창작이나눔이
Changjagi & Nanumi
KOREA COPYRIGHT COMMISSION

저작권기술 동향 2019-02호

저작권기술 동향

이슈 분석

최신 동향

인공지능 기술과 저작권

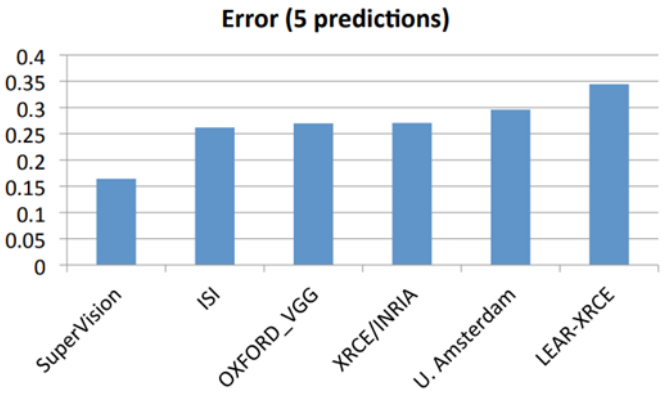
인공지능 기술과 현황

인공지능(Artificial Intelligence) 기술은 인간의 학습능력과 추론능력, 지각능력 등을 실현하는 기술을 뜻한다. 즉 컴퓨터 시스템 스스로 상황에 따라 판단, 의사결정 및 행동을 하는 기술을 의미한다. 1950년 앨런 튜링이 발표한 논문[1]에서는 인공지능에 대한 근본적인 시각을 정의한다. 기계가 '생각'을 할 수 있는지에 대한 판단을 위해 튜링은 문자로 사람이 기계의 의사결정 내용을 받아보고, 기계의 판단인지 사람의 판단인지 구분하지 못한다면 기계의 판단을 '생각'에 의한 것으로 정의할 수 있다고 말했다 [1]. 흔히 '튜링 테스트'로 알려진 이 실험 방법은 지금도 인공지능에 대한 판단의 근간으로 활용되고 있다.

1950년대 이후 인공지능은 수많은 시행착오를 통해 더딘 발전을 이루고 있었다. 그러던 중 2010년을 전후하여 발전한 딥 러닝[2]은 인공지능의 획기적인 발전을 가져왔다. 빅 데이터 처리기술의 발달, 정보처리(연산 및 저장) 능력의 향상, 클라우드 기반 환경의 조성 등이 딥 러닝 알고리즘을 인공지능에 적절히 활용할 수 있는 기반이 되었다. 인공지능은 특히 시각, 음성, 언어, 행동 지능 측면에서 비약적인 발전을 이뤄냈다 [2].

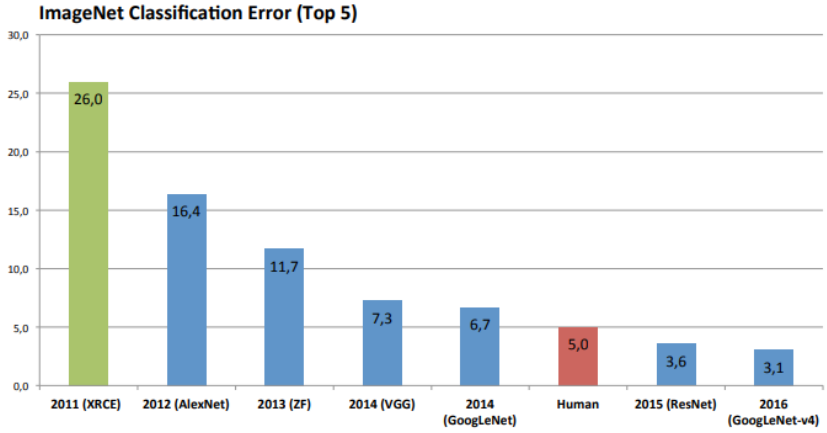
첫째 인공지능의 발전은 컴퓨터 시스템의 영상을 인식하는 능력을 크게 향상시켰다. 인공지능을 통한 영상 인식은 영상 데이터를 축적하여 데이터베이스화하고 데이터별 특징을 뽑아낸 후 그 특징과 유사한 특징을 가지는 영상을 찾는 과정을 통해 이루어진다. 이 과정의 정확도를 획기적으로 향상시킨 사건이 2012년 ImageNet[3]에서 일어났다. (그림 1)에 나타난 바와 같이 다른 팀보다 매우 낮은 오류율을 보이는 캐나다의 SuperVision 팀이 딥 러닝 알고리즘의 일종인 CNN(Convolutional Neural Networks) 알고리즘을 사용하였다는 사실이 알려지고, 이 알고리즘을 활용한 영상처리가 증가하면서 영상 인식률이 비약적으로 발전하였다.

1) A. M. Turing, Computing Machinery and Intelligence. Mind 49: 433-460, 1950
2) 기계학습 알고리즘의 일종으로 심층학습을 활용하여 인공지능의 성능을 크게 향상시킨 알고리즘
3) 사진 데이터에서 물체를 인식하는 성능을 겨루는 세계적인 경진대회, http://www.image-net.org/challenges/LSVRC/2012/



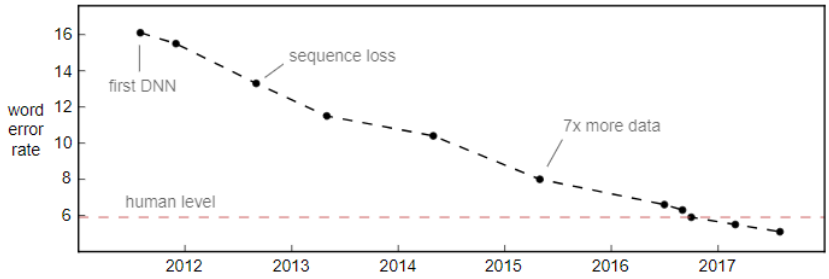
(그림 1) 2012년 ImageNet 팀별 영상인식 오류율 [4]

2012년 이후 ImageNet 경진대회는 딥 러닝 기법 기반의 시스템이 1위를 독식하게 되었다. 영상인식과 관련해 주요 애플리케이션인 얼굴인식 분야에서도 딥 러닝 알고리즘의 적용을 통해 비약적으로 발전하였다. (그림 2)에 나타난 바와 같이, ImageNet 우승팀들의 영상인식 성능은 2015년 처음 인간의 인식 능력을 넘어섰으며, 현재까지도 비약적으로 발전 중이다 [3].



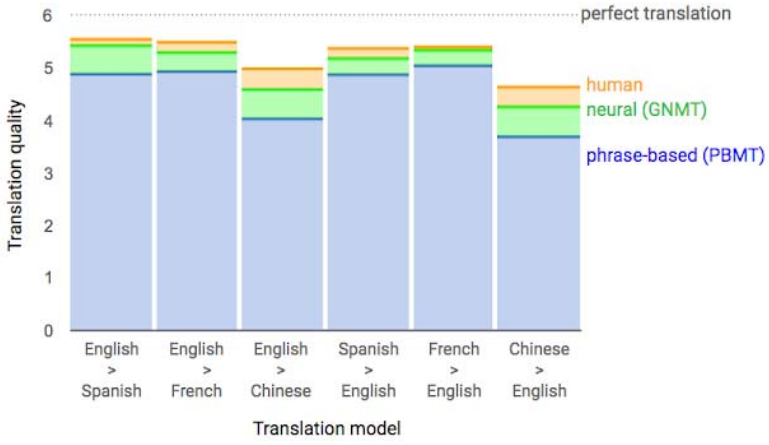
(그림 2) ImageNet 우승팀들의 영상인식을 통한 분류 오류율[4]

두 번째는 음성인식 지능영역이다. 음성인식에 대한 성능평가는 미국 표준 연구소(NIST, National Institute of Standards and Technology)에서 구축한 표준 데이터인 'Switchboard'라는 데이터의 활용을 통해 이루어지고 있다. 음성인식 오류율은 2009년까지 답보상태였으나, 2011년을 기점으로 크게 내려갔다. 딥 러닝 알고리즘의 적용이 성능 향상의 큰 계기가 되었는데, 적용 후 오류율이 2010년 15%, 2011년 9%까지 줄었다. 그리고 현재는 인간의 인식 오류율인 6% 이하로 그 성능이 향상되었다(그림 3).



(그림 3) 딥 러닝 알고리즘 사용을 통한 음성인식 오류율 향상 [5]

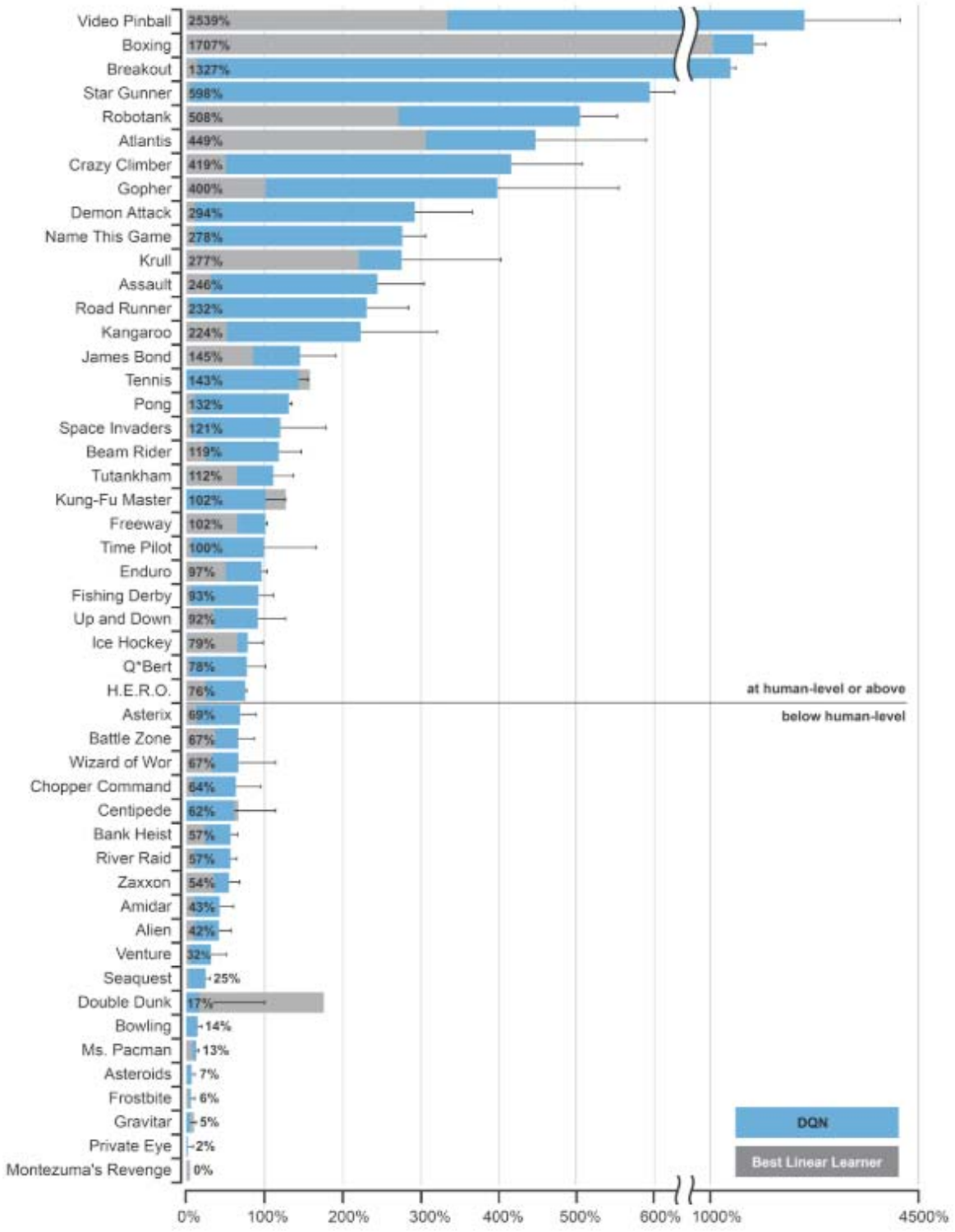
세 번째는 언어지능 분야이다. 문서 자동생성, 자동 번역 등에 활용할 수 있는 언어지능 역시 2010년 이전에 비해 크게 발전했다. 구글이 AI 블로그[4]를 통해 발표한 바에 따르면 인공지능에 의한 자동 번역 성능은 몇 가지 언어에 대해 인간의 수준에 다다르고 있다 [6]. 2019년에는 진짜 뉴스와 유사한 내용으로 가짜 뉴스를 만들어 내는 인공지능이 개발되어 사회 문제가 된 바도 있다 [7]. 인공지능이 생성한 문서는 이미 인간이 구분해 내기 어려운 수준에 가까워졌고, 인공지능을 활용한 소설, 시 등의 문학 창작도 시도되고 있다. 언어지능의 수준은 단순히 단어를 문법에 맞게 조합하는 수준을 넘어서, 단어 간의 의미 수준과 문법적 관계를 파악하여 의미가 있는 구문을 생성해 내는 수준이다. 이는 특정 분야에서는 인간이 언어를 사용하는 것과 유사한 수준으로 발전하였음을 의미한다.



(그림 4) 구글 번역기의 언어 번역 수준 [6]

마지막으로 행동지능의 발전은 불가능하다고 여겨졌던 바둑에서 구글의 '알파고'가 세계적인 바둑 기사들을 이기며 대중들에게 크게 회자된 바 있다. 구글에 인수된 인공지능 업체인 DeepMind의 2016년 글에 따르면, DQN(Deep Q-Networks) 알고리즘을 활용한 인공지능이 50여 개 게임에서 인간의 능력을 앞지르는 성능을 보여줬다 [8]. 인공지능의 결정능력은 이미 많은 분야에서 인간의 결정능력보다 우수한 성능을 보인다.

4) https://ai.googleblog.com/



(그림 5) 인간(회색)과 인공지능(파란색)의 게임 플레이 점수비교 [8]

인공지능 기술과 저작권

시각, 음성, 언어, 행동지능의 발전으로 크게 구분되는 인공지능 기술의 발전은 각 분야 저작권과 밀접한 관련을 가진다. 저작권법상에서 '저작물'은 인간의 사상 또는 감정으로 표현한 창작물로 정의하고 있다 [9]. 그러나 최근 인공지능의 발전으로 인간의 창작물과 인공지능 창작물의 구분이 어려워지고 있다. 예를 들어, 언어지능의 발전은 인간이 창작한 글과 인공지능에 의해서 창작된 글 사이의 구분이 모호해 문제를 일으킨다. 인공지능에 의해 창작된 글의 저작권이 존재하는지, 존재한다면 프로그램 소유자, 구동자 중 누구의 것인지 등에 대한 법·제도적 문제가 있다.

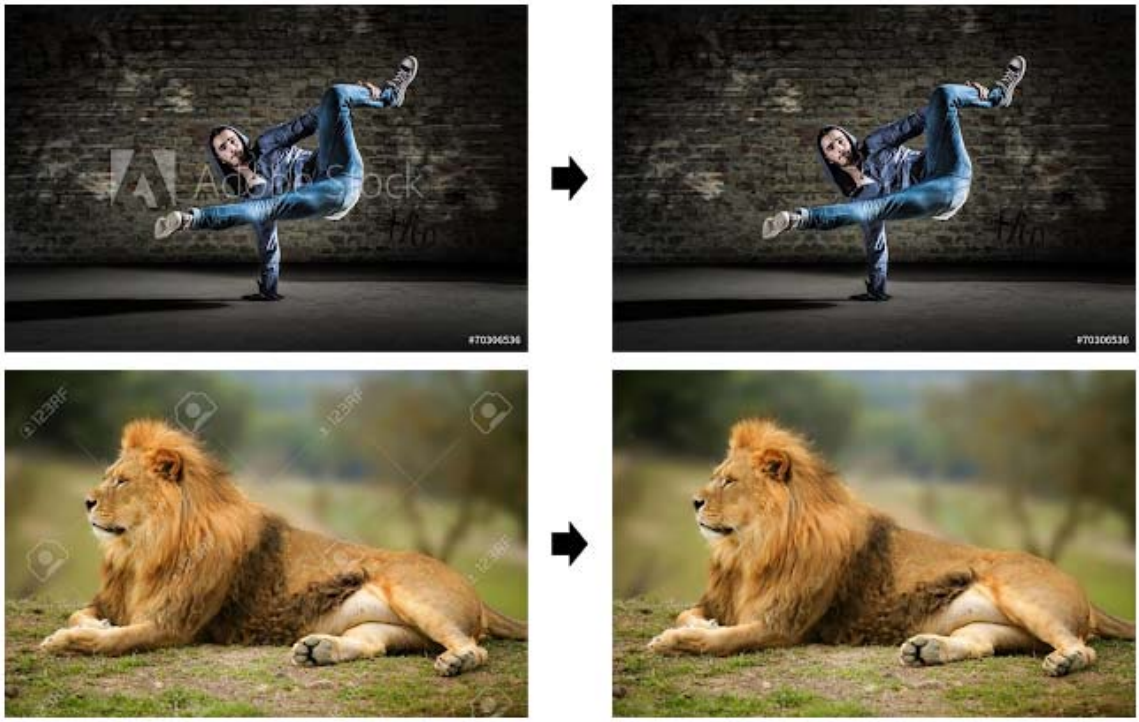
기술적으로도 인공지능 기술의 발전은 저작권과 밀접한 관계를 가진다. 기존의 저작권 보호 기술을 회피하기 위해 인공지능을 활용한 무력화 기술들이 속속 등장하고 있으며, 이 무력화 기술을 방어하기 위한 기술도 등장하고 있다. 방어 기술은 다시 무력화 기술의 등장을 유도하기 때문에 이는 지속적으로 문제가 될 수 있다.

구글의 워터마크 무력화 기술

2017년 구글은 기존 워터마크의 취약점에 대한 논문을 발표했다 [10]. 논문은 가시적으로 표시된 워터마크를 인식하여 이를 자연스럽게 삭제하는 기술에 대한 내용이다. (그림 6)에서 보이는 것처럼 기존의 워터마크는 구글이 개발한 기술에 의해 깔끔하게 지워진다[5].

일반적으로 인터넷에 공개되는 유료 사진들은 가시적인 워터마크를 표시하는 방법으로 무단 사용을 사전에 방지한다. 워터마크는 흰색이나 희미한 색이지만 육안으로 확인할 수 있다. 원본 사진에 추가되어 저작권으로 보호받는 사진임을 알리며, 구입을 통해서만 워터마크가 제거된 사진을 사용할 수 있다. 구글은 이러한 워터마크 사용 방식이 모든 사진에서 동일한 패턴으로 적용된다는 문제점을 발견했다. 그리고 수천 장의 사진을 수집하여 워터마크의 패턴을 찾아낸 후 이를 활용하여 워터마크가 포함된 사진에서 워터마크만 삭제할 수 있는 기술을 개발해냈다. 구글은 Adobe Stock, CanStock, Fotolia 등 유명 온라인 원본 사진 판매 업체들의 워터마크를 무력화할 수 있음을 실험을 통해 증명했다.

5)Making Visible Watermarks More Effective,
https://ai.googleblog.com/2017/08/making-visible-watermarks-more-effective .html, 2017



(그림 6) 구글의 워터마크 무력화 기술의 예 [10]

패턴을 분석하고 사진을 처리하는 과정에서 인공지능 기술이 활용되는데, 구글은 워터마크 제거 방법과 함께 제거하기 어렵게 만드는 방법도 소개했다. 그 방법은 워터마크 삽입 패턴을 찾기 어렵게 미세하게 차이가 나게 워터마크를 생성해서 삽입하는 것이다.

마인즈랩의 영상 저작권 보호 기술

국내에서도 영상 저작물을 보호하기 위한 인공지능 활용 기술이 개발되었다. 마인즈랩은 인공지능을 활용하여 불법 영상물 탐지를 수행하는 maum Visual[6]이란 솔루션을 개발했다. 기존에는 영상 주파수와 색상, 화소 정보 등을 이용해 필터링하는 방법이 있었으나, 화면의 좌우를 바꾼다거나, 화질을 변환시키거나, 임의의 마크를 영상에 추가하는 등의 방법을 통해 필터링을 피할 수 있었다. 마인즈랩은 이러한 한계를 극복하기 위해 딥 러닝 기반의 영상 기술을 활용했다. 인공지능 스스로 영상물의 분류 및 식별을 설정하여 영상의 변형과 무관하게 저작권 보호 대상인 영상을 찾아낼 수 있는 기술을 개발했다.

6) https://mindslab.ai/solution

(그림 7) 국내 인공지능 기반 영상 저작권 탐지 시스템 (maum Visual)

인공지능은 이외에도 유튜브 영상 저작권 체크를 위해 ContentID의 핵심 기술 중 하나로 활용되고 있으며, 국내 네이버 등에서도 영상인식을 활용한 영상 필터링 기술 등이 활용되고 있다. 또한, 음성인식 등을 활용하여 유사 음악 체크 등의 저작권 관리 기술 연구가 진행되고 있다.

턴인잇(Turninit)의 인공지능을 활용한 표절탐지

기사, 논문 등에서 표절은 오래전부터 문제가 되어왔다. 저작자의 견해나 의견이 담긴 글을 무단으로 표절하는 것은 저작권을 심각하게 침해하는 행위이다. 기존 표절탐지 기술의 한계를 넘어 인공지능을 사용해 표절탐지의 정확도를 높인 연구가 진행되고 있다. 기존 텍스트 매칭 기반 표절탐지의 경우 오탐지(False positive, 표절이나 표절 아닌 것으로 탐지되는 경우), 위음성(False negative, 표절이 아니나 표절로 탐지되는 경우)이 많고 표현 바꿔 쓰기, 복사 후 편집하기 등의 방법으로 시스템 탐지를 벗어나는 경우가 많았다.

인공지능을 활용한 표절탐지의 경우, 단순한 텍스트 매칭 만을 고려하는 것이 아니라 학습을 통해 문맥, 문단의 구성 등을 고려하기 때문에 기존 택스트 매칭 기반 표절탐지 시스템보다 훌륭한 성능을 보여준다. 턴잇인 이외에도 Copyleaks 등의 해외 기업이 인공지능을 활용한 표절탐지 기술 등을 개발하고 있다.

(그림 8) Copyleaks의 인공지능 기반 표절탐지 시스템

시사점

인공지능 기술은 최근 10년간 다른 어떤 분야보다 빠르게 발전하고 있는 분야이다. 딥 러닝 알고리즘과 빅 데이터 처리 시스템의 발전은 이러한 인공지능의 빠른 발전을 견인하고 있다. 인공지능 기술이 발전되면서 콘텐츠 저작권 보호와 저작권 침해 관련 문제가 동시에 대두되고 있다. 기존에 오랜 처리 시간이 소요되던 영상, 음성의 인식 영역에서 인공지능 기술을 활용하여 보호하고자 하는 콘텐츠의 저작권 침해 자료를 빠르게 찾아낼 수 있어, 콘텐츠를 보호할 수 있다. 동시에 인공지능 기술을 활용해 콘텐츠 보호를 위한 워터마크를 제거하는 등 콘텐츠 침해가 일어나기도 한다.

인공지능의 인식률과 처리 속도는 일부 분야에서는 이미 사람의 능력을 넘어섰다. 인공지능에 의한 창작, 표절 등이 발생해도 사람이 이를 인지하기 어려운 단계에 들어섰다. 이러한 문제를 해결하기 위해서는 사람이 인지하기 어려운 침해를 보다 강력하게 찾아낼 수 있는 기술 개발이 필요하다.

참고문헌

[1] A. M. Turing, Computing Machinery and Intelligence. Mind 49: 433-460, 1950

[2] 김병희, 장병탁, 딥 러닝: 인공지능을 이끄는 첨단 기술, Technical Report: BI-17-001, 2017

[3] Gustav von Zitzewitz, Survey of neural networks in autonomous driving, ADVANCED SEMINAR SS 2017: SURVEY OF NEURAL NETWORKS IN AUTONOMOUS DRIVING, pp 1-8, 2017

[4] http://www.image-net.org/

[5] Speech Recognition Is Not Solved, https://awni.github.io/speech-recognition/

[6] A Neural Network for Machine Translation, at Production Scale, https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html

[7] '진짜 같은 가짜뉴스' 만드는 AI 나왔다, http://www.zdnet.co.kr/view/?no=20190218113741&from=Mobile

[8] 6) Deep Reinforcement Learing, https://deepmind.com/blog/deep-reinforcement-learning/

[9] http://www.law.go.kr/법령/저작권법

[10] Tali Dekel Michael Rubinstein Ce Liu William T. Freeman, On the Effectiveness of Visible Watermarks, CVPR, 2017

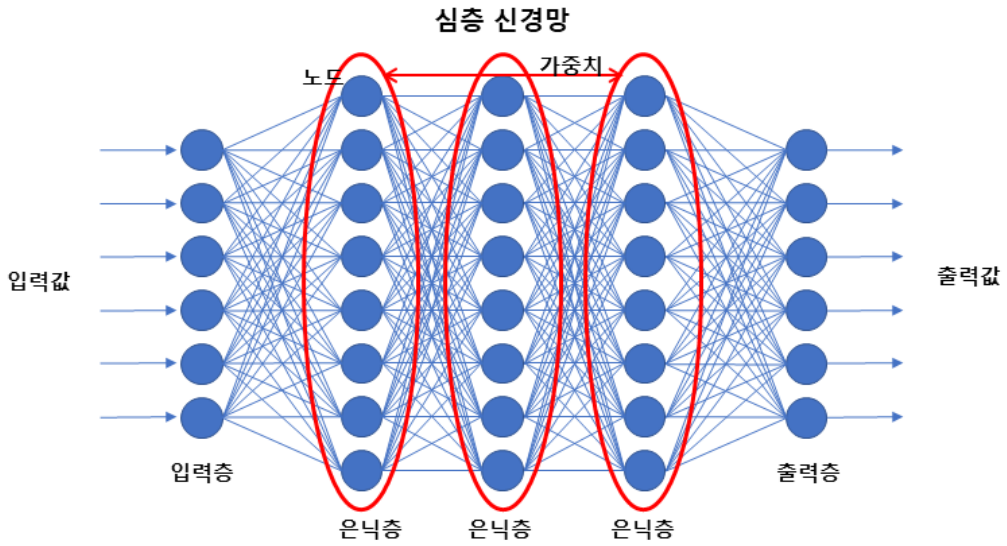
미국, 인공지능 시스템 모델 보호를 위한 디지털 워터마킹 기술

기술의 배경

지식 재산권(Intellectual Property, IP)은 인간의 지적 활동으로 인하여 발생하는 모든 재산권을 말한다. IP는 특허, 저작권 및 상표 등의 형태로서 법률로 보호되며, 법률은 사람들이 자신이 창작한 것에 대한 이익을 얻을 수 있도록 돕는다. 그러나 지적 재산을 보호하기란 쉽지 않다. 오늘날과 같이 정보의 유통이 급속하게 이루어지고 있는 시대에는 상당한 시간과 인력 및 비용을 투입하여 얻은 각종 정보, 기술 등이 쉽게 유출될 수 있다. 절차상의 허점이나 저작권 회피 기술을 활용해 권리를 침해하고자 하는 경우도 있다.

인공지능의 성능을 크게 향상시킨 심층 신경망(Deep Neural Networks, DNN)은 최근 정보, 의료, 자율주행차, 자동화 제조 등 다양한 분야에서 활용된다. 인공신경망(Artificial Neural Network, ANN)은 데이터 입력층(Input Layer)과 출력층(Output Layer) 사이에 여러 개의 은닉층(Hidden Layer)들로 이루어져 있다. 입력층에서 입력된 값은 은닉층을 구성하고 있는 각 노드들의 입력값으로 입력되고 노드별로 출력 결과를 얻어 다음 노드로 전달하는 과정을 거쳐 최종적으로 판단결과를 출력값으로 출력하는 방식이다(그림 1). 층별 구조가 복잡하고 가중치 설정이 어려워 개발에 오랜 시간과 비용이 소요된다. 최근 이러한 신경망 모델을 저작권 측면에서 보호하려는 움직임이 있다.

심층 신경망의 각 은닉층에서는 입력층에서 입력된 데이터를 판단하여 다음 은닉층으로 입력 값을 전달하는 작업이 일어난다. 각 노드(그림 1의 동그라미 부분)는 각각 다른 가중치를 가지고 판단을 수행한다. 데이터가 들어오면 각 노드에서 판단하여 다음 노드로 전달하는 작업이 반복적으로 수행되면서 사람과 유사한 방법으로 판단이 일어난다. 심층 신경망을 활용하기 위해서는 각 노드의 가중치를 변경하여 심층 학습 모델을 만들기 위한 학습 과정이 필요하다. 가중치 변경은 특정 판단의 중요도를 변경하는 과정이다. 심층 신경망 구조와 가중치의 설정은 인공지능이 정확한 판단 결과를 도출할 수 있도록 하는 가장 중요한 작업이다.



(그림 1) 심층 신경망의 구조

심층 학습 모델은 처음부터 올바른 출력값을 출력하는 것은 아니다. 반복적인 학습을 통해 노드의 설계, 노드의 가중치를 조절하며 올바른 출력값을 출력한다. 심층 신경망을 교육하려면 대규모 라벨링 된 교육 데이터[1] 세트 준비, 모델 교육을 위한 대량의 컴퓨터 리소스[2], 매개 변수 조정[3] 및 모델 아키텍처 설계[4]에 많은 비용이 소요된다. 데이터를 라벨링 하기 위해서는 초기에는 사람이 직접 라벨링을 수행하여야 한다. 심층 학습 모델의 학습을 위해서는 최소 수천 건의 학습 데이터가 필요하기 때문에 충분한 데이터를 확보하기 위해서는 라벨링 작업이 고비용의 작업이 된다. 또한, 학습 과정에 GPU 연산이 매우 많이 필요하기 때문에 고성능 시스템이 필요하다. 이러한 이유로 교육된 심층 신경망 모델 판매는 수익성 높은 비즈니스 모델이 될 수 있고, 해당 심층 신경망 모델은 지식 재산으로 보호해야 한다. 하지만 개발된 심층 신경망 모델은 쉽게 복사되고 재배포 될 수 있다는 문제가 있다.

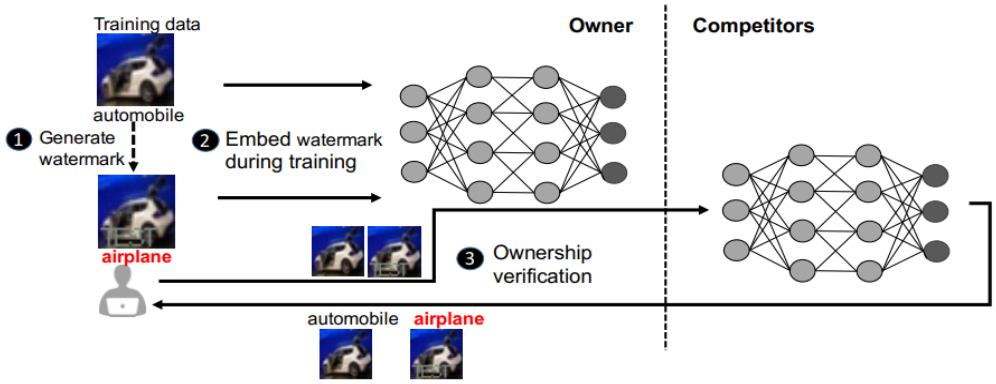
디지털 워터마킹 기술은 이미지와 같은 디지털 콘텐츠에 인간의 지각으로는 식별할 수 없도록 저작자의 정보를 삽입하는 기술로, 삽입된 워터마크를 통해 저작자의 정보 등을 확인할 수 있다. 디지털 워터마킹 기술의 프로세스는 크게 삽입과 검출 단계로 나눌 수 있다. 삽입 단계에서는 다양한 알고리즘에 따라 저작권자가 자신의 콘텐츠에 워터마크를 삽입한다. 검출 단계에서는 삽입된 워터마크를 추출하여 저작권자의 권리를 확인하고 보호한다. 이러한 워터마크를 심층 신경망에 적용하고자 하는 연구가 최근 진행되고 있다.

1) 원본 데이터에 사람이 먼저 판단을 해서 라벨을 입력해 놓은 데이터로 일일이 사람이 수작업을 통해 생성해야 함
2) 학습을 위해서는 복잡한 계산이 일어나는 신경망을 동작시키기 위해 다수의 GPU, 메모리 등이 필요
3) 각 심층 신경망 노드의 가중치를 변경하는 작업
4) 심층 신경망 배치를 조정해 가는 작업

심층 신경망 워터마킹 기술

최근 IBM 연구소에서는 심층 신경망 모델에 워터마크를 삽입하는 프레임워크를 제안했다 [1]. 워터마크를 삽입하는 과정을 요약하면 (그림 2)와 같다. 학습 데이터 셋을 구성하고 여기에 워터마크를 추가한다. 워터마크 추가 시 보호하고 싶은 모델의 라벨을 미리 설정된 다른 키워드로 교체한다. 그리고 워터마크와 라벨이 바뀐 데이터를 활용해 원래대로 트레이닝을 수행한다.

IBM의 심층 신경망 모델 보호를 위한 워터마크 활용기술을 요약하면 다음과 같다. 첫째, 저작권자만 알 수 있는 방법으로 라벨링된 학습 데이터를 이용하는 것이다. 예를 들어, 자동차를 비행기로 라벨링 해서 학습시키면 학습된 심층 신경망 모델은 자동차 사진을 비행기로 인식한다. 저작권자는 이를 알고 있기 때문에 출력값 비행기를 자동차로 바꿔주면 된다. (그림 2)는 IBM의 기법을 자세히 설명한다. (그림 2)에서 automobile의 라벨이 airplane으로 바뀐 것을 볼 수 있다. 라벨이 바뀐 상태로 트레이닝을 진행하면 자동차는 비행기로 인식될 것이고, 학습된 심층 신경망은 자동차 사진을 비행기로 인식한 결과물을 보여줄 것이다. 바뀐 라벨 정보는 심층 신경망의 저작권자만 알고 있기 때문에, 경쟁자(Competitors)가 심층 신경망을 불법적으로 사용하고자 해도 결과물이 정확하게 나오지 않는다.



(그림 2) IBM의 심층 신경망 워터마킹 동작 방식 [1]

둘째, 워터마크가 삽입된 학습 데이터를 사용하는 것이다. 심층 신경망 모델은 미세한 데이터 변화에도 완전히 다른 출력값을 도출한다. 따라서 워터마크가 삽입된 학습데이터를 충분히 확보하지 않으면 심층 신경망 모델을 재학습시킬 수 없다. 불법적 사용을 막기 위해 심층 신경망 저작권자는 일부러 특정 이미지를 입력하면 저작권 위반이라는 정보를 출력하도록 심층 신경망 모델을 학습시킬 수 있다. 심층 신경망이 저작권을 위반하여 사용한 경우 저작권자는 잘못 나오는 결과물을 알고 있으므로, 신경망 모델을 불법적으로 도용한 것인지 파악할 수 있다.

시사점

인공지능의 핵심이 되는 심층 신경망 모델도 보호받아야 할 저작물로 본다는 점에서 관련 연구는 다음과 같은 시사점을 준다. 첫째, 인공지능 자체를 창작물로 보고 있다는 점이다. 인공지능을 개발하기 위해서는 많은 사람의 노력과 큰 비용이 들어가지만, 만들어진 결과는 복사만 하면 되는 프로그램에 불과하다. 임의로 복제해서 사용한다면 이는 개발자의 노력으로 생산된 이익을 침해하는 것이다. 소프트웨어 개발 시 프로그램의 저작권 문제와 동일한 차원으로 생각하면 보다 이해가 쉽다. 둘째, 실험을 통해 얻어진 가중치나 프로그램의 흐름도 보호받아야 하는 정보라는 점이다. 심층 신경망의 각 노드가 가지는 가중치와 구조를 조절함으로써, 심층 신경망은 원하는 대로 동작하게 된다. 현재 이 작업은 실험을 통해 원하는 결과물이 나올 때까지 동작시켜보며 얻는 경우가 대부분이다.

심층 신경망 모델을 개발하는 비용과 노력을 고려했을 때 보호가 필요하지만, IBM의 기법을 활용해도 문제점을 완전히 해결하지는 못한다. 예를 들면, 심층 신경망 모델을 도용하기 위해서 학습 과정에서의 라벨링 정보와 워터마크 생성방법만 알면 심층 신경망 모델을 도용할 수 있다. 이를 근본적으로 해결할 수 있는 연구가 필요하다.

참고문헌

[1] Zhang, Jialong, Gu, Zhongshu, Jang, Jiyong, Wu, Hui, Ph. Stoecklin, Marc, Huang, Heqing , Molloy, Ian, Protecting Intellectual Property of Deep Neural Networks with Watermarking, pp. 159-172, 2018

WIPO, 2019년 인공지능 기술 관련 세계 특허 트렌드 분석

지식 재산권 분석과 인공지능의 발전

인공지능 기술은 최근 10년간 눈부시게 발전하고 있다. 인공지능은 다양하고 풍부한 디지털 데이터를 기반으로 의료, 자율주행, 첨단생산 등 다양한 분야에서 활용되고 있다. 인공지능의 최대 강점은 빅 데이터의 빠른 처리와 판단이다. 인간이 처리하기 어려운 많은 양의 데이터에서 다양한 패턴을 찾아내고 이를 기반으로 판단의 기준을 제공한다.

WIPO(World Intellectual Property Organization)는 최근 이러한 인공지능의 발전과 사용 증가에 대한 2019년 기술 동향 보고서[1]를 발간했다[1]. 보고서에는 인공지능에 대한 특허, 기술 동향 분석 등에 대한 내용이 포함되었다. WIPO 분석에 따르면 1950년대 인공지능의 개념이 소개된 이래로, 34만 건의 인공지능 관련 특허가 출원되었으며, 160만 건 이상의 과학 논문이 발표되었다. 이러한 추세는 최근 가속화되고 있는데 지금까지 출원된 특허와 논문의 절반 이상이 2013년 이후에 발표된 것이다. 인공지능 논문이 급증한 시기는 2001년경이며, 최근에는 논문 기반의 이론적 발전에서 상업적 사용으로 기술개발 동향이 바뀌고 있다. 이는 2010년 논문과 출원 특허 비율이 8대 1에서 2016년에 3대 1로 변화된 것을 보면, 이론적 연구보다는 실제 제품 개발을 위한 연구가 최근 증가되고 있다는 것을 알 수 있다.

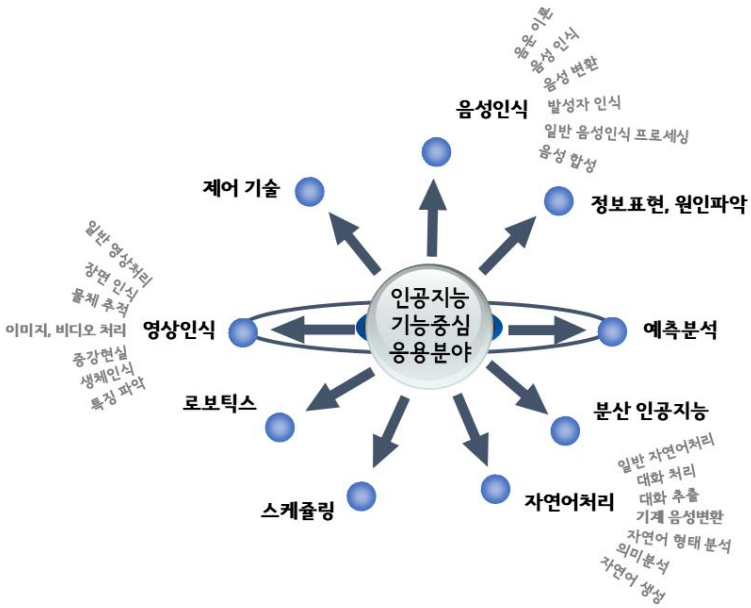
WIPO는 인공지능 관련 기술영역을 크게 세 가지 영역으로 분류하고 동향 분석을 진행하였다. 기계학습(Machine learning)과 같이 인공지능에 사용된 기술, 음성인식이나 영상인식과 같이 인공지능을 직접 활용하는 기술, 통신이나 운송과 같이 직접적으로 관련은 없으나 인공지능을 활용한 응용 프로그램 개발 기술이 그 세 가지 영역이다. 그리고 각각의 영역에 대해 지적재산권 동향, 주요 기술 개발주체 등에 대해 분석하였다.

1) 이 글은 WIPO 기술 동향 보고서 데이터를 바탕으로 작성되었음

인공지능 관련 기술동향

인공지능 기술의 발전의 핵심인 기계학습 분야는 최근 10년간 다른 인공지능 분야보다 빠르게 발전하고 있다. 기계학습 분야는 전체 인공지능 관련 특허 출원의 1/3 이상(13만 4천여 건)을 차지하고 있다. 기계학습 관련 특허는 2013년에 비해 28% 늘어나(2016년 기준) 20,195건의 출원이 이루어졌다. 기계학습 분야 중에서도 심층 학습과 심층 신경망 분야 특허 출원이 가장 빠르게 늘고 있다. 심층 신경망 분야는 1950년 인공지능의 개념이 발표된 이후 2010년까지 크게 발전하지 않았던 인공지능의 발전을 견인한 분야이다. 심층 학습 관련 특허는 2013년 대비 175% 성장률을 보이며, 2016년에만 2,399건의 특허가 출원되었다. 심층 신경망 기술에 대한 특허 성장률은 46%(2013년 대비)이며 2016년에는 6,506건의 특허가 출원되었다. 이렇듯 빠른 성장세는 심층 학습 관련 기술에 대한 연구가 활발하게 이루어지고 있으며, 인공지능 분야의 핵심이 심층 신경망 분야라는 것을 보여준다.

(그림 1)은 기능 중심으로 분류한 인공지능 응용 분야를 보여준다. 영상인식, 음성인식, 자연어처리를 중심으로 다양한 기술 분야가 존재하는 것을 볼 수 있다. 이 중 가장 많은 특허가 출원된 분야는 영상인식이다. 전체 인공지능 특허의 약 절만 가량을 차지하는 49%의 특허(167,038건)에 영상인식에 대한 내용이 등장하며, 매년 평균 24% 비율로 영상인식 관련 특허가 증가하고 있다. 2016년 한 해만 봐도 영상인식 특허는 21,011건이 출원되었다. 이는 음성이나 자연어처리보다는 영상 데이터를 활용한 인공지능 응용 프로그램이 많이 개발되고 있다는 점을 확인할 수 있는 부분이다.



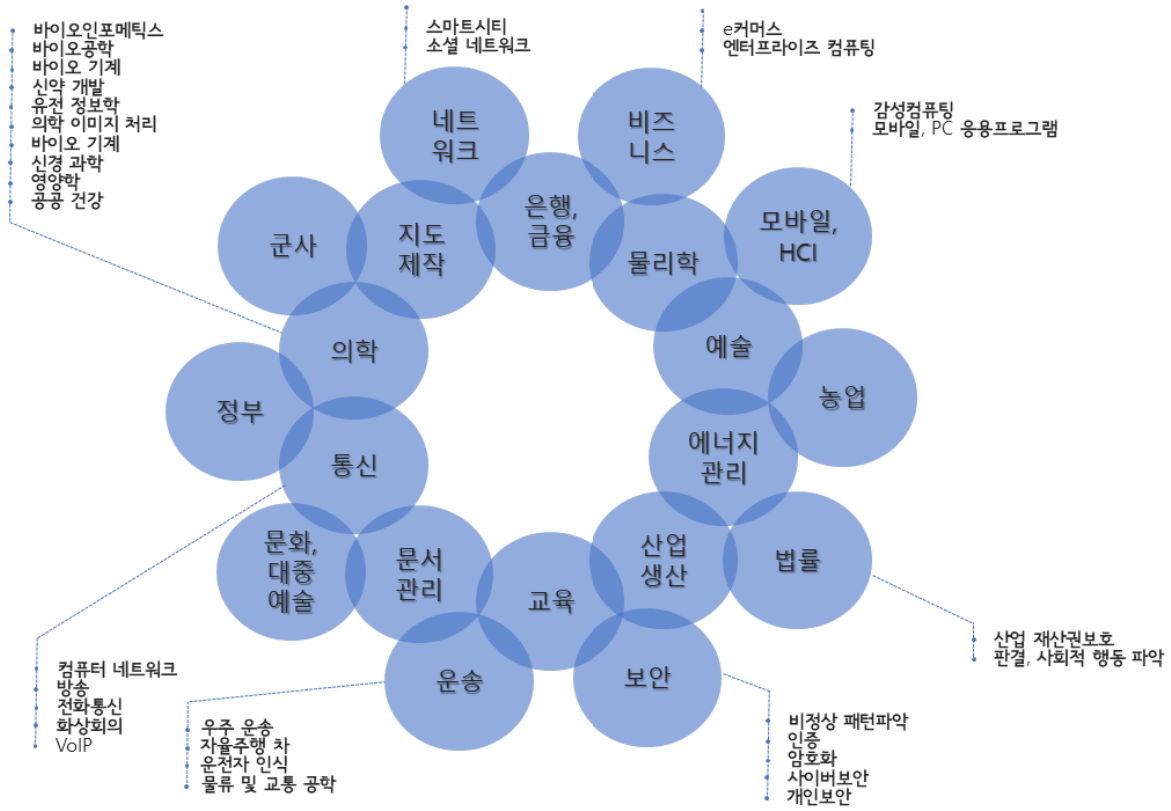
(그림 1) 기능 중심으로 본 인공지능 응용 분야 분류

인공지능을 활용한 응용 프로그램 개발 기술 분야는 총 20개 분야로 구분된다. 전체 응용 분야 중 통신과 운송 분야가 각각 15%로 가장 높은 비율을 차지하고 있으며, 의학과 생명공학 분야(12%), 개인 모바일 컴퓨팅 장치와 HCI(Human Computer Interaction) 분야 (11%)가 높은 비율을 차지하고 있다. 그 외의 분야로는 은행 관련 업무, 연예 오락 분야, 보안, 산업과 생산 분야, 농업, 소셜 네트워크, 스마트시티, IoT(Internet of Things) 분야 등이 있다.

이 중 운송 분야의 특허, 논문이 가장 큰 비중을 차지하고 있는 것은 흥미로운 부분이다. 이는 항공우주 관련 분야와 최근 각광받고 있는 자율 주행 차량 관련 분야의 연구가 활발히 진행되고 있기 때문으로 분석된다. 전체 운송 분야의 인공지능 기술 활용 특허는 2013년에서 2016년까지 연간 33%의 성장률을 보이고 있다. 그중에서도 항공우주 운송 분야가 연간 67%(2016년 8,764건)으로 가장 높은 성장률을 보이며, 그다음으로 자율 주행 차량 관련 인공지능 특허가 연간 42%(2016년 5,569건)의 성장률을 보이고 있다. 항공우주 분야의 경우 기존에도 인공지능의 활용이 높았던 분야로 최근 심층 신경망 등 관련 기술의 발달이 특허, 논문 발표의 증가세를 이끌고 있다. 자율 주행 차량 관련 분야의 경우는 향후 5년간의 발전이 더욱 기대된다. 차량 자체의 상용화 및 관련 서비스의 증가가 예상되는 가운데, 자율 주행 차량에서 활용되는 영상 데이터, 차량 자체 데이터, 인포테인먼트 데이터 등에 인공지능 활용이 증가할 것으로 예상되기 때문이다.

인공지능 관련 특허의 특징은 특허 가운데 70% 이상이 인공지능을 다른 분야와 결합하여 응용한다는 점이다. 자주 결합되는 분야는 컴퓨터 비전과 심층 학습, 컴퓨터 비전을 활용한 인공지능을 사용하는 운송, 통신, 보안, 자연어처리, 의학 파생 분야 등이다. 전체 결합 분야를 살펴보면 (그림 2)와 같다. 활용 분야 중 가장 높은 비중을 차지하는 분야는 운송이었고, 다음이 통신 분야였다. 통신 분야 중 컴퓨터 네트워크/인터넷 분야(17%), 라디오 및 TV 방송(17%)이 인공지능 활용에 대해 높은 성장률을 보였다. 그다음으로 높은 활용을 보인 의료 분야에서는 의료 정보학(18%), 공중 보건(17%)의 성장률이 높은 편이었다. (그림 2)에서 보면 사회 전반에 인공지능을 활용하기 위한 기술개발이 이루어지고 있음을 확인할 수 있다.

법률, 농업, 정부, 지도 제작 등 인공지능과 크게 상관없을 것 같은 분야에서도 인공지능의 활용은 진행되고 있다. 인공지능을 활용한 판결, 인공지능에 의한 자동화된 스마트 팜, 정책 결정에 활용하기 위한 인공지능 개발 등 다양한 분야에서 응용 프로그램에 대한 연구가 이루어지고 있음을 알 수 있다. 이는 향후 인공지능이 사회 전반에 걸쳐 모두 활용될 것이며, 그 활용 비율 또한 높아질 것으로 유추할 수 있다.



(그림 2) 인공지능 활용 분야

인공지능 관련 기술은 현재 미국, 일본, 중국이 시장 지배적인 위치를 차지하고 있다. 인공지능 관련 특허를 가장 많이 보유하고 있는 기관 30곳을 분석했을 때, 대학은 4곳, 기업은 26곳으로 분석되었다. 미국, 일본, 중국 중 가장 인공지능이 발전한 국가는 일본으로, 특허를 많이 보유하고 있는 기업 20곳을 뽑으면 그 중 일본이 12곳, 미국이 3곳, 중국이 2곳이다. 전체 85%가량의 다수 보유 업체가 미국, 일본, 중국에 분포할 정도로 세 국가에 기술개발이 편중되어 있다.

미국의 인공지능 선두 기업은 IBM과 마이크로소프트(Microsoft)로 각각 8,290건, 5,930건의 특허를 보유하고 있다. 상위 5개 업체를 살펴보면, 도시바(5,223건), 삼성(5,102건), NEC(4,406건)이다. 중국의 State Grid Corporation은 2013년부터 2016년까지 연평균 70%의 특허 출원 수를 늘려 상위 20위권에 포함되었다. 중국의 바이두는 심층 학습 분야 특허 출원에서 두각을 나타내고 있으며, 운송 분야에서는 도요타와 보쉬, 생명과 의학 분야에서는 지멘스, 필립스, 삼성 등이 두각을 나타낸다. 인공지능을 활용한 소셜 네트워크 관련 특허는 페이스북과 텐센트가 선두 그룹이다.

대학 분야에서 인공지능 연구의 선두 그룹은 중국으로 파악된다. 중국 대학은 인공지능 특허 분야의 상위 20명 중에서 17명을 보유하고 있다. 인공지능 관련 출판물 측면에서도 중국이 상위 20개 출판물 중 10개를 보유하고 있다. 중국은 심층 학습 관련 기술에 특히 강하다. 가장 선두에 있는 중국 과학 아카데미(CAS, Chinese Academy of Sciences)의 경우 2,500개가 넘는 특허와 20,000편 이상의 논문을 발표했다. 또한, 성과물의 양적 성장 측면에서도 매년 20% 이상의 성장률을 보이고 있다.

우리나라의 한국전자통신연구원(ETRI, Telecommunications Research Institute)은 중국의 CAS 다음으로 많은 인공지능 분야 특허를 출원하고 있다. 상위 500인의 특허 출원자를 분석한 결과 167개국의 연구기관이 포함되어 있으며, 이 중 110개 기관은 중국, 20개는 미국, 19개는 대한민국, 4개는 일본의 연구기관이었다.

시사점

인공지능은 새로운 콘텐츠 생산 분야에서 중요한 역할을 할 것으로 예상된다. 인공지능의 응용 분야를 봤을 때, 기능적인 측면에서 영상인식, 음성인식, 자연어처리 분야의 연구가 차지하는 비중이 매우 큰 것을 볼 수 있는데, 이 기술들은 모두 콘텐츠와 연관된 기술들이다. 활용분야 측면에서 보면 통신, 운송, 의학, 네트워크, 보안, HCI 등의 분야에서 인공지능을 많이 사용한다는 점이 눈에 띈다. 기능적 측면과 활용분야 측면을 연계해서 생각해보면 자율 주행차를 위한 콘텐츠 보호 기술이나, 의학 분야 생성 콘텐츠 보호, IoT 등에서 활용되는 콘텐츠 보호, 감성 콘텐츠 보호 등이 앞으로 새로운 이슈로 떠오를 확률이 높다는 것을 예측할 수 있다.

콘텐츠 생산뿐만 아니라 인공지능을 활용한 콘텐츠 보호도 앞으로 많이 활용될 것으로 예상된다. 인공지능을 활용한 예측, 원인 파악 등이 많이 연구되고 있는 기능으로 조사되었는데, 이를 활용한 콘텐츠 저작권 보호도 연구가 진행되어야 할 분야이다.

비록 미국, 일본, 중국이 인공지능 기술 분야를 이끌면서 전 세계 관련 기술 시장의 약 80%가량을 점유하고 있으나, 국내 ETRI가 중국의 CAS 다음으로 많은 인공지능 연구 결과를 내고 있다는 점은 고무적이다. WIPO의 리포트에서도 미국, 일본, 중국을 제외하면 유럽 전체보다 많은 인공지능 연구 결과물을 내는 나라로 한국을 꼽고 있다.

참고문헌

[1] WIPO Technology Trends – Artificial Intelligence, https://www.wipo.int/tech_trends/en/artificial_intelligence/

인공지능을 활용한 딥 페이크 기술의 문제와 저작권기술

딥 페이크 기술의 문제

미국의 콘텐츠 제작사인 버즈피드(BuzzFeed)[1]는 2018년 4월 버락 오바마 미국 대통령의 동영상을 유튜브에 게재했다[1]. 영상 속 버락 오바마는 백악관 집무실을 배경으로 트럼프 대통령, 미국 정책 등을 비난한다. 해당 영상은 조작된 영상으로 (그림 1)의 오른쪽 사람[2]이 하는 말을 오바마 대통령의 영상과 합성하여 생성한 영상이다. 버즈피드 측은 영상 합성을 위해 '딥 페이크(Deepfake)' 기술을 이용했다고 밝혔다. 이는 시청자가 영상의 조작을 알아채지 못할 정도로 입 모양, 표정, 동작 등이 자연스럽다.

(그림 1) 딥 페이크에 의해서 생성된 버락 오바마의 발언 조작 화면 [1]

1) https://www.buzzfeed.com/
2) 영화감독 조던 필(Jordan Peele)

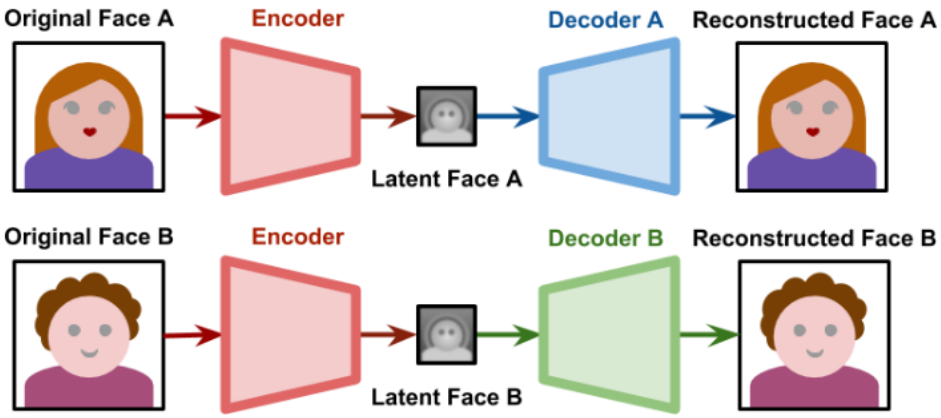
딥 페이크 기술은 영상에 등장하는 사람의 얼굴을 조작하고 음성을 입혀 가짜 영상을 만드는 기술이다. 보통 입 모양과 하는 말이 다르면 어색하고 조작된 영상으로 인식되는데, 딥 페이크 기술을 이용하면 입 모양, 목소리, 억양 등까지 조작이 가능하다. 딥 페이크 영상을 만들기 위해서는 조작하고자 하는 사람의 영상이 담긴 15초 분량의 원본 영상과 대중에게 이미 공개된 오픈소스 소프트웨어[3]만 있으면 되기 때문에 문제는 더욱 심각하다. 유명인의 포르노 영상이나 외교적으로 민감한 발언에 대한 영상 등이 딥 페이크 기술을 이용해 등장하면서 이 문제는 더욱 심각한 문제로 대두되고 있다.

딥 페이크 기술 분석

딥 페이크 기술은 일종의 이미지 조작 기술이다. 기계학습 알고리즘의 일종인 딥 러닝(Deep learning)을 사용해 원본 이미지나 동영상에 다른 영상을 결합하여 조작 영상을 만들어 낸다. 딥 페이크 기술을 적용하기 위해서는 크게 추출, 학습, 병합의 세 단계가 필요하다[2]. (그림 1)에 등장하는 오바마 대통령과 조던 필 감독의 영상을 각각 추출하여 얼굴 부분을 찾고, 눈, 코, 입 등 학습에 필요한 위치를 찾는다. 더욱 정교한 가짜 영상을 위해서는 조작 대상이 되는 사람의 영상을 더 많이 수집하면 된다. 이 경우 더 많은 학습을 진행할 수 있기 때문에 정교한 조작 영상을 생성할 수 있다. 지금까지 공개된 딥 페이크 영상들이 대부분 유명 정치인, 연예인 등을 대상으로 하는 이유가 이것이다. 정치인이나 연예인의 영상은 이미 인터넷에 많이 공개되어 있어서 이들의 영상 조작을 위한 데이터 확보가 용이하다.

(그림 2)는 확보된 영상의 학습 내용을 보여준다[3]. A의 원본 얼굴과 B의 원본 얼굴을 다양한 각도, 조명, 표정 등에 대해 인공지능을 학습시킨다. 원본 영상은 인코더를 통해 부호화되고, 부호화된 데이터는 분석이 가능한 디지털 데이터로 변환되어 처리된다. 학습을 통해 인공지능은 부호화된 얼굴의 잠복면(Latent face) 데이터를 찾아내게 된다. 잠복면 데이터는 딥 페이크 기술에서 중요한 역할을 한다. (그림 2)에서는 잠복면 데이터가 얼굴처럼 표현되어 있기는 하지만, 실제 잠복면 데이터는 인코딩을 거친 데이터이기 때문에 의미 없어 보이는 데이터일 수도 있다. 그러나 잠복면 데이터는 사람의 얼굴이 가지는 기본적인 특징을 모두 포함한다. 얼굴의 잠복면 데이터를 활용해서 같은 사람의 얼굴을 디코딩하면 원래 사람의 얼굴이 나온다. 인코딩과 디코딩 과정에서 데이터 손실이 발생할 수 있기 때문에 재생성된 데이터는 원본과 동일한 화질의 영상은 아닐 수 있다.

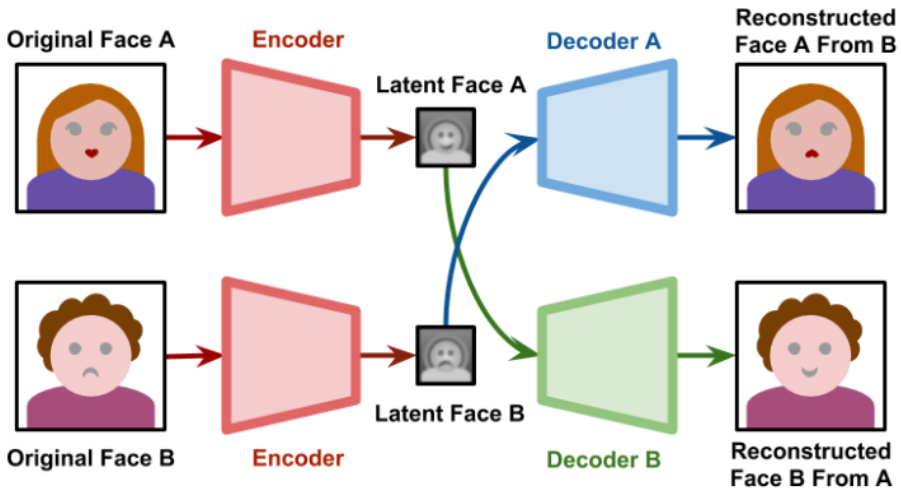
3) 누구나 사용하고 수정할 수 있도록 개발자가 공개한 소프트웨어



(그림 2) 딥 페이크를 위한 학습 과정 [3]

잠복면 데이터가 확보되면 (그림 3)과 같이 디코더를 바꿔서 잠복면 데이터를 입력한다. A라는 사람의 특징을 가지고 B라는 사람의 얼굴을 복원하고 B라는 사람의 얼굴을 가지고 A라는 사람의 얼굴을 복원하는 과정이다. 이 과정을 거치면 두 사람의 얼굴이 바뀌게 된다. 이 과정이 많은 학습을 통해 이루어지고 잠복면 데이터의 정확도가 올라가면 얼굴이 바뀐 것을 사람이 파악하기 힘들어진다.

마지막으로 병합 과정을 거치게 되는데 병합 과정은 합성한 이미지를 더욱 자연스럽게 보일 수 있도록 결합시키는 과정이다. 색상 보정이나 경계선 보정 등 후 작업이 이 단계에서 이루어진다.



(그림 3) 딥 페이크 기술을 활용한 얼굴 합성 방법 [3]

현재 딥 페이크 기술은 몇 가지 한계를 지니고 있다. 첫째, 인공지능 학습에 기반을 두고 있기 때문에 충분한 양의 학습 데이터가 필요하다. 일반적으로 최소 300개에서 2,000개 이상의 대상에 대한 이미지 데이터가 필요하다. 둘째, 학습된 대상에 대한 이미지 조작이 가능하지만 잠복면을 추출하지 않은 데이터에 대한 조작은 불가능하다. 셋째, 원본 영상에 얼굴 이외의 물체가 없어야만 원하는 영상을 얻을 수 있다.

딥 페이크 기술에 대한 대응

딥 페이크 기술에 대한 대응으로 연구되고 있는 내용은 주로 인공지능을 활용한다. 딥 페이크 영상이 가지는 특징을 검출하기 위해 인공지능을 사용하는 것이다. 딥 페이크 영상의 경우 눈 깜빡임 같은 행동이 잘 나타나지 않거나 지나치게 불규칙하게 발생하는 경향이 있는데, 이를 찾아내는 방법이 지금까지 주로 보고된 내용이다. 미국 올버니대학교(State University of New York at Albany)의 시웨이 류(Siwer Lyu) 교수는 딥 페이크 영상에서 조작된 사람의 얼굴은 정지 이미지를 활용하여 학습하기 때문에 눈 깜빡임이 부자연스럽다는 내용을 발표했다[4]. 기사에 따르면 시웨이 류 교수는 이 방법을 활용해 조작된 영상을 95%까지 탐지했다고 한다.

학계 뿐만 아니라 기업에서도 딥 페이크 탐지를 위한 기술이 개발되고 있다. 미국의 영상 데이터 처리업체인 지피캣(Gfycat)[4]은 딥 페이크에 의해 생성된 불법 콘텐츠를 찾아내는 인공지능 솔루션을 개발했다[5]. 지피캣은 딥 페이크가 인코딩, 디코딩을 거치면서 해상도나 화질이 나빠지는 특징을 이용했다. 지피캣은 낮은 해상도의 영상을 찾아내는 앙고라 프로젝트(Angora project)와 사람의 얼굴을 자동으로 인식하여 태깅하는 마루 프로젝트(Maru project)의 기술을 활용했다. 만약 특정 영상의 해상도가 낮고, 웹상에서 더 고해상도의 영상을 찾는다면 그 영상은 딥 페이크로 조작된 영상일 수 있다. 또한, 특정인의 얼굴로 인식된 영상이 다른 그 사람의 얼굴을 포함한 영상과 차이가 난다면 조작 영상으로 판별하는 것이다.

블록체인을 활용한 딥 페이크 탐지 기술도 등장했다. 미국 샌프란시스코의 스타트업 앰버 사에서는 영상이나 이미지가 생성될 때 동시에 해당 영상이나 이미지에 대한 해시(Hash)를 생성한다[6]. 생성된 해시 데이터는 블록체인에 기록되고 원본 여부를 파악하기 위해 사용된다. 원본 데이터가 변형되면 원본 데이터 정보와 새롭게 생성된 해시값이 블록체인에 다시 기록되어 원본 비교에 활용되는 방식이다. 만약 영상이 변형되면 변형된 영상 테두리에 색을 (그림 4)처럼 바꿔 조작된 영상임을 인지하도록 알려준다. 일반적인 개인 촬영에서는 활용하기 어렵겠지만, (그림 1)에서처럼 백악관에서 촬영된 대통령 영상 같은 경우에 활용하면 원본 영상임을 보장할 수 있다.

4) https://gfycat.com/ko/

(그림 4) 앰버 사의 딥 페이크 영상 검출 결과 [6]

지금까지 다양한 탐지 기술이 개발되고 있지만, 각각의 기술들은 모두 단점이 있다. 눈 깜빡임 문제는 영상 재생시간 조정이나 얼굴의 옆모습만 바꾸거나 하는 방법이 활용될 수 있다. 그리고 원본 영상의 해상도가 낮을 경우에는 지피캣의 방법으로도 탐지가 불가능하다.

시사점

딥 페이크 기술은 원본 영상의 조작 결과를 사용자가 인지하기 어렵다는 점에서 영상 데이터 위변조에 쉽게 악용될 수 있는 기술이다. 이는 저작권 문제, 초상권 침해 문제 등 다양한 사회 문제를 야기시킬 수 있다는 점에서 우려할 만하다. 단순히 개인의 피해나 콘텐츠의 피해뿐만 아니라 영상 콘텐츠에 대한 신뢰의 문제를 유발할 수도 있다.

최근 급증하고 있는 유튜브, 트위치, 아프리카 TV 등 개인 방송이나 스트리밍 플랫폼 콘텐츠들의 경우 딥 페이크 문제에 더욱 취약할 수 있다. 이러한 콘텐츠의 저작권 위반을 모니터링하기 위해서 인공지능을 활용한 얼굴인식이 많이 사용되는데, 얼굴인식 기반 저작권 모니터링 기술이 무력화될 수도 있다. 딥 페이크를 위한 소프트웨어나 하드웨어가 누구나 손쉽게 구할 수 있다는 점에서 문제는 더욱 심각해진다.

지금까지 발생한 딥 페이크 관련 문제들은 가짜 뉴스 전파 등이 있다. 기술이 보급되기 시작한 것이 2년 남짓 되었다는 것을 생각하면 앞으로 발생할 문제는 더욱 다양하고 복잡해질 것이다. 딥 페이크 기술이 단점만 있는 것은 아니지만, 이를 활용한 콘텐츠 위변조 문제는 해결하기 어려운 저작권 문제와 함께 다양한 사회 문제를 양산할 수 있으므로, 이를 기술적으로 검출하고 방지하기 위한 노력이 필요하다.

참고문헌

[1] You Won't Believe What Obama Says In This Video!, https://www.youtube.com/watch?v=cQ54GDm1eL0

[2] 최순욱, 이소은, 딥페이크와 사실의 위기: 어떻게 대응할 것인가?, 한국언론진흥재단, 2019.

[3] https://www.alanzucconi.com/2018/03/14/understanding-the-technology-behind-deepfakes/

[4] https://academicminute.org/2018/12/siwei-lyu-university-at-albany-detecting-deepfake-videos/

[5] https://www.vice.com/en_us/article/ywe4qw/gfycat-spotting-deepfakes-fake-ai-porn

[6] https://ambervideo.co/#