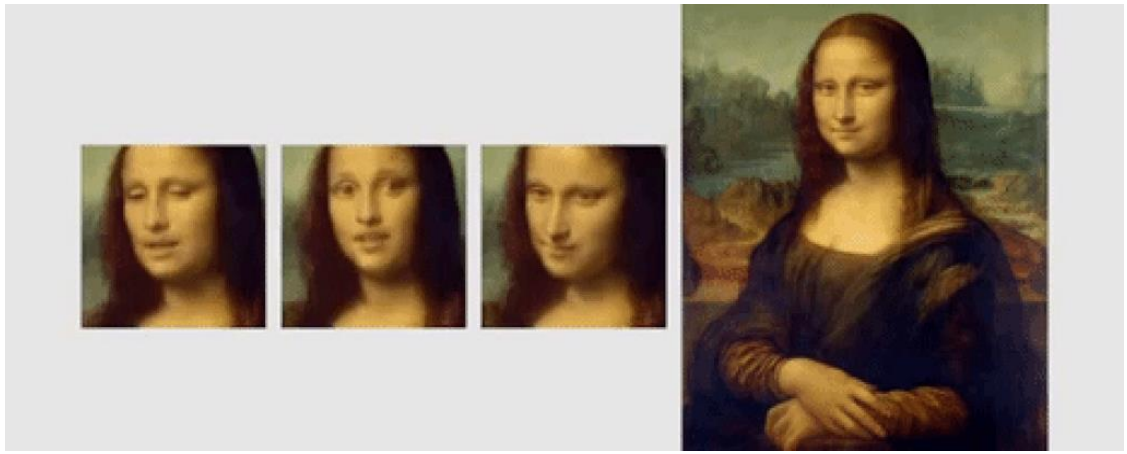
인공지능 기술에 대한 IP 보호

2020. 9. 4.

전정현 변호사
국가지식재산위원회 AI-IP특별전문위원회 전문위원

기존 논의 - Creative AI에 초점

- 인공지능에 의한 발명이 특허 대상이 되는지?
- 인공지능 창작물이 저작권 보호 대상이 되는지?

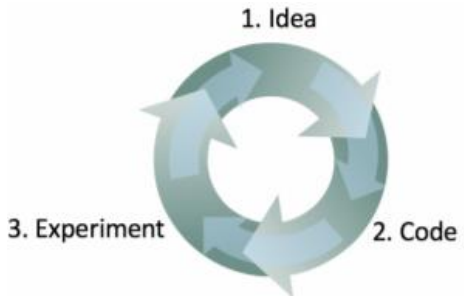


인공지능의 창작물, 지식재산권은?

인공지능 개발 과정과 그 산출물

데이터 수집

학습 전 모델
(network structure, hyperparameters)



개발 과정

학습된 모델

개발 결과

인공지능 기술에 대한 IP 보호 - 실무상 쟁점

저희 연구소가 개발한 AI 알고리즘이 특허 대상이 되나요?

경쟁사가 저희 학습된 AI 모델을 도용하지 못하게 하려면 어떻게 하나요?

인공지능 기술에 대한 IP 보호 - 실무상 쟁점

저희 연구소가 개발한 AI 알고리즘이 특허 대상이 되나요?

경쟁사가 저희 학습된 AI 모델을 도용하지 못하게 하려면 어떻게 하나요?

알고리즘 특허

Google - Dropout 특허

- Dropout - 인공신경망 학습에서의 과적합(overfitting)을 막기 위해, 학습단계에서 선별적으로 일부 노드의 학습을 막는 것
- 제프리 힌튼 등 저명 AI 연구자에 의해 제안되어 널리 활용 중
- AI Community는 Dropout 특허에 대해 큰 우려를 제기

System and method for addressing overfitting in a neural network

Abstract

A system for training a neural network. A switch is linked to feature detectors in at least some of the layers of the neural network. For each training case, the switch randomly selectively disables each of the feature detectors in accordance with a preconfigured probability. The weights from each training case are then normalized for applying the neural network to test data.

Images (3)

US9406017B2
United States

Download PDF | Find Prior Art | Similar

Inventor: Geoffrey E. Hinton, Alexander Krizhevsky, Ilya Sutskever, Nitish Srivastva

Current Assignee : Google LLC

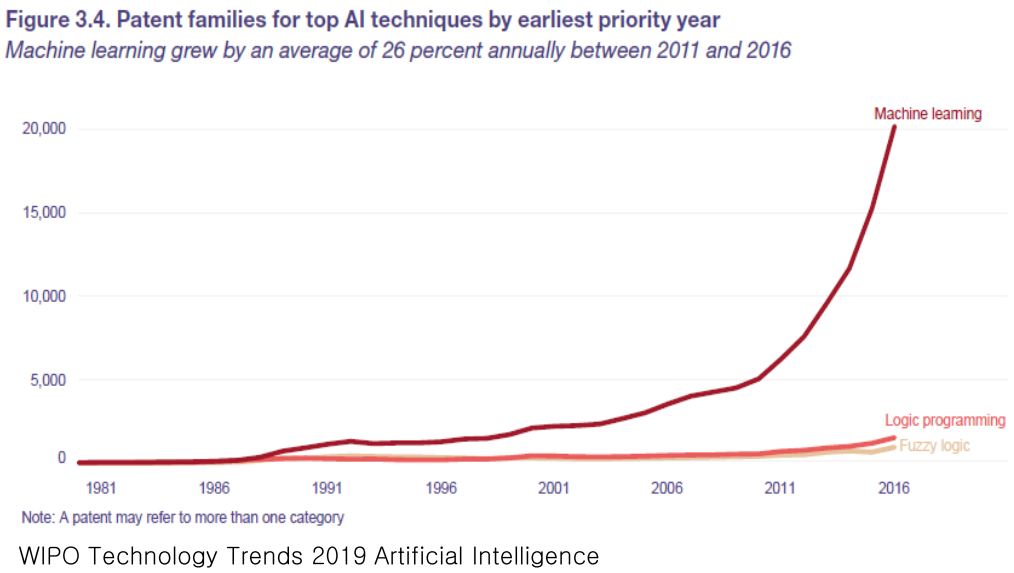
Worldwide applications

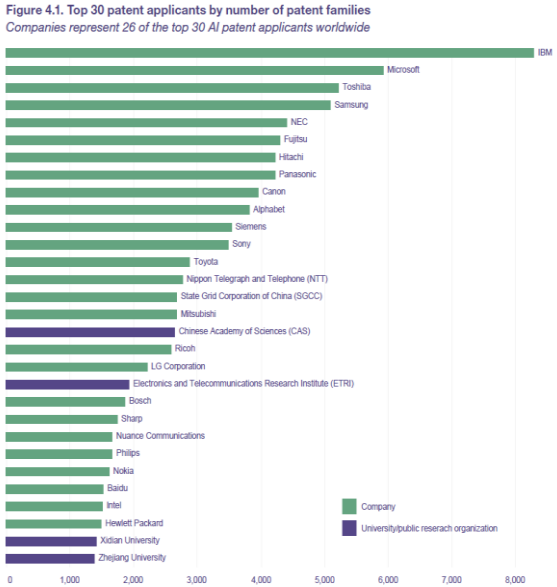
2013 • US WO BR ~~AU~~ 2016 • US

인공지능 특허의 문제

인공지능 특허의 급속한 증가(WIPO 2019)

- 2011년 이후 인공지능 특허는 매년 평균 26%씩 증가
- 민간 기업에 의한 특허가 대다수를 차지





AI 핵심 기술에 대한 특허

- Recurrent Neural Networks (RNN) (WO2018083669), Word2Vec (US 9037464)등 핵심 기술에 대한 특허 존재 → 향후 AI 특허 분쟁이 예상됨

특허와 오픈소스 소프트웨어

AI 연구자의 질문

- “저희는 구글에서 제공하는 TensorFlow API를 사용합니다. TensorFlow는 오픈소스로 공개되어 있으니, 구글이 개발한 API를 사용하면 Google 특허 침해 문제는 없게 되는 것인가요?”

TensorFlow 오픈소스 라이선스

- TensorFlow 오픈소스 라이선스는 코드 작성자(Contributor)가 제공한 부분에 관해 코드 작성자가 보유한 특허에 대한 라이선스를 부여
- 따라서 TensorFlow API 사용시 구글의 특허 침해 문제는 없게 됨

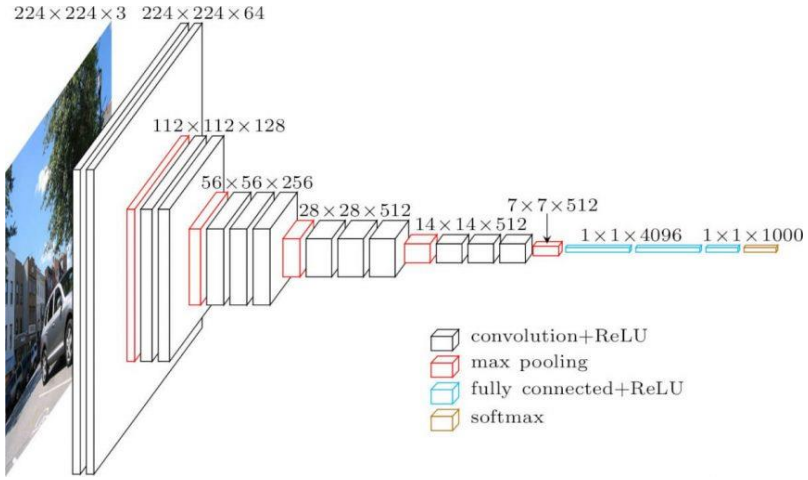
3. Grant of Patent License. Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable (except as stated in this section) patent license to make, have made, use, offer to sell, sell, import, and otherwise transfer the Work, where such license applies only to those patent claims licensable by such Contributor that are necessarily infringed by their Contribution(s) alone or by combination of their Contribution(s) with the Work to which such Contribution(s) was submitted.

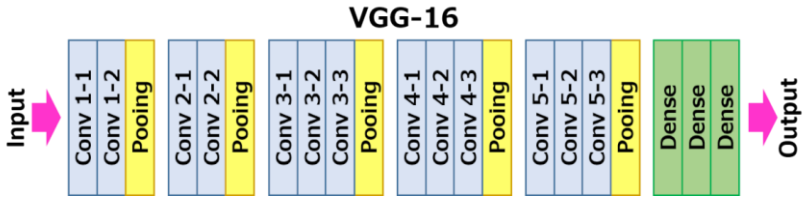
인공지능 기술에 대한 IP 보호 - 실무상 쟁점

저희 연구소가 개발한 AI 알고리즘이 특허 대상이 되나요?

경쟁사가 저희 학습된 AI 모델을 도용하지 못하게 하려면 어떻게 하나요?

학습된 모델 (Trained Model)





학습 데이터

⇩

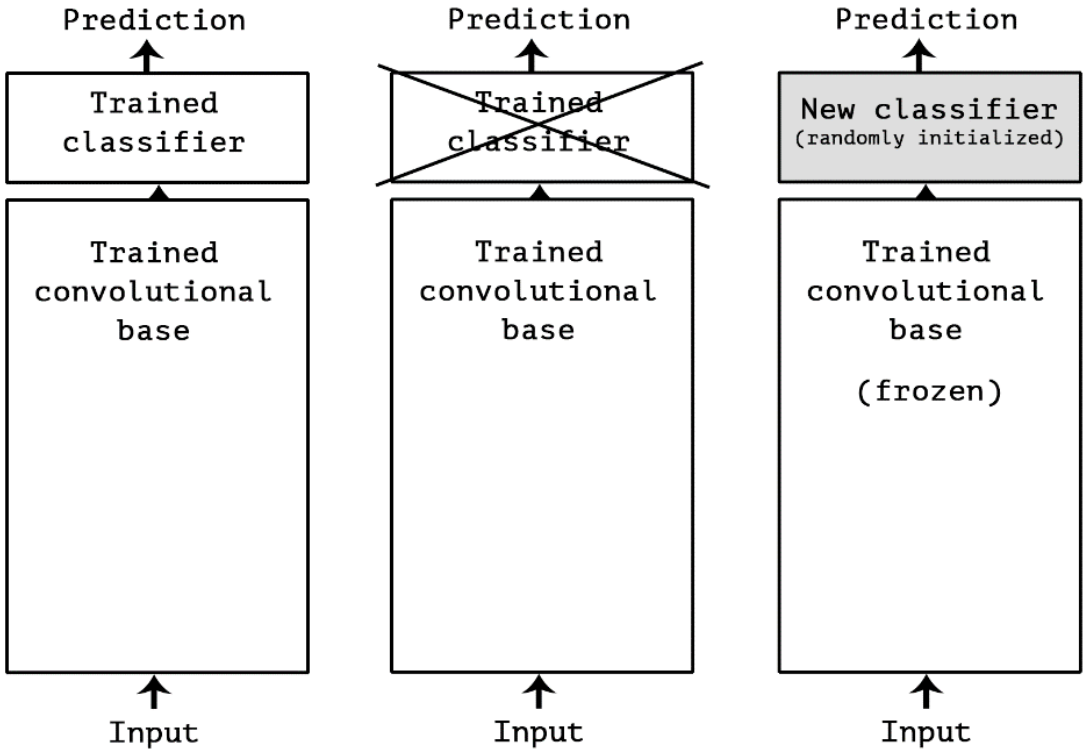
인공지능 네트워크

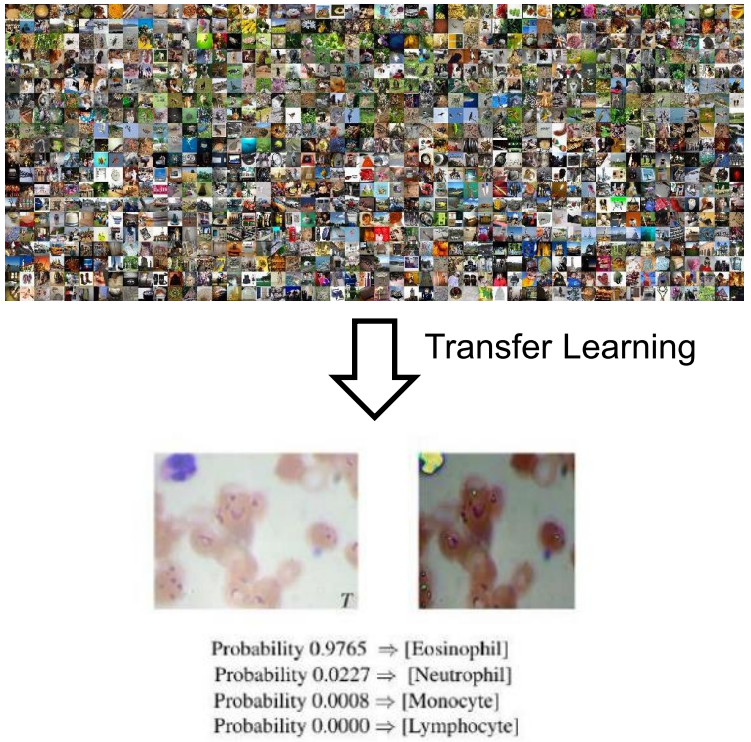
⇩ 수십 시간~수 주일 학습

학습된 모델 (.h5 파일 등)

학습된 모델의 가치

- 특정 용도로 잘 훈련된 AI Network는 관련 용도로 쉽게 재활용(Transfer Learning) 가능
  - 사물 인식 AI 모델 → 얼굴 인식 / 의료영상판독
  - 문장 연관성 판독 AI 모델 → 기계번역
- 학습된 모델(AI Brain)의 경제적 가치가 높음





학습된 모델의 보호 (AI-Stealing Attack)

가상 사례 (1) - 학습된 모델 파일이 이용자에게 배포되는 경우

- 자율주행 소프트웨어 개발 업체인 A 회사는 운전자 수십 명의 수 년간 운전 기록을 확보한 다음, 여러 달의 학습 기간을 거쳐 다양한 환경에 대처할 수 있는 인공지능 모델을 개발함
- 위 인공지능을 탑재한 자율주행차가 출시되자 경쟁사인 B회사는 위 인공지능 모델 파일을 추출하여 A회사의 모델을 파악한 다음(구조 및 Decision Boundary등), 자사의 인공지능 개발에 활용함
- A회사는 B회사의 인공지능 역분석이 컴퓨터프로그램저작권 침해라고 주장함
- B회사는 (1) 학습된 모델은 인간의 창작물이 아니므로 저작권의 보호 대상이 아니고 (2) 만약 저작권의 보호 대상이라고 하더라도 A회사 제품의 기능을 모방하였을 뿐이므로 정당한 역설계라고 주장함

누구의 주장이 맞을까?

학습된 모델의 저작권 보호

학습된 모델은 컴퓨터프로그램저작물인가?

저작권법 제2조(정의)
16. "컴퓨터프로그램저작물"은 특정한 결과를 얻기 위하여 컴퓨터 등 정보처리능력을 가진 장치(이하 "컴퓨터"라 한다) 내에서 직접 또는 간접으로 사용되는 일련의 지시·명령으로 표현된 **창작물**을 말한다.

- **긍정설**
  - **학습된 모델을 생성하기 위해 학습 데이터의 선정, 학습 방법의 선정 등 인간의 창작 행위가 개입됨**
  - **기계만이 독해 가능한 실행파일이나 폰트 파일도 컴퓨터프로그램저작권의 대상이 되므로 마찬가지로 보아야 함**

- **부정설**
  - **저작권의 보호 근거는 인간의 창작 행위에서 기인**
  - **그런데 학습된 모델 자체(.h5 파일 등)은 프로그램의 실행 과정에서 파생된 데이터에 불과하고 인간이 창작한 것이 아님**
  - **다만, 비밀로 관리된 경우에 한해 학습된 모델은 영업비밀로서 보호**

역분석과 (AI-Stealing Attack)

프로그램코드역분석

- 배포된 컴퓨터 실행파일을 원래 소스코드로 변환하는 행위
- 다른 프로그램과의 호환에 필요한 부분에 한하여 허용

AI-Stealing Attack

- 배포된 컴퓨터 실행파일을 원래 소스코드로 변환하는 행위
- 다른 프로그램과의 호환에 필요한 부분에 한하여 허용

저작권법
제101조의4(프로그램코드역분석) ① 정당한 권한에 의하여 프로그램을 이용하는 자 또는 그의 허락을 받은 자는 호환에 필요한 정보를 쉽게 얻을 수 없고 그 획득이 불가피한 경우에는 해당 프로그램의 호환에 필요한 부분에 한하여 프로그램의 저작재산권자의 허락을 받지 아니하고 프로그램코드역분석을 할 수 있다.

② 제1항에 따른 프로그램코드역분석을 통하여 얻은 정보는 다음 각 호의 어느 하나에 해당하는 경우에는 이를 이용할 수 없다.
1. 호환 목적 외의 다른 목적을 위하여 이용하거나 제3자에게 제공하는 경우
2. 프로그램코드역분석의 대상이 되는 프로그램과 표현이 실질적으로 유사한 프로그램을 개발·제작·판매하거나 그 밖에 프로그램의 저작권을 침해하는 행위에 이용하는 경우

학습된 모델의 보호 (AI-Stealing Attack)

가상 사례 (2) - 공개된 온라인 서비스

- 로보어드바이저 업체인 A 회사는 인공지능을 이용하여 이용자의 나이, 소득, 예상 은퇴 시기, 위험 성향을 바탕으로 최적화된 자산 투자 포트폴리오를 구성해 주는 서비스를 제공하고 있음
- 후발 주자인 B회사는 자사 직원들과 그 지인을 통해 A회사 서비스에 다양한 입력값을 넣어보는 방법으로 A 회사 인공지능의 동작 원리를 파악한 다음, A회사보다 더 우월한 서비스를 구현해 냄
- B 회사가 이러한 방식으로 A회사의 인공지능 원리를 파악하는 행위가 허용될까?

부정경쟁방지법에 의한 보호

가상 사례 (2)에서의 B 회사의 행위

- 온라인 서비스의 동작 방식을 파악한 것이므로 프로그램의 복제 행위가 일어난다고 볼 수 없음

정당한 경쟁 행위 vs 부정경쟁행위

- 경쟁사 서비스를 분석하는 행위는 원칙적으로는 정당한 경쟁행위에 해당
- 하지만, 통상적이지 않은 해킹 방법을 통해 학습된 모델의 세부사항을 획득할 경우 부정경쟁행위에 해당할 수 있음

부정경쟁방지 및 영업비밀 보호에 관한 법률

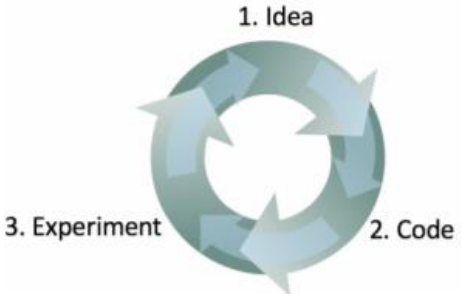
제2조(정의) 1. "부정경쟁행위"란 다음 각 목의 어느 하나에 해당하는 행위를 말한다. (중략)

차. 그 밖에 타인의 상당한 투자나 노력으로 만들어진 성과 등을 공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 자신의 영업을 위하여 무단으로 사용함으로써 타인의 경제적 이익을 침해하는 행위

정리

데이터 수집

(저작권법)
개별 데이터의 저작권
데이터베이스 제작자 권리
오픈소스 데이터



학습 전 모델
(network structure, hyperparameters)

(특허법)
소프트웨어, 학습방법의 특허 보호

(영업비밀)
개발 노하우 보호

개발 과정

(저작권법)
컴퓨터프로그램저작물 보호

(부정경쟁방지법)
부정경쟁행위

학습된 모델

개발 결과