러한 Ontology 기반의 언어 인식 모델은 언어의 확장성이 낮다는 큰 단점을 갖고 있다. 새로운 언어가 추가될 때마다 사람이 직접 모델을 다시 수정해야 하거나, 의학/법률/금융 등과 같이 정확한 언어 이해를 위해 특정한 분야의 전문지식이 바탕이 되어야 하는 경우 각 분야의 전문가가 언어 모델의 작성에 개입해야 하는 등의 한계가 있다.



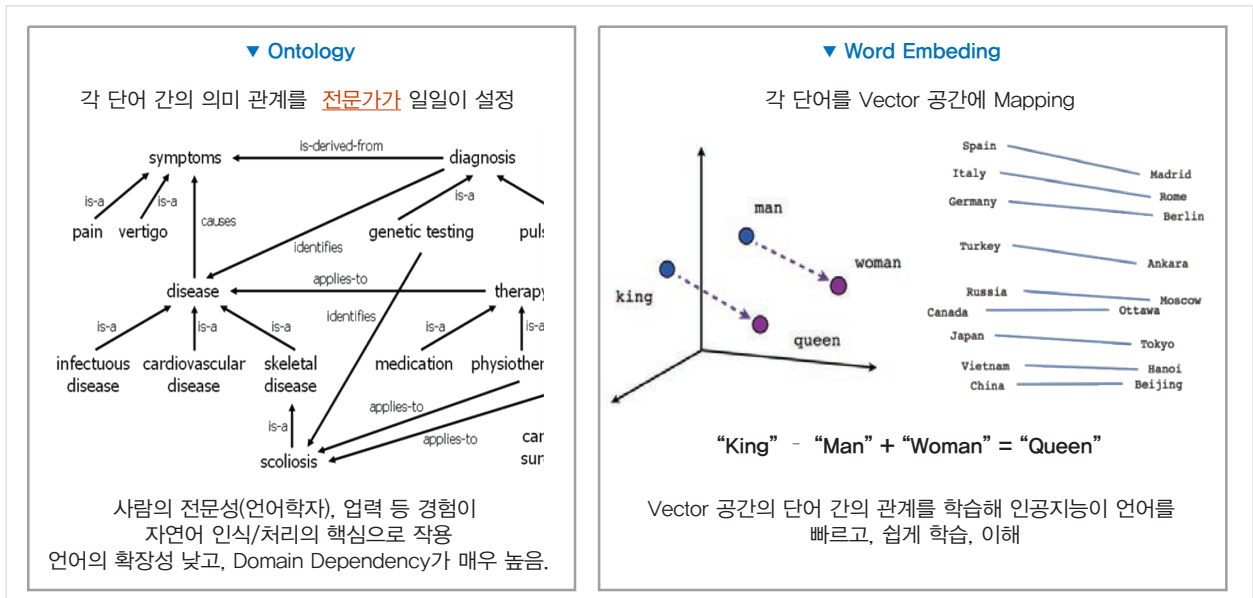
12 언어 인식 지능의 발전

하지만 최근 딥러닝이 적용되면서 과거와 달리 사람(전문가)에 의존하지 않고 인공지능이 데이터에 기반한 학습을 통해 스스로 언어를 이해하게 하는 방식으로 전환되고 있다. 구글은 웹 서비스 과정에서 축적한 데이터를 기반으로 ‘word2vec’²⁰이라는 언어 모델을 구현하였다. 수년간 뉴스 서비스를 통해 확보한 텍스트 정보에서 약 1000억개에 이르는 단어²¹를 기계학습에 활용하였다. 개별 단어가 아닌 구문 단위로 이해된²² 각각의 단어를 수 백개의 차원으로 구성된 벡터 공간에 위치 시켰다. <그림 13>과 같이 각 단어를 벡터 공간에 개별적으로 위치 시키면, 결과적으로 관련성이 높은

20 Y. Glodberg, et al, word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method, 2014

21 단어를 구문 단위로 학습해 동일 단어라도 문맥적 의미에 다른 벡터 공간에 위치 시킴

22 개별 단어가 아닌 구문 내의 단어의 의미를 분석해 벡터 공간에 위치... 동음이의어의 경우 앞/뒤 단어와 관계를 통해 같은 문자의 단어라도 다른 벡터 공간에 위치하게 함(He 'may' want vs. 'May' 15th)



13 Ontology 방식과 Word Embedding 방식 비교

Source: J. Pennington(2014)²³

단어들이 벡터 공간 내 서로 유사한 위치에 존재하게 되며 단어 별 상관 관계가 자동적으로 정의된다. ‘Spain/Italy/Germany...’ 와 같은 국가 이름과 ‘Madrid/Rome/Berlin...’과 같은 수도가 서로 유사한 공간에 그룹된 형태로 위치하게 되고 국가 이름과 수도 간에는 일정한 거리를 두고 위치하게 되는 것이다. 최근 구현되고 있는 언어 인식/이해 지능들은 대부분 이렇게 데이터를 기반으로 구성된 언어 모델에 의해 구현되고 있다. 구글의 word2vec, 스탠포드대의 GloVe²³ 등이 대표적인 예다. 이들은 모두 구현된 World Embedding 형태의 모델을 공개해 누구나 쉽게 이들 모델을 기반으로 언어 인식 지능을 연구·개발할 수 있게 하고 있다.

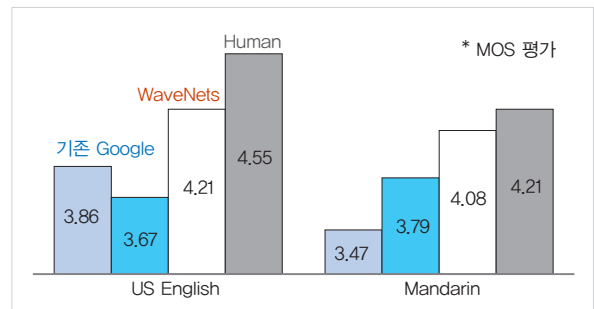
이렇게 데이터를 기반으로 구성된 언어 모델은 과거 Ontology 방식에 비해 확장성이 높고 특정 분야에 종속(Domain Dependent)되지 않는다. 새로운 언어가 추가되거나, 전문성이 필요한 특정 분야에 활용하려 할 경우 관련 데이터를 기계에 학습 시켜 모델을 업데이트 하면 된다. 또한 전혀 새로운 언어도 충분한 데이터만 확보된다면 과거에 비해 매우 빠르고 쉽게 인식 가능한 언어로 확장될 수 있다. 예를들어 한국어의 경우 한국어에 대한 전문성과 경험을 갖고 있는 국내 기업들의 언어 인식/이해의 정확도가 글로벌 기업 대비 높았었지만 최근 구글, 페이스북 등은 데이터를 기반으로 종전과 비교가 안될 속도로 한국어 인식 성능을 높이고 있다. 구글은 언어 인식/이해 분야에 딥러닝 기술을 적용한지 2년 만에 인식 가능 언어를 32개(2017.3)까지 확장하였고 Microsoft, 바이두 또한 매우 다양한 언어에 대해 인간 수준(Human-level)의 언어 이해 지능을 구현해 내고 있다. 바이두의 경우 딥러닝 기반의 언어 인식 기

23 J. Pennington, et al., GloVe: Global Vectors for Word Representation, 2014

술인 ‘Deep Speech’ 논문을 2014²⁴, 2015년²⁵ 차례로 발표하며 사람의 개입을 최소화하면서도 높은 성능을 갖는 음성 인식 기술을 구현하고 있다.

인간 수준의 언어 인식/이해 지능을 갖게 된 인공지능은 사람의 목소리를 자유롭게 생성해 내기도 한다. 과거 기계에 의해 생성된 사람의 목소리는 개별 단어/구문을 조합하는 형태였기 때문에 문장 단위로 인식할 때 발음, 억양 등이 매우 자연스럽지 못했다. 하지만 딥러닝이 적용되면서 음성 생성은 개별 단어 단위의 발음, 악센트(Accent) 뿐만 아니라 문장 단위에서의 억양(Intonation)까지 매우 정교한 수준으로 구현되고 있다. 딥마인드는 기존 최고 수준이었던 구글의 음성 생성(TTS: Text-to-Speech) 기술을 획기적으로 발전시킨 WaveNet 논문²⁶을 발표하였다. 기존 구글의 방식도 딥러닝에 기반해 구현되었지만 딥마인드는 알고리즘 고도화와 학습 데이터의 다양화를 통해 성능 향상을 이루었다.

약 100여개의 문장에 대해 사람의 목소리와 비슷한 정도를 정량화해 테스트한 결과에서 딥마인드가 발표한 WaveNet은 사람의 목소리(4.55점)에 근접한 4.21점을 기록했다(그림 14). 실제 WaveNet이 생성한 음성을 들어보면 인공지능이 생성했다는 사전 정보가 없을 경우 기계가 생성한 가상의 목소리라는 것을 판별해 내기가 매우 어려울 정도다. 바이두 또한 딥러닝을 음성 생성 분야에 적용하고 있다. DeepVoice²⁷라 불리는 이 기술은 특정 사람의 목소리를 반복 학습해 그 사람의 목소리의 특징을 완벽히 분석, 모델링 한다. 이렇게 만들어진 모델을 기반으로 생성된 음성은 마치 그 사람이 말을 한 것과 매우 유사한 수준으로 특정 인물의 음성을 생성해 낸다. 국내 Naver도 유명 연예인의 목소리를 학습 하고 특징을 모델링해 가상으로 생성된 음성으로 동화책을 읽어 주는 서비스를 선보이기도 했다.



14 인공지능 기반의 언어 생성 기술 발전: MOS 평가 점수 비교
(평가자가 인공지능으로 생성된 100개의 문장을 듣고 인간의 목소리와 비슷한 정도를 1~5점 척도로 점수 부여)
Source: A.Oord(2016)²⁶

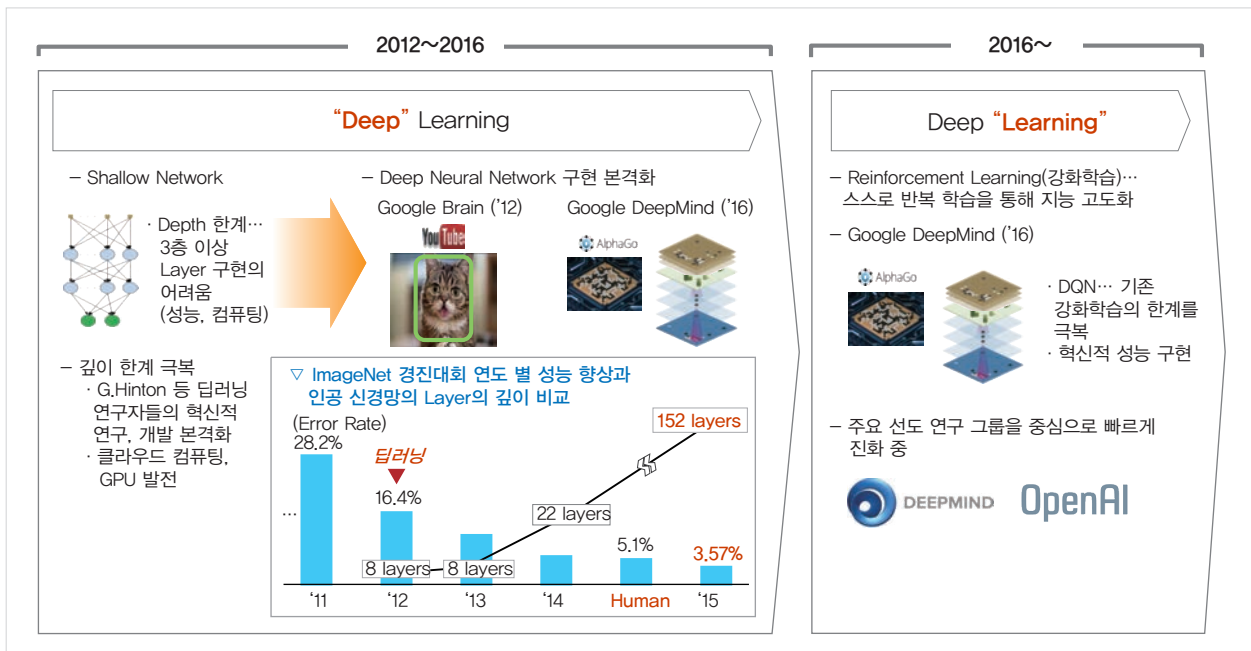
이러한 언어 인식/이해 기술의 혁신적인 발전은 애플 Siri, 아마존 Alexa와 같은 지능형 비서 서비스를 더 활성화시킬 것으로 전망된다. 지금까지 음성 인식 기반의 서비스들은 낮은 인식정확도와 제한적인 기능으로 인해 크게 상용화되지 못했다. 하지만 최근 딥러닝을 활용한 음성 인식/이해 기술을 구현하는 주요 기업 및 Startup 들은 매우 높은 수준의 정확도와 인식률을 보이고 있다. 특히 Viv Labs, SoundHound와 같은 Startup 들의 언어 인식/이해 지능은 단문 뿐만 아니라 두 개 이상의 단문이 연결된 복문/혼합복문 등도 높은 정확도로 인식하며 실시간으로 반응하는 수준으로 구현해

24 A. Hannun, et al., Deep Speech: Scaling up end-to-end speech recognition, 2014

25 D. Amodei, et al., Deep Speech 2: End-to-End Speech Recognition in English and Mandarin, 2015

26 A. Oord, et al., WaveNet: A Generative Model for Raw Audio, CoRR 2016

27 S. Arik, et al., Deep Voice: Real-time Neural Text-to-Speech, ICML 2017



15 딥러닝의 경쟁 핵심 변화: 깊이 경쟁에서 학습 경쟁으로 진화

내고 있다. 이러한 언어 인식 지능의 발전은 구글/애플/아마존 등 주요 글로벌 기업들이 최근 출시하고 있는 스피커 형태의 비서형 음성 인식 서비스 경쟁과 맞물려 향후 새로운 혁신을 만들어 낼 수 있을 것으로 전망된다.

(2) 학습 지능의 발전: 강화학습

2012년을 시작으로 본격화된 딥러닝의 발전은 2016년 알파고의 출현을 기점으로 전환점을 맞이하였다. 과거 알고리즘, 컴퓨팅, 데이터의 한계로 인해 얇은 신경망(Shallow Net)에 그쳤던 딥러닝이 깊이(Depth)²⁸의 한계를 극복하며 이제 시각/청각 지능에 대해서는 Human-level 혹은 그 이상의 인공지능을 구현하고 있다. 2016년 초까지 진행되었던 딥러닝의 깊이(Deep) 경쟁, 성능 경쟁은 알파고의 출현 이후 이제 학습(Learning)의 경쟁으로 전환되고 있다.

① 강화학습(Reinforcement Learning)의 등장: 스스로 지식을 키워가는 인공지능

강화학습 기반의 인공지능의 학습 과정은 과거의 방식과 전혀 다르다. 기존 기계학습 기반의 인공지능에 있어서는 목표 달성을 위해 사람(전문가)이 일일이 모델링하고 구현해야 했다. 또한 환경, 목표가 달라지게 되면 모델을 매번 변경하거나 모델을 전혀 새롭게 설계해야 했다. 하지만 강화학습 방법에서는 인공지능이 스스로 현재의 환경을 인식하고 행동하며 목표를 달성해 나간다. 게다가 이러한 방식은 범용적으로 활용

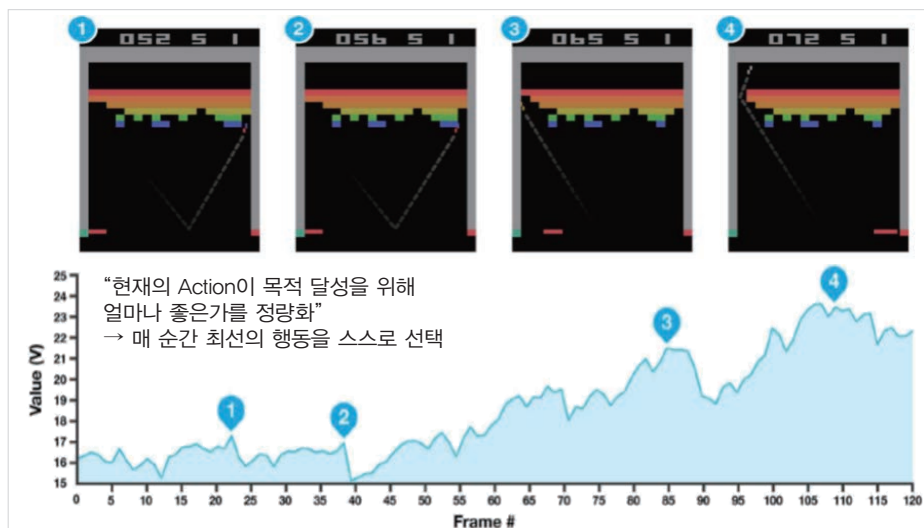
²⁸ 인공신경망 구조의 층(Layer)의 개수

가능해 새로운 환경에서도 학습을 반복하게 되면 하나의 알고리즘을 가지고 매우 다양한 환경에 적용 가능한 인공지능을 구현해 낼 수 있다.

이러한 강화학습 기반의 인공지능으로 인해 인공지능이 해결 가능한 문제의 범위 또한 빠르게 확대되고 있다. 이미지 인식, 언어 인식(이해)과 같은 인지 지능의 발전으로 인해 인공지능이 다양한 문제에 적용되게 되었지만, 단순한 인지 범위를 넘어 선 문제에 대해서는 고도화된 인공지능을 구현하기 매우 어려웠다. 예를들어 자율주행과 같은 기술 구현 시 전방의 사물을 단순히 인식하는 것을 넘어 상황에 따라 속도를 조절하고, 정지하는 등 차량 주행과 관련된 전 과정을 인간의 개입을 최소화하며 인공지능으로 구현하고 싶었지만 그러지 못해 왔던 것과 같다.

구글은 이러한 강화학습의 폭발적인 잠재력을 매우 빨리 인지했다. 딥마인드를 약 4500억원에 인수했던 2014년 당시만해도 딥마인드가 보유한 핵심 기술은 인공지능이 반복 학습을 통해 주어진 목적을 달성하는 방법을 스스로 깨우치게 하는 강화학습 알고리즘이 전부였다. 이후 딥마인드는 더욱 고도화된 강화학습 알고리즘을 통해 1년만에 알파고(AlphaGo)를 구현해냈다.

물론 기본적인 이론들은 오래전부터 제안되어 왔었지만 딥마인드는 그것을 실제 구현해 내고 인간 수준 혹은 그 이상의 성능으로 검증해냈다. 강화학습에 딥러닝을 접목한 ‘Deep Reinforcement Learning²⁹⁾’을 구현하면서 딥마인드는 우선 게임 환경에서 인공지능을 구현했다. 딥마인드가 초기 강화학습을 구현하며 검증을 위해 공개한 영상³⁰⁾에서는 학습 초기 과정에서는 게임을 전혀 진행시키지 못하지만 수시간에 걸친



16 강화학습 기반의 인공지능: 벽돌깨기 게임 상황(상), 매 순간 행동에 대한 정량화된 가치값(하)

Source: V. Minih(2015)²⁹⁾

29 V. Minih, et al, Playing Atari with Deep Reinforcement Learning, NIPS 2013

V. Minih, et al, Human-level control through deep reinforcement learning, Nature 2015

30 <https://www.youtube.com/watch?v=V1eYniJ0Rnk>

시행착오를 통해 인간 수준 이상으로 게임을 능숙하게 진행하는 것을 볼 수 있다. <그림 16>과 같이 매 순간 인공지능은 자신이 처한 환경과 행동 가능한 옵션들을 인지하고 각 행동에 대해 정량화된 보상값을 최고로 달성할 수 있는 행동을 반복적으로 선택하며 최종 목적을 달성해 나가게 된다. 알파고의 경우도 매 수를 둘 때마다 다양한 착점 중 가장 승률이 높을 것으로 계산된 수를 선택한다. 매일 128만번³¹에 이르는 반복학습을 통해 바둑을 두는 과정을 깨우쳐 나간 것이다.

알파고를 시작으로 강화학습에 대한 연구가 활발히 진행되며 2016년 이후 빠르게 발전하고 있다. 이를 가능하게 한데에는 딥마인드, OpenAI와 같은 선행 연구 기관들이 공개한 오픈소스가 큰 역할을 했다. 인공지능 연구자들은 자신들이 개발한 강화학습 알고리즘을 실험하고 검증하기 위한 환경이 필요하다. 알고리즘 검증을 위해 매번 게임 자체를 개발할 수는 없기 때문이다. 이러한 어려움을 해결하기 위해 OpenAI는 자신들의 연구결과물을 모두 공개하고 있다. OpenAI는 약 200개 이상의 게임 환경을 오픈소스로 공개³²하고 있다. OpenAI는 Tensorflow, Theano등 인공지능 구현에 주로 사용되는 개발 환경과 연동되기 때문에 강화학습 개발자와 연구자들은 단 몇 줄의 코드만 사용하면 다양한 환경에서 자신의 알고리즘을 테스트 하는 것이 가능해졌다. 연구자들이 자신이 구현한 인공지능을 동일한 환경에서 성능을 검증하고 경쟁하는 것이 가능해졌다. 이러한 경쟁의 결과로 강화학습 분야의 연구 논문은 매우 빠르게 발표되고 있다.

② 강화학습의 진화: 3차원 및 현실 세계(Physical World)로의 적용

강화학습 기반의 인공지능이 빠르게 발전해왔지만 게임과 같은 2차원 가상환경에서의 지능 구현이었기 때문에 실제 환경과는 큰 차이가 있어왔다. 이러한 한계를 극복하기 위해 최근 강화학습 분야의 주요 연구들은 3차원 혹은 실제 물리적 환경을 고려한 상황에서 이루어 지고 있다.

3차원 환경의 강화학습

2차원 환경과 달리 3차원 환경의 강화학습 연구는 단순한 공간의 개념 확장을 넘어 문제의 복잡도가 매우 높아진다. 우선 2차원에서는 매 순간(Snapshot) 현재의 상태, 환경, 보상(Reward)과 같은 모든 정보가 완벽하게 파악된다. 따라서 인지된 정보를 종합적으로 분석해 현재 상황에서 최선의 선택을 해 나가며 목적을 달성하는 것이 가능하다. 하지만 이와 달리 3차원 환경에서는 매 순간 인공지능은 전체 정보 중 매우 제한된 부분정보(Partial Information)만 갖게 된다. <그림 17>과 같이 3차원 미로의

31 16만개의 기보, 3000만개의 착점 정보, 128만번/1일

32 OpenAI Gym: gym.openai.com

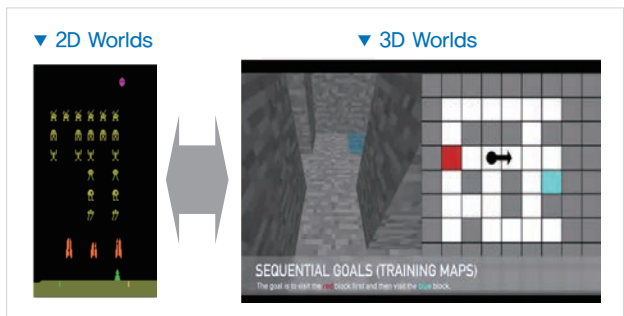
경우와 같이 인공지능은 현재의 시점에서 직선 거리에 공개된 경로만 보게 될 뿐 앞쪽에 있는 좌/우 방향의 보이지 않는 길에 대한 정보는 갖지 못한다. 이 경우의 강화학습 기반의 인공지능은 제한된 부분정보와 지금까지 반복을 통해 학습한 기억을 지식화해 3차원 환경에서 목적을 달성해야 한다.

주요 연구자들은 기존 Deep Reinforcement Learning에 새로운 기술을 접목시키며 3차원 환경에 강화학습을 구현하고 있다. 인공신경망 내에 과거의 경험을 기억할 수 있는 별도의 메모리를 설계³³하거나 학습을 통해 축적된 지식을 일반화해 새로운 환경(unseen environments)에 적용³⁴하기도 한다. OpenAI는 3차원 환경에서 강화학습 알고리즘을 검증할 수 있는 환경으로 Minecraft, Doom과 같은 3차원 게임 환경을 공개해 개발자/연구자들이 활용하게 하고 있다. 또한 동일한 환경에서 개발자들이 서로의 알고리즘의 성능을 경쟁하며 발전할 수 있도록 각종 경진대회가 진행 중이다. Doom 게임의 경우 Vizidoom³⁵이라는 대회에서 개발자들은 학습 시간을 최소화하거나 최소 시간에 목표 도달, 고득점 달성 등과 같은 다양한 목적에 대해 서로 경쟁하며 알고리즘을 고도화 시킨다.

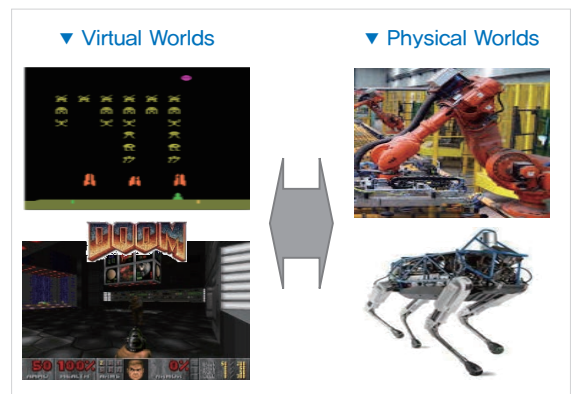
강화학습의 현실 세계 적용

게임과 같은 가상의 환경에서만 구현되었던 강화학습 알고리즘은 이제 현실 세계를 반영한 실제환경에서 연구, 개발되고 있다. 현실 세계에 강화학습 기반의 인공지능을 구현하는 것은 게임 등 가상의 환경에 인공지능을 구현하는 것에 비해 매우 복잡하다. <그림 18>과 같이 현실 세계에서는 게임에서와 같이 가상의 캐릭터가 하나의 객체 단위로 구동되는 것이 아니라 수 많은 하드웨어가 물리적으로 조합되어 연동한다. 또한 물리적 마모, 고장 등으로 인해 가상 환경과 달리 수 만에서 수 십만번 이상의 반복 학습이 불가능하다.

최근 이러한 한계를 극복하기 위한 연구들이 빠르게 진행되고 있다. 캐나다 UBC 연구팀은 가상의 캐릭터를 실제 물리환경을 최대한 반영하여 모델링 후 강화학습을 적용³⁶



17 2차원 환경 : 매 순간 모든 정보가 파악 가능(좌) vs. 3차원 환경 : 매 순간 전체 정보 중 제한된 일부 정보만 파악 가능(우)



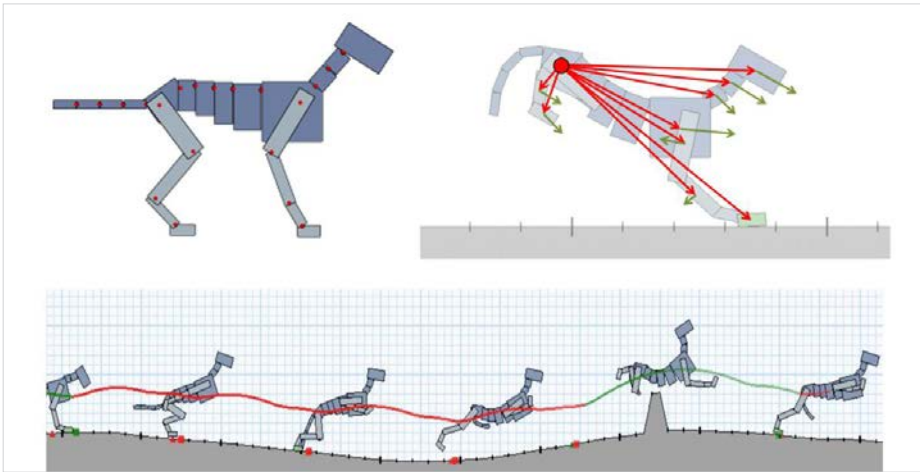
18 가상 환경 : 하나의 객체로 매우 단순화된 가상의 캐릭터 vs. 실제 환경 : 다수의 물리적 객체가 조합되어 연동

33 J. Oh, et al., Control of Memory, Active Perception, and Action in 3D world, ICML 2016

34 D. Chaplot, et al., Transfer Deep Reinforcement Learning in 3D Environments: An Empirical Study, NIPS 2016

35 <http://vizidoom.cs.put.edu.pl/competition-cig-2016>

36 X. Peng, et al., Terrain-Adaptive Locomotion Skills Using Deep Reinforcement Learning, SIGGRAPH 2016

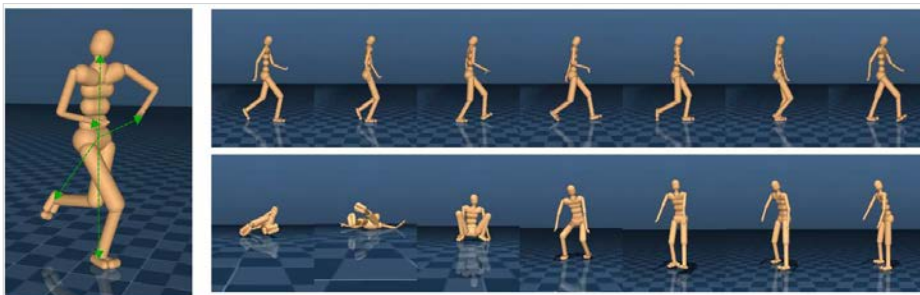


19 물리적 환경을 반영한 강화학습 구현: 동물 모양의 캐릭터가 정교하게 모델링되어 매 순간 약 300개 이상의 물리적 변수들이 함께 고려됨.
Source: X. Peng(2016)³⁶

하였다. <그림 19>와 같이 강아지 모양의 동물 캐릭터가 스스로 달리며 오르막을 오르거나 장애물을 통과하는 방법을 강화학습으로 구현할 때, 강아지 모양을 단순히 하나의 객체로 처리하지 않는다. 실제 동물의 몸이 수 많은 관절로 연결되어 있는 것과 같이 가상환경의 대상을 매우 정교하게 모델링했다. 논문에 따르면 달리기를 학습하는 과정에서 매 순간 약 300개 이상의 물리적 변수들이 종합적으로 분석되고 약 30만번의 반복학습을 통해 스스로 뛰는 방법을 터득하게 했다.

딥마인드 또한 최근 세편의 논문³⁷을 연이어 발표하며 현실 환경을 정교하게 반영한 강화학습 알고리즘을 구현하였다. <그림 20>과 같이 사람 모양의 캐릭터가 스스로 걷고, 뛰는 방법을 강화학습을 적용해 스스로 터득하게 한 것이다. 사람을 머리, 몸통, 팔/다리로 구성하였고 신체의 각 부분의 관절까지 반영하여 움직임에 따른 힘의 분배 등의 관계가 고려되게 하였다.

또한 이러한 복잡한 물리법칙을 고려해 만든 실제 로봇에 강화학습을 적용하기도 한



20 인간의 신체 구조를 반영한 강화학습 구현: 신체 구조를 반영해 모델링 후 스스로 걷고, 뛰고, 일어서는 방법을 터득하게 함.
Source: J. Merel(2017)³⁷

37 N. Heess, et al., Emergence of Locomotion Behaviours in Rich Environments, 2017, 7

J. Merel, et al., Learning human behaviors from motion capture by adversarial imitation, 2017, 7

Z. Wang, et al., Robust Imitation of Diverse Behaviors 2017, 7



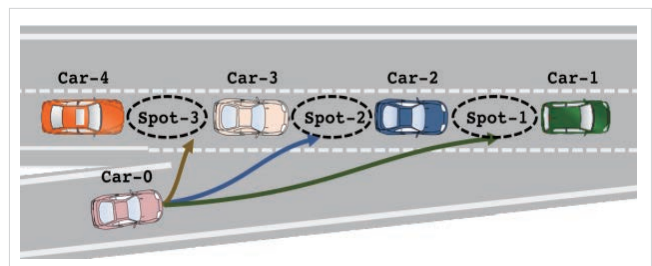
21 강화학습 알고리즘의 로봇 적용

Source: M. Tesch(2013)³⁸, M. Katharina(2013)³⁹, S. Gu(2017)⁴⁰

다. SnakeBot³⁸, Pingpong Robot³⁹등과 같은 로봇 연구에서는 로봇이 스스로 반복 학습을 통해 장애물을 넘거나 탁구 치는 방법을 익혀간다. 또한 UC 버클리대의 세르게이 레빈교수팀은 강화학습을 적용해 로봇이 스스로 학습해 물체를 집어 올리거나⁴⁰, 문을 열어가는 과정⁴¹을 터득하게 한다. 다양한 모양의 물체에 대해 약 80만번의 시행착오를 반복하며 물체를 집는 방향과 힘을 조절한다. 이때 한대의 로봇이 아닌 약 6~14대의 로봇이 동시에 학습을 수행하게 하며 개별 로봇의 학습 과정이 서로 공유되게 함으로써 모든 로봇이 같은 실수를 반복하지 않고 매우 빠르게 지능을 발전시킨다.

답마인드의 연구로 본격화된 강화학습 알고리즘의 연구는 약 2년만에 엄청난 발전을 이루고 있다. 간단한 2차원 게임 환경을 넘어서 3차원, 물리 환경에 기반한 연구들이 빠르게 진행되며 현실 세계에 강화학습 기반의 인공지능 구현을 앞당기고 있다. 실제로 주요 기업들의 연구소에서는 강화학습을 실제 제품, 서비스에 적용하려는 연구, 개발이 활발히 진행 중이다.

실제 주요 자동차 제조사의 선행연구 기관에서는 강화학습을 적용한 지능형, 자율주행 기능을 시도하고 있다. <그림 22>와 같이 자동차가 램프에 진입할 때 옆 차선에서 진행 중인 다른 차들의 속도, 거리 등을 고려해 진입 위치를 결정해야 한다. 사람의 경우에도 초보 운전 시 매우 어려운 과정이지만 수 많은 반복 경험을 통해 직관적으로 판단하여 진입 위치를 결정하고 속도를 조절해 진입하는 것과 마찬가지로 강화학습을 통해 인공지능이 스스로 상황을 판단해 매 순간 최적의 결정을 내릴 수 있도록 한다. 이 과정을 기존 모델링 기반의 기계학습 방식으로 완벽하게 구현하기는 매우 어렵다. 하지만 강화학습에 기반해 충분한 학습 과정만 반복할 수 있다면 인간 수준의 성능으로 구현이 가능해 질 수도 있다. 자율주행 외에도 스마트

22 강화학습 기반의 지능형/자율주행 기능 구현: 반복학습을 통해 스스로 진입 위치를 판단
Source: VW Research

38 M. Tesch, et al., Learning Stochastic Binary Tasks using Bayesian, Optimization with Shared Task Knowledge, ICML 2013

39 M. Katharina, et al., Learning to select and generalize striking movements in robot table tennis, The International Journal of Robotics Research, 2013

40 S. Gu, et al., Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection, IJRR, 2017

41 S. Gu, et al., Deep Reinforcement Learning for Robotic Manipulation with Asynchronous Off-Policy Updates, ICRA 2017

팩토리 내 다양한 제조 공정, 로봇 등에 강화학습 기반의 인공지능이 앞으로 빠르게 적용 될 것으로 전망되며 이렇게 되면 현실 세계의 인간처럼 학습하는 기계가 인간과 협력 혹은 경쟁하는 것이 가능하게 될 것으로 예상된다.

(3) 추론/행동 지능의 발전

① 추론의 시작

추론 영역은 오랫동안 인간의 고유 영역이라고 여겨지며 기계적으로는 구현하기에 매우 어려운 분야 중 하나였다. 수 많은 정보들을 단순히 조합해 검색, 추천 기능으로 구현하는 것은 가능했으나 주어진 정보들을 기초로 하여 전혀 새로운 명제를 도출하는 추론(Inference/Reasoning) 과정은 기계적으로 구현하기 매우 어려웠다. 텍스트, 이미지 등으로 주어지는 정보를 인식해 정보의 문맥적 의미를 이해해야 할 뿐 아니라 같은 정보라도 상황 별 변화하는 문맥적 관계를 고려하며 이해하는 과정이 요구되기 때문이다.

하지만 최근 발전된 딥러닝, 그 중 학습 알고리즘의 진화로 인해 추론 분야의 인공지능도 빠르게 연구, 개발되고 있다. 인공지능 분야의 혁신 Startup 중 하나인 MetaMind⁴²는 추론 문제를 해결하는 인공지능을 구현한 논문⁴³을 2016년 발표했다. ‘Ask me Anything’이라는 논문 제목처럼 MetaMind의 인공지능은 텍스트로 제공된 다양한 정보들을 이해하고 조합해 추론 유형의 질문에 답한다. <그림 23>과 같이 우유(Milk)의 위치가 전제들 속에 직접적으로 주어지지 않지만 우유를 가지고 있는 사람이 위치한 곳에 우유가 위치한다는 정보를 유추해 우유의 위치를 추론해 낸다. 사람의 관점에서는 매우 단순한 문제로 보일 수 있다. 하지만 인공지능의 관점에서는

<p>I : Jane went to the hallway. I : Mary walked to the bathroom. I : Snadra went to the garden. I : Daniel went back to the garden. I : Sandra took the milk there. Q: Where is the milk? A: garden</p>	<p>Task 19 : Path Finding The kitchen is north of the hallway. The bathroom is west of the bedroom. The den is east of the hallway. The office is south of the bedroom. How do you go from den to kitchen? A: West, north How do you go from office to bathroom? A: north, west</p>
	<p>Task 17 : Positional Reasoning The triangle is to the right of the blue square. The red square is on top of the blue square. The red sphere is to the right of the blue square. Is the red sphere to the right of the blue square? A: yes Is the red square to the left of the triangle? A: yes</p>

23 추론 문제 유형

Source: A. Kumar(2016)⁴³, FaceBook bAbI Project(2015)

42 딥러닝의 3대 핵심 연구자 중 하나인 Andrew Ng의 제자, Richard Socher가 설립, Salesforce에 인수(약 350억원, 2016년)

43 A. Kumar, et al, Ask me Anything: Dynamic Memory Networks for Natural Language Processing, ICML 2016

텍스트로 주어진 수 많은 정보중 질문과 연관성이 있는 정보만을 선별적으로 인지하고 문맥적 관계를 유추해내야 하기 때문에 매우 어렵다.

물론 이러한 추론 기능을 수행하는 인공지능은 과거부터 구현되어 왔지만 상당한 수준의 인간의 개입이 요구되면서도 성능 또한 한계를 가지고 있었다. 구글의 지식 그래프(Knowledge Graph), 애플 Siri의 기반이 되는 울프람 알파(Wolfram Alpha), IBM의 Deep Q&A 등과 같은 주요 IT 기업들의 지능형 서비스들이 그 예이다. 이러한 시스템들은 인간의 질문에 대해 방대한 데이터 속에서 질문과 연관성 있는 정보들을 분석하고 조합해 답을 한다. 질문에 대한 답이 직접적으로 데이터에 존재하지 않더라도 부분 정보들의 조합을 통해 답을 찾아간다는 점에서 추론에 해당한다고 할 수 있다. 하지만 이러한 시스템들을 구현하기 위해서는 상당한 정도의 인력투입이 필요하다. 데이터가 구조화된 형태로 저장되어야 하며 이들 간의 관계 등이 전문가에 의해 사전에 정교하게 정의되어야 한다. 따라서 복잡한 수준의 문제나 미리 정의되지 않은 문제의 유형에 대해서는 답을 하기 어려웠다.

하지만 최근 딥러닝 기반의 추론 방식은 과거와 달리 인간의 개입을 최소화하며 인공지능이 스스로 정보의 문맥적 의미를 이해해 추론해 낸다는 점에서 크게 다르다. 이것이 가능하게 된 것에는 크게 세가지 요인이 작용했다. 딥러닝을 통한 인공지능의 언어 분야에 대한 ‘인지/이해’ 수준의 진전이다. 특히 개별 단어 단위가 아닌 구문 단위의 함축된 의미가 word2vec, GloVe, Fasttext와 같은 단어 세트(Word Embedding)를 통해 분석된다. 또한 Attention, LSTM과 같은 딥러닝 분야의 알고리즘 고도화를 통해 인공지능은 이제 긴 문장, 여러 문단으로 구성된 텍스트를 처리하면서도 답을 찾기 위해 중요한 정보를 선별적으로 집중하거나 기억해 추론과정에 활용하게 된다.

마지막으로 추론 과정을 학습할 수 있는 학습용 데이터가 주요 연구기관 및 기업들에 의해 공개되고 있다. Microsoft에 의해 인수된 Maluuba는 CNN 뉴스를 활용해 추론형태의 Q&A 데이터 세트를 만들어 공개하고 있다. 텍스트 형식의 CNN뉴스와 추론형 문제를 하나의 세트로 제공하며 약 12만개에 이르는 문제로 구성되어 있다. 또한 Facebook은 약 20개에 이르는 다양한 형태의 추론형 Q&A 데이터 세트를 공개하고 있다. 단순한 연역법, 귀납법에서부터 경로/위치 추론 등과 같이 다양한 난이도의 문제를 포함하며 약 2만개의 데이터 세트로 구성되어 있다. 이외 스탠포드대의 SQuAD(10만 개), Microsoft의 MACRO 등 다양한 기관의 추론형 데이터 세트가 공개되고 있다. 특히 스탠포드, Microsoft는 데이터 세트를 공개하고 연구자, 개발자들이 자신들의 인공지능 성능을 경쟁할 수 있는 경진대회⁴⁴를 운영 중이다. 매우 다양한 연구 단체들이 참여하며 실시간으로 순위가 바뀌면서 인간의 추론 능력에 가까워지고 있다.

44 <https://rajpurkar.github.io/SQuAD-explorer/>

② 관계 추론이 가능한 인공지능의 등장: Relational Network

딥마인드는 지난 6월 관계 네트워크(Relational Network)라는 논문⁴⁵을 발표했다. 관계형 추론(Relational Reasoning)이 가능한 인공지능을 구현한 논문으로서 인공지능 학계에서는 올해 가장 혁신적인 논문 중 하나로 꼽히고 있다. 인공지능이 인식된 객체(영상/이미지, 텍스트)에 대해 서로 간의 상대적인 관계를 추론하는 것으로서, 인공지능의 추론 과정이 이제 사람의 추론 방식과 유사하게 구현 가능하다는 것을 의미한다.

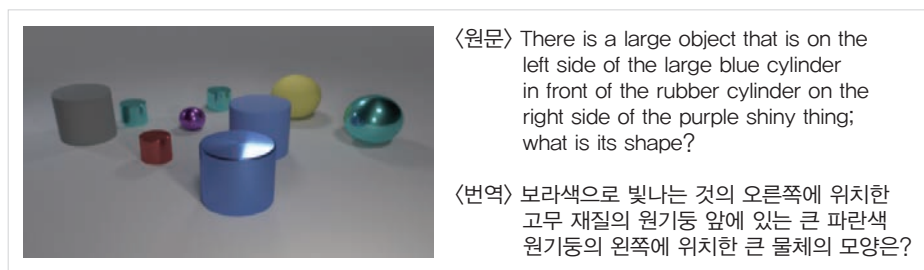
관계형 추론을 할 수 있는 인공지능을 구현하기가 어려운 것은 단순히 개별적인 정보들을 인식하고 이해하는 것을 넘어 각 정보들 사이의 상대적 관계를 파악해 논리적 결론에 도달해야 하기 때문이다. 예를 들어 <그림 24>



24 관계형/비관계형 추론

림 24>와 같은 이미지를 보았을 때 기존 이미지 인식 지능에 기반한 인공지능은 단순히 나무의 개수는?(네 그루), 색깔은?(초록/갈색), 모양은?(타원형, 마름모형)과 '비관계형 질문(Non-relational Question)에만 높은 정확도로 응답이 가능하다. 하지만 "나무간의 높이 차이가 가장 큰 나무들 중 오른쪽에 위치한 나무의 모양은?"과 같은 '관계형 질문(Relational Question)'에는 쉽게 답을 하지 못한다. 인식된 객체들 간의 관계가 명확히 이해되어야만 답을 찾을 수 있기 때문이다.

이러한 관계형 질문에 대해 딥마인드가 발표한 인공지능은 단번에 인간의 정확도를 뛰어 넘었다. CLEVR 데이터 세트⁴⁶는 <그림 25>와 같은 이미지 내 사물들에 대한 매우 복잡한 관계에 대해 답을 하는 Q&A 데이터 세트이다. 딥마인드 이전 방식의 인공지능으로 구현한 알고리즘은 정확도가 최고 68.5%에 그쳐 인간의 정답률인 92.6%에 크게 미치지 못했다. 하지만 딥마인드의 관계형 네트워크 기반의 인공지능



25 관계형 추론 테스트를 위한 문제(CLEVR)

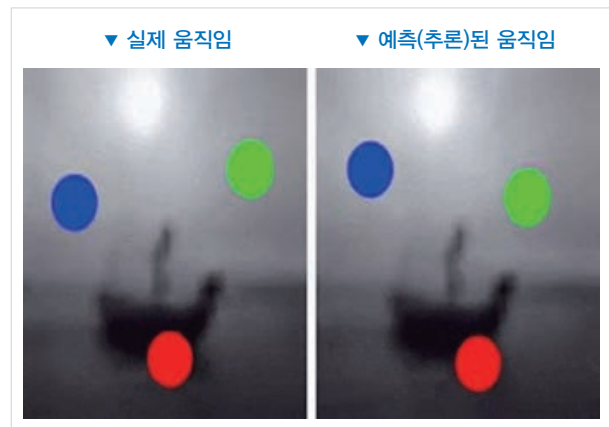
Source: J. Johnson(2017)⁴⁶

45 A. Santoro, et al., A simple neural network module for relational reasoning, 2017, 6

46 CLEVR Data Set: 70,000개의 이미지와 699,989개의 관계/비관계형 추론 QA를 담은 Dataset(Stanford Univ.), J. Johnson, et al., CLEVR: A Diagnostic Dataset for Compositional Language and Elementary Visual Reasoning, CVPR 2017

은 단 한번에 정답률 95.5%를 달성하며 인간 수준을 추월했다.

답마인드의 또 다른 논문⁴⁷에서는 이러한 관계형 추론에서 한 발 더 나아가 예측까지 가능한 인공지능을 제안했다. 논문에서 제안된 인공지능은 <그림 26>과 같이 사물의 움직이는 패턴을 학습한다. 사물들이 서로 부딪히며 움직이는 각도, 속도가 지속적으로 변화하는 패턴을 학습하여 향후의 움직임을 추론하는 것이다. 논문에 따르면 약 6 프레임의 움직임을 학습해 향후 200 프레임의 향후 움직임을 예측했을 때 150 프레임까지는 실제와 거의 일치하는 수준으로 예측하는 것으로 나타났다.



26 상호 관계의 관찰에 기반한 예측/추론

Source: N. Watters(2016)⁴⁷

답마인드의 최근 이 두 논문은 인간 수준의 추론이 가능한 인공지능 구현의 첫 시작으로 평가 받고 있다. 단순히 방대한 데이터 속에서 인간보다 빠르고, 정확하게 답을 찾는 것이 아니라 인간처럼 사물, 정보 등에 대해서 사람처럼 생각하며 추론하는 것이 가능해진 것을 의미한다. 미리 정의된 관계를 학습하는 것에서 그치지 않고 인간과 같이 유연한 사고가 가능해지는 것이기도 하다. 이러한 추론 분야의 발전은 향후 다양한 산업에 큰 영향을 미칠 것으로 전망된다. 예를들어 지능형/자율주행 관련 인공지능의 경우 기존에는 단순히 차간 거리, 속도, 장애물의 위치, 표지판/신호등 등과 같은 비관계형 정보를 기반으로 주행 관련 지능이 구현되었다. 하지만 관계형 지능이 적용된다면 차선을 급격하게 변경하는 자동차의 향후 주행 패턴을 추론/예측하여 주의하거나, 초보 운전자와 같이 주의가 요구되는 차량의 움직임을 추론해 회피하는 등의 기능이 구현 가능하게 되는 것이다. 이러한 관계형 지능은 자동차 산업 뿐만 아닌 제조 공정, 금융, 보안 등 매우 다양한 분야에 적용 가능할 것으로 예상된다.

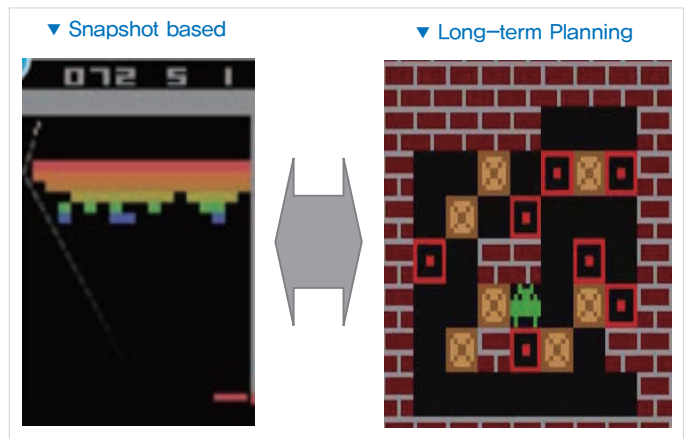
③ 미래 상황을 예측하고 계획하는 인공지능의 등장:

Imagination and Long-term Planning

행동(Action)은 추론과 마찬가지로 인공지능이 인간 수준으로 구현되기 매우 어려운 분야로 생각되어 왔다. 인간 수준의 행동이란 단순히 현재만을 고려해 행동하는 것이 아니라 현재의 행동이 미래에 미치는 영향을 고려하고 동시에 최종적인 목적 달성을 위해 매 순간 계획(Planning)과 결정(Decision)이 동반되어야 하기 때문이다. 그렇기 때문에 때로는 현재 시점에서 최선의 선택이 아니더라도 장기적 관점에서 목적 달성에 도움이 된다면 차선책을 선택해 행동하는 것도 필요하다. 이러한 모든 과정이 고려되어야 하기 때문에 인간처럼 행동하는 인공지능을 구현하는 것은 최근까지도 쉽지 않았다.

47 N. Watters, et al, Visual Interaction Networks, 2017.6

하지만 최근의 발표된 연구는 이러한 행동 방식을 인공지능으로 구현해 내고 있다. 딥마인드가 발표⁴⁸한 인공지능은 인간처럼 미래를 예측하며 장기적 관점에서 계획하며 행동한다. 완벽한 수준은 아니지만 인간의 행동 패턴을 닮은 인공지능 구현의 시작이라는 점에서 큰 의미가 있다. 앞서 연구된 강화학습과 같은 경우 인공지능은 매 순간 최고의 보상(Reward)을 받을 수 있는 행동(Action)을 반복하며 목표를 달성해 간다. 물론 순간 순간 최선의 선택을 반복하는 것이 최종 목적을 달성하기 위한 가장 이상적인 행동일 수도 있다. 딥마인드가 2년전 구현한 게임환경(Atari Game)의 인공지능 문제가 이에 해당할 것이다.



27 현재 상황만을 고려한 최선의 행동 선택(좌) vs. 미래 상황을 예측 후 계획해 최선의 행동을 선택(우)

반면, 딥마인드의 이번 연구는 인공지능이 매 순간의 행동에 대해 미래에 미치는 영향을 상상(Imagination)해 최적의 행동을 선택한다. 이 과정에서 최종의 목적을 달성하기 위해 인공지능은 장기적 관점에서 계획(Long-term Planning)하고 행동한다. 현재 시점에서는 손해가 동반되는 선택이라 하더라도 이러한 행동이 최종 목적 달성을 위한 과정이라고 판단된다면 행동하는 것이다. 이러한 점에서 기존 강화학습 연구들과는 큰 차이가 있다.

딥마인드는 이러한 과정을 Sokoban이라는 게임에 적용해 증명하였다. <그림 28>과 같이 다수의 벽돌을 지정된 위치에 모두 옮기게 되면 해당 레벨을 완료하는 게임이다. 이 게임은 벽돌을 옮길 때 한번의 실수가 게임 전체에 영향을 미칠 수 있다는 점



28 매 순간 미래에 발생 할 상황을 상상하고 계획 후 최선의 행동을 선택

Source: T. Weber(2017)⁴⁸

48 T. Weber, et al., Imagination-Augmented Agents for Deep Reinforcement Learning, 2017.7

(벽돌을 모서리에 이동시키면 회복할 수 없음)과 최소한의 움직임으로 벽돌을 옮기는 것이 항상 옳은 선택은 아니라는 점에서 게임의 레벨이 상승할수록 매우 어려워진다. 즉 게임을 진행하면서 앞서 설명된 매 순간 앞으로 일어날 상황을 상상하고 계획해야 하는 인간의 행동의 특성을 모두 반영하고 있다. 게임을 진행 하며 답마인드의 인공지능은 현재의 행동이 향후 미치는 영향을 상상한다. 상상한 결과에 따라서 현재의 선택이 최선은 아니더라도 최종 목적 달성에 도움이 된다고 판단되면 행동한다. 예를들어 벽돌을 바로 옆칸으로 한 번 이동시키면 될 경우도 더 먼 거리에 있는 벽돌을 선택해 해당 칸으로 벽돌을 이동시키기도 하는 것이다.

그 동안 매우 어렵고 오랜 시간이 걸릴 것으로 예상되었던 행동 분야의 인공지능 구현의 시작은 추론 분야의 지능 발달과 함께 향후 인공지능 분야 연구에 큰 혁신을 만들어 갈 것으로 전망된다.

3. 한계와 극복

딥러닝으로 인해 인공지능 분야가 빠르게 혁신되고 있지만 딥러닝은 엄청난 양의 데이터와 컴퓨팅 파워를 요구한다. 2012년 구글이 구현한 인공지능은 유튜브 영상 속의 고양이를 스스로 구분해 내며 혁신을 시작하였으나 이것은 약 1,000만개의 동영상 학습한 결과였다. 2016년 이세돌 9단과 대결에서 승리한 알파고는 약 3000만개의 착점 정보와 16만개의 프로 바둑 기사의 기보를 필요로 했다. 이러한 데이터 학습과정을 거쳤을 뿐 아니라 실제 바둑 대결에는 클라우드 기반으로 연결된 약 1202개의 CPU와 176개의 GPU⁴⁹를 동시에 가동하여 약 3억 4천번의 반복 학습을 통해 구현한 결과물이었다. 따라서 현재까지의 인공지능 혁신을 이루어 내고 있는 기업들은 엄청난 데이터와 컴퓨팅 파워를 확보한 거대 IT 기업이 중심이 되고 있다. 학계, 선도 연구단체를 시작으로 이러한 한계를 극복하기 위한 노력이 시도되고 있다. 다양한 선행 연구가 진행 중으로 크게 두 가지로 나뉜다.

(1) 인공지능 학습 환경의 인위적 생성

① 학습 데이터 확보의 한계 및 극복

인공지능 구현에 필요한 데이터를 확보하는 것은 매우 어렵다. 방대한 양의 빅데이터를 확보하는 것 뿐만 아니라 데이터의 질적(Quality)인 측면까지 동시에 고려되어야

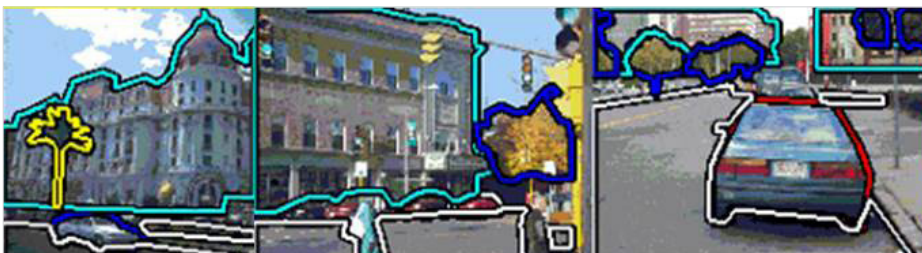
49 D. Silver, et al., Mastering the game of Go with deep neural networks and tree search, Nature 2016

하기 때문이다. 데이터의 질(Quality)은 크게 두 가지 의미를 갖는다. 첫째, 데이터의 다양성이다. 정제되고 완벽한 상황을 반영한 데이터가 많다고 해서 이를 기반으로 학습된 인공지능의 성능이 높아지는 것은 아니다. 음성인식의 경우 정확한 발음으로 녹음된 음성 데이터 보다는 각종 소음, 다양한 억양/말투 등의 데이터가 함께 학습되어야 실제 제품이나 서비스로 출시 될 때 완성도를 높일 수 있다. 자율주행 기능의 경우에도 각 종 사고, 위험 상황을 반영할 수 있는 데이터가 충분히 학습되어야 실제 주행 환경에서 다양한 상황에 대응 가능하게 된다.

둘째, 확보된 데이터가 기계 학습이 가능한 형태로 준비되어야 한다. 과거에는 빅데이터를 분석하는 주체가 사람이었지만 이제는 기계가 데이터를 직접 학습하고 분석한다. 사람에게서는 단순하게 보이는 데이터라 할 지라도 기계가 이해하기 위해서는 데이터의 전처리 과정이 필수적으로 요구된다. 예를들어 <그림 29>와 같은 이미지의 경우 이미지 내 각 종 사물을 알아보고 경계선을 구분하는 것은 인간에게는 매우 쉽다. 하지만 동일한 이미지를 기계가 인간처럼 알아보기 위해서는 이미지 속의 건물, 나무, 자동차에 해당하는 부분들을 사람이 일일이 경계선으로 구분 짓고 해당 사물의 명칭을 이미지와 함께 기록해주어야 한다. 이미지 어노테이션(Annotation)이라 불리는 전처리 과정을 거쳐야 해당 이미지를 기계가 이해할 수 있게 되는 것이다.

이러한 데이터의 전처리 과정에 소요되는 비용은 매우 높다. 스탠포드대학의 ImageNet 경진대회에 사용된 약 1400만장의 이미지를 전처리 하기 위해서 약 1000여명이 6년에 걸쳐 전처리 작업을 수행했다고 한다. 딥러닝 기반의 자율주행 인공지능을 구현하는 drive.ai의 창업자는 약 1시간 가량의 주행 영상 데이터를 전처리(Preprocessing)하는데 약 800 시간이 소요된다고 한다.

구글, 페이스북 등과 같은 거대 기업들이 최근의 인공지능을 선도하고 있는 가장 큰 요인은 바로 이들 기업의 데이터가 단순히 양이 많은 빅데이터가 아닌 질적 측면이 함께 고려된 빅데이터였기 때문이다. 기업들은 데이터 확보를 위해 막대한 자금 투자와 노력을 병행하고 있다. 대표적인 기업으로 IBM은 자사 인공지능 서비스인 Watson Health의 고도화를 위해 최근 수년에 걸쳐 약 4조원 이상을 투자했다. 동종 업계의 IT 기업에 대한 투자가 아닌 병원, 의학 연구기관과 같은 의료 기관을 인수하

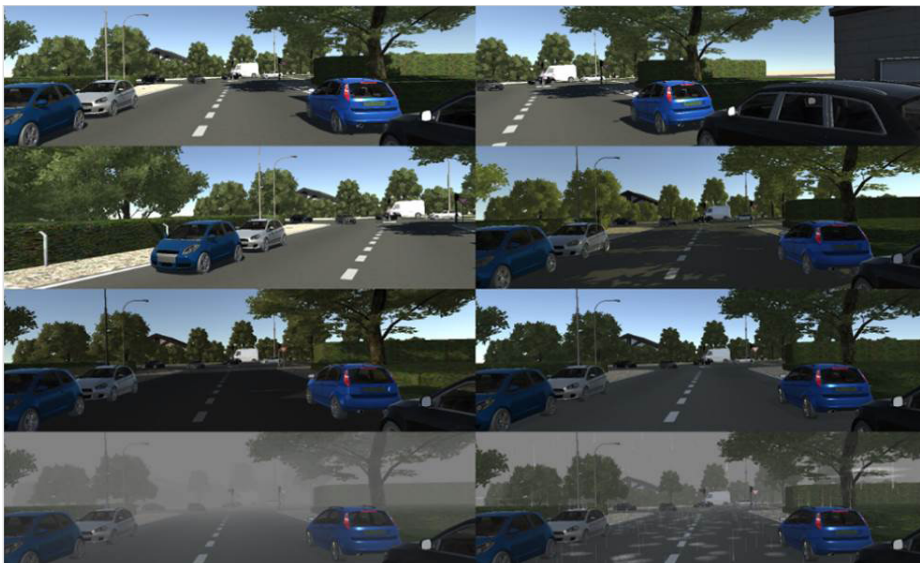


29 데이터 Annotation 과정: 이미지 내 각 사물의 경계선을 사람이 일일이 표기하고 명칭을 기록함. Source: MIT LabelMe project

거나 투자하였다. 이를 통해 IBM은 약 150만건의 환자기록과 200만 페이지에 이르는 의학 전문 자료를 확보하고 Watson의 의학분야의 전문성을 고도화하는데 활용하고 있다.

이러한 한계들을 극복하는 방법중 하나로 인공지능 자체를 이용하는 방법이 있다. 인공지능을 통해 가상의 데이터를 생성하고 이를 새로운 인공지능의 학습 과정에 활용하는 것이다. 가장 대표적인 연구가 GAN⁵⁰이다. GAN 기반의 인공지능은 세상에 존재하지 않는 전혀 새로운 가상의 데이터를 생성한다. GAN은 사람얼굴, 자동차, 꽃, 동물 등 데이터를 생성하려는 대상의 종류만 입력하게 되면 해당하는 데이터를 자유롭게 생성한다. 이렇게 GAN 방식으로 생성된 데이터는 데이터에 대한 정보가 이미 주어져 있기 때문에 어노테이션(Annotation)과 같은 전처리 과정에 소요되는 시간과 비용을 획기적으로 줄일 수 있다.

게다가 인공지능은 하나의 실제 데이터를 여러가지로 변형하기도 한다. 예를들어 <그림 30>과 같이 맑은 낮시간에 촬영된 주행 사진을 변환해 흐린 날씨 혹은 밤에 주행된 사진으로 변환하거나 사람의 정면 사진을 기반으로 좌, 우측 면의 이미지를 생성⁵¹해 내기도 한다. 이렇게 변형되어 생성된 데이터는 데이터의 다양성 측면의 한계를 해결한다. 즉 인위적으로 생성되거나 변형된 데이터는 데이터의 양적인 문제를 해결함과 동시에 데이터의 질적인 문제도 동시에 해결해 주고 있다. 또한 관련 분야의 연구가 빠르게 발전하면서 단순한 이미지 데이터 뿐만 아닌 사람의 음성(언어) 등에 이르기까지 가상으로 생성 가능한 데이터의 종류 또한 확장되고 있다.



30 주행 사진 변환: 한 장의 실제 사진(좌상)을 가지고 날씨, 시간, 위치 등을 변형해 새로운 이미지를 생성 Source: A. Gaidon(2016)⁵¹

50 Generative Adversarial Networks

51 A. Gaidon, et al., Virtual Worlds as Proxy for Multi-object Tracking Analysis, 2016



물론 가상으로 생성된 데이터의 한계는 존재한다. <그림 31>과 같이 가상의 데이터를 학습해 구현된 인공지능은 실제 데이터를 학습해 구현된 인공지능 보다 더 낮은 성능⁵²⁾을 보이기도 한다. 이는 가상으로 생성된 데이터 보다 실제 데이터에 더욱 많은 데이터가 함축되어 있고 인공지능의 학습 과정에 이러한 함축된 정보가 활용되기 때문이다. 하지만 이러한 가상의 데이터를 활용하는 방법은 인공지능을 빠르게 구현해 시장에 출시할 수 있다는 점에서 큰 의미를 갖는다. 최근 상용화되고 있는 인공지능들이 출시 초기에는 완벽한 수준을 보이지 못하더라도 향후 사용자들이 인공지능 제품, 서비스를 사용하며 발생하는 데이터를 학습하면서 지능을 발전시켜가기도 하기 때문이다.

② 현실적용의 한계 및 극복

소프트웨어로 구현된 인공지능을 실제 물리적 환경에 곧바로 적용하는 것은 매우 어렵다. 수 많은 반복학습을 통한 지능 고도화 과정이 현실 세계에서는 거의 불가능하다. 딥러닝, 강화학습 등에 기반한 인공지능은 수십, 수백만번 이상의 반복학습을 통해 스스로 지능을 발전시켜 나간다. 하지만 로봇, 드론 등과 같은 물리 환경에 사용될 인공지능 구현 과정에 수십만 번의 학습 과정을 반복하는 것은 매우 어렵다. 반복 과정 동안 발생하는 물리적 마모, 온도/습도 등의 변화 등으로 인해 학습 과정 동안 주변 환경 변수가 고정되지 않으며 기계적 고장, 오류 등으로 인해 수십만 번에 이르는 지속적인 반복 학습이 사실상 불가능하다.

따라서 주요 기업들은 이러한 한계 극복을 위해 현실 세계를 매우 정교하게 모델링한 시뮬레이션 환경을 구축해 인공지능을 구현한다. Nvidia는 자율주행 자동차를 구현하는 데 있어 시뮬레이션 환경을 활용했다. 자율주행의 경우 사고 등의 위험으로 인해 실제 차량이 주행되는 일반 도로에서 성능을 테스트하고 검증하는 과정이 매우 어렵다. 특히 시스템의 완성도가 높지 않은 인공지능의 학습 과정 초기에는 실제 도로에서

52 M. Johnson-Roberson, et al., Driving in the Matrix: Can Virtual Worlds Replace Human-Generated Annotations for Real World Tasks, ICRA 2017

성능 테스트를 하는 것은 거의 불가능하다. 따라서 Nvidia는 게임 환경을 통해 자율주행의 기본적인 성능을 학습시킨다. 사람이 게임을 통해 자동차를 제어하는 과정을 통해 인공지능이 차선 유지, 속도 조절, 안전 거리 유지 등과 같은 기본적인 주행 방법을 학습하게 된다. 약 100시간의 주행 시뮬레이션 과정을 통해 학습된 지능을 실제 자율주행 자동차에



32 시뮬레이션에 기반한 자율주행 기능 구현



Source: Nvidia

적용하였고, 실험 결과 고속도로/일반도로, 맑은 날씨/비오는 날씨 등 다양한 환경에서 차선 유지, 속도 조절 등과 같은 기본적인 주행이 가능하다고 한다⁵³.

마이크로소프트는 AirSim 이라는 드론에 활용 가능한 자율주행 시뮬레이터를 발표⁵⁴했다. 정교하게 모델링된 시뮬레이션은 현실의 환경을 매우 정확하게 정교하게 반영한다. 시뮬레이션을 통해 자율 주행 기능이 학습된 드론을 실제 환경에서 테스트한 결과 거의 유사한 경로와 속도로 이동한다고 밝혔다. 즉 시뮬레이션 환경에서 반복 학습된 지능이 실제 환경에서도 바로 적용이 가능한 수준으로 구현된다는 의미이다.

이와 같은 시뮬레이션을 통한 학습 환경 구현은 크게 세가지 측면에서 장점을 갖고 있다. 첫째, 물리적 기계를 통한 학습 과정에 소요되는 시간, 비용을 획기적으로 줄일 수 있다. 법적, 안전성 등 기술 외부 요인으로 인해 실제 환경에 기술을 적용하지 못하는 제약을 극복할 수 있다. 특히 법, 제도적 제약이 많은 자율주행 자동차, 드론과 같은 경우 이러한 시뮬레이션 기반의 지능 학습 과정이 필수로 작용할 것이다.

둘째, 학습 환경을 무한대로 생성해 다양한 환경에 적응 가능한 인공지능 구현이 가능하다. 최근의 가상 환경 시뮬레이터는 매우 정교하고 정밀하게 구현되고 있다. 또한 고도화된 인공지능 기반의 이미지/영상 변형, 생성 기술의 발달로 다양한 시뮬레이션 환경의 생성이 가능하다. 자동차 주행 환경의 경우 날씨를 변형하거나 낮, 밤 등 시간대를 변경하거나 하는 등의 환경 변수를 다양화하고 변형된 환경 속에서 인공지능이 주행 과정을 학습하는 것이 가능한 것이다.

셋째, 다양한 알고리즘 실험 및 검증이 가능하다. 인공지능의 학습 초기 단계의 시행착오 과정을 시뮬레이션으로 검증한다. 데이터를 반복적으로 학습해 지능을 고도화하는 동시에 다양한 알고리즘을 시도하며 검증하는 단계도 요구된다. 실제 물리적인 환경에서 다양한 알고리즘을 직접 적용하는 것은 시간, 비용 측면에서 매우 비효율적이지만 가상의 시뮬레이션을 통해서만 알고리즘을 검증하고 검증된 알고리즘을 데이터

53 M. Bojarski, et al., End to End Learning for Self-Driving Cars, 2016

54 S. Shah, et al., AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles, 2017

를 통해 학습하는 과정 전반이 가능해 지기 때문에 매우 효율적으로 인공지능 구현이 가능하다.

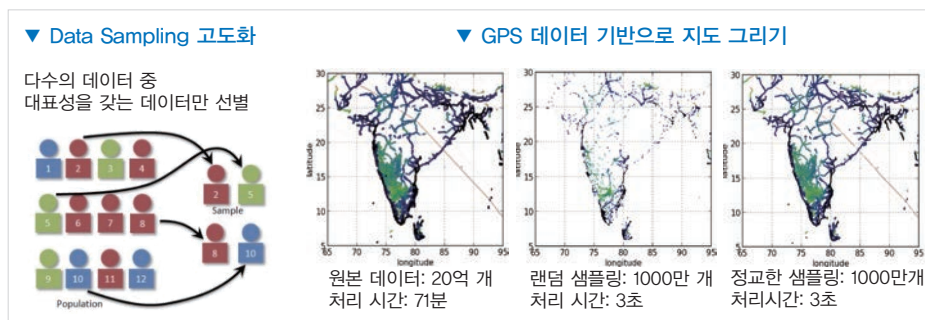
(2) 사람처럼 ‘생각’하는 인공지능(Computing Like Human)

엄청난 양의 데이터와 연산 과정에 의해 구현되는 인공지능은 인간의 학습 과정, 사고 방식과 비교하면 매우 비효율적이다. 사람은 대부분의 경우 직관과 일반화를 통해 상황을 판단하고 매우 빠르고 즉각적으로 결정한다. 그 답이 반드시 정답이 아닐지라도 근사치에 가까운 답을 빠르게 찾아내는 것이다. 예를들어 3.14×3.14 의 값을 정확히 계산하지 않아도 3×3 의 결과인 9 보다 크다는 것을 사람은 즉시 알 수 있는 것과 같다.

이렇게 인간처럼 계산하고 사고하는 방식으로 인공지능도 구현된다면 데이터와 컴퓨팅의 비효율성을 개선할 수 있을 것이다. 물론 정교하고, 정확한 작업을 요구하는 분야에는 적용이 어려울 수 있다. 하지만 사람들의 일상 생활에서도 매우 정확한 계산과 정교한 판단이 요구되는 상황은 제한적이다. 반대로 단지 몇 번의 경험에 의존해 새로운 상황에 대응하거나, 변화하는 환경에 직관적이며 빠르게 대응해야 하는 상황이 더 빈번하게 발생한다. 따라서 이러한 측면을 고려해 다음과 같이 크게 두 가지 측면에서 인공지능 구현을 위한 연구, 개발이 진행 중이다.

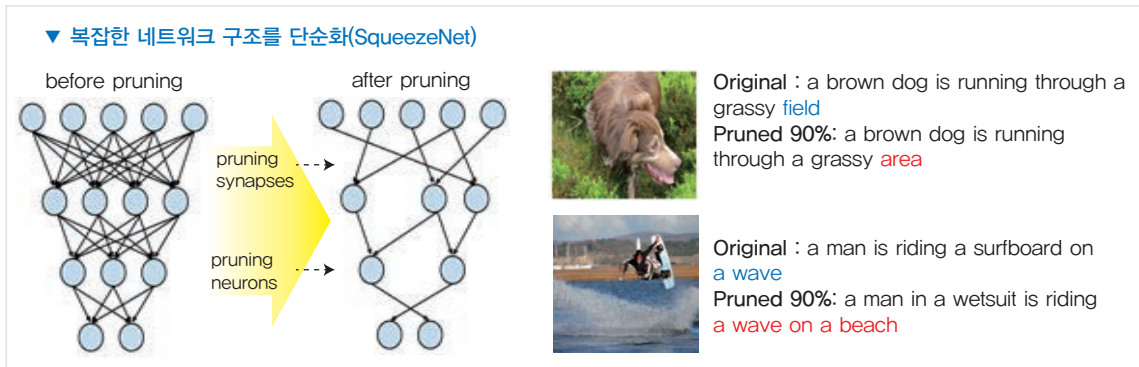
① Approximate Computing

첫번째는 적은 양의 데이터와 컴퓨팅 파워를 사용하면서 일정 수준 이상의 지능을 매우 빠르게 구현하는 것이다. 방대한 빅데이터를 모두 학습하는 것이 최선이지만 학습 시간, 비용 등 측면에서는 소량의 데이터를 학습 시키는 것이 더 효율적이다. 만약 방대한 빅데이터 중 대표성을 갖는 데이터를 정교하게 선별해 소량의 데이터만 학습 과정에 활용한다면 매우 효과적인 것이다. 실제 미시건 대학의 연구팀은 이러한 데이터 샘플링 방식의 고도화에 대한 논문⁵⁵을 발표하였다. <그림 33>에서와 같이 논문에



33 데이터 샘플링 고도화 기법: 정교한 알고리즘을 통해 데이터를 선별해 소량의 데이터 만으로도 높은 성능의 시스템을 구현
Source: Y. Park(2015)⁵⁵

⁵⁵ Y. Park, et al., Visualization-Aware Sampling for Very Large Databases, 2015



34 SqueezeNet: 인공 신경망의 네트워크 구조를 단순화해 컴퓨팅 비용을 최소화하면서도 기존과 동등한 성능을 구현

Source: F. Iandola(2017)⁵⁶

서 연구된 샘플링 방식을 활용해 데이터 처리 속도를 약 200배(400초→2초) 이상 단축 시켰다. 물론 정확도 측면에서는 약 2%~5%의 하락이 있었지만 속도 측면의 향상과 비교했을 때 의미있는 결과라고 할 수 있을 것이다.

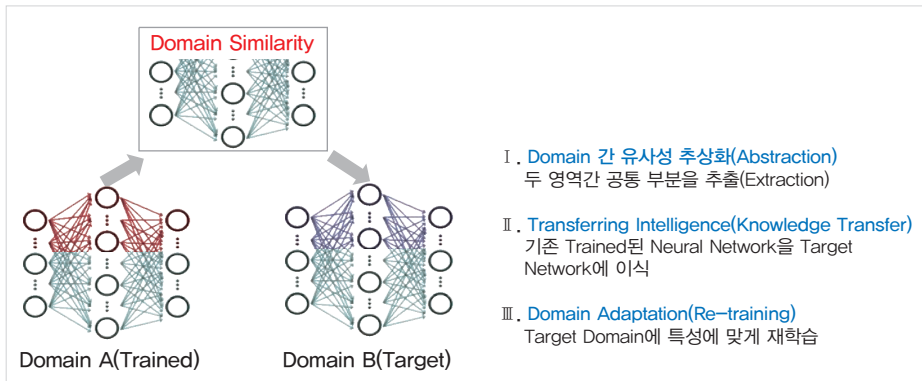
또한 인공신경망 구조를 단순화해 컴퓨팅 비용을 최소화하려는 연구도 진행 중이다. SqueezeNet⁵⁶이라는 이 연구는 <그림 34, 좌>와 같이 매우 복잡한 형태로 연결된 인공 신경망을 매우 단순화 시킨다. 논문에 따르면 비슷한 성능을 유지하면서도 신경망을 약 30~50배로 압축시키는 것이 가능하다고 한다. 물론 압축된 인공 신경망의 결과는 원래의 복잡한 신경망의 결과와는 다를 수 있다. <그림 34, 우>와 같이 같은 사진을 보고 인공지능의 해석이 조금씩 차이가 있다는 것을 알 수 있다. 하지만 그 해석의 의미가 크게 어색하지 않은 수준이라는 것을 알 수 있다.

이러한 인공 신경망의 압축 방식은 인간의 뇌 발달 과정과도 유사한 측면이 있다. 인공 신경망의 연결과 유사한 인간의 뇌의 시냅스의 연결 개수를 보면 2세까지는 매우 빠르게 증가하지만 이후 다시 감소하는 것으로 알려져 있다(50조: 1세 → 1000조개: 2세 → 500조: 10세). 이것은 인간의 지능이 점차 발달하면서 불필요하거나 중복된 시냅스의 연결이 약해지거나 끊어지는 과정을 의미한다. 즉 인공 신경망의 구현도 마찬가지로 중복되거나 성능에 큰 영향을 미치지 않은 연결들을 삭제(Pruning)함으로써 성능을 최적화 하는 것이다.

② 지능의 이식(Transferring Intelligence)

인공지능 구현 시 기존에 학습된 지능을 활용한다면 학습 과정에 요구되는 데이터와 컴퓨팅 비용을 크게 줄일 수 있다. 즉 유사한 기능을 수행하는 인공지능이 이미 존재한다면 기존의 지능을 새로운 인공지능에 활용하는 지능의 이식(Transferring Intelligence)이 가능한 것이다. 실제 딥마인드 등 연구 기관에서는 지능의 이식, 재

⁵⁶ F. Iandola, et al., SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size, ICLR 2017



35 Transferring Intelligence: 학습 지능의 이식 과정

활용에 대한 연구가 활발히 진행 중⁵⁷⁾이다. 인공지능이 새로운 영역에 활용될 때 적용 분야가 서로 다르더라도 기존 지식을 최대한 활용하기 때문에 단시간에 성능을 발휘하는 것이 가능하다. 새로운 영역에 대한 학습 과정이 매우 짧다는 의미에서 ‘One-shot Learning’ 혹은 ‘Zero-shot Learning’이라 불리기도 한다. 지능의 이식 과정은 <그림 35>와 같이 크게 세 단계로 나뉜다. 서로 다른 두 영역 간의 유사성을 추출하고 학습된 지능을 새로운 지능에 이식한다. 지능이 이식된 새로운 인공지능은 새로운 데이터를 통한 재학습 과정을 거쳐 목적인 영역에 최적화된 형태로 구현된다.

이러한 지능의 이식 관련 연구는 매우 초기 단계이지만 실제 산업 영역에 활용될 가능성을 두고 연구, 개발이 진행 중이다. 예를들어 언어 인식의 경우 각 언어가 갖는 특성으로 인해 매우 다르다. 하지만 기본적인 언어의 속성들은 유사한 측면이 많다. 주어/동사/목적어 등의 문장의 구성 요소 혹은 서술문/의문문/감탄문 등 문장의 속성 등이 이에 속한다. 따라서 이러한 언어의 기본 속성들과 관련된 지능은 공통적으로 활용 가능할 것이다. 실제 인공지능 분야에서 상당한 성과를 내고 있는 기업인 바이두는 영어와 중국어에 대한 언어 인식 관련 인공지능 구현 시 이러한 언어의 유사성을 고려했다고 한다. 영어 인식을 학습한 인공지능에서 언어의 기본적인 속성과 관련된 지능을 중국어 인식 지능에 이식한다. 그 후 중국어만이 갖는 문장 구조/어순 등의 특성을 중국어 데이터를 통해 학습한 것이다.

폭스바겐의 선행 연구팀에서도 자율주행 기능 구현 시 유사한 개념을 활용한다. 기본적으로 자동차를 주행하는 방식은 공용화한다. 차선 유지, 서행, 급정지 등과 같은 일반적으로 모든 나라에 적용 가능한 주행 기능은 범용적인 지능으로 구현하는 것이다. 이후 각 국가별 차이가 있는 주행 방식은 개별적으로 재학습 과정을 통해 맞춤화한다. 교통신호, 주행 우선 순위, 표지판 등이 이에 해당한다.

57 O. Vinyals, et al., Matching networks for one shot learning, Advances in Neural Information Processing Systems, 2016,

G. Koch, Siamese neural networks for one-shot image recognition, Diss.University of Toronto, 2015,

L. Bertinetto, et al., "Learning feed-forward one-shot learners," Advances in Neural Information Processing Systems, 2016



36 자율주행 지능의 이식

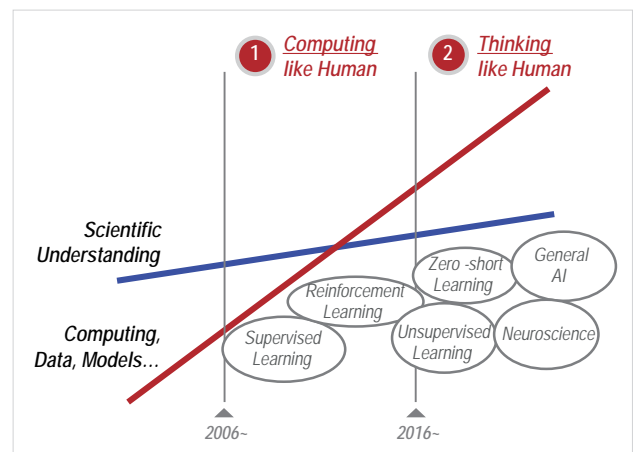
지능의 이식, 활용은 인공지능의 학습 과정에 요구되는 비용을 혁신적으로 줄일 수 있다. 한 발 더 나아가 모든 영역에 활용 가능한 범용 인공지능(General Artificial Intelligence)을 구현하는 초기 연구 단계로서의 의미도 있기 때문에 많은 인공지능 연구소에서 연구활동이 활발한 분야다.

4. 새로운 시도의 시작

최근 5년간 인공지능은 엄청난 발전을 이루었다. 하지만 현재 구현되고 있는 인공지능 방식으로 인간의 지능을 완벽하게 구현하는데는 한계가 있다. 인공지능 구현에 엄청난 양의 데이터와 연산 과정이 요구된다. 인식/학습 분야에서는 인간 수준의 지능을 갖추었지만 추론/행동과 같은 분야에서는 매우 초기적 단계에 머무르고 있다. 자율적인 상황 판단과 능동적인 행동을 기반으로 하는 인간의 지능 수준과 큰 차이가 있는 것이다. 현재 구현되고 있는 인공지능 기술이 오래 전부터 제안되어 온 수학/과학 분야의 이론과 모델링에 기반한 ‘인간처럼 계산(Computing like Human)’하는 지능을 구현한 것이기 때문이다. 이와 달리 ‘인간처럼 생각(Thinking like Human)’하는 지능을 구현하기 위한 다양한 연구가 요구되는 상황이다<그림 37>.

이런 한계를 극복하기 위한 노력은 크게 두 분야로 나뉜다. 첫째는 기존 인공지능 구현 방식을 하드웨어, 소프트웨어적으로 고도화 시키려는 노력이다. 인간의 신경망을 추상화한 인공신경망을 더욱 정교하게 모델링하거나 컴퓨터의 하드웨어적 구조의 한계 극복을 위해 전혀 새로운 개념의 컴퓨터 구조를 제안하기도 한다.

둘째는 기존 과학, 공학 분야의 연구가 아닌 신경과학



37 인공지능 이론 연구와 구현 환경 간 발전 관계

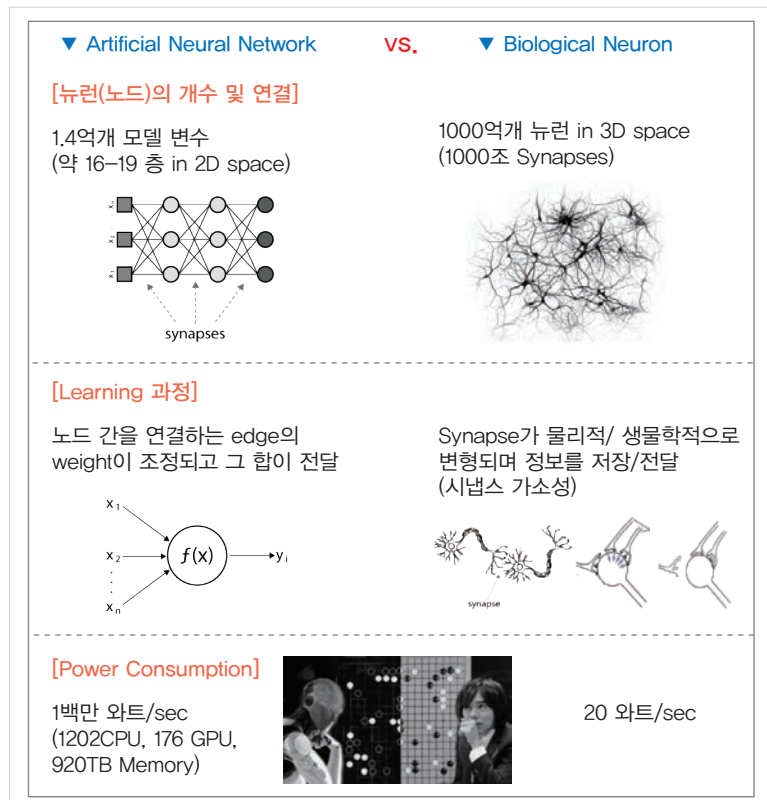
(Neuroscience), 뇌과학(Brain Science)과 같은 분야의 연구를 기반으로 인공지능을 구현하려는 시도이다. 오래전부터 인간의 신경망 원리를 이해하고, 인간과 컴퓨터를 연결하려는 시도들이 지속되어 왔으며, 최근 발전한 인공지능 기술을 접목해 새로운 혁신을 만들어 가려는 움직임이 시작되고 있는 것이다.

(1) 인공신경망의 진화 (HW/SW)

인간의 뇌 신경망을 모델링한 인공신경망 기반의 인공지능이 혁신적인 발전을 이루고 있지만 인간의 실제 뇌와 인공신경망은 근본적으로 큰 차이가 있다. 인공신경망은 인간의 신경망을 추상화, 단순화를 통해 소프트웨어적으로 구현한 것으로 <그림 38>과 같이 인간 신경망과 매우 큰 차이가 있다. 이와 같은 근본적인 차이는 인공지능이 인간 수준의 지능으로 구현되는데 한계 요인으로 작용하기도 한다.

이러한 한계를 극복하기 위해 소프트웨어, 하드웨어 측면에서 현재의 인공신경망 구조를 개선하기 위한 연구, 개발이 진행 중이다. 첫째 소프트웨어 측면에서는 현재 매우 단순화된 구조의 인공신경망을 더욱 정교화하기 위해 노력하고 있다. 현재 주로 구현되고 있는 인공지능의 인공신경망 구조는 2세대 모델이다. <그림 39, 좌>와 같이 연관된 뉴런들이 서로 연결되며 학습 과정을 반복하며 연관성의 정도가 수치 값으로 계산된다. 이 수치 값들의 합(weighted sum)이 활성화 함수(activation function)를 통해 다음 단계의 뉴런으로 전달된다. 뇌의 뉴런과 시냅스간 연결 구조를 모델링한 것이다. 이러한 2세대 모델은 모델링이 비교적 단순하면서도 높은 성능의 인공지능으로 구현될 수 있다는 장점을 갖지만 인간 뇌 신경망의 구동 원리를 정교하게 반영하지는 못한다.

따라서 현재의 인공신경망 구조를 더욱 정교하게 모델링한 3세대 인공신경망 모델이 연구되고 있다. Spiking Neural Networks(SNN)⁵⁸은 인간의 뉴런과 시



38 인공 신경망과 인간의 신경망과 차이

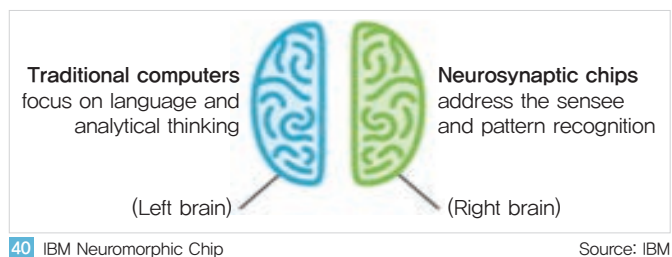
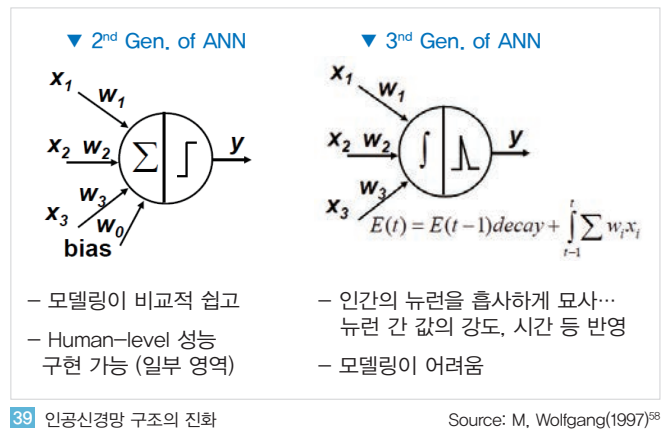
58 M. Wolfgang, et al., Networks of spiking neurons: The third generation of neural network models, Neural Networks, 1997

냅스간 정보 전달 과정을 2세대 모델보다 더욱 정교하게 반영한다. 인간의 뉴런 간 정보 전달은 전기적 신호에 기반한다. 발생된 전기적 신호가 시간의 흐름에 따라 강도가 조절되며 시냅스를 통해 다음 뉴런으로 전달되는 것이다. 이 과정을 2세대 뉴런에서는 단순히 연결 가중치의 합으로 단순화 하였지만 SNN에서는 <그림 39, 우>와 같이 시간에 따라 변화하는 전기적 신호의 강도를 계산 해 뉴런 간 정보 전달 과정에 구현함으로써 인간 신경망 구조와 매우 흡사하게 모델링하려고 하는 것이다.

이러한 모델은 향후 인공지능 기반의 인공지능의 새로운 혁신을 만들어 낼 것으로 기대되지만 모델 자체의 복잡도 및 높은 구현 난이도로 인해 아직은 선행 연구단계에 머무르고 있다.

소프트웨어 뿐만 아닌 하드웨어 측면에서도 기존 한계 극복을 위한 연구가 진행 중에 있다. IBM, 스탠포드, HBP(Human Brain Project)⁵⁹ 등과 같은 연구 기관에서는 컴퓨터의 구조적 한계를 혁신하려한다. 현재 사용되는 거의 모든 컴퓨터는 폰노이만 구조(Von Neumann Architecture)에 기반하고 있다. CPU를 통해 연산된 결과를 메모리에 기억하고 하드디스크와 같은 저장장치에 기록하는 구조인 것이다. 하지만 이러한 컴퓨팅 구조는 연산, 기억, 저장이 하나의 공간에서 이루어 지는 인간의 뇌 구조와 비교해 볼 때 매우 상이하다. 특히 CPU, 메모리, 하드디스크의 정보가 버스(Bus)라고 불리는 연결 부위를 통해 전달되는 구조로 이루어져 있기 때문에 개별 부품이 하드웨어적으로 급속도로 발전하고 있어도 버스의 한계로 인한 성능 저하가 전체 성능을 결정할 수 있다.

따라서 주요 선도 기관에서는 인간의 뇌 구조를 하드웨어적으로 구현하는 뉴로모픽 컴퓨팅칩(Neuromorphic Computing Chip)에 대한 연구를 진행 중이다. IBM이 제안한 TrueNorth Chip은 인간의 뇌 구조가 좌뇌, 우뇌로 구분되어 역할을 하는 것과 같이 컴퓨터의 CPU구조를 분리하려 한다. 언어, 분석적 사고와 같이 인간의 좌뇌에 해당하는 부분은 기존 컴퓨팅 구조로 구현하고 감각, 패턴 인식 등과 같은 우뇌에 해당하는 부분은 뉴로모픽 칩에 기반해 구현하고 서로를 연결한다<그림 40>. 현재까지 IBM의 연구결과에 따르면 단일 TrueNorth Chip에 약 1000만개의 뉴런과 2억 5,600만개의 시냅스의 연결을 구현하였으며



59 Human Brain Project: 유럽연합이 중심이 된 10년 장기 뇌 연구 프로젝트

연구 그룹	개발연도	뉴런개수	시냅스 개수	공정(nm)
Stanford Neurogrid	2009	1x10 ⁶	8x10 ⁹	180
HBP SpiNNaker	2012	2x10 ⁷	2x10 ¹⁰	130
HBP HICANN	2012	1.2x10 ⁷	3x10 ⁸	65
HRL Neuromorphic chip	2014	2304	292000	90
IBM TrueNorth	2014	1.6x10 ⁷	4x10 ⁹	28

41 주요 뉴로모픽 컴퓨팅 칩 비교

약 4,096개의 개별 칩을 서로 연결해 40억개의 뉴런과 1조개의 시냅스로 구현했다고 한다. <표 41>과 같이 다양한 연구기관이 서로 경쟁하며 하드웨어에 기반한 뇌 구조를 구현하려는 노력이 진행 중이지만 설계, 공정, 구현 등 하드웨어적 구현의 난이도로 인해 소프트웨어 분야의 연구에 비해 혁신의 속도가 빠르지는 않다.

하지만 이러한 하드웨어적, 소프트웨어적인 연구 결과가 함께 융합되어 진보된 인공지능 구현에 활용된다면 현재의 성능을 월등하게 뛰어넘는 인공지능으로 구현될 가능성도 있다.

(2) 뇌과학 기반의 연구들

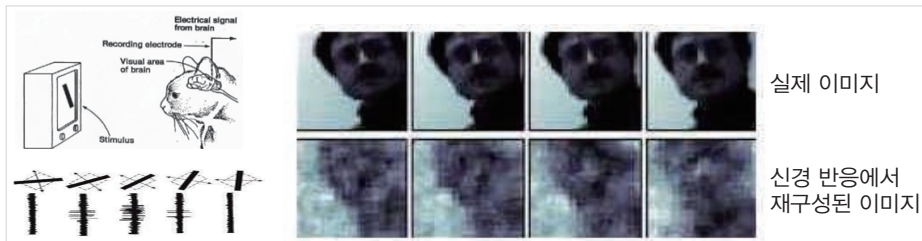
딥마인드의 창업자 하사비스는 최근 발표한 논문⁶⁰에서 뉴로사이언스와 컴퓨터과학 분야의 융합을 통해 인공지능이 한 단계 더 진보할 수 있다고 강조하고 있다. 오랜 시간 동안 두 분야에서 인간의 뇌, 지능과 관련된 연구를 진행해 왔었지만 각 분야의 전문가들이 함께 논의하고 서로의 기술을 융합 시키려는 노력이 부족했음을 지적하고 있는 것이다.

하지만 최근 딥러닝을 시작으로한 인공지능 분야 급속한 발전에 힘입어 컴퓨터과학 분야의 연구자들의 신경과학, 뇌 과학 분야 연구에 대한 관심이 급속하게 높아지고 있다. 실제 테슬라의 CEO 일론머스크는 2017년 3월 뉴럴링크(Neuralink)라는 스타트업을 발표하며 인간의 뇌와 기계를 연결해 인간의 지능을 자유롭게 저장하거나 이식하는 기술을 개발하겠다고 한다. 즉 오래전부터 연구되어 온 신경과학, 뇌 과학 분야의 연구에 IT/전자 기술분야를 접목하려는 시도로서 아직은 SF 수준으로 보일 수도 있지만 새로운 방식으로 인공지능이 발전되는 계기가 될 수도 있을 것이다.

일찍이 신경과학, 뇌과학 분야에서는 인간이 정보를 인식하고 이해해 지식화 하는 과정에 대한 연구를 시작해 왔다. UC버클리대의 연구팀은 1999년에 이미 고양이 뇌 신경망을 분석해 시각 영상을 재구성하는데 성공⁶¹했다. 연구팀은 우선 특정 시각 자

60 D. Hassabis, et al., Neuroscience-Inspired Artificial Intelligence, Neuron 2017, 6

61 G. Stanley, et al., Reconstruction of Natural Scenes from Ensemble Responses in the Lateral Geniculate Nucleus, 1999



42 고양이 시신경 세포의 반응 분석

Source: G. Stanley(1999)⁶¹

극에 대해 특정한 신경 세포가 반응하며 사물을 인식한다는 것을 발견하고, 역으로 반응하는 신경세포를 분석하면 시각 자극을 구성하는 것이 가능하다는 것을 논문을 통해 발표했다. <그림 42>와 같이 신경세포를 통해 재구성한 이미지와 원본 이미지가 유사한 윤곽선을 보이는 것을 알 수 있다.

이러한 뇌 신경의 반응에 대한 연구는 이후 지속되고 있으며 최근에는 매우 높은 수준으로 실현 되고 있다. 지난 5월 캘리포니아 공대(CalTech)의 연구팀이 발표한 논문⁶²은 뇌 신경 신호의 분석 만으로 사람의 얼굴을 재구성 하는데 성공했다. <그림 43>과 같이 실제 얼굴과 뇌 신경 신호를 통해 재구성된 얼굴이 거의 일치하는 것을 알 수 있다. 마치 언어, 기호적으로 표현해 몽타주를 그리 는 과정이 향후에는 뇌파 분석만으로도 가능해 질 수도 있는 것이다.



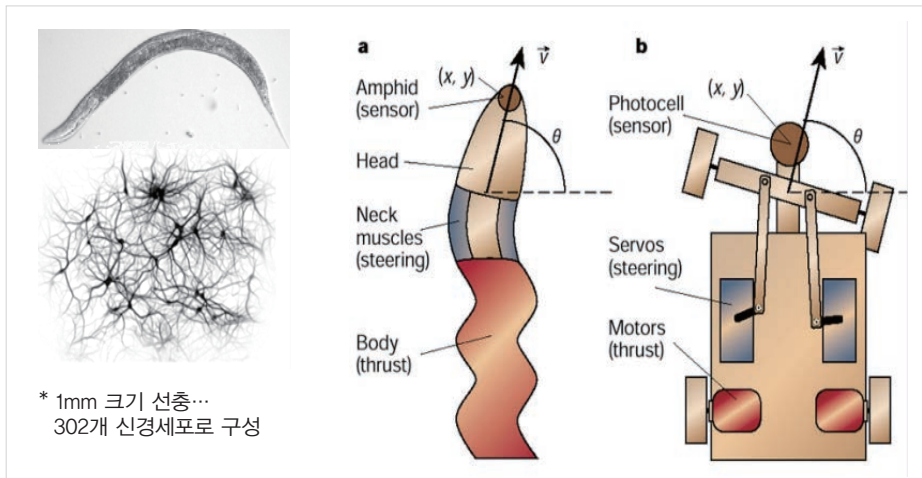
43 뇌 신경 반응에 기반한 사람 얼굴 재구성

Source: L. Chang(2017)⁶²

또한 생물체의 신경 세포 전체를 분석해 신경망의 작동 원리를 분석하고 인공적으로 구현하기도 한다. 예쁜 꼬마 선충(C. Elegans)은 약 1mm의 선충으로 약 302개의 신경 세포로 구성되어있다. 단일 생물체의 신경세포가 모두 파악된 유일한 생명체로 뇌, 신경계 분야의 연구에 자주 활용된다. 뉴로로보틱스(Neurorobotics) 분야의 전문가인 Timothy Busbice는 예쁜 꼬마 선충의 움직임에 따른 신경망의 변화를 관찰 했다. 특히 장애물을 발견하거나 충돌 시 반응하는 신경 세포간의 관계를 분석했다. 그 후 분석된 모델을 실제 로봇에 적용해 로봇이 스스로 움직이고 동작하는 기능으로 구현해 냈다. 매우 초기 단계이지만 구현된 로봇은 스스로 방향을 전환해 장애물을 피하거나 충돌 후 원래의 진행 방향으로 복귀하기도 한다<그림 44>.

이와 같은 신경 과학계의 선행 연구로 뇌과학(Brain Science)에 기반한 인공지능의 구현 가능성이 점차 가시화되면서 주요 혁신가들의 스타트업 및 주요 기업들 또한 이 분야에 뛰어들고 있다.

⁶² L. Chang, et al., The Code for Facial Identity in the Primate Brain, Cell, 2017. 5



44 C.Elegans 신경망의 로봇 적용

Source: Lockery Lab, University of Oregon

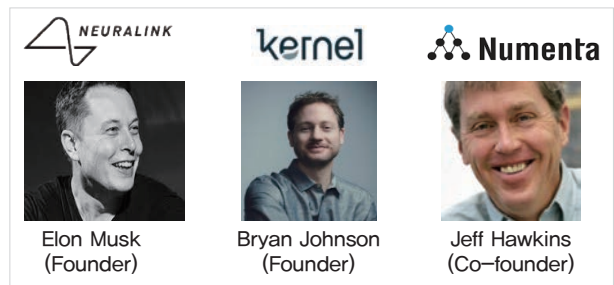
인간의 뇌와 컴퓨팅 칩의 연결을 구현하려고 하는 뉴럴링크는 의학전문 법인으로 등록된 데서 나타나듯이 단기적으로는 인간의 뇌와 관련된 질병의 치료에 목적을 두고 있다. 그러나 장기적으로는 인간의 지능, 기억 등을 컴퓨팅 칩에 저장하거나 혹은 반대로 저장된 정보를 인간의 뇌에 주입해 손상된 뇌 기능을 복원하고 다른 사람의 지능을 이식하는 것도 구현할 것이라고 한다. 여러 사람의 지능을 하나의 칩에 저장하며 높은 지능을 갖는 칩으로 구현해 다양한 사람들에게 이식하는 것도 가능할지 모른다. 뉴럴링크는 실제 이러한 목표를 달성하기 위해 관련 분야의 최고 전문가를 영입하고 있다. BMI(Brain Machine Interface)의 전문가인 Flip Sabet, MIT 신경외과 및 컴퓨터공학 박사인 Ben Rapoport 등 뇌과학, 컴퓨터 공학, 전자 공학 분야의 전문가를 영입해 연구를 시작 중에 있다.

또 다른 스타트업인 커널(Kernel)은 뉴럴링크와 유사한 개념을 목표로 한다. 실리콘 밸리의 사업가 Bryan Johnson⁶³이 설립한 이 기업은 인간의 뇌와 컴퓨팅 칩을 연결해 뇌 손상과 관련된 치료를 단기적인 목표로 한다. 신경과학계 및 소프트웨어 분야의 전문가 20여명으로 구성된 연구팀은 연구, 개발을 시작한지 10개월 만에 프로토타입을 개발해 실제 뇌 손상 환자에 실험 중에 있다.

뉴멘타(Nument)는 인간의 대뇌의 신피질(Neocortex)을 소프트웨어적으로 모델링해 인공지능을 구현하려 한다. 2005년 일찍이 설립된 이 기업은 뇌 관련 분야의 선도 기업 중 하나이다. 인간의 뇌 중 가장 바깥 부분에 해당하는 신피질은 고등 생물의 지능을 관장하는 것으로 알려졌다. 의지/의욕/판단과 같은 주로 지능과 관련된 기능의 수행과 관련된 부분으로서 포유류 이상의 동물에서만 신피질이 발달하며 고등 동물일수록 넓은 범위를 갖는다고 한다. 특히 뉴멘타는 인공지능이 스스로 문제를 인식

63 온라인 Payment 관련 스타트업 Braintree 창업자, Paypal에 약 9000억원에 매각(2013)

하고 판단/행동하는 강 인공지능(Strong AI)의 단계로 발전하기 위해 신경질의 동작 원리의 구현이 필수적이라고 말한다. 실제 이 기업은 신경질의 구동원리를 모델링한 HTM(Hierarchical Temporal Memory)을 오픈소스⁶⁴화 하며 다양한 개발자들과 함께 기술을 발전시켜 가고 있다.



45 뇌 연구 기반의 혁신 스타트업

페이스북도 인간의 뇌 신호를 분석해 인간과 기계간의 인터페이스를 혁신하겠다고 한다. 키보드, 터치와 같은 매개체를 통해 인간이 컴퓨터에 입력을 하는 것이 아니라 컴퓨터가 인간의 뇌 신호를 통해 의도를 분석해 내는 것이다. 약 80여명에 이르는 연구자들이 페이스북의 선행 연구소인 Building 8에 참여하고 있으며 UC버클리, 존스홉킨스 대학과 같은 주요 대학과 협업 중에 있다. 페이스북은 지난 4월 개발자 컨퍼런스인 F8에서 인간의 뇌 신호에 기반해 타이핑을 하는 ‘Brain Typing’을 시연하였다. 신체가 자유롭지 못한 환자의 뇌를 초당 약 100회로 스캔해 뇌 신호만으로 타이핑 기능을 구현한 것이다. 향후 2년 내 분당 100단어 이상 타이핑이 가능한 수준까지 구현하는 것을 목표로 하고 있다. 이러한 페이스북의 인간의 뇌와 기계 사이의 인터페이스(BMI)에 관한 연구는 단순히 인간의 의사 표현 방식이나 입력 수단을 대체하는 것 이상의 의미를 갖는다. 장기적인 관점에서 인간의 지능을 컴퓨터로 저장하거나 반대로 이식하는 등의 인공지능 구현을 위해서는 인간과 기계 간의 연결이 우선적으로 해결 되어야 하기 때문이다.

5. 맺음말

인공지능은 매우 빠르고 급진적으로 진화하고 있어서 인공지능의 미래, 인류의 미래를 예측하기는 점점 더 어려워지고 있다. 수학/컴퓨터 과학 분야를 중심으로 제안되어 온 이론이 2010년 이후 기하급수적으로 발전해 온 전자/IT 기술에 힘입어 빠르게 구현되고 있다. 게다가 최근에는 뉴로사이언스, 뇌과학 등과 같은 새로운 기술들이 함께 융합되며 인공지능 분야의 새로운 전환기가 도래할 가능성도 높아지고 있다.

더욱이 최근의 이러한 연구 결과들은 오픈소스의 활성화를 통해 더욱 빠르게 구현되며 확산되고 있다. 누구나 쉽게 이해하고 활용 가능한 형태로 연구 결과물들이 공개되면서 많은 사람들의 참여로 인공지능의 성능 향상은 더욱 가속되고 있다. DeepMind, OpenAI 등에서 혁신적인 논문으로 새로운 연구 분야를 개척하면 다양

64 <https://github.com/numenta>

한 연구 기관들이 후속 연구를 통해 단지 몇 달 만에 높은 완성도의 인공지능으로 구현해 내고 있는 상황이다. 뿐만 아니라 주요 기업들은 이러한 연구 결과들을 자신들의 제품과 서비스에 빠르게 적용해 상용화 해 내고 있다.

이러한 측면에서 선도 기업들과 우리나라의 격차는 더욱 심화되고 있는 상황이다. 국내 기업들은 실리콘밸리의 기업들에 비해 상대적으로 소프트웨어 역량과 데이터 측면에서 상당히 열위에 있다. 아무리 인공지능의 연구 결과물이 오픈소스로 공개되고 인공지능 구현의 난이도가 낮아진다 하더라도 인공 신경망 자체를 설계하고 학습 알고리즘을 구현하는 것이 인공지능 성능의 결정적 요인으로 작용하기 때문이다. 하지만 국내에는 글로벌 경쟁력을 갖춘 연구 논문을 발표하거나 제품/서비스를 출시하는 연구자, 기업이 매우 소수에 불과하다.

또한 기업들이 축적한 데이터의 양도 글로벌 선도 기업들에 비해 매우 적을 뿐만 아니라 데이터의 질적 측면(다양성, 기계학습 가능 형태)도 매우 미흡한 상황이다. 구글, 페이스북, 아마존과 같은 빅데이터를 확보한 기업들이 최근 인공지능 분야를 선도하고 있는 것이 방대한 양의 데이터에 기반한 측면도 있지만 데이터의 초기 수집 단계부터 질적인 측면이 함께 고려되어 축적되었기 때문이다. 향후 이러한 양질의 데이터를 확보하는 것은 단순히 많은 양의 데이터를 확보하는 것 보다 더욱 많은 시간과 노력이 필요 할 지도 모른다.

단기적으로는 Tensorflow 등과 같은 오픈소스 기반의 개발 및 참여를 통한 역량 축적이 시급하며 보다 근본적으로는 중장기적인 관점의 양질의 데이터 확보, 경쟁력있는 인공지능의 개발 역량을 높이기 위한 노력이 병행되어야 할 것이다. www.lgeri.com



본 보고서에 게재된 내용이 LG경제연구원의 공식 견해는 아닙니다. 본 보고서의 내용을 인용할 경우 출처를 명시하시기 바랍니다.