ETRI Insight

Insight Report 2018-10

딥러닝 이후, AI 알고리즘 트렌드

이승민

본 보고서는 ETRI 기술경제연구본부 주요사업인 “ICT R&D 경쟁력 제고를 위한 기술경제연구”를 통해 작성된 결과물입니다.

ETRI Insight

Contents

표목차

그림목차

ETRI Insight

요 약

요 약

연구 배경

- 딥러닝이 본격적으로 주목받기 시작한 2012년 이후 AI 알고리즘의 특징과 적용 가능성을 살펴보고 향후 인공지능의 발전 방향을 살펴보고자 함
- AI 알고리즘은 지난 60여 년 동안 두 번의 침체기와 두 번의 전성기를 겪었고 현재 인공지능은 세 번째 전성기를 맞고 있음
- 3차 전성기를 주도한 딥러닝은 범용성을 지닌 인간의 두뇌와 비교하여 정보입력, 학습과정, 도출결과 등에서 한계점을 보이며 도전적 연구의 필요성을 제기함

구분		한계점
입력	데이터 수집범위	- 인터넷 상에 공개된 이미지, 음성 데이터 위주로 활용, 적용 분야에 따라 학습데이터 양이 부족
	데이터 입력방식	- 지금까지 대부분 비즈니스에서 성능을 보장하는 인공지능은 인간이 정답을 부여한 라벨링된 학습데이터를 사용
과정	학습과정 투명성	- 딥러닝 등 인공신경망은 우수한 성능에도 불구하고 학습과정에서 도출결과에 대한 설명력을 제공하지 못함
	학습모델 최적화	- 인간이 사전에 설계한 알고리즘 구조 안에서 파라미터 최적화를 추구, 향후 스스로 진화하고 생성하는 방식으로 확장 필요
결과	도출결과 타당성	- 학습모델의 성능이 통계적으로 우수한 성능을 보장하지만 특정 상황에서 인간이라면 범하지 않을 비상식적인 결과를 도출
	도출결과 재활용	- 한 영역에서 만든 학습모델을 이와 유사한 다른 영역에서 일부분이라도 재활용하지 못함

AI 알고리즘 주요 트렌드

① **[모방화]** 학습모델 내부 또는 학습모델 간 경쟁을 통해 현실 데이터와 유사한 데이터를 생산하여 제한된 학습 데이터량을 늘리려는 트렌드

② **[자동화]** 데이터 수집에서 학습모델 구성에 이르는 일련의 과정에서 인간의 개입을 최소화하여 End-to-End 자동화하려는 트렌드

③ **[통합화]** 특정 알고리즘의 장점을 극대화하고 단점을 보완하기 위해 다양한 AI 알고리즘을 통합적으로 사용하려는 트렌드

④ **[범용화]** 기존 학습모델을 부분 최적화하거나 학습모델이 인간의 인식 과정과 유사하게 데이터를 처리하여 범용성을 높이려는 트렌드

AI 알고리즘별 주요 특징

- 딥러닝과 관련하여 현재 가장 활발히 연구되고 있는 다섯 가지 종류의 AI 알고리즘의 주요 특징을 요약하면 다음과 같음

AI 알고리즘 종류	특징과 의미
생성적 적대 신경망 (Generative Adversarial Networks)	- 적대적으로 경쟁하는 생성기와 판별기를 통해 진본데이터와 매우 유사한 위조데이터를 생성 - 현실에 없는 새로운 데이터 생성, 새로운 형태로 데이터 변환, 데이터 품질 향상 등 새로운 기회 가능성을 제시
심층강화학습 (Deep Reinforcement Learning)	- 복잡한 실제 환경에서 반복적인 경험(데이터)의 시행착오를 통해 최적의 학습모델을 스스로 발전시킴 - 지금까지 PC안에서 이뤄졌던 인공지능을 현실 세계의 다양한 객체에 적용하기 시작한 계기이며 감각기관의 확장을 가져옴
전이학습 (Transfer Learning)	- 학습데이터 확보가 현실적으로 어려운 분야에서 기존에 학습이 완료된 모델의 일부를 재사용하여 학습시간을 단축하고 성능을 보장 - AI 알고리즘이 인간과 같이 학습효과를 가지고 발전 가능성 제시
설명가능 인공지능 (Explainable AI)	- 기존 설명력이 높은 알고리즘의 일부를 활용하거나 개선하여 학습모델이 도출한 결과의 근거를 제공 - AI 알고리즘 사용 시 법과 제도적 문제로 인해 비즈니스 활용 범위의 한계를 극복할 수 있는 가능성 제시
캡슐망 (Capsule Networks)	- 외부세계를 인식하는 과정이 3차원적 벡터방식의 인간의 뇌 인식 과정과 유사하게 알고리즘 구조를 설계 - 현재 연구 최기단계이나 보다 범용적인 알고리즘 혁신을 이끌 차세대 AI 알고리즘으로 주목

AI 알고리즘의 미래를 위한 제언

- **[범용 AI 알고리즘 연구]** 현재 비즈니스 활용 시 경제적 가치를 생산할 수 있는 AI 알고리즘뿐만 아니라 보다 범용성을 추구하는 기초·원천 알고리즘 연구 강화
- **[몸체를 가진 AI 연구]** 실제 환경에 노출된 기계들과 결합한 알고리즘이 현실의 데이터(행동 데이터)를 이용하고 상호작용하는 몸체를 가진 AI 연구
- **[생활 속의 AI 연구]** 지금까지 보고 듣는 영역의 활용을 넘어 소비자와 직접 소통하고 인간을 보조·협업하는 생활 속의 AI 연구
- **[AI 부작용 및 위험 대비]** 인간을 압도하는 우수한 성능을 지닌 AI 알고리즘을 악용하여 가짜뉴스를 생산하거나 사회를 통제하는 수단으로 사용하고 나아가서 국제정치와 전쟁의 패러다임을 바꿀 미래 위험에 대비한 기술적 방안 고민 필요

ETRI Insight

Ⅰ

연구 개요

1. 연구 목적
2. 주요 내용

Ⅰ 연구 개요

1 연구 목적

● 본 보고서는 딥러닝이 본격적으로 주목받기 시작한 2012년 이후의 AI 알고리즘 연구 동향을 살펴보고 적용 가능성 및 R&D 방향 설정을 위한 기초자료로 활용하고자 함

- AI 알고리즘은 지난 60여 년 동안 두 번의 침체기와 두 번의 전성기를 겪었고 현재 인공지능은 세 번째 전성기를 맞고 있음

  * 1차 전성기(1950년대 후반~1960년대 초)는 추론과 탐색 기법 중심이었으나 간단한 문제를 푸는 것 외에 뚜렷한 가능성을 제시하지 못하고 곧바로 1차 침체기를 경험함

  * 2차 전성기(1980년대 후반~1990년대 초)는 특정 분야에서 정해진 규칙에 따라 전문가시스템을 구축하는 방식이었으나 확장성 측면에서 한계를 보이며 2차 침체기를 맞이함

- 3차 전성기를 주도한 딥러닝은 인공지능 역사에서 가장 혁신적 기술로 평가받으며 영상 데이터를 시작으로 음성과 행동 데이터로 적용 범위를 넓혀가고 있음

- 2012년 ImageNet Challenge 대회에서 딥러닝이 압도적 성능으로 우승한 이후 DNN/CNN/RNN 등 알고리즘이 비약적으로 발전하며 산업에 본격 활용

- 그러나 범용성을 지닌 인간의 두뇌와 비교하여 정보입력, 학습과정, 도출결과 등에서 한계점을 보이며 이를 극복하기 위한 연구의 필요성을 제기함

표 1 딥러닝 알고리즘 활용 시 한계점

구분		한계점
입력	데이터 수집범위	- 인터넷 상에 공개된 이미지, 음성 데이터 위주로 활용, 적용 분야에 따라 학습데이터 양이 부족
	데이터 입력방식	- 지금까지 대부분 비즈니스에서 성능을 보장하는 인공지능은 인간이 정답을 부여한 라벨링된 학습데이터를 사용
과정	학습과정 투명성	- 딥러닝 등 인공신경망은 우수한 성능에도 불구하고 학습과정에서 도출결과에 대한 설명력을 제공하지 못함
	학습모델 최적화	- 인간이 사전에 설계한 알고리즘 구조 안에서 파라미터 최적화를 추구, 향후 스스로 진화하고 생성하는 방식으로 확장 필요
결과	도출결과 타당성	- 학습모델의 성능이 통계적으로 우수한 성능을 보장하지만 특정 상황에서 인간이라면 범하지 않을 비상식적인 결과를 도출
	도출결과 재활용	- 한 영역에서 만든 학습모델을 이와 유사한 다른 영역에서 일부분이라도 재활용하지 못함

2 주요 내용

- 본 보고서에서는 딥러닝 관련하여 가장 활발히 연구되고 있는 다섯 가지 종류의 AI 알고리즘의 한계와 특징, 비즈니스 적용 가능성 등을 중점적으로 분석

그림 1 AI 알고리즘 종류 및 보고서 범위

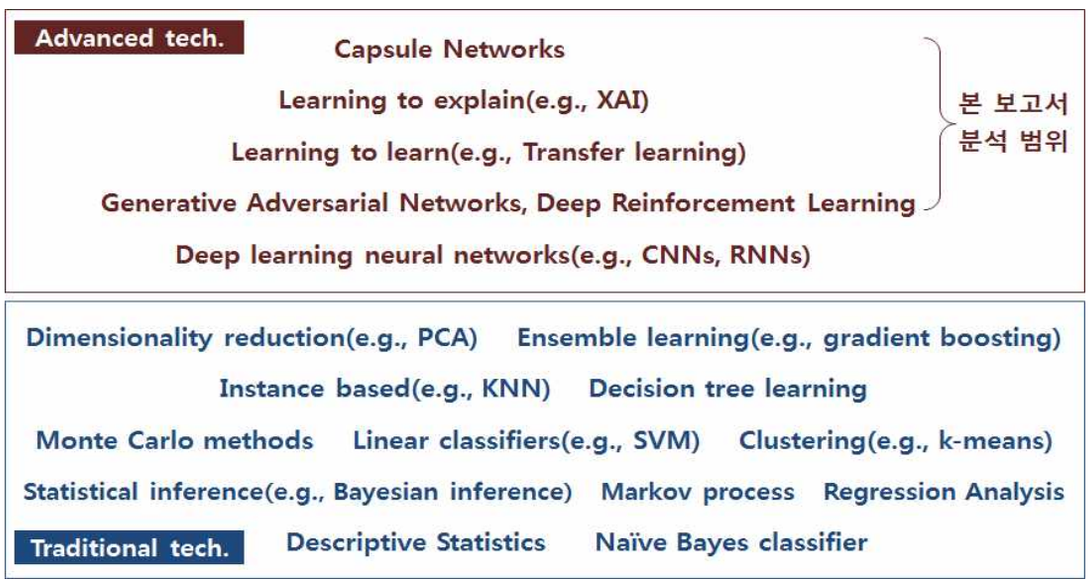


표 2 분석 대상 AI 알고리즘 현황

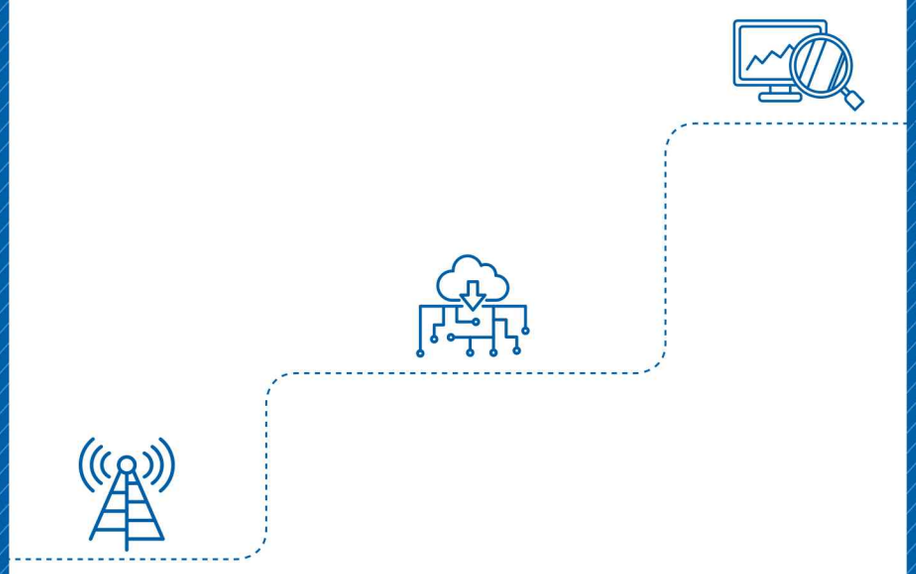
구분	주요 현황
생성적 적대 신경망 (Generative Adversarial Networks)	- 적대적으로 경쟁하는 생성기와 판별기를 통해 실제 데이터와 매우 유사한 위조 데이터를 생성하는 원리로 2014년 제안된 이후 수많은 GAN의 변형들이 개발되고 있음
심층강화학습 (Deep Reinforcement Learning)	- 특정 환경에서 정의된 에이전트가 반복적인 시행착오를 통해 현재의 상태를 탐색하고 정량화된 보상을 최대화하려는 강화학습에 딥러닝을 결합한 것으로 AlphaGo에 적용된 DQN이 대표적
전이학습 (Transfer Learning)	- 특정 분야에서 학습한 결과를 유사 분야에서 재사용하기 위한 것으로 DeepMind의 PathNet, Google의 Inception 모델 전이학습, DARPA의 L2M 등이 대표적
설명가능 인공지능 (Explainable AI)	- 딥러닝 등 인공신경망 알고리즘이 도출한 결과에 대한 설명 가능한 근거나 해석력을 보장하여 알고리즘의 투명성과 신뢰성을 강화하려는 것으로 DARPA에서 제안한 XAI 프레임워크가 대표적
캡슐망 (Capsule Networks)	- 외부 세계를 인식하는 과정이 2차원적 스칼라 방식인 딥러닝 알고리즘의 근본적인 한계를 극복하기 위해 2017년 Hinton이 새롭게 제안한 3차원적 벡터 방식의 인간의 뇌 인식과정과 유사하게 설계

ETRI Insight

Ⅱ

주요 AI 알고리즘 트렌드

1. 생성적 적대 신경망
2. 심층강화학습
3. 전이학습
4. 설명가능 인공지능
5. 캡슐망



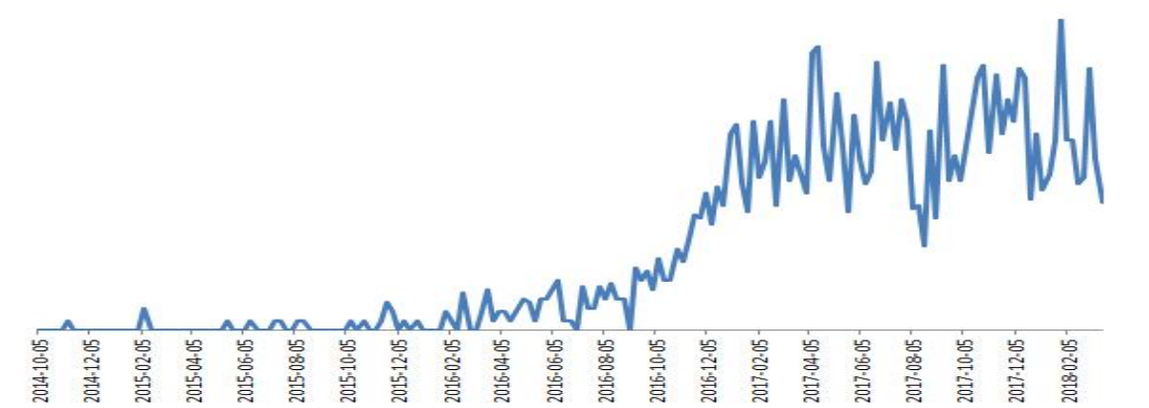
Ⅱ 주요 AI 알고리즘 트렌드

1 생성적 적대 신경망 (Generative Adversarial Networks)

가. 등장 배경

- GAN[1]은 대립하는 두 시스템이 서로 경쟁하는 방식으로 학습이 진행되는 비교사(Unsupervised) 학습 알고리즘으로 2014년 이안 굿펠로우(Ian J. Goodfellow)가 NIPS(Neural Information Processing System)[2]에 소개한 이후 학계의 관심이 증가하고 산업계 활용이 본격화되고 있음[3]
  - CNN, RNN 등은 이미지와 음성을 인식하지만 새로운 이미지와 음성을 생성하지는 못하고 대부분 지도학습 기반으로 활용되고 있다는 한계가 있음
  - GAN의 혁신성은 기존 딥러닝 알고리즘과 달리 비교사 학습 방식으로 스스로 이미지와 음성을 생성한다는 데 있음
  - 즉 현실 세계에서는 라벨링 된 데이터에 비해 라벨링 되지 않은 데이터가 대부분이며 정답이 없는 데이터가 훨씬 많기 때문

그림 2 GAN 키워드의 구글 트렌드 추이



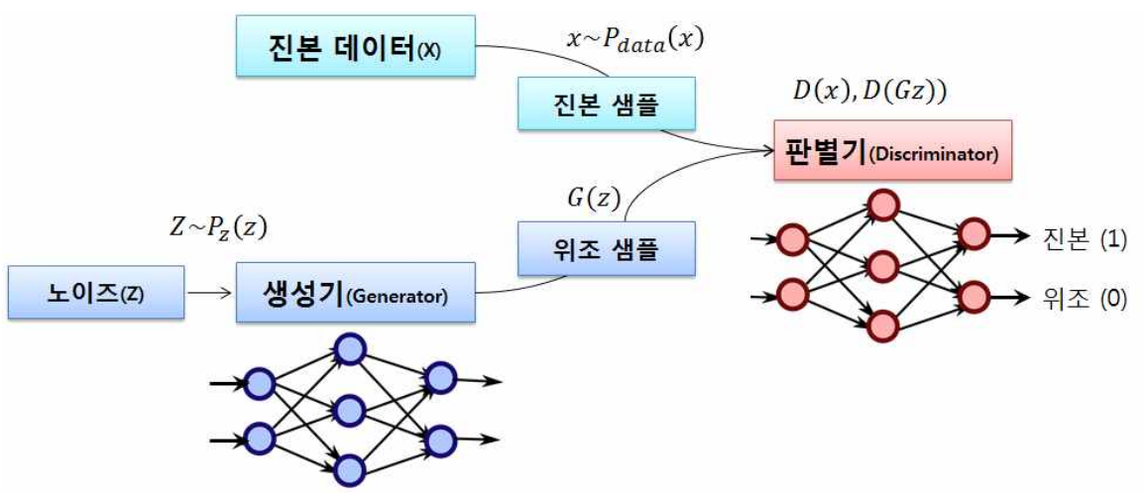
_출처 : Google Trend (2018)

1) Hinton, Yann LeCun, Andrew Ng 등은 GAN과 같은 비교사 학습모델이 딥러닝의 미래를 이끌 것으로 전망하고 GAN을 지난 10년 동안 가장 혁신적인 알고리즘이라고 평가
2) Ian J. Goodfellow et al., "Generative Adversarial Networks", Neural Information Processing Systems(NIPS) 2014.
3) 2014년 이후 지금까지 DCGAN, iGAN, StackGAN, CycleGAN, DiscoGAN, LAPGAN, BIGAN, EBGAN 등 200여 개의 GAN 변형들이 계속 발표되고 있음.

나. 구조 및 원리

- GAN은 생성기와 판별기로 구성된 서로 다른 주체가 적대적으로 경쟁하며 자신의 성능을 강화하는 과정을 통해 진본 데이터에 가까운 위조 데이터를 생성하는 원리
  - 즉 생성기에서는 임의의 분포로부터 위조 데이터를 생성하고, 판별기는 진본 데이터와 위조 데이터를 구별하기 위해 학습을 진행
  - 이 과정에서 생성기를 판별기를 최대한 잘 속이기 위해 노력하고 판별기는 진본 데이터와 위조 데이터를 최대한 구별하기 위해 경쟁적으로 학습
  - 결과적으로 GAN은 진짜와 같아지는 학습을 통해 사용자가 입력한 조건에 가장 가까운 샘플을 만들어 보다 생생한 데이터(이미지, 음성 등)를 생성할 수 있음

그림 3 GAN의 구조 및 원리



➔ $min_G\ max_D V(D,G) = E_{x \sim P_{data}(x)}[logD(x)] + E_{z \sim P_z(z)}[1 - logD(G(z))]$

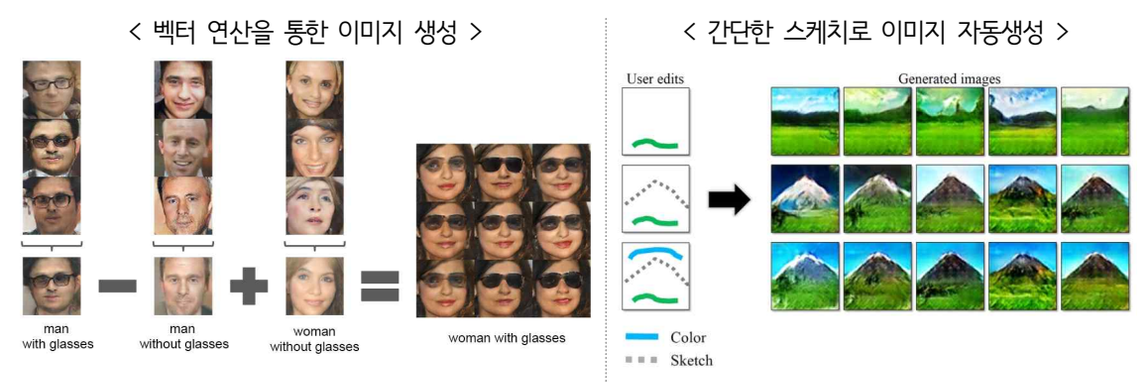
- 생성기 G는 가지고 있는 진본 데이터 x의 분포를 알아내려고 경쟁
- 만약 G가 정확히 진본 데이터 분포를 모사할 수 있다면 이로부터 추출한 샘플은 진본 데이터 분포의 샘플과 구분할 수 없음
- 판별기 D는 자신이 판별하려는 샘플이 생성기 G가 만든 위조 샘플인지 혹은 진본 데이터로부터 만들어진 진본 샘플인지 구별하여 각각의 경우에 대한 확률을 계산
- 판별기 D는 진본 데이터로부터 추출한 샘플 x의 D(x)=1이 되고, 생성기 G에 임의의 노이즈 분포로부터 추출한 z로 만들어진 샘플에 대해서는 D(G(z))=0가 되도록 경쟁
- 즉, D는 실수할 확률을 낮추기(min) 위해 경쟁하고 반대로 G는 D가 실수할 확률을 높이기(max) 위해 경쟁하는 minimax 문제임

_출처 : 이승민 작성

다. 연구 동향

- (이미지 생성) 다양한 형태의 이미지를 생성·조작하고, 간단한 스케치만으로 완전한 이미지를 만들거나, 텍스트를 읽고 이에 해당하는 이미지를 생성하는 것도 가능
  - (DCGAN: Deep Convolution GAN) 가장 대표적인 GAN기법으로 유명인, 동물, 자동차 등 다양한 현실 객체를 생성할 뿐 아니라 이미지 연산 조작이 가능[4)]
  - (iGAN: Interactive GAN)[5)] 간단한 스케치만으로 자동으로 이미지를 생성
  - (StackGAN: Stacked GAN)[6)] 입력된 문장과 단어를 이용하여 가상의 이미지를 생성하며 저해상도 이미지 생성 단계와 고해상도 이미지 생성단계로 구성됨

그림 4 DCGAN, iGAN, StackGAN을 이용한 이미지 생성



_출처 : A. Radford and L. Metz (2016), https://github.com/junyanz/iGAN, Zhang et al. (2017)

4) A. Radford and L. Metz, "Unsupervised representation learning with deep convolutional generative adversarial networks", ICLR(International Conference on Learning Representations) 2016.
5) iGAN(interactive GAN): https://github.com/junyanz/iGAN
6) Zhang et al., "StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks", ICCV(International Conference on Computer Vision) 2017.

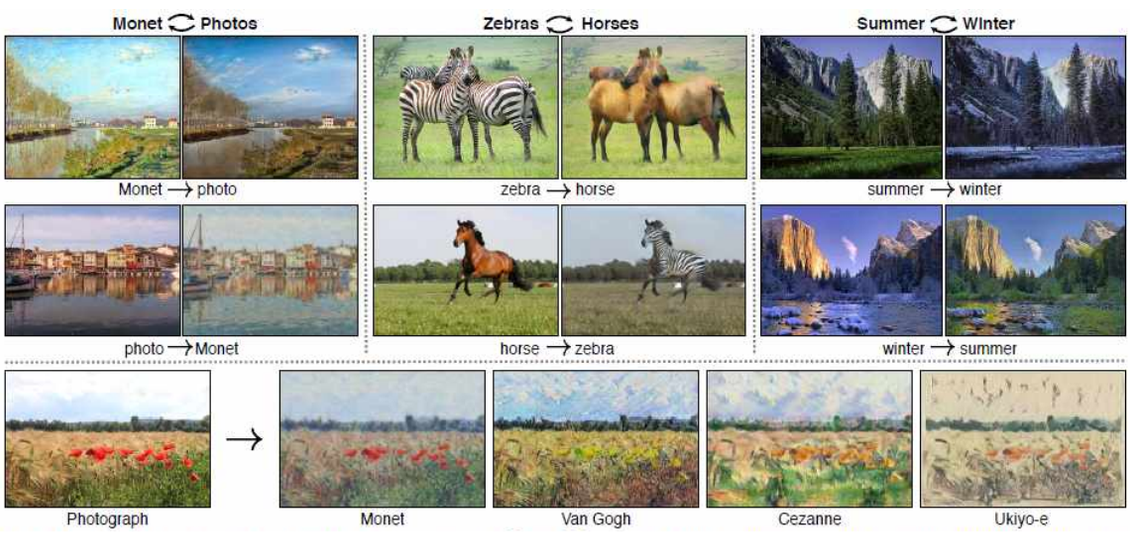
● (이미지 변환) GAN은 동일한 객체를 스타일이 다른 이미지로 변환하거나 유사한 스타일의 새로운 객체를 생성할 수 있음

- (CycleGAN: Cycle-consistent GAN)[7] 두 도메인의 데이터 쌍이 없는 상황에서 한 도메인에 있는 이미지를 다른 도메인의 이미지로 스타일을 바꿈

  * 이와 유사한 기존 논문에서는 변환하고자 하는 두 도메인에 대응하는 데이터 쌍(예: 고흐 그림과 이에 대응하는 동일한 풍경 사진)이 있어야 가능

  * CycleGAN에서는 두 개의 생성기를 사용하여 한 도메인에 있는 데이터를 다른 도메인으로 변환하고 다시 원래의 도메인으로 재변환하는 순환 일관성(Cycle Consistency) 개념을 제안

그림 5 CycleGAN을 이용한 이미지 변환



- 위: 모네 그림과 사진 간 스타일 변환(왼쪽), 얼룩말과 말 이미지 간 스타일 변환(가운데), 요세미티 여름과 겨울 사진 간 스타일 변환(오른쪽)
- 아래: 유명 화가 그림으로 학습한 후, 풍경 사진을 각 화가 스타일의 이미지로 변환

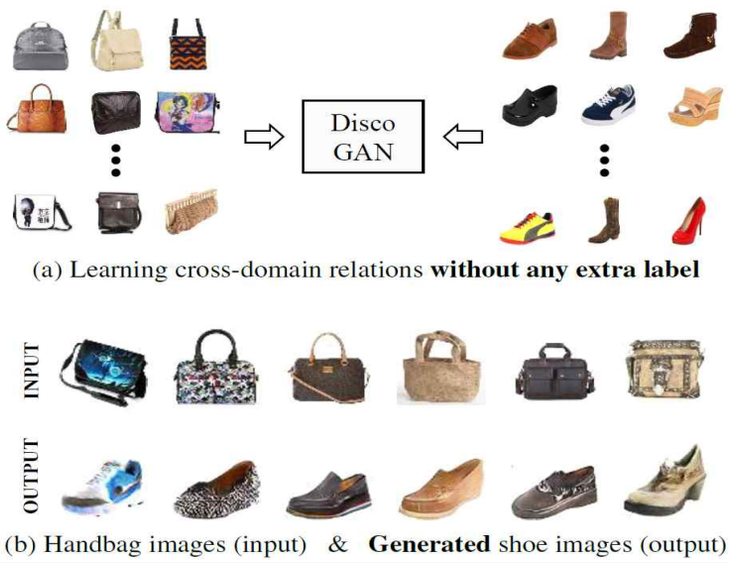
_출처 : Jun-Yan Zhu et al. (2018)

- (DiscoGAN: Discovery GAN)[8] 두 도메인에 있는 이미지의 관계를 발견하여 한 도메인의 객체(가방) 스타일을 유지한 채로 다른 도메인의 객체(신발)로 변환

  * 기존 GAN과 달리 DiscoGAN에서는 명시적으로 라벨링되지 않은 데이터를 가지고 서로 다른 도메인 간의 관계를 매핑하는 함수를 찾으려는 아이디어가 핵심

7) Jun-Yan Zhu et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks", arXiv:1703.10593v4, 2018.02.19.
8) Taeksoo Kim et al., "Learning to Discover Cross-Domain Relations with Generative Adversarial Networks", arXiv:1703.05192v2, 2017.05.15.

그림 6 DiscoGAN 학습 및 적용 결과



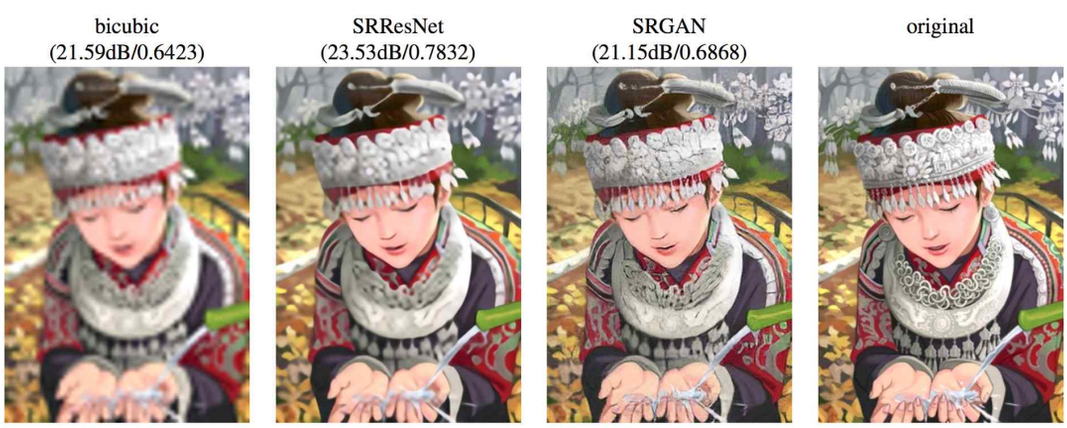
(a) Learning cross-domain relations **without any extra label**

(b) Handbag images (input) & **Generated** shoe images (output)

_출처 : Taeksoo Kim et al. (2017)

- (이미지 복원) SRGAN(Super-Resolution GAN)을 이용하면 낮은 해상도의 이미지를 높은 해상도의 이미지로 복원하는 것 또한 가능

그림 7 GAN을 이용한 이미지 복원과 생성[9]



- (bicubic) 기존에 사용하던 2차원 외삽법에 의해 해상도를 높인 이미지
- (SRResNet) 딥러닝 손실함수에 최적화된 SRRestNet을 이용해 해상도를 높인 이미지
- (SRGAN) GAN을 이용한 이미지 복원 결과로서 원본 이미지(original)에 가장 가까움

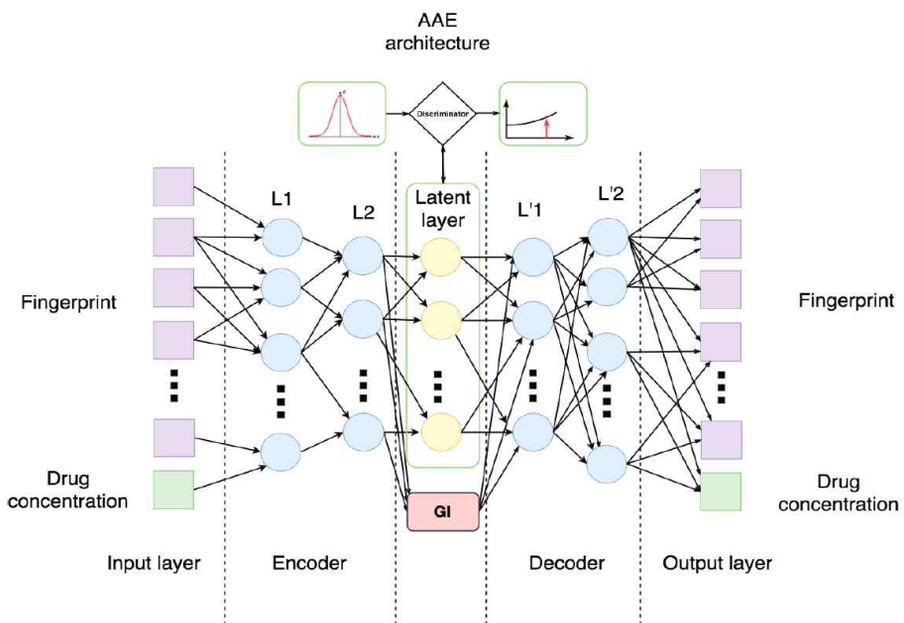
_출처 : C. Ledig et al. (2016)

9) C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network", NIPS(Neural Information Processing Systems) 2016.

- (패션 디자인) Amazon은 GAN을 사용해 패션 트렌드를 학습하여 사람 대신 새로운 패션을 디자인하고 이용자 패션 취향에 따른 개인 맞춤형 패션 제작 사업을 추진[10)]
  - GAN은 페이스북, 인스타그램 등에 게시된 사진으로 최신 패션 트렌드를 학습하여 새로운 패션을 스스로 디자인함
  - 이를 통해 Echo Look은 이용자의 이미지를 통해 개별 패션 취향을 이해하고 추천한 다음 GAN이 만든 주문형 옷을 이용자에게 판매하는 단계까지 수행
- (시설 점검) 일본 도시바는 송전선 점검을 자동화하기 위해 GAN 모델을 시스템으로 구현하여 시험용 철탑 등을 통해 실증실험, 그 효과를 확인함[11)]
  - 현재 낙뢰[12)]로 인한 송전선 보수 작업이나 정기 점검은 작업자가 철탑에 올라가 직업 확인하거나 헬리콥터를 이용해 송전선을 촬영하는 방식으로 진행
  - 그러나 높은 철탑에 직접 올라가서 육안으로 확인하는 것은 매우 위험하고 헬리콥터의 공중 촬영은 송전선과 철탑에 접근하기 위해 숙련된 인력과 많은 비용이 소요
  - 이를 해결하기 위해 무인 자율비행 드론으로 촬영한 방대한 사진을 CNN 등 딥러닝 기술을 이용하여 자동 검출하려는 방안이 제시됨
  - 이때 가장 큰 문제는 낙뢰 등으로 손상된 이미지 데이터량이 학습하기에는 절대적으로 부족, 이 점을 극복하기 위해 GAN 모델을 사용함
  - 즉 GAN 모델을 이용하여 인공적으로 낙뢰 등으로 손상된 비정상적인 이미지를 생성하고 이를 CNN 학습용 데이터로 활용함
- (신약 개발) 인실리코 메디슨(Insilico Medicine)은 딥러닝과 GAN 등 인공지능 기술을 활용하여 신약개발 기간을 10년에서 3년으로 단축시킬 수 있을 것이라 주장
  - 2017년 발표된 논문[13)]에서 GAN과 Autoencoder를 이용하여 항암치료 후보물질을 찾는 과정을 소개, 구체적으로 생성기는 항암 속성을 가진 새로운 분자들을 만들고, 판별기는 기존 치료법을 기반으로 새로 만든 분자가 적절한 지를 판별
    * 실험 결과 72,00만 가지 화학 물질에서 GAN 판별기를 통해 신약 후보 물질을 제시하였고 이 가운데 특허 받은 항암제가 60가지 포함되었음을 확인

10) Will Knight, “Amazon has developed an AI fashion designer”, MIT Technology Review, 2017.08.24.
11) Nikkei Robotics, “東芝が送電線点検に新種のディープラーニング、“知能”より“創作者”と呼ぶべき新AI「生成モデル」”, 2017.04.
12) 도쿄전력그룹의 관할하는 송전선 길이는 2015년 말 기준으로 약 5,000km, 철탑 수는 5만개 이상이며 연간 수 백회의 낙뢰가 발생하는 것으로 보고됨
13) Arur Kadurin et al., “The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology”, Oncotarget, Vol. 8, 2017.

그림 8 GAN과 Autoencoder를 이용한 신약 후보물질 탐색 구조



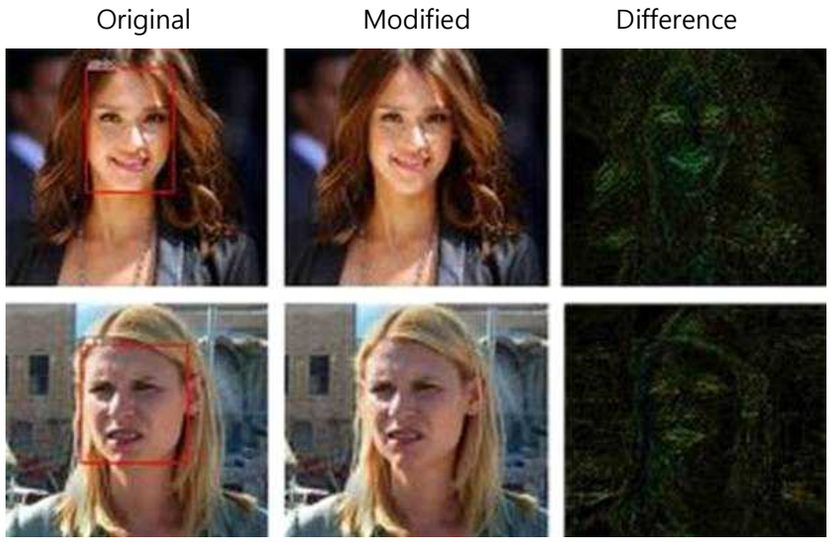
Encoder consists of two consequent layers L1 and L2 with 128 and 64 neurons, respectively. In turn, decoder consists of layers L'1 and L'2 comprising 64 and 128 neurons. Latent layer consists of 5 neurons one of which is Growth Inhibition percentage (GI) and the other 4 are discriminated with normal distribution.

_출처 : Arur Kadurin et al. (2017)

- (프라이버시보호) 캐나다 토론토대 Parham Aarabi 교수팀은 GAN을 이용하여 소셜 미디어나 인터넷에 올린 얼굴 사진을 미세하게 바꿔 기존 얼굴 인식 시스템을 속일 수 있는 프라이버시 필터 기술을 개발[14)]
  - 연구결과에 따르면 실내외에서 촬영한 다양한 인종의 600장 이상의 얼굴 사진을 이용하여 실험한 결과, 기존 얼굴 인식 시스템의 정확도를 100%에서 0.5%로 떨어뜨려 얼굴인식을 무력화시킬 수 있음
  - 그러나 이 기술이 프라이버시보호에 사용될 수 있다는 긍정적 측면 외에 자신을 감추거나, 타인의 모습을 교묘히 변형하여 악용하는 수단으로 활용 가능성이 있음
  - 나아가서 GAN을 이용한 가짜 사진과 동영상 생성이 용이해짐에 따라 인공지능의 위험성이 더욱 높아지면서 진짜 같은 가짜 콘텐츠로 인한 사회 혼란이 심각해질 것이라는 우려 또한 증가

14) https://www.utoronto.ca/news/u-t-ai-researchers-design-privacy-filter-photos-disables-facial-recognition-systems

그림 9 GAN을 이용한 프라이버시 보호



_출처 : https://www.utoronto.ca/news/u-t-ai-researchers-design-privacy-filter-photos-disables-facial-recognition-systems (2018)

라. 의미와 전망

- GAN은 인공지능이 '수동적 인식'에서 '능동적 생성'으로 활용의 가능성을 한 단계 끌어올리면서 '지능' 보다 '창작자'로서 새로운 가능성을 제시
  - 현재 200 여개에 달하는 GAN 변형 기술은 딥러닝이 안고 있는 비교사학습과 라벨링되지 않은 데이터 활용 문제를 해결할 수 있는 돌파구로 인식
  - GAN은 스스로 새로운 지식과 경험을 축적할 수 있는 방향으로 진화할 수 있는 가능성을 확인시켜 주었고 향후 산업적 활용 가능성이 매우 클 것으로 기대
  - 그러나 GAN을 이용한 딥페이크(Deep Fake) 등 가짜 콘텐츠 생성으로 인한 심각한 사회 문제를 발생시킬 것에 대한 부작용 우려 또한 증가

- GAN은 데이터 생성이 용이하고 스스로 새로운 모형을 생성할 수 있다는 특별한 장점을 지니고 있으나 생성기와 판별기 간 성능 불균형이 큰 경우 학습이 매우 어렵다는 기술적 한계가 있음
  - GAN은 최소최대화(minimax) 문제를 기반으로 두 네트워크를 경쟁적으로 학습시켜 새로운 데이터를 생성하는 데 우수하나 최소최대화의 근본적인 문제가 존재
  - 이를 해결하기 위한 다양한 기법이 제시되고 있으나 아직까지 생성적 적대 문제를 완벽하게 해결할 수 있는 방안은 없으며 지속적인 연구 필요

2 심층강화학습 (Deep Reinforcement Learning)

가. 등장 배경

- 강화학습(Reinforcement Learning)은 2016년 AlphaGo와 이세돌 바둑 대결 이후 본격적인 관심을 받기 시작하였으며 2017년 MIT 10대 혁신 기술 중 하나로 선정
  - 강화학습은 특정 환경에서 정의된 에이전트가 현재의 상태를 탐색하고 보상을 최대화하는 행위를 선택하며 스스로를 개선해 나가는 기계학습의 한 영역
  - 강화학습은 반복적인 시행착오를 통해 문제를 해결하는 방법을 스스로 터득한다는 점에서 기존 지도학습, 비지도학습의 정적인 학습 방식과는 다름
- 심층강화학습은 심층학습(Deep Learning)과 강화학습(Reinforcement Learning)을 결합한 기술로서 기존 강화학습의 한계를 극복
  - 현재의 심층강화학습은 단순한 게임 분야를 넘어 최근 자율주행자동차, 로보틱스 등 잘 정의되지 않는 복잡한 현실 분야에서 적용 가능성을 확인하는 단계

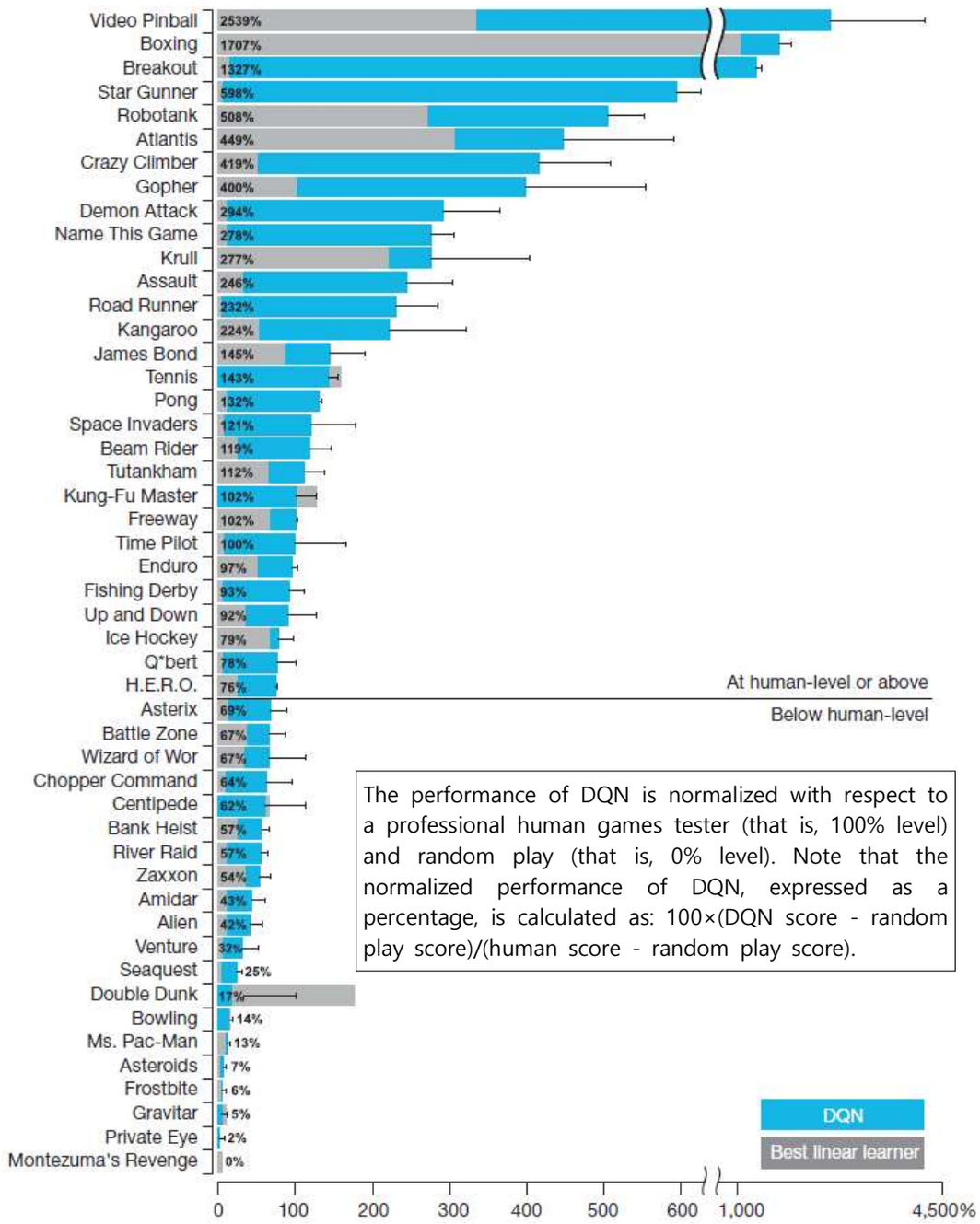
◎ 강화학습 vs. 심층강화학습
- 기존 강화학습은 게임과 같이 현재의 상태와 보상이 잘 정의된 환경에서 매우 뛰어난 성능을 보이나 현실 세계에서는 다양한 상태와 행위에 따른 보상을 정의하기 어렵기 때문에, 결국 에이전트 스스로 이를 이해하고 판단하여 행동해야 한다는 문제가 있음
- DeepMind의 업적은 딥러닝과 강화학습을 결합한 심층강화학습을 통해 기존 강화학습의 한계를 극복하고 다양한 현실 문제에 적용할 수 있다는 가능성을 제시했다는 점

나. 연구 동향

- DeepMind는 DQN(Deep Q-Network)이라는 심층강화학습을 개발하여 2013년 Atari Breakout 게임에 적용[15] 이후 알파고에 이르기까지 놀라운 성과를 보여줌
  - Deepmind는 Breakout 게임 외에도 2015년 49종류의 게임에서 인간 수준 이상의 능력을 보인 결과를 Nature지에 발표하고 심층강화학습의 가능성을 제시[16]
  - 2016년 AlphaGo와 이세돌 간 바둑 대결은 DQN 심층강화학습의 응용 가능성과 기존 알고리즘의 확장 가능성을 보여준 결정적 계기
  - AlphaGo에 적용된 DQN은 사람의 개입 없이 바둑 AI 끼리의 대국을 통해 스스로 데이터를 만들고 높은 점수를 획득할 수 있는 행동 패턴을 스스로 학습

15) Volodymyr Mnih et al., "Playing Atari with Deep Reinforcement Learning", arXiv:1312.5602, 2013.12.19.
16) Volodymyr Mnih et al., "Human-level control through deep reinforcement learning", Nature, 2015.02.26.

그림 10 DQN와 기존 강화학습 결과 비교[17]



_출처 : Volodymyr Mnih et al. (2015)

17) Volodymyr Mnih et al., "Human-level control through deep reinforcement learning", Nature, 2015.02.26.

- 심층강화학습은 게임 분야 외에 충분한 주행 데이터 확보가 어려워 학습이 어렵거나 시간이 많이 소요되는 자율주행 기술 구현에 본격적으로 활용되기 시작
  - 자율주행을 학습하기 위해 모든 주행 상황을 고려한 학습 데이터를 수집하거나 실제 주행 환경을 완벽히 재현한다는 것은 불가능하기 때문
  - 강화학습은 다양한 주행 데이터를 반복적으로 미세 조정하여 수백만 번의 주행 상황을 재현함으로써 매우 효과적으로 학습모델을 생성할 수 있음
    * 완성차 업체 중심으로 센서와 딥러닝 등 기존 방법과 함께 강화학습을 보완적으로 적용 연구
  - PFN(Preferred Networks)은 CES 2016에서 심층강화학습을 적용하여 자동차의 속도와 방향 등을 고려하여 개발한 자동주행 시뮬레이터를 소개
    * PFN이 개발한 기술은 병렬처리 방식으로 다수의 신경망을 동시에 학습하기 때문에 실제 주행 데이터를 수집하여 학습하는 것보다 100만 배 빠른 속도를 보인다고 설명
  - 게임 개발자가 사용하는 게임 제작 엔진 Unity를 만든 기업 Unity Technologies는 최근 자율주행자동차 분야에서 심층강화학습을 통해 가상현실에서 주행시험을 할 수 있도록 Unity 엔진을 적용, 글로벌 자동차 업체와 협력 진행 중
    * Unity Technologies는 심층강화학습으로 주행성능을 높이기 위한 '데이터 생산 역량'을 보유, 게임 외 자동차, 건설, 제조업, 영향 등 다양한 산업으로 적용 계획

- 심층강화학습은 복잡한 실제 상황을 이해하고 스스로 최적의 행위를 수행하는 로봇 분야의 제어 및 동작 등의 연구에 활발히 적용 중
  - 심층강화학습은 상태와 행위, 보상 등을 명확히 정의하기 어려운 환경에서 제한된 데이터만 제공되는 경우에도 매우 효과적으로 학습모델을 생성
  - 로봇의 물리적 제어뿐 아니라 인간의 표정과 행동을 통해 인간과 교감하는 로봇의 감성적 분야에서도 적용 가능성이 기대됨
  - DeepMotion 연구팀은 심층강화학습을 이용하여 사람과 같은 동작을 하는 아바타 생성 기술을 개발하고 향후 에니메이션, 로봇 동작 등의 구현 가능성을 제시[18)]
  - NVIDIA는 제조, 물류, 농업 등 다양한 산업 분야에 활용할 수 있도록 차세대 자율로봇 개발 플랫폼 아이작(Isaac)을 공개[19)]
    * 아이작에는 젯슨 자비에(Jetson Xavier) 로봇 프로세스가 포함되어 있으며 심층강화학습 등 인공지능 알고리즘 SW 개발 툴을 제공

18) Libin Liu et al, "Learning Basketball Dribbling Skills Using Trajectory Optimization and Deep Reinforcement Learning", ACM Transactions on Graphics, 2018.08.
19) Dean Takahashi, "Nvidia launches Isaac robot platform with Jetson Xavier robot processor", VentureBeat, 2018.06.04.

그림 11 NVIDIA의 아이작 및 젯슨 자비에

_출처 : NVIDIA (2018), https://www.nvidia.com/en-us/autonomous-machines/

다. 의미와 전망

- 심층강화학습은 영상, 음성 데이터를 넘어 시행착오를 통해 행동 데이터를 학습, 딥러닝의 활용 가능성을 획기적으로 확장시킬 수 있다는 측면에서 큰 의의가 있음
  - 특히 게임과 같이 보상이 확실한 분야뿐 아니라 학습에 필요한 충분한 데이터를 구하기 어렵거나 현실적으로 재현이 어려운 경우 매우 효과적으로 적용 가능
    * 데이터가 부족하면 인간과 기계의 인지과정의 구조적 차이로 인해 학습모델 생성에 한계가 존재하기에 심층강화학습과 함께 생성적 적대 신경망, 전이학습 등이 주목받고 있음

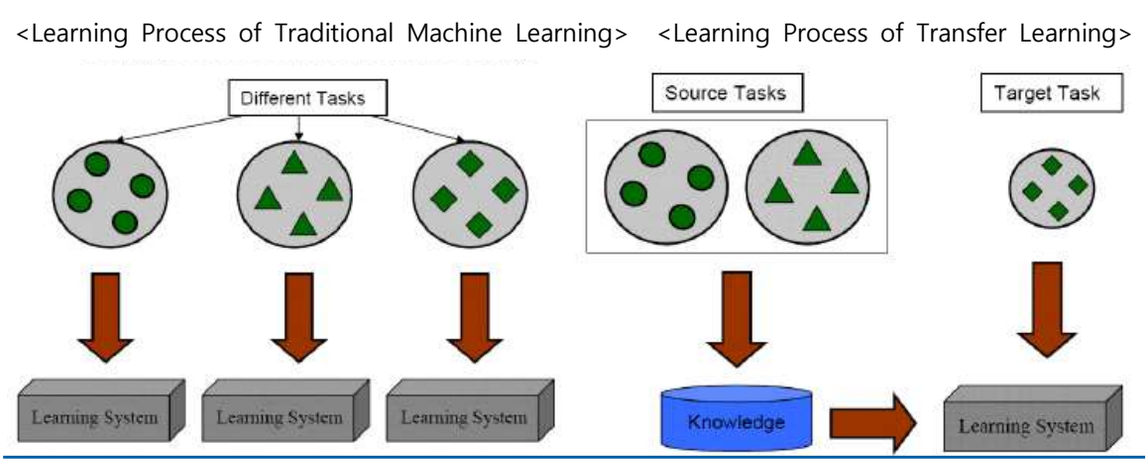
- 최근 자율주행 분야에서 심층강화학습이 활발히 연구되고 있으며 점차 다양한 로봇에 적용되어 보다 복잡한 환경에 적용되어 인간의 작업을 자연스럽게 모사할 수 있는 기술로 발전할 것으로 전망됨
  - 심층강화학습은 복잡한 현실세계에서 충분한 학습데이터를 구하기 어렵고 명확한 조건을 정의하기 어려운 상황에서 가상데이터를 만들어 시행착오를 통해 스스로 배우고 개선해 나간다는 점에서 앞으로 활용 가치가 크게 기대됨
  - 나아가서 성능이 입증된 딥러닝 기반 영상인식 기술 등과 함께 사용되어 보다 복잡하고 위험한 환경에서 사람처럼 행동하는 에이전트를 구현할 수 있음
  - 뿐만 아니라 전이학습(Transfer Learning) 등과 결합하여 심층강화학습으로 생성한 지식을 다른 분야에 재활용하는 연구도 활발히 전개될 전망

3 전이학습 (Transfer Learning)

가. 등장 배경

- 기존에 학습한 지식을 재사용할 수 있는 인공지능 알고리즘은 1995년 NIPS에서 'Learning to Learn'의 이름으로 주목받기 시작
  - 이후 전이학습(transfer learning)[20], 생애학습(life-long learning), 메타러닝(meta learning) 등 다양한 용어가 등장
  - 전이학습은 사전에 학습이 완료된 모델을 활용하여 새로운 모델을 생성하는 방법으로 Google의 Inception 모델 활용과 DeepMind의 PathNet 등이 대표적
  - 인간의 평생학습 개념과 유사하게 새로운 환경에 적응하여 학습하는 알고리즘 연구로 DARPA의 L2M(Lifelong learning Machines)[21] 또한 주목

그림 12 기존 머신러닝과 전이학습 비교



_출처 : Sinno Jialin Pan and Qiang Yang (2010)

나. 연구 동향

- 전이학습은 특정 분야에서 만든 학습모델을 이와 유사한 분야에서 재사용함으로써 학습 데이터가 부족한 경우에 빠르고 효과적으로 학습모델을 생성할 수 있음
  * 고양이에 대한 데이터가 많고 개에 대한 데이터가 적을 경우, 개를 인식하는 모델을 만드는 데 고양이를 인식하도록 학습한 모델을 사용
  - 전이학습에서 지식의 재활용이 가능한 이유는 딥러닝의 은닉계층 가운데 하위 계층이 상위 계층에 비하여 상대적으로 보다 기초적인 지식을 제공하기 때문

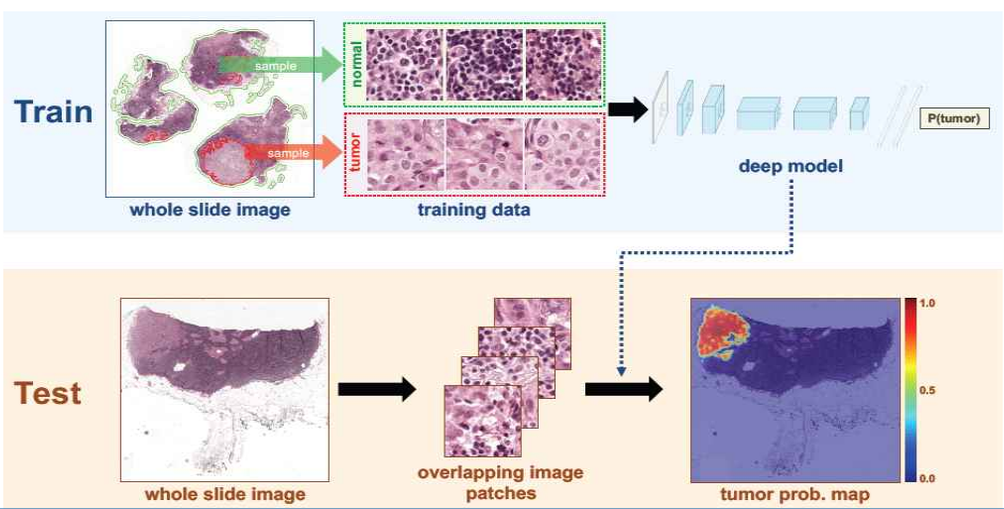
20) Sinno Jialin Pan and Qiang Yang, "A Survey on Transfer Learning", IEEE Transactions on Knowledge and Data Engineering, Vol. 22, Issue 10, 2010.10.
21) Hava T. Siegelmann, "Lifelong Learning Machines (L2M)", DARPA, 2017.04.26.

- 하위 계층의 지식을 재활용하기 위해 상위 계층을 재학습하는 방법(파라미터 미세 조정)이 주로 사용되고 있음

● Google의 Inception 모델에 전이학습을 통한 실험 결과 유방암, 피부암, 당뇨성 망막변성 등 의료영상 판독에 전문의 수준과 비슷하거나 능가하는 성능을 보임

- Camelyon16 대회에서 Andrew Beck 교수팀은 유방암 병리 슬라이드 판독을 위해 Google Inception 모델을 전이학습시켜 7.5% 에러율로 우승[22)]
- 유사한 방식으로 전이학습한 결과 피부암[23)]과 당뇨성 망박변성[24)] 진단에서 매우 높은 정확도를 보였고 제한된 실험에서 인간 전문의 수준에 근접

그림 13 Andrew Beck 교수팀의 딥러닝 모델



_출처 : Dayong Wang (2016)

● Stanford대 연구진은 인공위성이 촬영한 데이터를 이용하여 아프리카 국가들의 빈곤지도를 생성하는 데 전이학습을 활용[25)]

- 현실적으로 빈곤 데이터는 매우 부족한 상황이며 본 논문에서는 야간에 촬영한 조명 데이터를 부의 분포 분류 지표로 사용함
- 연구진은 풍부한 야간 조명 데이터를 활용한 전이학습 모델과 고해상도의 주간 위성 데이터를 결합하여 빈곤 지역을 예측할 수 있는 알고리즘을 개발

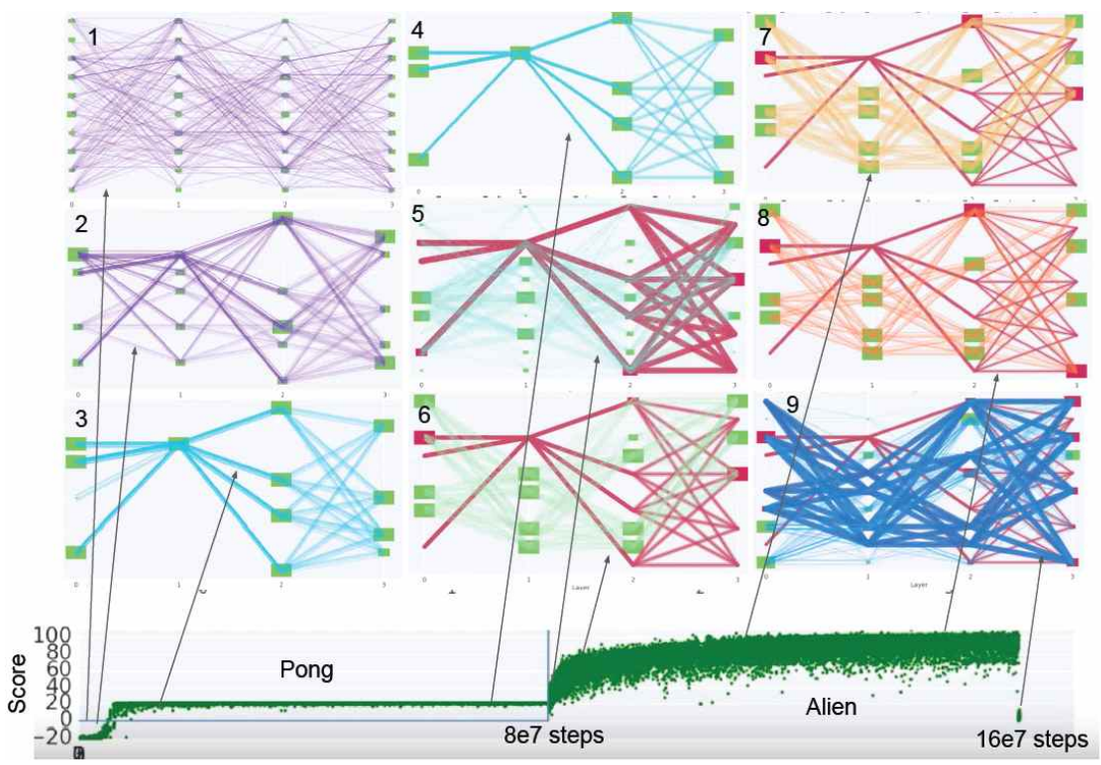
22) Dayong Wang et al., "Deep Learning for Identifying Metastatic Breast Cancer", arXiv: 1606.05718v1, 2016.06.18.
23) Andre Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks", Nature 542, pp.115-118, 2017.02.02.
24) Varun Gulshan et al., "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs", JAMA, 2016.11.29.
25) Michael Xie et al., "Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping", arXiv:1510.00098v2, 2016.02.27.

● (DeepMind PathNet) DeepMind는 학습한 신경망에서 재사용 Path가 있다고 가정하고 이를 찾아내서 재활용할 수 있는 전이학습 알고리즘 PathNet을 발표[26)]

- 주목할 점은 기존 전이학습이 신경망의 하위 계층을 재사용하는 관점(수평)이라면 PathNet에서는 모든 계층에서 특정 Path를 찾아내서(수직) 재사용
- PathNet에서 task A(재활용 부분)를 학습시킨 후 task B(새로운 부분)를 학습시켰을 때가 task B를 처음부터 학습하는 것보다 학습 속도가 빠름

* 논문에서는 PathNet을 분산처리 가능한 강화학습 기법인 비동기(Asynchronous) 강화학습에 적용하여 Breakout 게임과 Labyrinth 게임에 실험한 결과, 학습 속도가 매우 빨라졌음을 확인

그림 14 PathNet 학습 과정



- Task A(그림의 Pong)를 정확도가 100%될 때까지 학습시킨 후, 학습한 신경망 모델에서 유전 알고리즘을 사용하여 best path(그림의 3번)를 찾아냄
- best path에 해당하는 파라미터에 대해서는 더 이상 학습하지 않고(fix) 다른 부분은 랜덤 초기화하여 Task B(그림의 Alien)를 재학습하여 새로운 path(그림의 9번)를 찾음

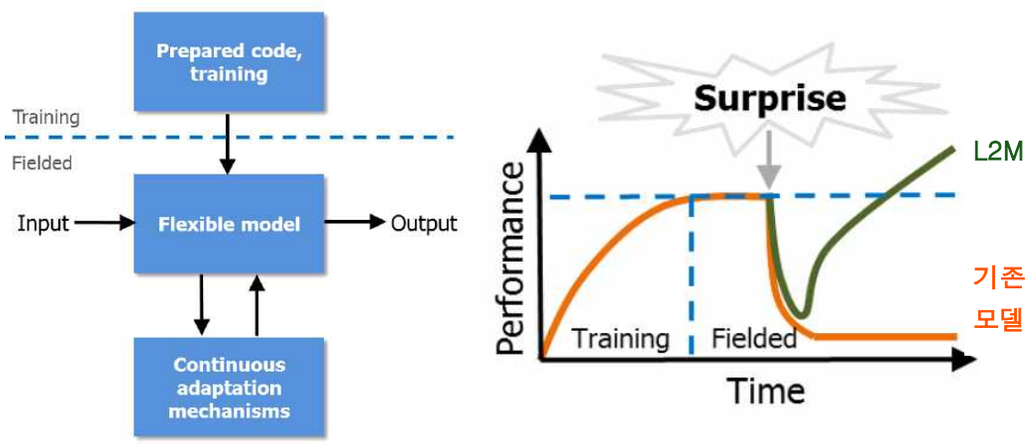
_출처 : Chrisantha Fernando (2017)

26) Chrisantha Fernando, "PathNet: Evolution Channels Gradient Decent in Super Neural Networks", arXiv:1701.08734, 2017.01.30.

● (DARPA L2M) 2017년 DARPA는 인간과 같이 항상 배우고 적응하는 기계를 개발할 목적으로 새로운 환경에 적응하면서 학습 모델을 지속적으로 학습하는 L2M 연구 프로젝트를 시작

- 본 프로젝트는 향후 4년 동안 약 6,500만 달러가 투자되어 16개 그룹에서 진행될 것이며 자율주행자동차, 사이버보안 등 안전을 강화하는 분야와 게임, 의료 등에서 폭넓게 적용될 것으로 기대됨

그림 15 DARPA L2M 개념도



_출처 : Hava T. Siegelmann (2017)

다. 의미와 전망

● 기존 인공지능 알고리즘은 특정 작업에 맞게 학습시킨 결과를 다른 작업에서 재사용할 수 없기 때문에 새로운 모델 만들기 위해서는 처음부터 학습과정을 다시 수행해야 하는 근본적인 문제가 있음

- 즉 기존 방식은 새롭게 학습하는 과정에서 학습된 결과를 버릴 수밖에 없고 이것은 인간이 직관성을 가지고 학습한 내용을 자연스럽게 응용하는 것과는 상이

● 학습한 결과를 학습하는 전이학습은 지식의 이전을 통해 인간이 가진 적응력과 유연성을 가진 인공지능을 만들 수 있다는 점에서 향후 큰 발전이 기대됨

- 이러한 점은 학습이 필요한 분야의 데이터가 부족한 경우에 상대적으로 풍부한 데이터가 있는 유사분야에서 학습한 결과를 활용할 수 있는 장점이 있음
- 또한 적용분야에 따라 기존 학습모델을 재사용함으로써 학습 시간을 현저히 단축시키면서 정확도를 보장할 수 있기에 향후 활발한 연구가 진행될 전망

4 설명가능 인공지능 (Explainable AI)

가. 등장 배경

- 최근 딥러닝이 금융, 의료 등 다양한 분야로 확산되고 산업적 가능성을 인정받기 시작하면서 학습 결과에 대한 신뢰성과 도출 과정의 타당성을 확보하려는 요구가 증가하고 있음
  - 사람의 생명과 관련 있는 의료, 자동차 등의 분야와 기업의 의사결정 프로세스 분야를 중심으로 인공지능 알고리즘의 투명성 보장을 위한 기술적, 법적 요구 증가
  - 기술적 대응으로는 2017년 DARPA의 XAI(Explainable AI) 프로젝트를 계기로 설명 가능한 인공지능 알고리즘 기술 개발이 본격적으로 전개됨[27)]
  - 또한 제도적 관점에서 2018년 EU의 GDPR(General Data Protection Regulation) 규제 조항 마련이 설명 가능한 알고리즘 개발 요구를 강화시키는 기폭제로 작용

◎ GDPR(General Data Protection Regulation)[28)]

- 최근 EU의 GDPR(General Data Protection Regulation) 규정 준수는 인공지능 알고리즘의 신뢰성고 투명성 제고를 위한 기술적 대응책 마련의 필요성을 증가시킴
- GDPR은 2015년 5월 EU 내에서 개인 데이터의 자유로운 이동과 보호 기능을 강화하는 한편 EU 외부로 개인 데이터 반출에 대한 문제를 다루기 위해 제정
- 이후 GDPR은 2018년 5월 25일부터 데이터를 다루는 기업의 데이터 프로세스까지 확대하고 위반 시 강력한 금전적 책임을 부과
- GDPR에서는 글로벌 회사를 포함(3조1항)하여 모든 EU 국가에 적용되고 규정 위반 시 전 세계 매출의 4% 또는 2,000만 유로 중 큰 금액을 패널티로 책정(제83조제5항)
- 이와 같은 데이터 처리과정의 투명성에 대한 규정이 강화된 배경에는 데이터 기반 글로벌 기업들의 개인 데이터 독점과 무분별한 사용 특히 고의적인 편향된 알고리즘으로 인한 사회적 위험에 대한 우려의 영향
- 알고리즘의 투명성을 제고하기 위해 '설명을 요구할 권리'와 '자동화된 의사결정을 제한할 권리' 등을 규정한 GDPR은 향후 설명 가능한 인공지능 알고리즘에 대한 기술 개발의 요구를 한층 강화시킬 전망

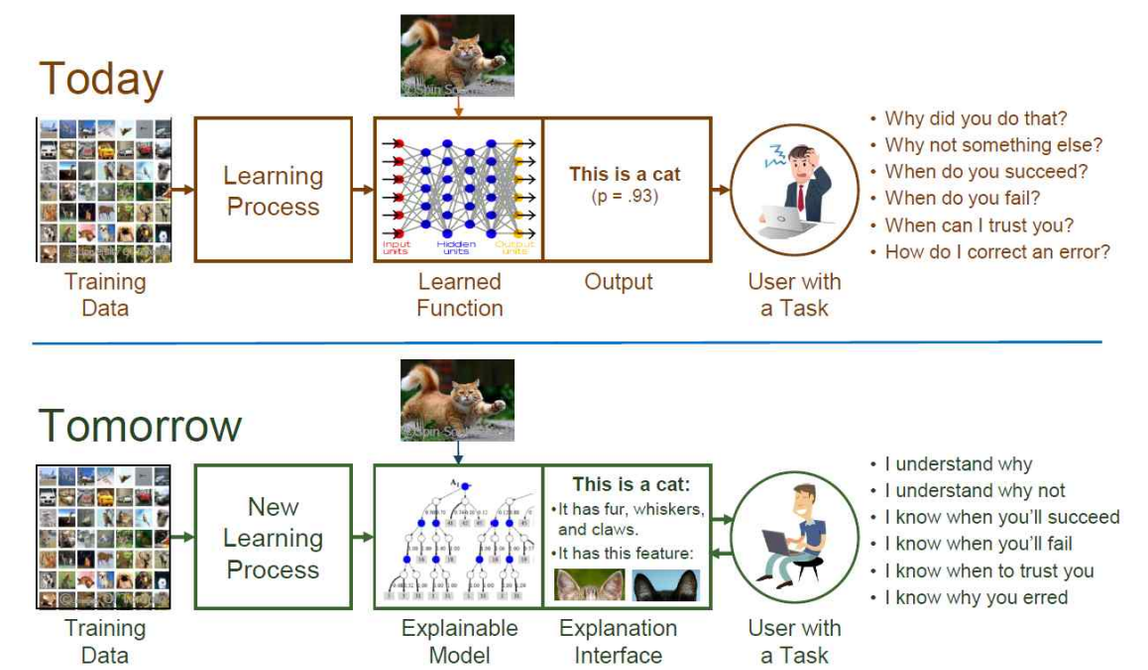
27) David Gunning, "Explainable Artificial Intelligence(XAI)", DARPA/I2O Program Update, 2017.11.
28) GDPR, https://www.eugdpr.org/

나. 연구 동향

● DARPA에서는 2017년부터 사용자가 인공지능 알고리즘의 작동 과정을 이해하고 판단 결과에 대한 이유를 설명할 수 있는 기술 개발을 추진

- 예를 들어 고양이 이미지에 대한 인식의 결과만 알려주는 딥러닝과 달리 XAI에서는 고양이라고 판단한 근거(수염, 털, 발톱 등)를 제시할 수 있어야 한다는 것
- DARPA의 XAI 프레임워크는 크게 '설명 가능한 모델'과 '사용자 인터페이스 개발'을 포함하고, 특히 '설명 가능한 모델'은 크게 딥러닝을 개선해 설명 가능한 특징값을 학습할 수 있는 기술(Deep Explanation), 결과 도출 과정을 해석할 수 있는 모델(Interpretable Model), 모델 추론 (Model Induction) 등의 방법을 제시

그림 16 DARPA XAI 프레임워크



_출처 : David Gunning (2017)

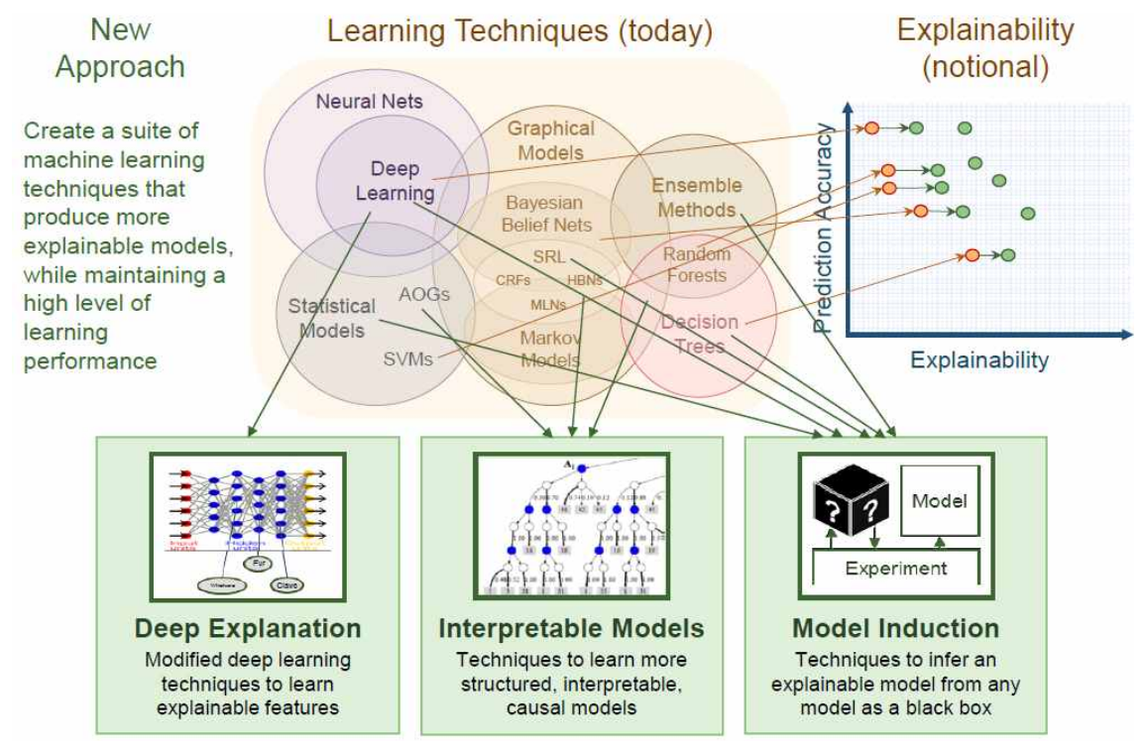
① (Deep Explanation) 설명 가능한 특징값을 학습할 수 있도록 기존 딥러닝 알고리즘을 변형하거나 결합하는 방향으로 접근

- 딥러닝의 각 은닉계층의 노드에 의미 있는 속성(예: 고양이, 개 등의 발톱, 콧수염 등)을 연결하여 학습하여 분류 결과에 대한 근거를 제공[29)]

29) Hui Cheng et al., "Multimedia Event Detection and Recounting", SRI-Sarnoff AURORA at

- 기존 CNN 알고리즘은 이미지내의 객체를 인식하도록 학습하고 RNN 알고리즘은 CNN으로 학습한 특징값을 단어와 갭션으로 번역하도록 학습함으로써 이미지 캡션을 생성[30]

그림 17 Explainable Model의 세 가지 접근 방법



_출처 : David Gunning (2017)

② (Interpretable Model) 입력 데이터(문자, 이미지 등)의 특징값을 확률적 추론을 통해 학습하여 새로운 모델을 개발하는 방향으로 접근

- 사람이 하나의 예만으로 새로운 개념을 배워서 사용할 수 있는 것처럼 사람이 개념을 익히는 과정을 베이지안 방법(Bayesian Program Learning)으로 추론하여 필기체 인식에 적용[31]
- 이미지의 특징값(색, 선, 위치 등)의 관계를 AND-OR 그래프로 표현하여 최종 결과에 이르는 과정과 확률을 제공하여 원인과 결과를 이해할 수 있도록 함[32]

TRECVID 2014.

30) Hendricks et al., "Generating Visual Explanations", arXiv:1603.08507v1, 2016.03.28.

31) Brenden M. Lake et al, "Human-level concept learning through probabilistic program induction", Science, Vol. 350, p. 1332-1338, 2015.

32) Zhangzhang Si and Song-Chun Zhu, "Learning AND-OR Templates for Object Recognition

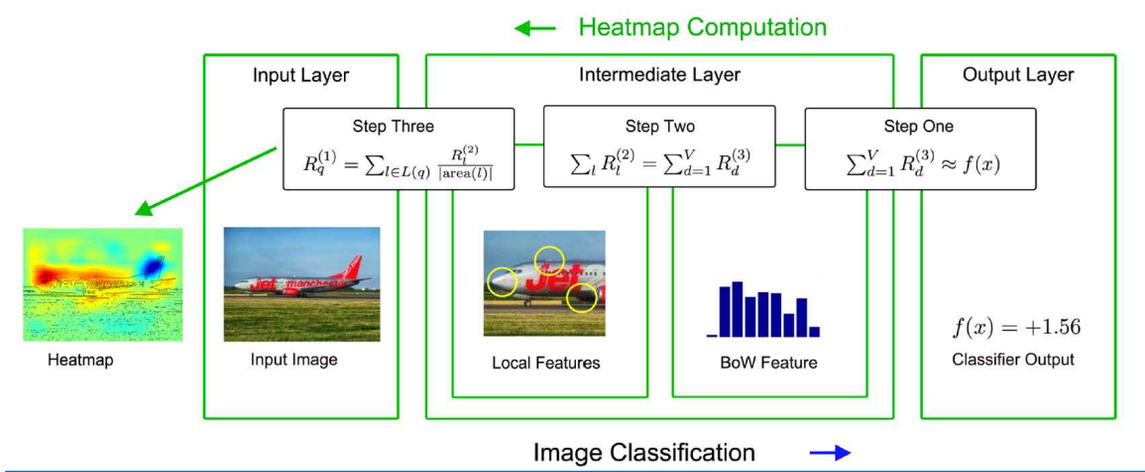
③ (Model Induction) 타 학습모델과 비교하여 예측결과에 대한 근거를 제시하거나 결정과정을 통해 설명가능 모델을 생성

- 기존 블랙박스 속성의 머신러닝 알고리즘을 설명 가능한 타 학습 모델과 비교하여 예측한 결과에 대한 근거를 제공[33]
- 기존 고차원의 특징값을 베이지안 룰(Bayesian Rule Lists)을 사용하여 단순하고 연속적인 결정과정으로 구분하여 인간이 이해할 수 있는 예측 모델을 생성[34]

● 계층적 상관성 전파(LRP: Layer-wise Relevance Propagation)[35] 모델은 신경망의 각 계층별 기여도를 측정하여 시각화함

- 각 계층의 기여도를 역전파 방식으로 열지도(Heatmap) 형태로 시각화하여 직관적으로 이해할 수 있도록 표현
- 즉 열지도로 표현된 기여도를 통해 입력 이미지의 어떤 부분이 결과 도출에 영향을 미쳤는지 직관적으로 알 수 있기에 의료분야의 질병진단 판단에 근거를 제시함으로써 유용한 정보를 제공할 수 있을 것으로 기대

그림 18 LRP를 이용한 입력 시각화 예



_출처 : Sebastian Bach et al. (2015)

and Detection", IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 35 No. 9, p. 2189-2205, 2013.

33) Marco Tulio Ribeiro et al., ""Why Should I Trust You?" Explainable the Predictions of Any Classifier", CHI 2016 Workshop on Human Centered Machine Learning, 2016.02.16.

34) Benjamin Letham et al., "Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model", Annals of Applied Statistics, Vol. 9, No. 3, 1350-137, 2015.

35) Sebastian Bach et al., "On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation", PLoS One, Vol. 10, No. 7, e0130140, 2015.06.10.

- 인과관계 모델[36]은 클래스 분류 문제를 속성(예: 줄무늬, 점박이, 털, 발 개수 등)과 슈퍼카테고리(예: 개, 말, 고양이, 육식동물 등)와 함께 선형 결합할 수 있도록 학습하여 예측 결과를 논리적으로 인관관계를 설명함
- 또한 금융분야, 제조분야 등 시계열 데이터에 대한 예측 결과의 근거를 제시하기 위해 시계열 분석 모델로부터 발견한 커널 조합을 자연어로 작성하여 보고서를 자동 생성하는 연구 등도 진행됨[37]

다. 의미와 전망

- 설명가능 AI 모델은 사람이 이해할 수 있는 근거를 제공함으로써 법적인 문제와 사회적 차별을 낳을 수 있는 기존 인공지능 알고리즘의 한계를 극복하려는 점에서 의의가 있음
  - 알고리즘 자체의 문제뿐 아니라 학습 데이터에 전적으로 의존하는 AI 알고리즘 특성에 비추어 볼 때 알고리즘의 프로세스를 볼 수 있다는 것은 결과에 설명력과 신뢰성을 높여 인공지능의 실질적인 활용 범위를 확장시킬 것으로 기대
    * 판단 결과가 일관성을 제공하지 못할 경우 데이터의 문제인지 알고리즘 설계의 문제인지를 명확히 밝힐 수 있는 근거를 밝힐 필요가 있음
  - 나아가서 설명가능 AI는 인공지능 알고리즘의 품질 인증과 적용 가이드라인 마련에 중요한 기준이 될 것으로 예상

- 설명가능 AI 모델 연구는 지금까지 다양한 접근 방법으로 진행되어 왔으나 학계 중심의 이론적 수준에 머물고 있으며 실용적 단계에 이르기까지는 많은 기술적 난관이 예상됨
  - 인간의 결정 과정에서 '先결정 後설명' 경우가 발생하듯이 XAI 연구의 필요성과 별개로 과연 XAI 연구의 기술적 가능성에 대한 의문이 존재
  - 그럼에도 불구하고 최근 EU의 GDPR을 계기로 인공지능 시스템이 산업과 실생활에 사용되었을 때 발생할 법적, 사회적 문제에 대응할 기술적 요구는 지속될 것
  - 나아가서 적용 영역에 따라, 제품과 서비스 제공 시 인공지능이 내린 결정 과정을 설명할 수 있는가의 여부는 향후 기업 경쟁력의 중요 변수로 작용할 전망

36) Sung Ju Hwang, Leonid Sigal, " A Unified Semantic Embedding: Relating Taxonomies and Attributes", NIPS(Neural Information Processing System) 2014.
37) Yunseong Hwang, Anh Tong, Jaesik Choi, "The Automatic Statistician: A Relational Perspective", arXiv:1511.08343, 2016.02.12.

5 캡슐망 (Capsule Networks)

가. 등장 배경

- Hinton 교수는 2017년 'Dynamic routing between capsules' 논문에서 캡슐망(Capsule Networks)[38]이라는 새로운 신경망 알고리즘을 제안
  - 이 논문에서는 1979년 힌튼 교수가 처음 제안한 아이디어를 알고리즘으로 구현함으로써 기존 CNN (Convolution Neural Networks)의 구조적 한계를 극복할 수 있는 가능성을 보여줌
  - 기본적인 아이디어는 눈으로 획득한 시각정보를 계층적으로 해체하고 이전에 습득한 지식과 비교해 객체의 종류와 위치, 방향 등의 정보를 역추론하는 것이며, 이러한 점은 사람이 사물을 인식하는 방식과 유사

나. 원리 및 특징

- 캡슐망에서는 각각의 뉴런이 독립적으로 작동하는 CNN과 달리 여러 뉴런들의 그룹을 캡슐이라는 단위요소로 정의하고 특정 개체가 존재할 확률과 성질을 벡터로 표현하여 출력값을 계산
  - CNN의 뉴런 출력값은 스칼라(scalar)이지만 캡슐의 출력값은 벡터(vector)이며 벡터의 크기는 특정 개체가 존재할 확률을, 벡터의 방향은 개체의 성질을 나타냄
  - 캡슐망에서는 개체가 존재할 확률을 효과적으로 표현하기 위해 새롭게 제안한 비선형 함수인 스쿼싱 함수(squashing function)를 사용하고 아래층 캡슐의 출력 벡터 가중치를 계산하기 위해서 맥스풀링(max pooling)이 아닌 동적 라우팅(dynamic routing) 방법을 사용
  - 동적 라우팅에서는 단순히 얼굴을 구성하는 눈, 코, 입 등 구성 요소의 존재(스칼라)만으로 얼굴을 인식하는 것이 아니라 각 요소 간 상관관계(벡터)를 계산

※ CNN vs. 캡슐망: scalar-out layer vs. vector-our layer, max-pooling vs. dynamic routing

◎ 맥스풀링(max pooling)
- CNN에서는 학습 시간을 줄이고 이미지 구성 요소의 위치에 상관없이 해당 이미지를 잘 인식할 수 있도록 주변 영역의 추론 결과값 중 최댓값만을 상위층에서 이용함
- 결과적으로 특징 탐색을 위한 시간이 1/4로 단축되고 특징 추출의 부하(load)가 줄어듦

38) Sara Sabour, Nicholas Frosst, Geoffrey E. Hinton, "Dynamic routing between capsules", NIPS(Neural Information Processing System) 2017.

- 얼굴 인식의 경우, 방향이나 각도에 따라 다른 특징이 추출되면 일반적으로 신경망의 상위 층에서 이를 제대로 인식할 수 없으나, 맥스풀링은 눈, 코, 입 등 구성 요소의 상대적인 위치에 무관하게 얼굴을 인식할 수 있음
- 그러나 맥스풀링에서는 구성 요소의 상대적인 위치와 방향을 고려하지 않고 특징을 추출하기 때문에 이미지의 각도나 크기가 변하면 해당 이미지를 다르게 인식하는 근본적 한계가 있음
- 이를 개선하기 위해 다양하게 변형된 이미지를 학습 데이터로 활용하는 방법이 사용되지만 학습시간이 증가하는 문제를 야기함

● (정확성) 캡슐망은 기존 CNN과 같은 딥러닝 알고리즘과 비교하여 사람이 사물을 인식하는 방식과 유사하게 3차원으로 세상을 바라봄으로써 보다 적은 데이터를 사용하더라도 우수한 인식 성능을 보임

- 일반적으로 CNN에서는 은닉 계층이 많을수록 사물의 특징을 잘 인식할 수 있으나 캡슐망에서는 벡터 자체에 충분히 많은 정보가 있어 많은 계층이 필요 없음
- 결과적으로 캡슐망에서는 CNN보다 적은 학습 데이터를 사용하여 보다 우수한 결과를 보이는 사례를 제시함

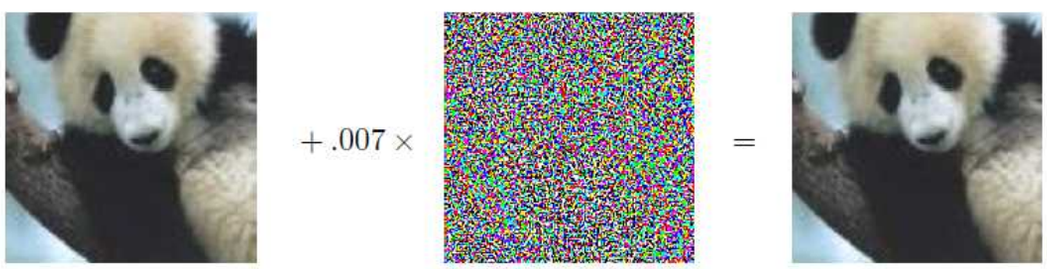
* CNN은 눈, 코, 입이 있으면 얼굴이라고 인식하지만 캡슐망은 두 눈이 가까이 있고 눈 아래에 코가 있고, 눈 아래 입이 있으면 얼굴이라고 인식

그림 19 CNN과 CapsNet 비교

_출처 : CB Insights (2018)

- (강건성) 캡슐망은 사람의 눈으로는 구분할 수 없도록 정교하게 조작된 이미지(예: 비트 단위에서 변화를 준 이미지)를 인식하는 성능이 CNN에 비해 우수
  - Google과 OpenAI 연구진이 수행한 연구에서 팬더 이미지에 작은 변화를 준 결과 CNN은 이를 긴팔 원숭이로 인식하는 결과를 보임[39)]
  - Hinton 등이 발표한 논문에 따르면 MNIST 데이터셋을 대상으로 실험한 캡슐망의 성능은 최신 CNN 알고리즘 대비 에러율을 45%까지 낮췄고 화이트박스 공격(white-box attack)에 대해서도 보다 우수하게 반응함

그림 20 CNN 교란 예



_출처 : Ian J. Goodfellow et al. (2015)

다. 의미와 전망

- 논문에서 제시된 MNIST 데이터셋(28×28) 보다 큰 이미지에 대해서도 캡슐망이 우수한 성능을 보장하는 지에 대해서 충분히 검증되지 않음
  - 또한 학습 소요시간이 기존 CNN보다 길다는 문제가 제기됨
- 그럼에도 불구하고 CNN에 비해 사람이 사물을 인식하는 과정에 더 가까운 캡슐망의 접근 방식이 딥러닝의 또 다른 혁신을 가져올 것이라는 기대가 높음
  - 1980년 대 오류역전파(Back Propagation) 기술이 30년이 지난 후 컴퓨팅 파워, 데이터 등의 발전으로 인해 딥러닝 기술을 혁신시켰듯이 지금의 캡슐(Capsule) 신경망은 인공지능 알고리즘의 연구방향에 큰 전환점이 될 것으로 기대[40)]

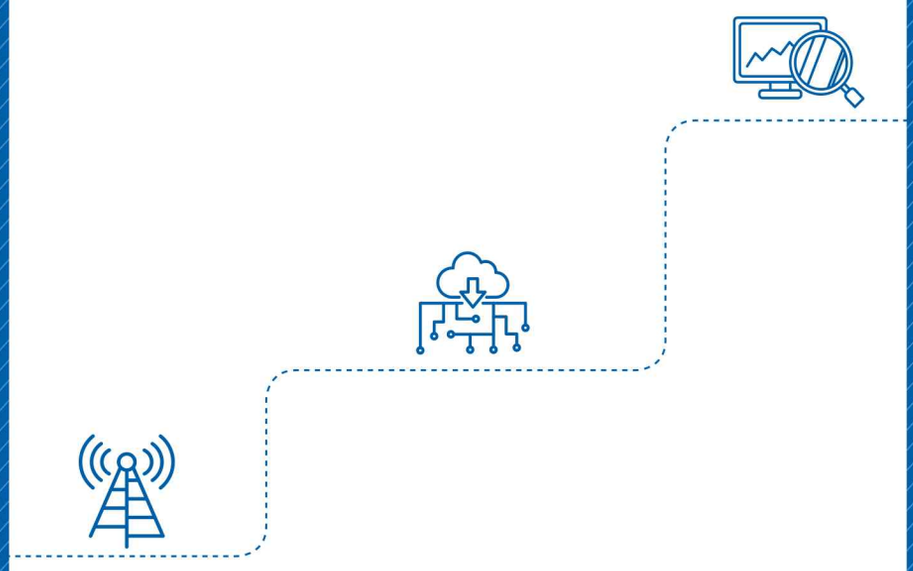
39) Ian J. Goodfellow et al., "Explaining and harnessing adversarial examples", ICLR(International Conference on Learning Representations) 2015.
40) James Somers, "Is AI Riding a One-Trick Pony?", MIT Technology Review, 2017.09.29.

ETRI Insight

V

요약 및 시사점

1. 주요 알고리즘 특징 요약
2. 시사점 및 향후 전망



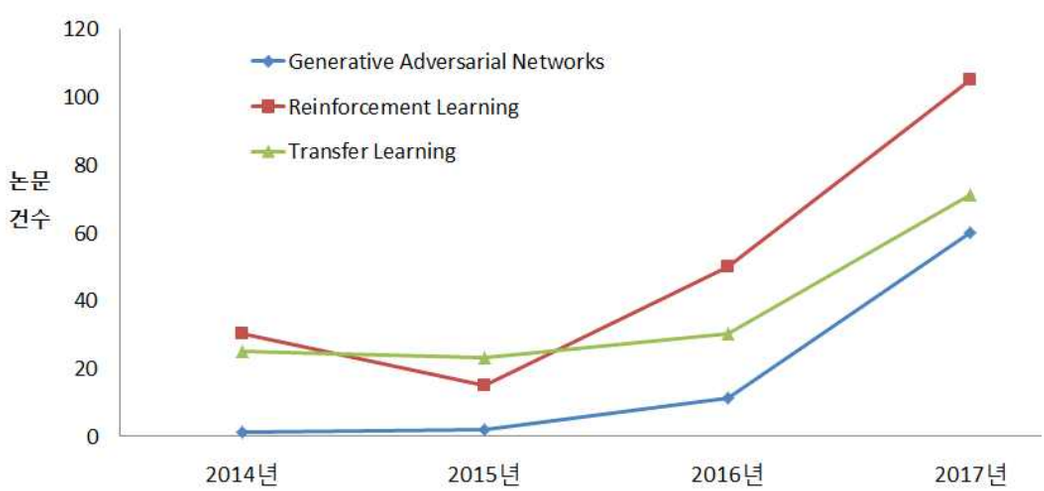
V 요약 및 시사점

1 주요 알고리즘 특징 요약

- 딥러닝 알고리즘이 본격적으로 관심을 받기 시작한 이후 이를 활용하는 과정에서 드러난 한계점을 극복하기 위한 다양한 알고리즘이 활발히 연구되고 있음
  - 인공지능 대표 학회인 NIPS/CPVR/ICML에 실린 논문만 보더라도 2014년 이후 생성적 적대 신경망, 강화학습, 전이학습 등의 키워드 증가세가 뚜렷
  - 또한, 주요 AI 논문이 arXiv.org[41] 사이트에 통해 배포되고 있으며 이러한 특징은 저널 게재나 주요 학회 참여와 관계없이 자신이 연구한 결과를 널리 알리는 경향이 빠르게 증가하고 있음을 보여줌

    * 2017년 arXiv.org에 실린 전체 AI 논문 건수는 13,325건이며 이 가운데 Computer vision and pattern recognition 분야가 4,895건(약 37%)으로 가장 많음[42]

그림 21 주요 AI 알고리즘 논문 건수 추이[43]



_출처 : https://www.scopus.com

41) arXiv.org는 수학, 물리학, 컴퓨터 과학, 통계학, 천문학, 계량 생물학 분야의 출판 전(preprint) 논문을 보관하는 웹사이트로 Cornell대에서 운영
42) Yoav Shoham et al., The AI Index 2018 Annual Report, Stanford University, 2018.12.
43) 인공지능분야의 대표적 학회 NIPS(Neural Information Processing Systems), ICML(International Conference on Machine Learning), CVPR(Conference on Computer Vision and Pattern Recognition)에 게재된 논문 가운데 Title, Abstract, Keyword에 Generative Adversarial Networks, Reinforcement Learning, Transfer Learning 이라는 단어가 포함된 연도별 논문 건수

- 본 보고서에서 다룬 AI 알고리즘은 기존 딥러닝의 처리 속도와 정확도 개선을 넘어 데이터처리와 학습 방식에 관한 새로운 시도들
  - 즉 딥러닝을 중심으로 제한된 데이터를 사용하여 정확도를 개선하거나 학습시간을 단축하기 위한 시도
  - 인간의 이성적 판단과 행동과 유사하게 인공지능 알고리즘이 스스로 학습모델을 개선하거나 새로운 결과물을 만드는 방안
  - 인간의 관점에서 알고리즘이 내린 판단 결과의 타당성과 도출 과정의 투명성을 보장하기 위한 노력
  - 딥러닝 중심의 인공신경망 구조와 근본적으로 다른 접근을 통해 또 다른 혁신을 준비하는 연구 등
  - 딥러닝은 새롭게 제안된 다양한 알고리즘을 통해 성능 개선뿐 아니라 데이터 수집과 활용, 처리 방식의 확장을 통해 범용성을 지닌 알고리즘으로 발전 기대

표 3 주요 AI 알고리즘 특징 요약

AI 알고리즘 종류	주요 특징
생성적 적대 신경망 (Generative Adversarial Networks)	- 적대적으로 경쟁하는 생성기와 판별기를 통해 진본데이터와 매우 유사한 위조데이터를 생성 - 현실에 없는 새로운 데이터 생성, 새로운 형태로 데이터 변환, 데이터 품질 향상 등 새로운 기회 가능성을 제시
심층강화학습 (Deep Reinforcement Learning)	- 복잡한 실제 환경에서 반복적인 경험(데이터)의 시행착오를 통해 최적의 학습모델을 스스로 발전시킴 - 지금까지 PC 안에서 이뤄졌던 인공지능을 현실 세계의 다양한 객체에 적용하기 시작한 계기이며 감각기관의 확장을 가져옴
전이학습 (Transfer Learning)	- 학습데이터 확보가 현실적으로 어려운 분야에서 기존에 학습이 완료된 모델의 일부를 재사용하여 학습시간을 단축하고 성능을 보장 - AI 알고리즘이 인간과 같이 학습효과를 가지고 발전 가능성 제시
설명가능 인공지능 (Explainable AI)	- 기존 설명력이 높은 알고리즘 일부를 활용하거나 개선하여 학습모델이 도출한 결과의 근거를 제공 - AI 알고리즘 사용 시 법과 제도적 문제로 인해 비즈니스 활용 범위의 한계를 극복할 수 있다는 가능성 제시
캡슐망 (Capsule Networks)	- 외부세계를 인식하는 과정이 3차원적 벡터방식의 인간의 뇌 인식 과정과 유사하게 알고리즘 구조를 설계 - 현재 연구 초기단계이나 향후 범용성을 지닌 알고리즘 혁신을 이끌 차세대 AI 알고리즘으로 발전 기대

● 최근 진행되고 있는 AI 알고리즘은 알고리즘별 독특한 특성을 가지고 데이터 활용, 학습 프로세스 개선 및 최적화, 새로운 구조의 학습모델 등의 관점에서 기존 AI 한계점의 극복 가능성을 제시

- 기존 AI 알고리즘은 무엇보다 학습모델을 생성하기 위해서 적용 분야별로 충분한 양의 라벨링된 데이터 즉, 정답이 있는 데이터가 필요하며
- 인공지능 알고리즘은 새로운 분야에 적용하기 위해 기존에 학습된 결과 또는 학습 과정을 재활용하고 응용할 수 있도록 학습효과를 가져야 하고
- 또한 의료, 법률, 금융 등 현장의 의사결정 과정에서 인공지능 알고리즘이 내린 판단 결과의 도출 과정의 투명성뿐 아니라 강건성이 보장되어야 함

표 4 기존 AI 알고리즘의 한계점 극복 가능성

구분	AI 알고리즘 활용 방안	기존 한계점 극복 가능성
▷ 데이터 양(量)과 질(質)	- GAN: 진본데이터와 유사한 위조데이터를 생성하여 학습데이터로 활용 - 강화학습: 준비된 데이터 없이 실제 환경에서 반복적인 경험(데이터)을 통해 학습	(한계) 적용 분야에 따른 대량의 정답이 있는 학습 데이터 부족 (극복) 대량의 학습데이터 생성
▷ 지식 전이	- 전이학습: 새롭게 적용하려는 분야에 기존 학습이 완료된 모델의 일부를 재사용 (특히 데이터 확보가 현실적으로 어려운 분야에 활용 필요)	(한계) 기존 학습 모델(지식)을 유사분야에 응용·재활용 못함 (극복) 제한된 분야에서 모델의 일부를 재사용
▷ 새로운 결과 생성	- GAN: 현실에 없는 새로운 데이터 생성, 새로운 형태로 데이터 변환, 데이터 품질 향상 - L2M: 인간의 평생학습 개념과 유사하게 기본적인 학습모델을 새로운 환경에서 지속적으로 개선 - 강화학습: 복잡한 현실 환경에서 시행착오를 통해 최적의 학습모델을 스스로 발전	(한계) 사람의 개입 없이 스스로 더 나은 학습모델과 결과를 만들지 못함 (극복) 새로운 결과 생성, 점진적 모델 개선, 무한 경쟁을 통한 최적 모델 발견
▷ 판단 과정 투명성	- XAI: 기존 설명력이 높은 알고리즘 일부를 활용하거나 개선하여 설명가능 모델과 인터페이스를 개발하여 도출 근거를 제시	(한계) 도출 과정 설명력 부족 (극복) 판단 결과의 근거를 다양한 방식으로 제공
▷ 알고리즘 강건성	- 캡슐망: 비상식적 결과 도출 오류 개선 및 간단히 변형하거나 정교하게 조작된 데이터에 대해 보안성 제고	(한계) 2차원적 데이터 해석, 데이터 조작 취약성 (극복) 3차원적 인식으로 알고리즘 강건성 제고

2 시사점 및 향후 전망

● 지금까지 살펴본 딥러닝 이후 활발히 연구되고 있는 AI 알고리즘의 트렌드는 모방을 통한 데이터 활용 극대화, 인간의 개입 최소화, 통합화, 범용화 등으로 요약

① **[모방화]** 학습모델 내부 또는 학습모델 간 경쟁을 통해 현실 데이터와 유사한 데이터를 생산하여 제한된 학습 데이터량을 늘리려는 트렌드

- CNN, RNN 등 딥러닝 등장 이후 최근 가장 주목받고 있는 생성적 적대 신경망이 대표적이며 기존에 존재하는 데이터를 모방하여 새로운 결과물을 창조
- 새로운 결과물(이미지, 음성 등)을 만들 수 있다는 것은 단순히 사물을 인식하는 수동적 수준에서 무엇인가를 생산하는 능동적 수준에 진입했다는 의미

  * 지금까지 200여개의 GAN의 변형 알고리즘은 비즈니스와 예술 등 창작 활동에 활용됨으로써 단순한 응용을 넘어 인간의 상상력 그 이상을 보여줄 수 가능성을 제시

- 결과적으로 모방한 데이터 자체가 진짜가 되는 최종 결과물로써 활용될 수 있고 새로운 모델을 생산하기 위한 입력 데이터로 사용될 수도 있음
- 그러나 진본과 유사한 결과물을 쉽고 풍부하게 만들 수 있다는 점을 악용해 가짜 콘텐츠를 생산하여 새로운 사회 문제를 유발

② **[자동화]** 데이터 수집에서 학습모델 구성에 이르는 일련의 과정에서 인간의 개입을 최소화하여 End-to-End 자동화하려는 트렌드

- 학습에 필요한 데이터를 구성할 때 사람이 일일이 정답을 부여하지 않고 현실 데이터를 그대로 사용하는 알고리즘 연구 활발
- 또한 데이터 수집 후에 특징 추출을 위한 전처리 과정에서부터 학습모델을 생성하는 각각의 단계를 딥러닝 알고리즘으로 일원화하려는 시도

  * 최근 음성인식, 대화 시스템 등의 분야에서도 End-to-End 딥러닝 학습 알고리즘이 적용 중

- 특히 심층강화학습은 정답이 없는 현실 데이터를 사용하여 End-to-End 학습 과정에서 무한 시행착오를 반복하며 최적 방안을 도출

  * 심층강화학습은 게임 등 가상환경을 넘어 현실 세계에서 적용 분야를 확대 중

- Google의 AutoML와 같이 학습이 완료된 인공지능이 인간의 개입 없이 새로운 형태의 학습모델을 자동 생성하는 연구 또한 주목 필요

③ **[통합화]** 특정 알고리즘의 장점을 극대화하고 단점을 보완하기 위해 다양한 AI 알고리즘을 통합적으로 사용하려는 트렌드

- 통합화 추세는 크게 하나의 적용 영역에서 단계별로 특화된 알고리즘을 통합적으로 활용하는 것과 이종 영역을 하나의 알고리즘에서 통합하려는 연구로 구분
- 하나의 적용 영역에서 부족한 데이터를 확보하기 위해서 또는 학습의 정확도를 높이기 위해서 심층강화학습, 생성적 적대 신경망, 전이학습 등을 통합 사용
- 한편 음성인식과 사물인식 등 별개의 영역을 딥러닝 알고리즘에서 통합적으로 처리하려는 시도가 진행

  * MIT 연구진은 오디오 파일과 이미지 파일을 딥러닝 알고리즘으로 연결하여 음성 재생과 동시에 문장의 각 부분을 사물 이미지에 표시하는 논문이 대표적[44]

④ **[범용화]** 기존 학습모델을 부분 최적화하거나 학습모델이 인간의 인식 과정과 유사하게 데이터를 처리하여 범용성을 높이려는 트렌드

- AI 알고리즘의 범용화는 기존 학습 결과를 응용하여 학습 효과를 만들거나 인간의 판단과정에 비추어 상식적이고 직관성을 갖춘 알고리즘 연구 등
- 전이학습은 지식의 전이를 통해서 인간의 적응력과 유연성을 가진 알고리즘을 만들 수 있다는 점에서 앞으로 큰 발전이 기대됨
- 무엇보다도 근본적으로 인간과 유사한 방식으로 세상을 인식하도록 알고리즘을 설계하여 보안성과 강건성을 보장하려는 연구도 진행

  * 캡슐망이 대표적으로 적대적 공격(Adversarial Attack)과 사물 인식 과정에 범용성을 추구

● 이와 같은 AI 알고리즘 트렌드는 인간의 뇌와 감각기관의 진화과정을 모방하고 기계 특화된 무한 최적화 방법을 접목하여 보다 범용성을 추구하며 발전

- AI 관점에서 감각기관의 진화는 최근 대화형 스피커 확산, 로봇의 지능화 등을 통해 현실 세계와 직접 소통하며 발전 중
- 인간의 뇌에 해당하는 학습 알고리즘 구조의 진화는 AI가 단순히 보고 듣는 것 이상으로 인간과 비슷한 이성을 갖고 판단하는 방식으로 연구가 진행
- 또한, 인간과 달리 기계이기 때문에 가능한 무한 반복적 시행착오와 경쟁을 통해 인간의 개입 없이 성능 극대화를 추구하여 인간을 능가하는 결과를 도출

44) David Harwath et al., "Jointly Discovering Visual Objects and Spoken Words from Raw Sensory Input", European Conference on Computer Vision(ECCV) 2018.

- 요컨대, 인간을 모방하되 기계만의 장점을 최대로 활용하여 인간을 보완하고 협업하는 다양한 AI 알고리즘으로 발전

● 단기적으로 AI 알고리즘은 다양한 접근법이 복합적으로 사용되어 자율주행자동차, 로봇 등 실생활과 다양한 산업에 적용될 것으로 기대

- 향후 자동차, 로봇, 엣지컴퓨팅 등에 딥러닝, 강화학습, 전이학습 등이 본격적으로 적용되면 몸체를 가진 AI 알고리즘이 세상 밖으로 나오는 시발점이 될 것
- 지금까지 대부분의 인공지능은 인간이 수집한 데이터를 PC 내부에서 모델을 생성하는 수준이었지만 앞으로 현실 세계의 다양한 물리적 객체에 AI가 스며들면서 또 다른 혁신 기대

● 그러나 이러한 기대 이면에 진짜와 같은 가짜데이터 생성, 인간 수준을 뛰어넘는 해킹머신 탄생, 전장(戰場)의 규칙을 재정의하는 등 미래 위험에 대한 대비가 필요

- 인공지능의 위험성은 크게 산업과 일상생활에서 인간이 정한 규칙 대신 기계에 의존하며 발생하는 사회 문제뿐 아니라 국가 안보 분야를 포함
- 사회 문제는 AI 알고리즘의 정확성과 별개로 인종과 민족, 그리고 사회경제적 편견이 반영된 알고리즘 사용의 문제, 가짜 정보 유포·확산의 문제 등이 대표적
- 보다 근본적인 위험성은 인공지능이 해킹에 악용되고 치명적인 공격의 가능성에 따른 AI 무기화 등 사회 안전과 국가 안보에 관련된 문제

● 향후 AI 알고리즘 연구는 기초·원천 강화를 위한 R&D뿐 아니라 광범위한 산업적 활용을 고민함과 동시에 미래 위험에 대한 준비가 필요

① **[범용 AI 알고리즘 연구]** 현재 비즈니스 활용 시 경제적 가치를 생산할 수 있는 AI 알고리즘뿐만 아니라 보다 범용성을 추구하는 기초·원천 알고리즘 연구 강화

② **[몸체를 가진 AI 연구]** 실제 환경에 노출된 기계들과 결합한 알고리즘이 현실의 데이터(행동 데이터)를 이용하고 상호작용하는 몸체를 가진 AI 연구

③ **[생활 속의 AI 연구]** 지금까지 보고 듣는 영역의 활용을 넘어 소비자와 직접 소통하고 인간을 보조·협업하는 생활 속의 AI 연구

④ **[AI 부작용 및 위험 대비]** 인간을 압도하는 우수한 성능을 지닌 AI 알고리즘을 악용하여 가짜뉴스를 생산하거나 사회를 통제하는 수단으로 사용하고 나아가서 국제정치와 전쟁의 패러다임을 바꿀 미래 위험에 대비한 기술적 방안 고민 필요

참고문헌

Andre Esteva et al., “Dermatologist-level classification of skin cancer with deep neural networks”, Nature 542, pp.115-118, 2017.02.02.

A. Radford and L. Metz, “Unsupervised representation learning with deep convolutional generative adversarial networks”, ICLR 2016.

Arur Kadurin et al., “The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology”, Oncotarget, Vol. 8, 2017.

Benjamin Letham et al., “Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model”, Annals of Applied Statistics, Vol. 9, No. 3, 1350-137, 2015.

Brenden M. Lake et al, “Human-level concept learning through probabilistic program induction”, Science, Vol. 350, p. 1332-1338, 2015.

C. Ledig et al., “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network”, NIPS 2016.

CB Insights, “Top AI Trends to Watch in 2018”, 2018.03.18.

Chrisantha Fernando, “PathNet: Evolution Channels Gradient Decent in Super Neural Networks”, arXiv:1701.08734, 2017.01.30.

Dayong Wang et al., “Deep Learning for Identifying Metastatic Breast Cancer”, arXiv:1606.05718v1, 2016.06.18.

David Gunning, “Explainable Artificial Intelligence(XAI)”, DARPA/I2O Program Update, 2017.11.

David Harwath et al., “Jointly Discovering Visual Objects and Spoken Words from Raw Sensory Input”, ECCV 2018.

Dean Takahashi, “Nvidia launches Isaac robot platform with Jetson Xavier

robot processor", VentureBeat, 2018.06.04.

GDPR, https://www.eugdpr.org/

Hava T. Siegelmann, "Lifelong Learning Machines (L2M)", DARPA, 2017.04.26.

Hendricks et al., "Generating Visual Explanations", arXiv:1603.08507v1, 2016.03.28.

https://www.utoronto.ca/news/u-t-ai-researchers-design-privacy-filter-photos-disables-facial-recognition-systems

Hui Cheng et al., "Multimedia Event Detection and Recounting", SRI-Sarnoff AURORA at TRECVID 2014.

Ian J. Goodfellow et al., "Explaining and harnessing adversarial examples", ICLR 2015.

iGAN(interactive GAN): https://github.com/junyanz/iGAN

James Somers, "Is AI Riding a One-Trick Pony?", MIT Technology Review, 2017.09.29.

Jun-Yan Zhu et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks", arXiv:1703.10593v4, 2018.02.19.

Libin Liu et al, "Learning Basketball Dribbling Skills Using Trajectory Optimization and Deep Reinforcement Learning", ACM Transactions on Graphics, 2018.08.

Marco Tulio Ribeiro et al., ""Why Should I Trust You?" Explainable the Predictions of Any Classifier", CHI 2016 Workshop on Human Centered Machine Learning, 2016.02.16.

Michael Xie et al., "Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping", arXiv:1510.00098v2, 2016.02.27.

Nikkei Robotics, "東芝が送電線点検に新種のディープラーニング、"知能"より"創作者"と呼ぶべき新AI「生成モデル」", 2017.04.

NVIDIA, https://www.nvidia.com/en-us/autonomous-machines/

Preferred Networks, https://www.preferred-networks.jp/en/whitepaper

Sara Sabour, Nicholas Frosst, Geoffrey E. Hinton, "Dynamic routing

between capsules", NIPS 2017.

Sebastian Bach et al., "On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation", PLoS One, Vol. 10, No. 7, e0130140, 2015.06.10.

Sinno Jialin Pan and Qiang Yang, "A Survey on Transfer Learning", IEEE Transactions on Knowledge and Data Engineering, Vol. 22, Issue 10, 2010.10.

Taeksoo Kim et al., "Learning to Discover Cross-Domain Relations with Generative Adversarial Networks", arXiv:1703.05192v2, 2017.05.15.

Varun Gulshan et al., "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs", JAMA, 2016.11.29.

Volodymyr Mnih et al., "Human-level control through deep reinforcement learning", Nature, 2015.02.26.

Volodymyr Mnih et al., "Playing Atari with Deep Reinforcement Learning", arXiv:1312.5602, 2013.12.19.

Will Knight, "Amazon has developed an AI fashion designer", MIT Technology Review, 2017.08.24.

Yoav Shoham et al., The AI Index 2018 Annual Report, Stanford University, 2018.12.

Yunseong Hwang, Anh Tong, Jaesik Choi, "The Automatic Statistician: A Relational Perspective", arXiv:1511.08343, 2016.02.12.

Zhang et al., "StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks", ICCV 2017.

Zhangzhang Si and Song-Chun Zhu, "Learning AND-OR Templates for Object Recognition and Detection", IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 35 No. 9, p. 2189-2205, 2013.

저자소개

이승민 ETRI 미래전략연구소 기술경제연구본부 산업전략연구그룹 책임연구원
e-mail: todtom@etri.re.kr Tel. 042-860-1775

딥러닝 이후, AI 알고리즘 트렌드

발행인 한 성 수
발행처 한국전자통신연구원 미래전략연구소 기술경제연구본부
발행일 2018년 12월 31일

www.etri.re.kr

비매품/무료

93560

9 788955 192629

ISBN 978-89-5519-262-9

ETRI 한국전자통신연구원 미래전략연구소

34129 대전광역시 유성구 가정로 218

TEL. (042) 860-6114 FAX. (042) 860-6504