

## I. 딥러닝: 인공지능을 이끄는 첨단 기술

### 1. 인공지능과 머신러닝, 딥러닝의 기본 개념

인공지능의 목표는 ‘사람처럼 생각하고 행동하는 기계/컴퓨터’를 만드는 것이다. 인공지능을 구현하고자 할 때, 지식과 규칙에서 시작해서 다양한 응용을 추구하는 top-down 방식과, 실세계의 데이터에서 시작해서 특정 응용문제를 해결하고 나아가 지식과 규칙의 발견까지도 기대하는 bottom-up 방식의 두 방향이 가능하다. 최근 주목을 받고 있는 머신러닝(machine learning)은 bottom-up의 대표적인 방식이고, 딥러닝(deep learning)은 머신러닝의 최신 기술/트렌드이다.

딥러닝을 포함한 머신러닝 과정에서는, 문제를 해결하는 기계/컴퓨터(에이전트라고 호칭)를 만들기 위해, 사람이 설정한 문제의 범주 안에서 데이터(대부분의 경우 사람이 모아서 정리)에서 시작하여 사람이 지정한 도구(알고리즘)를 이용하여 학습(learning)을 수행하고 문제를 풀게 된다. 아직은 이 과정의 대부분에 사람이 개입하여 가이드를 해주어야만 하는데, 데이터만 보고 에이전트를 만드는 핵심적인 과정의 대부분이 자동으로 이루어지도록 한 것이 인공지능 구현의 큰 발전이었다. 하지만 ‘스스로 학습한다’라고 표현을 하기에는 아직 많이 부족한 것이 인공지능의 현 주소이다.

‘에이전트를 만드는 핵심적인 과정을 자동화’하는 방법에 대해 조금 더 자세히 살펴보는 것이 현재의 인공지능을 이해하는데 도움이 된다. 딥러닝을 포함한 머신러닝의 기본

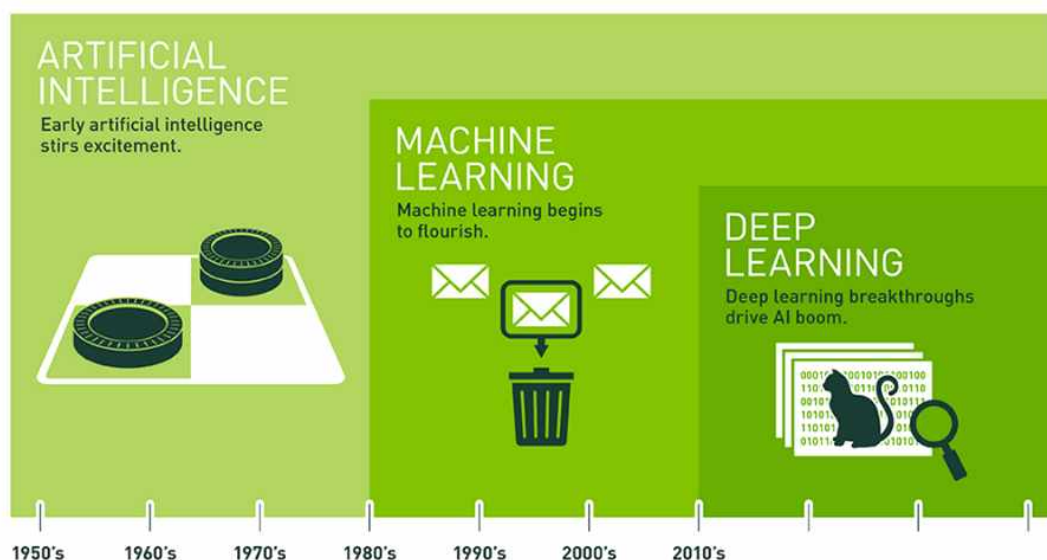


그림 1. 인공지능 기술은 1950년대에 태동하였으며, 1980년대 이후 머신러닝, 2010년 전후 딥러닝 기술이 인공지능을 구현하는 선도적 기술로 자리잡았다.

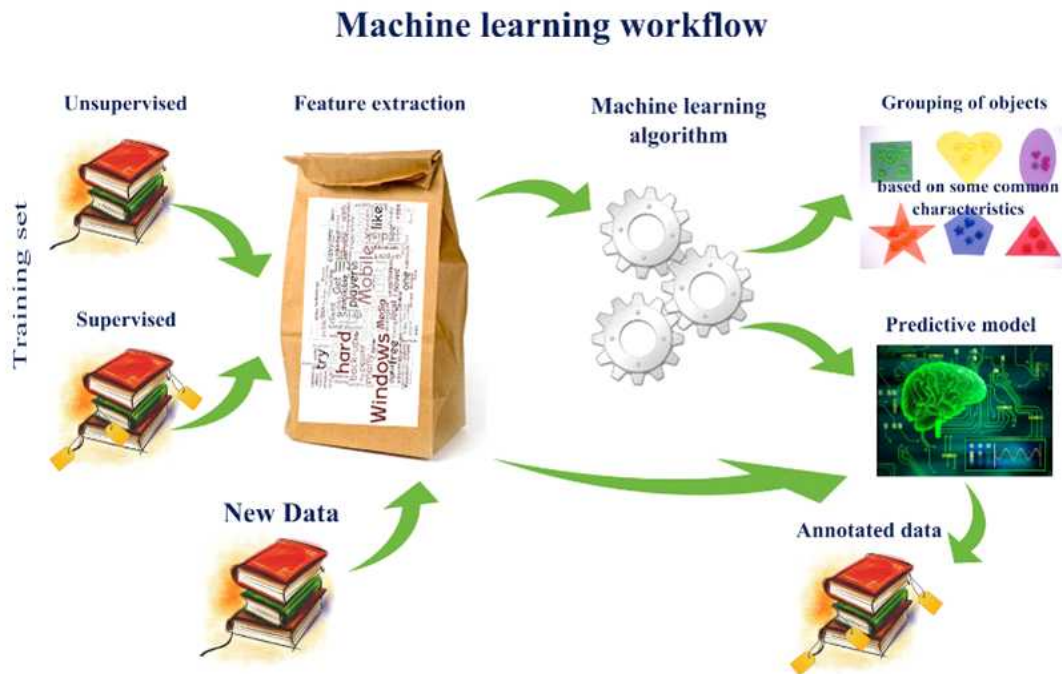


그림 2. 머신러닝의 핵심 절차는 [데이터 수집 및 정리] - [특징 추출(feature extraction)] - [머신러닝 알고리즘을 적용하여 학습 수행] - [학습한 모델을 적용]하는 네 단계로 정리할 수 있다. 학습의 종류는 데이터에 ‘표지/레이블’을 지정하는지 생략하는지 여부에 따라 비지도(unsupervised)와 지도(supervised) 학습으로 구분된다.

과정은 학습의 목표와 형태에 따라 크게 지도학습(supervised learning)과 비지도학습(unsupervised learning)으로 구분할 수 있다. 머신러닝 과정에서 가장 먼저 하는 일은, 연구자가 풀고자 하는 문제를 파악하고 명확하게 문제를 정의(define)하는 것입니다. 풀려는 문제가 인물사진을 보고 남녀를 구분하는 경우와 같이 입력받은 데이터(x)의 종류(y)를 답하는 문제로 정의 가능하면 x와 y를 이어주는 함수  $y=f(x)$ 를 데이터에서 찾는 지도학습 방식의 머신러닝 도구(알고리즘)를 사용하게 된다. 풀려는 문제가 데이터(x)에 내재된 규칙, 특성(예를 들면 인물 사진이 많이 있을 때 얼굴형태가 비슷한 사람들의 사진끼리 모으기)을 발견하려 하는 것이라면 비지도학습 방식으로 문제를 풀게 된다.

두 경우 모두 ‘문제를 푸는 데 필요한 데이터(x)’를 적어도 ‘컴퓨터가 알아볼 수 있는 형태’로 정리를 해서 입력해야 하고, 더 나아가서는 ‘문제를 잘 푸는데 도움이 되는 형태’로 변형/요약하는 과정이 필요하다. 이 ‘문제를 잘 푸는데 도움이 되는 형태로 변형/요약’하는 과정을 보통 ‘특징 추출(feature extraction)’이라고 한다. 문제에 따라, 연구자가 추구하는 해법에 따라 추출하는 특징(feature)은 얼마든지 달라지지만, 이러한 특징을 입력으로 하는 학습 과정에는 공통의 머신러닝 도구(알고리즘)를 쓸 수 있다. 지도학습의 경우에는 대표적인 도구로 결정트리, 인공신경망, 지지벡터기계(SVM) 등이 있고, 비지도학습 중 경우 비슷한 데이터를 모으는 문제의 경우는(‘군집화(clustering)’라고 부르

는 비지도학습 문제의 한 형태) 계층적 군집화, k-평균 등의 도구가 있다.

기존에는 문제의 종류에 따라 해당 분야의 전문가가 특징 추출을 하는 방법을 고안해내어야 했으나, 딥러닝에서는 ‘특징 추출’ 과정도 학습 과정에 포함을 하여 데이터만 보고도 문제를 표현하고 푸는 데 좋은 특징을 학습을 통해 찾는다. 그래서, 요즘 딥러닝을 이용해 지도학습 또는 비지도학습을 할 때는 주변에서 구할 수 있는 형태의 데이터(문서의 경우 단어 또는 알파벳, 사진의 경우 픽셀의 집합으로 표현된 형태의 컴퓨터 파일, 소리의 경우 spectrogram이라 부르는 가장 기본적인 파일)를 그대로 알고리즘의 입력값으로 사용한다. 이러한 학습 과정을 표현력 학습(representation learning)이라 하며, 데이터와 알고리즘이 충분히 준비된 문제에 대해 입력과 출력 사이의 전 과정을 기계가 학습을 통해 자동적으로 구축하는 종단학습(end-to-end learning) 해법이 등장하고 있다.

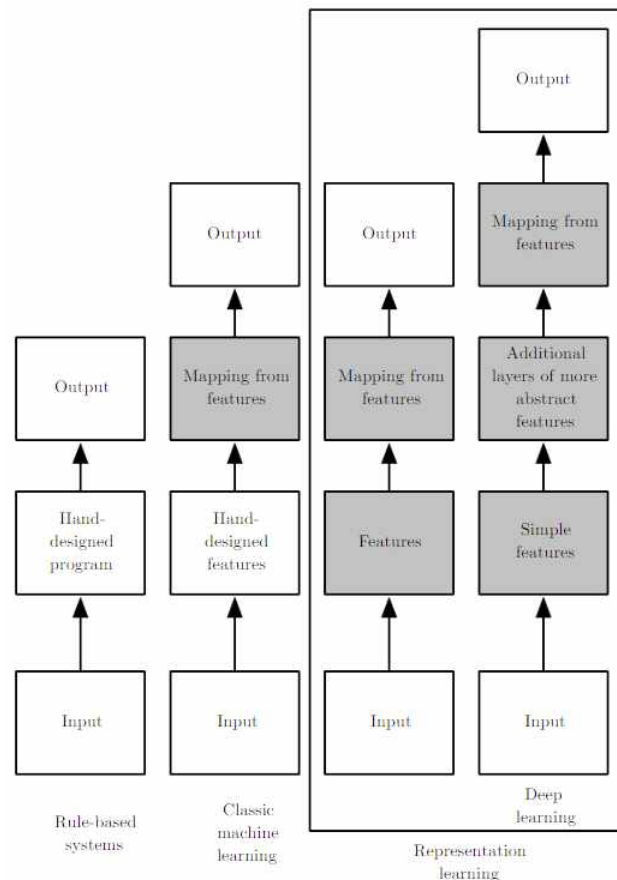


그림 3. 인공지능을 구현하는 대표적 방법론의 발전과 구성 비교. 회색 상자는 데이터에서 학습하여 구성할 수 있는 요소이다. 규칙 기반 시스템과 기존의 머신러닝과 대조적으로 딥러닝은 데이터에서 다양한 개념 수준의 특징값을 학습하여 문제를 스스로 해결하는 표현력 학습을 추구한다.

전통적인 방법



종단 학습



그림 4. 자율주행 문제 중 핸들 조정에 대한 종단학습의 예시. 머신러닝과 딥러닝이 발전하면서 특정 문제에 대한 해법을 단편적 모듈의 조합으로 구성하던 방법이 입력과 출력 사이의 전 과정을 단일한 학습 모듈로 해결하는 종단학습이 가능해지고 있다.

종단학습의 개념을 간략히 표현하자면, 학습에 필요한 데이터를 거의 원본 그대로(raw) 입력하여 지정한 형태의 출력이 산출되도록 학습하는 과정이다. 딥러닝이 등장하기 이전에, 문서의 종류를 분류하거나 사진 속에 있는 물체를 맞추는 전형적인 분류(classification) 문제의 해법은 종단학습으로 얻을 수 있었다. 딥러닝 기술이 발전함에 따라 종단학습으로 해법을 얻을 수 있는 문제의 종류와 복잡도가 크게 증가하였다. 예를 들면, 자율주행 자동차에 들어가는 필수 요소 중 하나로서 전방 카메라의 입력 정보를 기반으로 한 실시간 자동 핸들 조정 방법을 살펴보자. 전통적인 방법에서는 카메라 이미지에서 자동차를 검출하는 모듈과 보행자를 검출하는 모듈을 따로 구성한 후 두 정보를 취합하여 주행 계획을 세우고 최종적으로 핸들 방향을 조정하는 복잡한 과정을 구성하였다. 그러나 딥러닝을 중심으로 하는 종단학습 해법에서는 카메라 이미지만을 입력하여 바로 핸들 방향 조정을 출력하는 것이 가능하다.

시각, 언어, 음성지능을 통합한 흥미로운 종단학습 해법은 2010년 중반을 전후하여 급격히 개발되고 있다. 대표적인 사례를 다음 표에 정리하였다. 이러한 해법은 학습에 필요한 대규모 데이터가 있어야만 구현할 수 있다. 이 중 일부는 ‘1-라’ 절에서 사례를 중심으로 조금 더 살펴볼 수 있다.

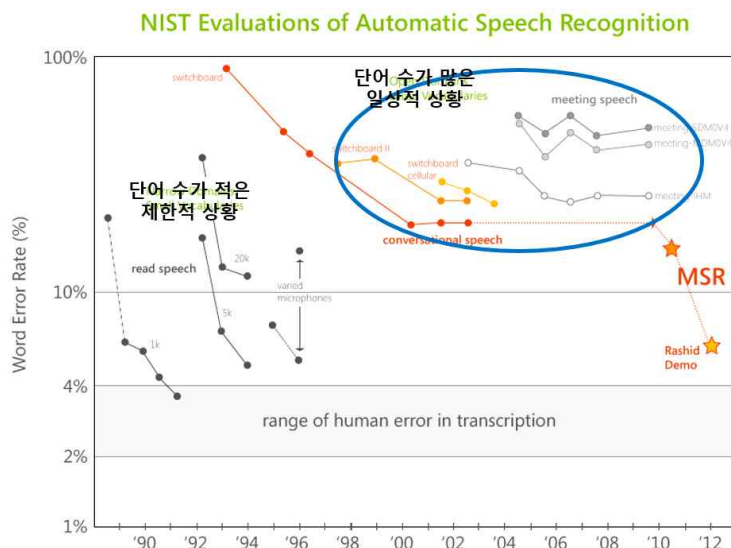
표 1. 딥러닝을 기반으로 한 ‘종단학습(end-to-end learning)’ 이 가능하게 된 문제의 사례. 기존의 해법에서는 입력과 출력 사이에 전문가의 경험을 바탕으로 구성된 여러 중간 단계가 필요하였으나, 종단학습을 통한 해법은 중간 과정이 한 종류의 딥러닝 모델 학습으로 얻게 된다. 이러한 해법이 가능하게 된 핵심 요인으로 대규모 데이터에서 개념적, 고차적 지식을 자동으로 추출하는 ‘표현력 학습(representation learning)’ 기법의 발전을 들 수 있다.

문제의 종류	입력	출력	설명
사진 태그 자동 생성 (image tagging)	사진	사진 설명 태그 집합	사진에 포함되어 있는 물체와 환경 등을 표현하는 단어의 집합을 생성
사진 설명 자동 생성 (image captioning)	사진	사진 설명 문장	사진을 설명하는 자연스러운 문장을 생성
동영상 자동 설명 (movie captioning)	동영상	동영상 설명 문장	동영상을 설명하는 자연스러운 문장을 생성
음성에서 글자 인식 (speech to text)	음성	스크립트	음성 인식. 특히, 하나의 딥러닝 모델로 영어와 중국어를 동시에 인식
자동 번역 (machine translation)	언어 #1 문서	언어 #2 문서	영어-프랑스어 등과 같이 두 언어 간의 번역을 위한 딥러닝 모델. 제한된 언어쌍 만으로 학습한 딥러닝 모델이 학습하지 않은 언어쌍 번역에도 활용된다.
사진/그림 자동 생성 (image synthesis)	표지, 태그	사진/그림	학습한 물체의 종류 중 하나를 지정하면, 학습 과정에서 본 적이 없지만 같은 표지를 달 수 있는 자연스러운 사진/그림을 생성

## 2. 딥러닝이 선도한 인공지능의 도약

인공지능의 목표는, 앞 절에서 언급하였듯이, ‘사람처럼 생각하고 행동하는 기계/컴퓨터’를 만드는 것이다. 간단히 말하자면, 똑똑한 컴퓨터 시스템을 만드는 것이다. 사람을 기준으로 똑똑하다는 판정을 받으려면 최소한 1) 주변 환경을 인식하고 이해할 수 있어야 하며, 2) 이러한 환경에서 적절한 행동을 취할 수 있어야 할 것이다. 요건을 약간 구체화하면 기본적인 음성 및 시각 능력, 언어 이해 능력, 행동 계획 및 주변 사물의 행동 예측 능력이 필요하다. 즉, 음성 지능, 시각 지능, 언어 지능, 행동 지능이 갖추어져야 한다. 1950년대 이후 인공지능의 역사는 이러한 지능을 구현하기 위한 수많은 시행착오와 더딘 발전으로 점철되어 있었으나, 2010년을 전후하여 딥러닝은 이러한 지능 구현에서 기존 방법론의 한계를 뛰어넘고 확장하며 다각화하는 급격한 변화를 이끌어내고 있다. ‘1-나’ 절에서는 각 지능별로 딥러닝이 이루어낸 획기적 도약을 살펴보고, ‘1-라’ 절에서는 이러한 개별 지능 요소가 결합된 다중 지능 구현 사례를 살펴보겠다.

첫 번째로 살펴볼 도약 영역은 음성지능이다. 사람의 말을 알아듣는 음성인식 연구의 역사에서 미국 표준연구소(NIST)에서 구성한 ‘Switchboard’라는 데이터는 음성인식의 척도를 재는 데 사용하는 표준적인 데이터이다. 아래 그림에서 볼 수 있듯이, 1990년대 초에 데이터가 공개된 후 1990년대 말까지 여러 연구자들의 노력으로 이 데이터에 대한 오류율은 빠르게 줄어들었으나 이후 10년 동안 오류율이 23%에서 더 이상 내려가지 않고 있었다. 그러나 2009년에 마이크로소프트사에서 딥러닝의 개척자 중 한 명인 토론토 대학교의 힌튼(Hinton) 교수를 초빙한 후, 딥러닝을 이용한 음성 인식 기법은 이 Switchboard 데이터셋에 대한 오류율을 2010년에 15%, 2011년에 7%대로 낮추었다. 이후 음성인식 기술은 눈에 띄게 향상되었으며, 음성인식 기능은 스마트폰을 거쳐 2017년 CES에서는 대부분의 가전에 기본적으로 탑재되기에 이르렀다.



두 번째로 시각지능을 살펴보자. 시각지능의 가장

그림 5. 딥러닝이 이루어낸 음성지능의 도약 사례. ‘switchboard’ 음성 인식 데이터에 대한 인식율의 10년간의 정체가 2010년 이후 딥러닝의 도입으로 크게 개선되었다.



기본적인 요소는 물체를 인식하는 능력이다. 서비스 측면에서는 특히 얼굴 인식 기능이 필수적이다.

머신러닝 연구자들은 대규모의 데이터베이스를 구축하고 공개하여 경쟁적으로 성능 향상을 도모하여왔다. 이러한 데이터베이스 중 시각지능 연구의 핵심 도구를 기존의 컴퓨터 비전 기법 춘추전국 시대에서 딥러닝 시대로 바꾸는데 기여한 ImageNet 데이터베이스<sup>1)</sup>가 2009년에 구축되었다. ImageNet의 사진 데이터에서 물체 인식 성능을 겨루는 대회가 2010년부터 매년 개최되었는데, 2012년도에 기존 컴퓨터 비전 연구자들을 경악시킨 결과가 나왔다. 컨볼루션망(CNN)이라는 딥러닝 기법이 다른 모든 컴퓨터 비전 기반 팀의 결과보다 월등히 우수한 성능으로 1위를 차지하였으며, 2013년 이후 매년 획기적인 딥러닝 기법이 1위 성능을 독식하는 한편 모든 대회 참가팀들이 딥러닝 기법을 적용하게 되었다.

얼굴 인식 분야에서도 딥러닝의 기여로 큰 도약이 계속되고 있다. 음성인식 사례와 유사하게, 1990년대 이후 미국 표준연구소에서 관리하는 얼굴 인식 대회와 관련 데이터셋<sup>2)</sup>을 중심으로 1993년부터 2011년까지 매 2년마다 오류율이 반으로 줄어드는 발전이 있었다. 딥러닝이 등장하며 오류율 감소의 가속화와 함께, 다양한 상황에서의 얼굴 인식과 인물 인식, 동일 얼굴 탐색 등의 다각화된 기술 발전이 이어지고 있다.

세 번째로 언어지능에서 딥러닝이 이루어낸 도약을 살펴보자. 음성과 시각지능 사례와 비교해볼 때 언어는 인류의 경험과 지식을 직접적으로 표현하는 단계의 정보라는 점에서 언어지능의 영향력은 매우 광범위하다. 문서 자동 생성, 의미 수준 단어 표현, 자동 번역

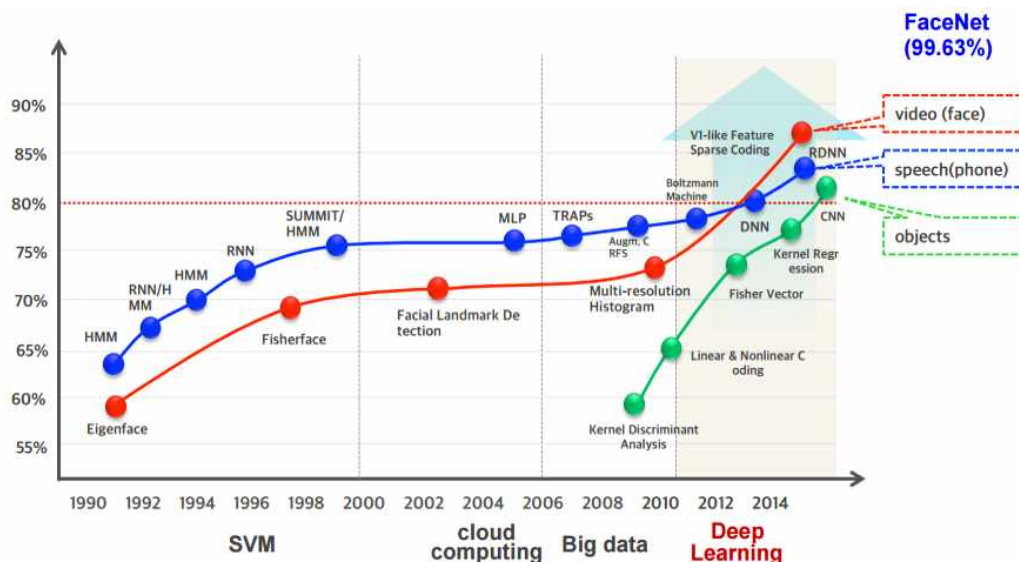


그림 6. 딥러닝이 이루어낸 시각지능의 도약 사례. 이미지에 있는 사물(object) 인식 정확도와 비디오 상에서의 얼굴 인식 정확도가 딥러닝의 도입으로 크게 향상되었다. (그림 출처: H. Choi, Recent Advances in Recurrent Neural Networks, 2015.04)

1) <http://image-net.org/>

2) <https://www.nist.gov/programs-projects/face-projects>

등과 같은 언어 단계의 지능뿐만 아니라 음성지능 및 시각지능과 결합된 사례, 예를 들면 음성 자동 번역, 사진 설명 자동 생성과 같은 기술이 딥러닝을 기반으로 발전하고 있다.

딥러닝을 핵심으로 한 문서 자동 생성 기술의 수준은 현재 보고서, 기사 등과 같은 기술적인 문서뿐만 아니라 시와 소설 등의 창작의 영역까지 넘보고 있다. 기본적인 딥러닝 기술만으로도 다양한 실질적 문서 생성을 시험해볼 수 있다. ImageNet 대회를 주관<sup>3)</sup>, 관리하고 2017년 1월 현재 OpenAI에서 재직 중인 Andrej Karpathy는 2015년도의 블로그 글<sup>4)</sup>에서, 알파벳 단위로 학습한 순환신경망을 이용하여 컴퓨터가 자동으로 생성한 다양한 문서를 소개하였다. 학습을 통해 자동 생성한 문서의 종류는 ‘셰익스피어의 희곡’, ‘위키피디아 문서’, ‘LaTeX 문서’, ‘Linux 소스 코드’를 포함하며, 이후 ‘가상 오바마’, ‘가상 트럼프’와 같은 수많은 확장 사례가 소셜넷에서 회자되고 있다.

다음으로 살펴볼 딥러닝 언어지능 사례는 의미 수준 단어 표현이다. 신경망을 이용하여 단어의 벡터(vector) 표현을 학습한 결과, 단어 간의 의미 수준에서의 연산, 문법적 관계, 단어의 조합을 통한 의미적 구문 생성 등과 같은 ‘언어를 이해하는’ 듯한 컴퓨터 프로그램을 구현하는 것이 가능해졌다.

마지막으로 보다 최근에 나타난 행동지능 구현의 도약을 살펴보겠다. 행동지능 도약의 배경에는 머신러닝의 기본 과정 중 하나인 강화학습(reinforcement learning)에 딥러닝이 도입이 있었다. 2015년 초에 Nature 지에 영국 스타트업 회사 DeepMind 연구진의 논문이 게재되었다. 이 논문에서는 Atari 사의 게임 화면만으로 학습하여 사람보다 게임을 더 잘 하게 된 사례가 소개되었으며, 이러한 딥 강화학습 기술을 높이 산 Google이 DeepMind를 고가에 인수하는 직접적인 계기가 되었다. Google DeepMind는 불과 1년 후에, AlphaGo를

- **semantics:**  $\text{vec(Beijing)} - \text{vec(China)} + \text{vec(Japan)} = \text{vec(Tokyo)}$
- **syntactic :**  $\text{vec(quick)} - \text{vec(quickly)} + \text{vec(slowly)} = \text{vec(slow)}$

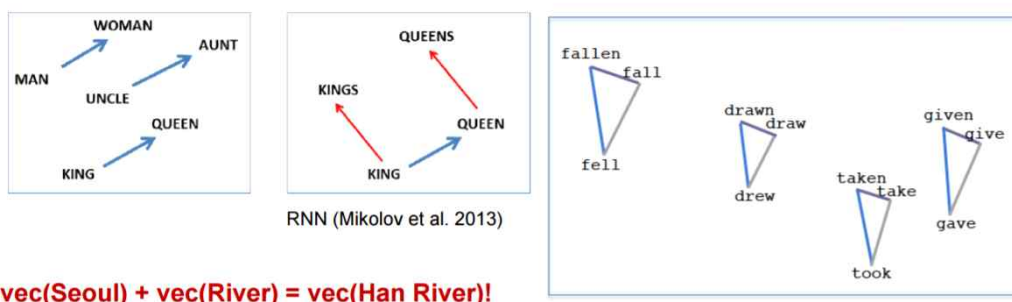


그림 7. 딥러닝이 이루어낸 언어지능 사례. 의미 수준의 단어 표현과 연산이 가능해졌다. 이러한 기술을 word2vec으로 통칭하며, sentence2vec, doc2vec 등과 같은 확장 연구로 이어지고 있다. (그림 출처: H. Choi, Recent Advances in Recurrent Neural Networks, 2015.04)

3) 스탠포드대의 Fei-Fei 교수진과 프린스턴대의 Li 교수진이 주도

4) <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>

통해 바둑의 최고수를 4:1로 꺾는 대 파란을 일으켰으며, 이는 전문가들의 예상보다 적어도 10년은 앞서 이루어낸 결과였기에 그 파급효과는 더욱 컸다.

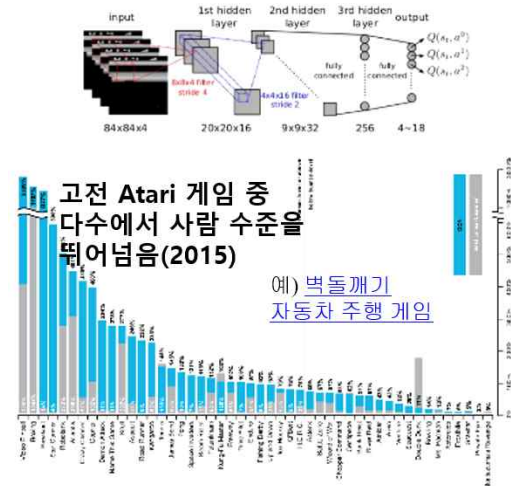


그림 8. 딥러닝이 이루어낸 행동지능의 도약 사례 1. 딥 강화학습을 통해 컴퓨터 게임의 최고수를 구현

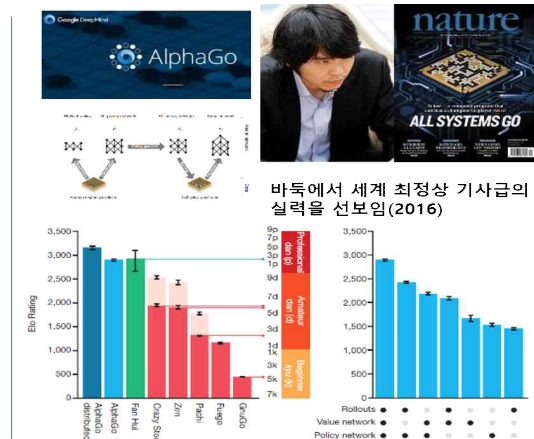


그림 9. 딥러닝이 이루어낸 행동지능의 도약 사례 2. 딥 강화학습을 발전시켜 바둑의 최강자를 이기는 이벤트를 일구어냄

### 3. 딥러닝의 구현 방법

인공지능 연구자들이 추구하는 가장 이상적인 인공지능 모델은 바로 인간의 두뇌이다. 뇌과학, 인지과학, 신경과학 등의 연구를 통해 밝혀진 뇌의 구조와 작동 원리는 인공지능 연구자들에게 다양한 영감을 불러일으키는 좋은 정보원이 되어왔다. 1940년대 초에 생물 신경세포를 모사하여 구성한 인공 신경세포(퍼셉트론)에 대한 연구가 시작되었으며, 이러한 신경세포로 구성된 신경망을 인공적으로 구현한 인공신경망은 1980년대 중반부터 본격적으로 인공지능 구현을 위한 도구로 사용되기 시작하였다. 뇌의 시각 기능에 관한 1960~70년대의 뇌과학 연구는 1980년대의 컨볼루션망(CNN)의 발명과 2010년대 중반의 ‘주의 집중(attention)’ 기반 딥러닝 기법 개발의 토대로 이어졌다.

딥러닝의 개념을 딥러닝 기법의 구성과 동작 방식을 기준으로 간략히 표현하자면, ‘인공신경망의 다층 구조가 깊어진 모델’로 설명할 수 있다. 정보가 인공신경망에 입력되어 여러 층을 거치며 처리됨에 따라, 마치 눈을 통해 들어온 정보가 여러 층을 거치며 물체로 인식되는 것과 같은 방식으로, 여러 단계의 정보 표현과 개념, 지식이 학습될 것을 기대하는 것이다.

깊은 인공신경망을 구성하는 기본적인 방법은 크게 세 가지가 있다. 가장 기본적인 구성은 퍼셉트론이라 불리는 기본적인 인공 신경세포의 다층 구조를 입력과 출력 사이에



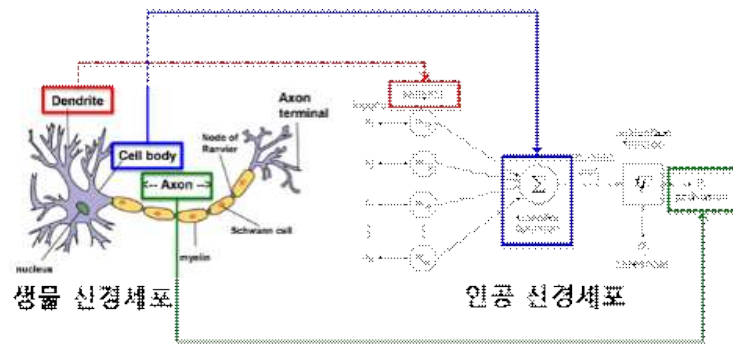


그림 10. 생물의 신경세포와 신경망을 본뜬 인공 신경세포와 인공신경망이 딥러닝 모델의 가장 기본적인 구성요소이다.

두는 다층 퍼셉트론(multilayer perceptron, MLP)이다. 다층 퍼셉트론의 이웃하는 층 간 신경세포는 모두 연결된(fully connected) 구조를 가지며, 데이터에서 학습하는 과정에서 연결선의 가중치가 전체적으로 조정되어 지능을 갖추게 된다. 두 번째 기본적인 구조는 사진과 같이 공간적인 구성을 가진 정보를 처리할 때 적합한 컨볼루션망(convolutional neural network, CNN)이다. 컨볼루션망은 이웃한 공간상에서 나타나는 패턴을 학습하고 요약하는 다층 구조의 망으로 구성하며, 복잡한 문제일수록 다층 구조를 깊게 구성하고 대규모의 데이터를 통해 학습하게 된다. 세 번째의 구조는 시간에 따라 변하거나 순서가 중요한 데이터에서의 학습을 위한 순환신경망(recurrent neural network, RNN)이다. 순환신경망에서는 시간 또는 단계를 따라 펼쳐지는 망의 구조로 인해 기본적인 딥 구조가 형성되며, MLP와 CNN의 경우와 같이 다층 구조를 구성하여 다양한 스케일의 순서 정보를 학습하는 것이 가능하다.

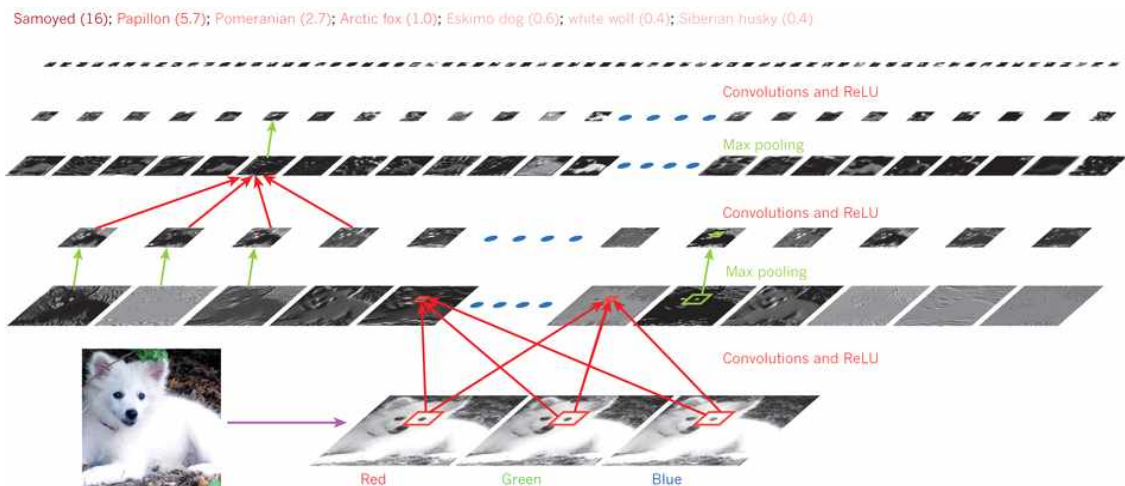


그림 11. 컬러 이미지를 위한 컨볼루션 신경망(CNN)의 작동방식. 계층별로 주어진 다수의 필터에 대해 컨볼루션과 ReLU 연산, 풀링 연산을 반복하면서 계층을 따라 올라간 후, 최종 계층에서 이미지에 나타난 물체의 후보군에 대한 가중치를 출력한다.

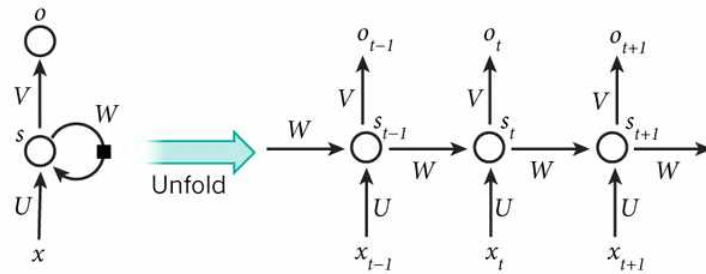


그림 12. 순환신경망 (recurrent neural network)의 개념도. 시간 또는 순서의 방향을 따라 심층 구조가 구성된다.

다음 그림에서는 인공신경망 구성에서 실제로 적용되는 다양한 사례를 종합적으로 살펴볼 수 있다.

이와 같은 다양한 방식으로 문제에 적합한 망의 구조를 정하고, 문제의 사례를 담은 대규모 데이터가 준비되면, 데이터를 살펴보고 반복학습을 통해 문제의 해법을 스스로 발견

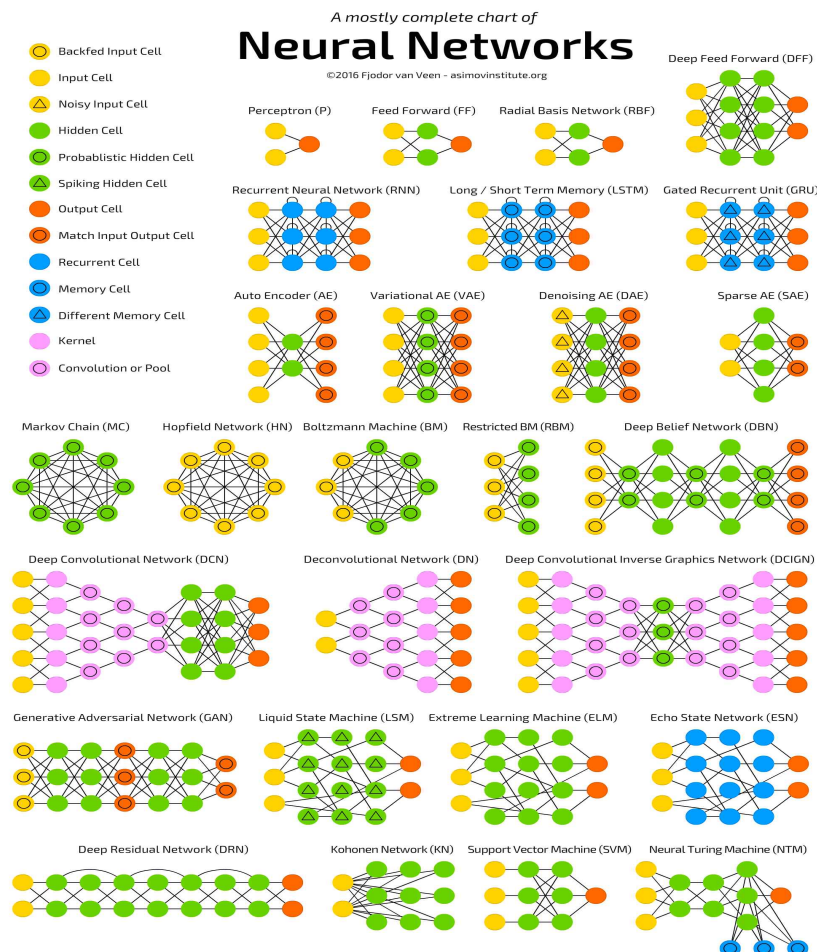


그림 13. 딥러닝의 핵심 구성 요소인 인공신경망의 다양한 사례. 인공뉴런은 원으로 표현이 되었으며, 그림에 표기된 바와 같이 다양한 종류의 인공뉴런과, 이들의 다양한 연결 방법이 가능하다.

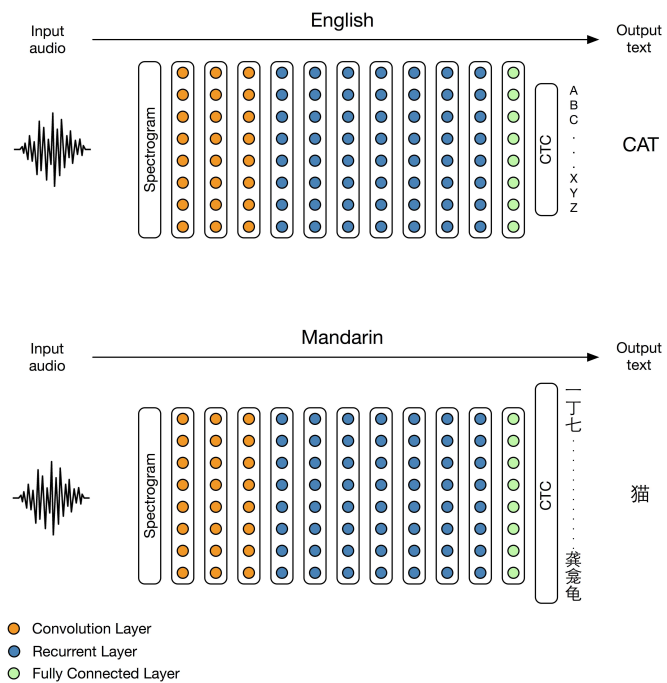
해나가는 ‘학습’ 단계와, 학습 결과를 점검하기 위한 ‘테스트’ 단계를 거치게 된다. 학습 과정에서는 학습 중인 딥러닝 모델이 내놓는 답의 오류를 층을 따라 순차적으로 줄이기 위해 오류 정보를 출력층에서 입력층 방향으로 ‘역전파(backpropagation)’를 하면서 신경망 내부의 연결선의 가중치를 조금씩 조정을 한다. 학습의 성공도를 점수화한 기준에 따라 반복학습을 통해 점수를 최대화하는 과정이 공통적으로 적용된다. 학습된 딥러닝 모델을 실전 테스트하는 단계에서는 학습에 반영하지 않은 데이터를 입력으로 주고 적절한 출력을 내놓는지를 정량적으로 평가하게 된다.

보통 학습의 궁극적 목표는 제한된 학습데이터만으로 학습한 모델이 풀고자 하는 다양한 실전 상황에서도 성공적이고 안정적으로 동작하는 것이다. 이러한 목표를 위해 보통 테스트 성능이 좋은 모델을 탐색하는 과정을 거치게 되며, 이 과정에서 수많은 계산 과정과 데이터가 필수적이다. 딥러닝 발전의 핵심 원동력은 CNN, RNN 등과 같은 새로운 알고리즘뿐만 아니라 이러한 알고리즘을 뒷받쳐줄 컴퓨팅 파워와 빅데이터가 포함된다.

## 4. 딥러닝의 최신 응용 사례

3절까지의 내용으로 딥러닝의 최신 응용 사례를 살펴보고 이해할 준비가 되었다.

먼저 음성지능의 사례로서 중단 학습 음성인식을 살펴보자. 다음 그림은 중국의 Google이라 불리는 Baidu의 실리콘밸리 연구진이 2015년 말에 논문으로 발표한 DeepSpeech 2의 소개이다. 입력 음성 신호와 인식한 글자 출력단 사이에 컨볼루션망, 순환신경망, 완전연결망이 각각 3층, 7층, 1층씩 순서대로 배치된 것을 확인할 수 있다. 연구진은 이 모델로 영어와 중국어를 모두 높은 성능으로 인식이 가능함을 보였다.



시각지능의 최신 사례로는 이미 지 자동 생성 사례를 살펴보겠다. 다음 그림은 2017년도 1월 현재 딥러닝 연구자들이 가장 주목하는

모델 중 하나인 생성대립넷(GAN)을 이용하여 얼굴을 생성하고, 의미적 수준에서 얼굴간

그림 14. 딥러닝을 적용한 음성지능의 최신 사례. Baidu에서 2015년 말에 발표한 DeepSpeech 2는 단일한 딥러닝 모델로 영어와 중국어 음성을 인식할 수 있음을 보였다.

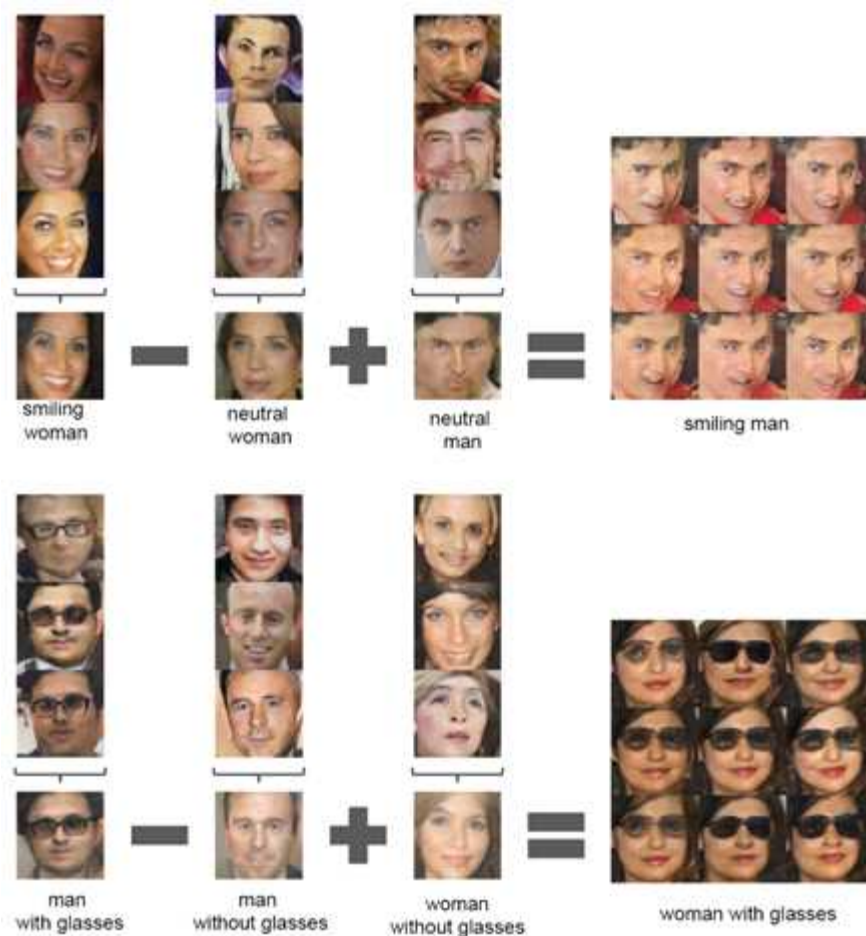


그림 15 생성대립넷(GAN)을 이용한 사람 얼굴 자동 생성 및 의미적 수준에서의 연산 사례 (Radford et al., 2015)

연상적 연산을 보인 사례이다. 컴퓨터가 생성한 얼굴의 모양은 완벽하지는 않으나 충분히 사람임을 인지할 수준이 되며, 나아가 얼굴에 나타난 고차적 정보의 가감이 가능함을 볼 수 있다.

언어지능의 사례로는 자동번역(machine translation) 시스템이 주목할 만하다. Google에 의해 2000년대 중반부터 널리 사용된 통계 기반의 기계번역은 딥러닝 기술에 힘입어 진일보하였다. 기존의 통계 기반의 기계 번역에서 사용된 메모리의 일부만 사용하면서도 깊은 표현의 학습을 통해 성능을 향상한 신경 기계 번역(Neural Machine Translation) 기법이 발전하였으며, 2016년도 하반기에 발표된 구글의 새로운 신경 기계 번역 시스템(다음 그림), 전문가의 번역에 버금갈 정도의 뛰어난 자동 번역 성능을 보인 점과, 일부 언어 쌍으로 학습한 사례를 학습하지 않은 언어쌍에도 적용 가능한 점에서 또다른 언어지능의 마일스톤을 정립하였다. 해당 시스템은 8개의 인코더, 8개의 디코더 레이어로 구성된 잔차(residual) 연결을 갖는 깊은 LSTM 네트워크이며, 인코더와 디코더는 주의 기작(attention)을 반영하여 연결된다.



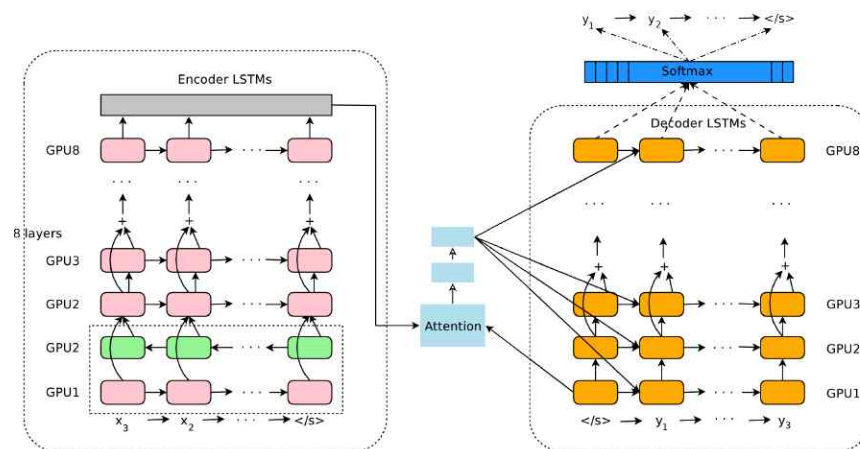


그림 16. 딥러닝을 이용한 기계 번역의 최신 사례, 구글의 신경 기계 번역 시스템(2016)

아래 그림에서는 언어지능 구현의 핵심 연구분야인 자연언어처리(natural language processing, NLP) 분야에서 딥러닝의 도입으로 인해 야기되는 해법의 변화를 소개하고 있다. 글의 종류 구분, 글에 나타난 감성 분석, 글이 다루는 주제 등과 같은 다양한 문제 해결을 위해 기존 NLP 연구자들은 특성이 다른 수많은 알고리즘과 프로세스를 만들어왔지만, 최근에는 딥러닝 기법을 공통적으로 적용하는 것이 표준적 방법으로 자리잡았다.

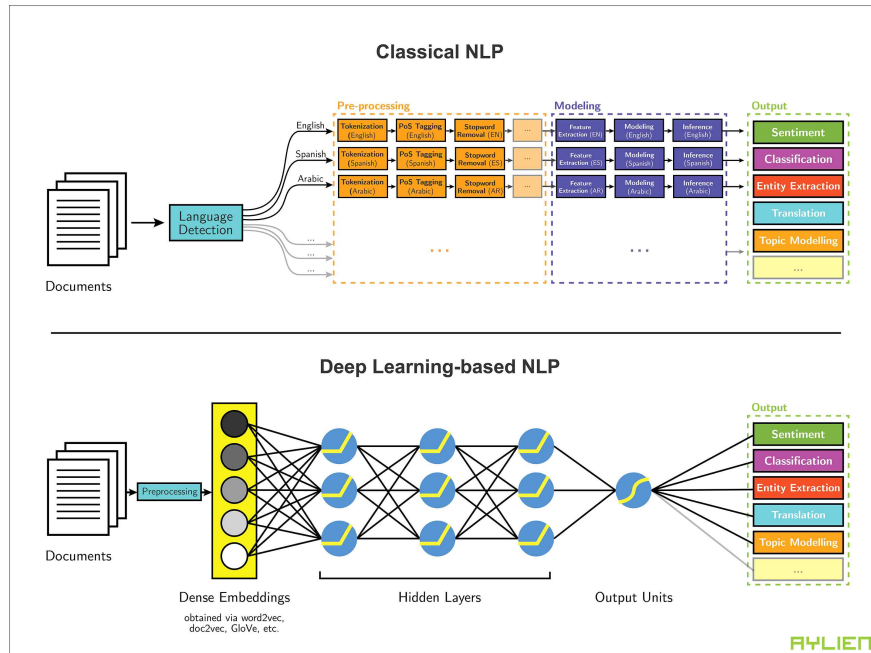


그림 17. 언어지능의 핵심인 자연언어처리(NLP) 기술에서 기존에는 출력에 따라 다양한 과정을 조합한 수많은 절차가 고안되었지만(상단), 최근에는 단일한 딥러닝 방법론으로 다양한 문제의 종단학습 기반 해법을 구성하는 것이 일반적인 해법으로 정립되고 있다(하단). (출처: <http://blog.aylien.com/leveraging-deep-learning-for-multilingual/>, 2017-1-21)

마지막으로 다중 지능 사례를 하나 살펴보도록 하자. 다음 그림은 사진을 설명하는 문장을 자동으로 생성하는 사례이다. 딥러닝 모델을 조합하여 시각지능과 언어지능을 모두 보인 경우다. 컨볼루션망(CNN)으로 사진의 정보를 분석한 후, 이 정보를 순환신경망(RNN)에 전달하여 글자 또는 단어를 순서대로 출력하는 방식으로 설명 문장을 생성한다. 특히, 단어의 선정은 입력된 사진에서 해당 단어에 해당하는 물체에 대해 주의 집중(attention)을 함으로써 이루어지는 점에서 딥러닝 이전의 인공지능과 비교하여 큰 도약이 있었음을 확인할 수 있다.

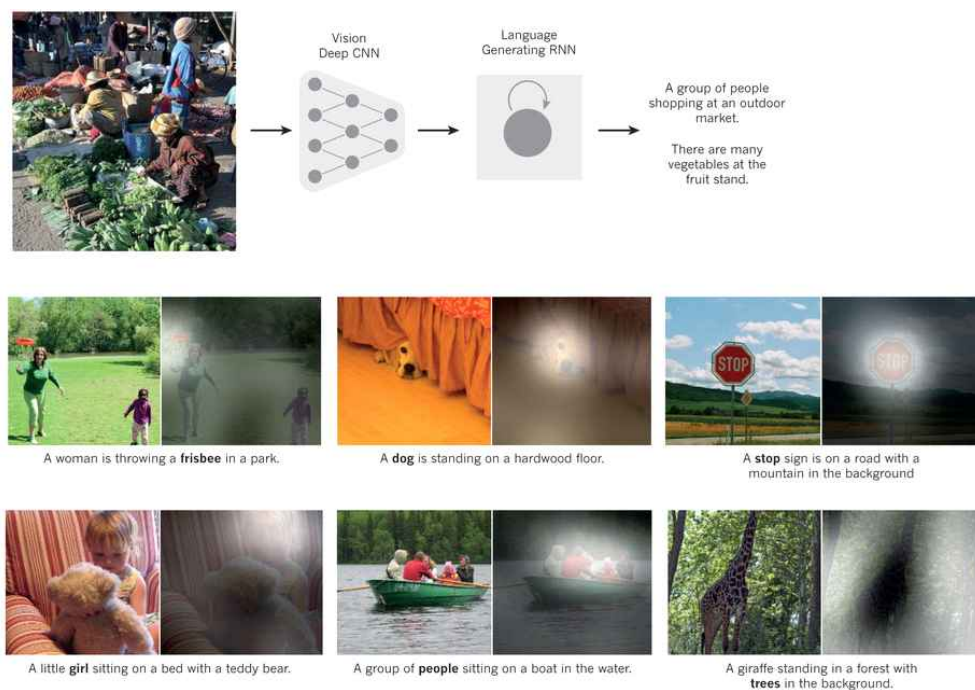


그림 18. 딥러닝 모델을 조합하여 사진을 설명하는 문장을 자동으로 생성하는 사례(2014, 2015)