

방송 미디어를 위한 AI 기술 - 1

- AI 기술의 개요와 방송 미디어 응용

글. 오주현 KBS 미디어기술연구소

연재 목차

1. AI 기술의 개요와 방송 미디어 응용
2. 다양한 방송 미디어 응용과 AI 기술 적용 방식

인공지능(Artificial Intelligence, 이하 AI) 기술은 최근 눈부시게 발전하여 이미지 인식을 포함한 일부 분야에서 일반 사람의 능력을 뛰어넘었다고 할 수 있다. 이 사실은 2016년 이세돌 대 알파고 바둑 경기와 같은 큰 이벤트를 통해 대중에게 널리 각인된 바 있다. 일반적으로 바둑 대국에서 나올 수 있는 경우의 수는 관측 가능한 우주 전체의 원자 개수(약 10^{80})보다 더 큰 것으로 알려져 있는데, 무차별 탐색(brute force search) 등 단순 계산 방식에 적합하다고 여겨지는 컴퓨터가 인간 프로그래머를 이길 수 없다는 기존 상식을 DeepMind의 알파고가 깬 것이다.

사실 알파고가 바둑을 통해 세계인을 놀라게 하기 전에 'ImageNet Moment'가 있었다. ImageNet이란 Wordnet(트리구조의 영어 어휘 데이터베이스) 기반으로 분류된 약 천 개의 사물 종류에 대해 약 천 이백만 개의 이미지를 모아 놓은 데이터셋이다. 2010년부터 이 ImageNet의 이미지를 학습하여 인식하는 경진대회 ILSVRC(ImageNet Large Scale Visual Recognition Challenge)가 개최되었고, 2012년부터 딥러닝이 도입되면서 인식능력이 급격히 향상되었다. 급기야 2015년에는 Microsoft Research의 ResNet이 무려 152개의 다층 레이어를 이용해서 비교군이라 할 수 있는 사람의 평균 인식오류율인 5.1%를 넘어서는 3.6%를 달성할 수 있었던 것이다.

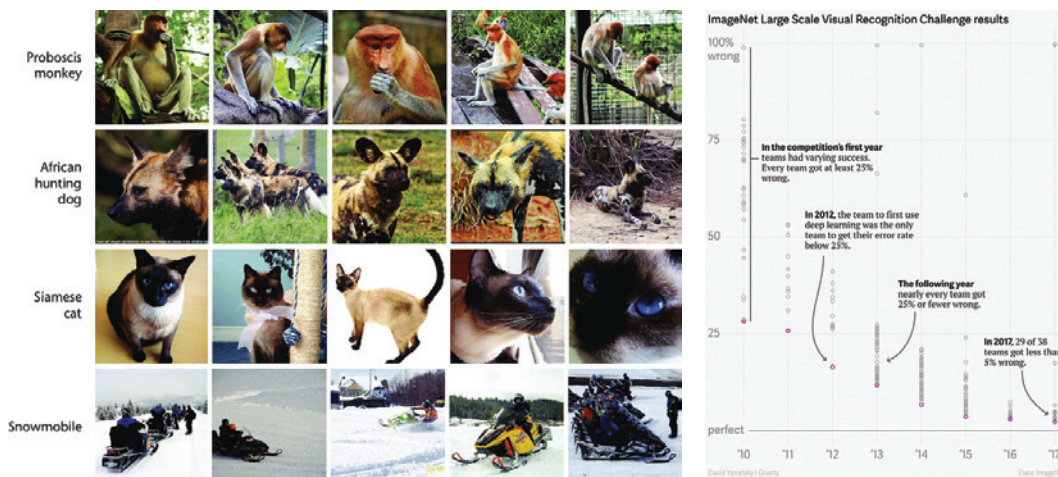


그림 1. ImageNet 데이터의 예(좌)와 ILSVRC 참여 알고리즘의 오류율 감소 그래프(우)

사실 기계학습이 아무리 발전했다 해도 인간의 분류능력을 뛰어넘는다는 것이 선뜻 와 닿지 않을 수 있는데, 이는 ImageNet에는 단순히 '개'나 '고양이' 정도가 아니라, '말라뮤트'나 '포메라니안' 같은 좀 더 세분화된 분류가 존재하므로 사람이 정확하게 판별하기 쉽지 않은 부분이 있기 때문이다. 그런데도 수치상으로 인간의 인식능력을 뛰어넘는 AI가 등장했다는 사실 자체가 큰 충격으로 받아들여졌고, 이후 패턴인식 등 관련 분야에서는 딥러닝을 빼고 연구하는 것은 상상하기 어려운 일이 되었다.

근래 AI가 화두가 되면서 딥러닝이 AI 그 자체인 것처럼 혼용해서 쓰이는 경우가 많은데, 본격적으로 방송 미디어를 위한 AI 기술에 대해 논의하기 전에 우선 AI에 대한 정의부터 확실히 하는 것이 좋을 것 같다. AI, 기계학습(머신러닝), 그리고 딥러닝은 그림과 같은 포함 관계가 있다.

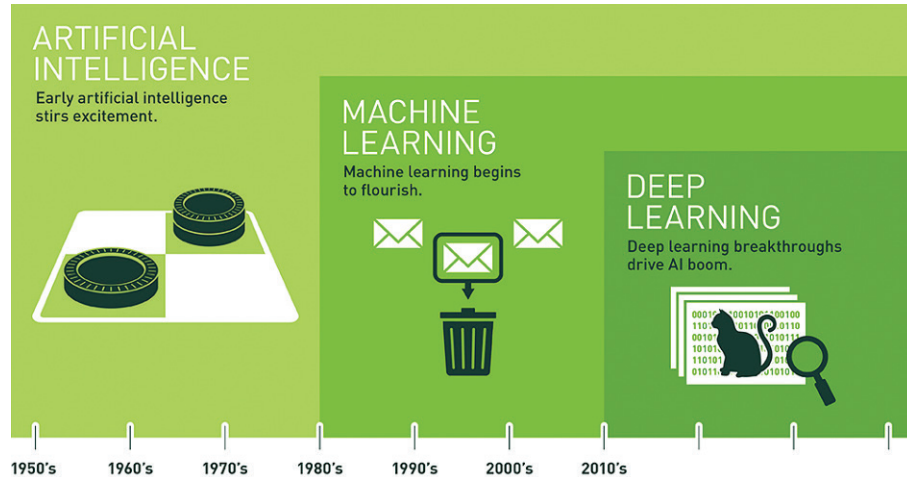


그림 2. AI, 기계학습, 딥러닝의 관계 (Nvidia)

우선 AI는 ‘인공지능’이므로 인간의 지능을 흉내 낸 어떤 프로그램이나 시스템에도 적용할 수 있는 용어이다. 예를 들면 사람이 하던 계산을 대신 수행하는 계산기도 AI라고 할 수 있다. AI 중에서도 머신러닝은 ‘기계학습’이므로 학습을 통해 개선되는 AI를 말한다. 스팸메일 필터와 같이 많은 데이터를 통해 학습 및 개선이 가능한 통계적 방식의 AI를 생각하면 된다. 앞서 예를 든 계산기의 경우 아무리 많은 계산을 수행하더라도 계산이 더 정확해지는 것은 아니므로 머신러닝이 아니다. 마지막으로 최근의 AI 발전을 이끈 딥러닝의 특징은 다층신경망(multilayered neural networks)을 사용하고 방대한 데이터로 이를 훈련시키는 것이다. 머신러닝과 딥러닝을 구분하는 또 하나의 차이점은, 머신러닝은 알고리즘 개발자가 인식과 분류를 위한 핵심 특징(features)을 결정하고 학습시키는 반면 딥러닝은 이를 방대한 데이터에 의해 네트워크 스스로 추출한다는 점이 될 것이다.

사실 이러한 AI는 역사적으로 전혀 새로운 개념이 아니며, 역사적으로 몇 번의(보통 두 번의 AI 겨울이 있었다고 본다) 부침을 거치고 최근 다시 부흥한 것이다. 비관적인 사람들에게는 지금 이미 세 번째 AI 겨울이 시작되었다고 보는 시각도 존재하지만, 어찌 됐든 여태껏 컴퓨터비전과 음성인식, 심지어 예술 등 많은 분야에서 도달할 것으로 상상하지 못했던 수준을 AI가 돌파한 것은 엄연한 사실이다.

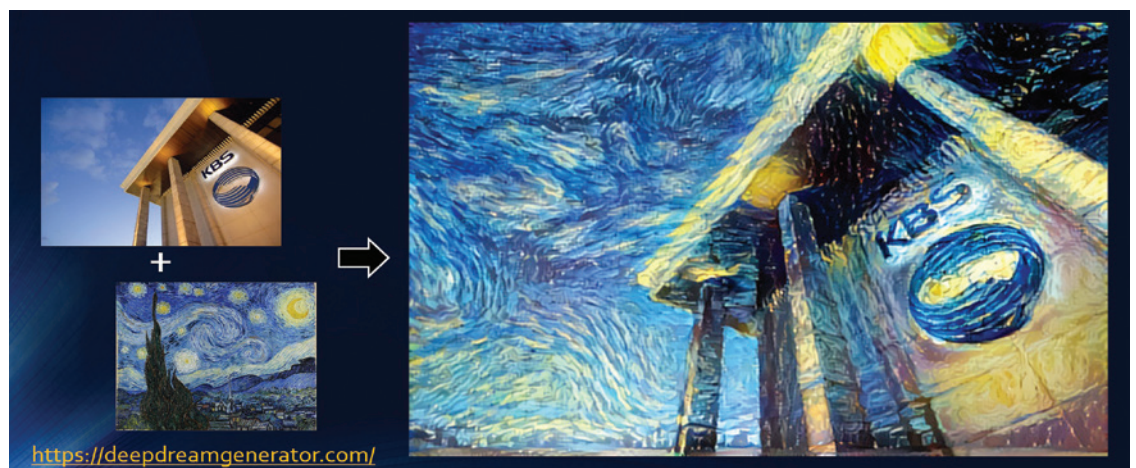


그림 3. AI 예술의 가능성을 연 구글의 ‘Deep Dream’

이와 같은 최근의 AI 발전은 크게 세 가지 요소에 의해 달성되었다고 본다.

첫째는 기존 인공신경망(artificial neural network)의 문제를 극복한 딥러닝 알고리즘이다. 생물의 뇌를 본떠 만든 최초의 퍼셉트론(perceptron)은 본격적인 AI의 시대를 열 기술로 많은 기대를 모았지만, 단순한 선형분류기로 서 XOR 논리연산 문제도 풀지 못하는 한계가 밝혀지면서 첫 번째 AI 겨울을 불러오게 되었다. 그 후 다층 퍼셉트론(multilayer perceptron, MLP)과 역전파(backpropagation) 방법이 등장하였고, Geoffrey Hinton 등이 ‘비지도 방식의 제한 볼츠만 머신’ 논문을 통해 종래 신경망의 국부최소점(local minimum)과 과적합(overfitting) 문제를 해소할 방법을 제시함으로써 본격적인 딥러닝의 시대가 열리게 되었다.

둘째는 인터넷을 기반으로 한 방대한 데이터이다. ‘Internet live stats’ 사이트에서는 실시간으로 발생하는 인터넷 데이터의 양을 어림할 수 있는데, 이에 따르면 1초마다 8,533회의 트윗, 921개의 인스타그램 사진 업로드, 74,131GB의 인터넷 트래픽, 75,186개의 구글 검색, 79,224개의 유튜브 비디오 시청이 이루어진다. 2014년의 한 논문에 따르면 인터넷 스토리지의 용량은 10^{24} 바이트 또는 백만 엑사바이트이다. 마찬가지로 2014년에 출간된 Mary Meeker의 연간 인터넷 동향 리포트에 따르면 매일 평균 18억 개(연간 6천5백억 개)의 이미지가 업로드된다. 이와 같이 방대한 인터넷 데이터를 기반으로 하여 사람이 일일이 레이블링 작업을 해 놓은 ImageNet과 같은 오픈 데이터셋이 딥러닝 발전에 큰 역할을 하였다.

셋째는 계산능력(processing power)의 발전이다. 집적회로의 성능이 일정 기간에 2배씩 증가한다는 무어의 법칙은 18개월 또는 24개월의 논란이 있긴 하지만 지금까지는 계산능력의 발전을 비교적 정확히 예측해 왔다. 또한 과거의 컴퓨팅에서는 하나의 CPU만을 사용하는 것이 당연한 상식이었지만 2005년부터 데스크톱용 멀티코어 CPU가 등장하고, 더 실감 나는 게임 그래픽을 위해 도입된 셰이더(pixel/fragment shader)가 GPGPU(general-purpose graphics processing unit)에 활용되면서 컴퓨팅 환경은 병렬컴퓨팅으로 패러다임이 바뀌었다. 지금 Nvidia의 최신 GPU인 RTX 2080 Ti의 코어 개수는 무려 4,352개에 달한다.

AI 발전의 중요 요소인 데이터의 증가와 프로세서의 발전이 모두 지수함수적인 모습을 보이는 데 주목할 필요가 있다. 신문을 10번 접으면 두께가 약 10cm에 불과하지만 50번을 접으면 1억km가 넘어 태양까지 닿는다는 얘기를 들어본 적이 있을 것이다. 그만큼 일정 시점을 지난 AI 발전속도는 우리가 상상하는 수준을 넘어설 것이다.

‘Super Intelligence’를 저술한 Nick Bostrom은 일단 AI가 인간의 수준에 오르기만 하면 스스로를 개량할 수 있기 때문에 특이점(singularity)이 발생하여 AI가 지구상에 유례없는 인공 초지능(artificial super intelligence) 레벨에 도달할 것이라고 보고 있다(물론 AI에게는 인간이 가진 게으름이나 사내정치 따위의 비효율적 속성이 없다는 가정 하에서). 그래서 인간 수준의 AI(human-level AI)는 순식간에 지나가 버릴 것이고, 일단 초지능이 발생한 후에는 인류에게는 ‘절멸’ 또는 ‘영생’ 두 가지의 미래만 존재한다고 보는 시각도 있으니, 그만큼 AI가 산업 전반과 인류의 미래에 큰 영향을 미칠 것으로 봐야 할 것이다.

그렇다면 AI가 인류를 지배하기 전에 방송 미디어 산업에 미칠 영향은 무엇일까?

Globecast의 Shakunt Malhotra는 방송 산업에서의 AI 응용 분야를 품질관리, 검색, 메타데이터, 규정 준수, 편집, 하이라이트, 구조 분할과 광고, 자막, 감독, 뉴스 프레젠테이션 등으로 제시한 바 있다. 이중 AI의 가장 직관적인 응용이라 할 수 있는 뉴스 프레젠테이션부터 한번 살펴보자.

뉴스 프레젠테이션

중국의 신화통신이 2018년 남성 AI 뉴스 앵커를 공개한 데 이어 2019년에는 여성 AI 앵커를 공개하였다. 각각 실존하는 남성과 여성 앵커를 본떠 만든 것이고, 일어선 채로 자연스럽게 뉴스를 진행할 수 있는 기술 수준을 보여준다. 신화통신은 AI 앵커가 웹사이트 및 다양한 소셜미디어 플랫폼에서 24시간 전문 앵커로 활동할 수 있다고 밝혔다.



그림 4. 신화통신의 AI 앵커들. 각각 실제인물(좌)-AI앵커(우) 순서

사실 AI는 뉴스에서 이미 상용화되어 쓰이고 있는데, 경제신문 인터넷사이트 등에서 주가지수 현황 등 정형화된 정보를 기사화하기 위해 사용되는 로봇저널리즘이 그것이다.

‘AI 앵커’와 ‘로봇 기자’는 AI의 미디어 응용을 가장 극명하게 보여주는 예이지만, 실제 방송사 입장에서 보면 그 파급력은 크지 않을 것으로 보인다. 방송사는 이미 아나운서와 기자 등 많은 프레젠테이션 전문인력을 보유하고 있고, 무엇보다도 이 기술들은 가짜뉴스가 큰 사회적 문제로 떠오르는 지금 뉴스의 신뢰성을 떨어뜨릴 수 있기 때문이다. 재난 등으로 인해 AI 앵커가 필요할 것으로 제시되는 긴급뉴스 송출 상황에도 자막 및 간단한 TTS를 제공하는 편이 나을 수 있다(물론 TTS 기술도 AI이긴 하다).

품질관리

품질관리(quality check, 이하 QC)는 콘텐츠가 송출이나 배포되는 단계 이전에 필요한 핵심 작업이다. 미디어와 플랫폼이 다양화되면서 예전 지상파만큼의 높은 QC가 요구되지 않는 경우가 대부분이지만, AI에 의한 QC는 두 가지 측면에서 더 중요해지고 있다. 첫째, 테이프가 아닌 파일로 재생 및 송출하는 송출/배포 환경 변화이다. 테이프는 실

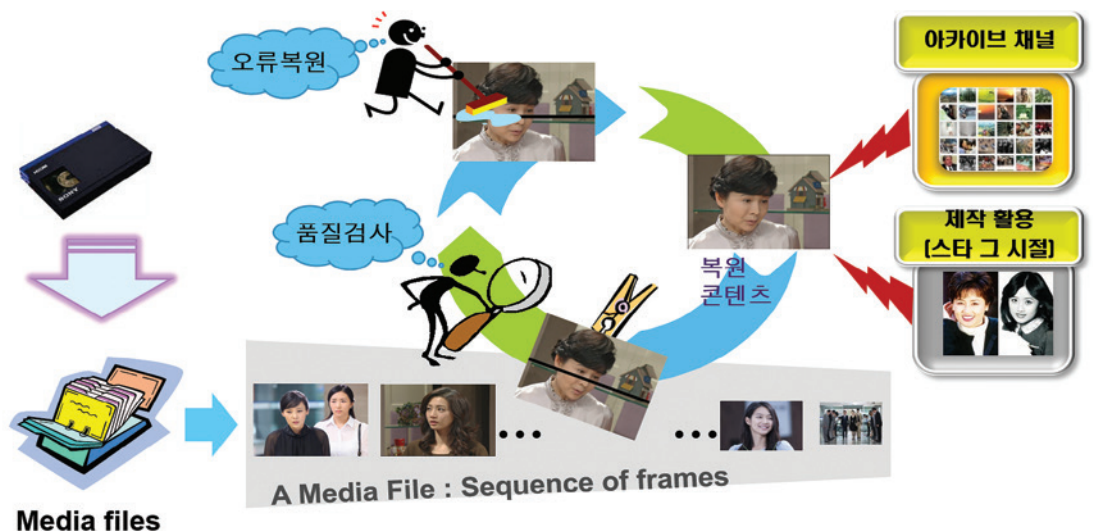


그림 5. 호크아이 시스템 개념도

시간을 넘어서 '배속 재생'을 할 경우 화질 유지가 어렵고, 사람이 하든 AI가 하든 영상 재생 시간만큼의 모니터링 시간이 소요된다. 그러나 파일로 콘텐츠를 송출/배포하는 환경에서는 파일을 한 번에 읽어 매우 빠른 속도로 검사하는 것이 가능하므로 AI가 QC 시간을 단축하여 효율성을 크게 높일 수 있다. 둘째, 디지털 미디어 시청환경에서는 기존의 아날로그나 HDTV와 같은 획일적인 수상기 외에 많은 모바일 디바이스가 존재할 수 있으므로, AI가 다양한 수신 디바이스와 네트워크 환경에서의 콘텐츠의 품질 보장을 도울 수 있을 것으로 생각된다. KBS 미디어기술연구소에서도 '호크아이'라는 이름으로 콘텐츠의 오류 검사와 복원을 할 수 있는 시스템을 개발한 바 있다.

최근 대세인 딥러닝을 이용한 잡음(noise) 제거를 위해서는 잡음이 있는 영상과 깨끗한 영상을 모두 포함한 데이터셋이 함께 확보되어야 하는데, Nvidia에서는 잡음 영상 데이터만 가지고도 영상 복원을 수행하는 기술을 개발하여 화제가 되기도 하였다. 이와 같이 기술 발전이 매우 빠르게 진행되고 있어 향후 AI 응용이 기대되는 분야이다.



그림 6. Nvidia의 영상 복원 기술

메타데이터와 검색

메타데이터와 검색은 불가분의 관계로, 메타데이터가 있다는 것은 곧 검색이 가능하다는 의미이다. AI는 예를 들면 영상에서 슬픈 순간과 기쁜 순간을 찾아낼 수 있다. 이렇게 영상 콘텐츠에 부가정보인 메타데이터를 부착함으로써 콘텐츠의 가치는 올라간다. 메타데이터가 없는 영상은 아카이브에서 잠자고 있을 뿐이지만 인물, 사물, 배경, 행동, 감정 등 상세하고 다양한 메타데이터가 있는 영상은 방송 후 사라지지 않고 검색과 추천에 의해 제작과 서비스에 끊임없이 재활용되고 수익을 창출할 수 있다.

그렇지만 이러한 메타데이터 생성 작업을 사람이 직접 한다고 가정해 보자. KBS의 경우 60만 시간이 넘는 방대한 아카이브 콘텐츠를 일일이 찾아서 등장인물과 사물, 배경 및 장소, 등장인물의 행동, 감정과 분위기 등 메타데이터를 장면(scene) 단위, 샷(shot) 단위, 또는 초나 프레임 단위로 부착한다는 것은 전 직원과 가족들까지 총동원한다고 해도 불가능한 일이다.

이를 해결하기 위해 KBS 미디어기술연구소에서는 AI 기반 메타데이터 생성 연구를 수행하고 있다. AI는 전력만 공급해 주면 지치지 않고 주당 52시간 제한도 초과수당도 따지지 않은 채 밤낮없이 메타데이터를 뽑아낼 것이기 때문이다. 물론 아직은 AI의 인식률이 완벽하지 않기 때문에 사람의 메타데이터 검수가 필요할 수 있으나, 검색과 추천 등을 통해 제작과 서비스에 사용하는 시나리오는 어느 정도 오류를 감내하는 응용이므로 실령 완벽하지 않은 메타데이터라도 그 활용도는 높을 것으로 생각된다.



그림 7. KBS 미디어기술연구소에서 개발중인 인물/사물 분석 AI 시스템

최근 많은 글로벌 기업들이 클라우드를 통해 AI 영상 분석 기능을 제공하고 있다. 이를 이용하면 개발 기간을 단축할 수 있지만, 비용이 많이 들고, 수많은 한국의 방송 출연자와 장소 등을 가려내기에는 아직 현지화가 부족한 한계가 있다. 또한 최첨단의 AI 구현들을 github를 통해 오픈소스로 내려받을 수 있는 지금, 고성능의 AI 구현을 위해 꼭 필요한 핵심요소는 ‘알고리즘’이 아니라 AI 학습을 위한 ‘데이터셋’이라는 것이 많은 연구자들의 판단이다. 그렇다면 아카이브에 방대한 콘텐츠를 보유하고 있는 방송사는 고성능 AI 구현을 위한 핵심요소를 확보하고 있는 셈이다. 이 논의에 대해서는 2부에서 다시 다루고자 한다.

이번 호에서는 이슈가 되고 있는 방송 미디어를 위한 AI 기술의 개요 및 현황과 응용 일부에 대해 살펴보았다. 다음 호에서는 AI 기술이 방송 미디어 산업에서 어떻게 사용될 수 있는지 좀 더 구체적으로 살펴보려고 한다. 📺