토론문

김종욱 교수
동아대학교 전자공학과
2019.07.24

인류는 유사 이래 수없이 많은 과학· 공학기술을 개발하고, 이 기술들을 통해 각종 어려움과 문제들을 해결하고, 도시를 건설 및 확장하고, 사회· 경제· 문화적 시스템들을 체계화하고 발전시켜 왔습니다.

현재까지 전기· 전자· 기계 공학 분야의 혁신적인 기술로는 전기, 내연기관, 반도체, 모터, 로봇, 인공지능(AI) 등이 될 수 있습니다. 오늘 토론회의 주제인 AI· 로봇은 이러한 다양한 기술이 총체적으로 집약되어 인간 수준의 지적 능력을 갖춘 로봇을 의미합니다.

인공지능과 로봇은 인공물인 소프트웨어와 하드웨어로 인간을 닮은 존재를 만들어 내고자 하는 인류의 오래된 염원을 담고 있습니다. 즉, 인간 수준으로 지혜롭고 똑똑하고 강해서 인간이 하기 싫거나 고된 일을 도와주거나 대신 해주길 원하는 것입니다. 로봇의 어원인 체코어 robota도 강제노동을 의미하는 것도 이런 이유입니다.

현재 자동차 분야에서 세계적 각축전이 벌어지고 있는 자율주행차만 봐도 이를 잘 알 수 있습니다. 전세계에서 매 23초마다 한 명씩 교통사고로 사망할 정도로 자동차는 현대 문명의 핵심적 이기(利器)이지만 일상적 재난 요소가 된 지 오래입니다. 만약 자동차가 사람처럼 똑똑해서 사고가 발생가능한 상황에서 운전자 실수(human failure)를 신속히 보완하거나 미연에 방지할 수 있게 된다면 교통사고 발생률은 현저히 낮아질 것입니다. 이것이 그 많은 기술적 난제에도 불구하고 공학자와 기업들이 자율주행 기술에 도전하는 이유입니다.

그렇다면 자율주행차에 들어가는 기술에는 어떤 것들이 있을까요? 아래 표를 보면 크게 환경인식, 위치인식 및 맵핑, 판단, 제어, 인터랙션 기술이 있습니다.

구성기술	내용
환경인식	• 레이더, 카메라 등의 센서 사용 • 정적장애물(가로등, 전봇대 등), 동적장애물(차량, 보행자 등), 도로표식(차선, 정지선, 횡단보도 등), 신호 등을 인식
위치인식 및 맵핑	• GPS/INS/Encoder 기타 맵핑을 위한 센서 사용 • 자동차의 절대/상대 위치 추정
판단	• **목적지까지의 경로 및 장애물 회피 경로 계획** • **차선유지, 차선변경, 좌우회전, 추월, 유턴, 급정지, 주정차 등 주행 상황별 행동 판단**
제어	• 운전자가 지정한 경로대로 주행하기 위해 조향, 속도변경, 기어 등 액추에이터 제어
인터랙션 (HCI)	• 인간자동차인터페이스(HVI, Human Vehicle Interface)를 통해 운전자에게 경고 및 정보를 제공, 운전자의 명령을 입력 • V2X(Vehicle to Everything) 통신을 통하여 인프라 및 주변차량과 주행정보 교환

출처: 자율주행자동차 기술개발의 특징 및 정책동향, 융합연구정책센터, 2017

자율주행차가 환경과 위치를 정확히 인식하기 위해서는 다양한 센서(레이더, 라이다, 카메라, GPS, 관성항법장치 등)들이 장착되어 측정값을 지속적으로 잘 감지해야 하고, 인공지능 기술이 이러한 센서 데이터들을 읽어 들여서 현재의 차량 외부환경을 종합적으로 판단해야 하며, 그 결과를 실행(" 전방에 사람이 지나가고 있으니 신속히 정지할 것")하는데 있어 가감속 페달과 조향각 제어도 잘 이뤄져야 합니다. 즉, 환경인식부터 제어까지 수많은 부품들과 소프트웨어들이 개별적, 통합적으로 완벽하게 동작해야한 정말 믿고 신뢰할만한(trustworthy) 인간 수준의 자율주행차가 될 수 있는 것입니다.

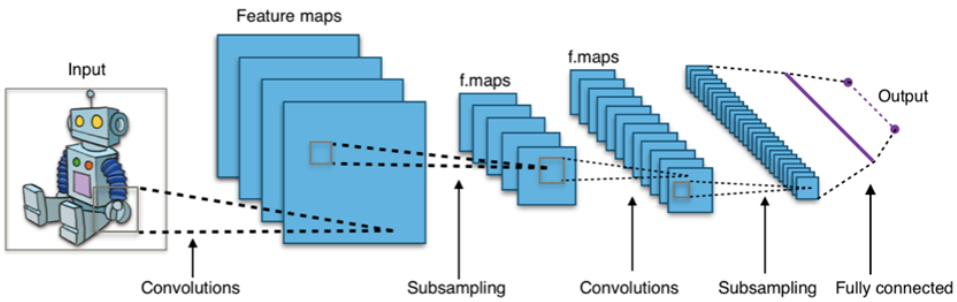
자동차 회사들은 사고로부터 탑승자와 보행자의 안전을 극대화하기 위해 수십년간 각고의 노력을 해왔기 때문에 판단 기술을 제외한 부분들은 기술적인 성숙단계에 이르렀다고 할 수 있습니다. 하지만 인공지능에 의한 인식과 판단 부분은 아직 갈 길이 멉니다. 2012년부터 급부상한 딥러닝(Deep Learning)이 ImageNet의 사진에 나오는 피사체들은 인간 수준으로 잘 인식하지만, 다양한 기후와 날씨 여건에서 찍히는 동영상을 실시간으로 인간처럼 정확히 분석하는 수준에는 아직 이르지 못했기 때문입니다.

Deep Learning은 방대한 데이터를 입력으로 넣어줄수록 학습이 잘 되는 속성이 있는데, 그와 반대로 학습되지 않은 데이터가 입력으로 들어오면 어떤 판단 결과가 나올지 불확실하다는 것이 가장 큰 위험요소입니다. 예를 들어 2017년 워싱턴 대학교의 타다요시 코노 박사 팀에서 수행한 실험에서 아래와 같이' 정지' 라고 적힌 교통표지판에 Love, Hate 같은 스티커를 붙이자 자율주행차 영상인식 소프트웨어의

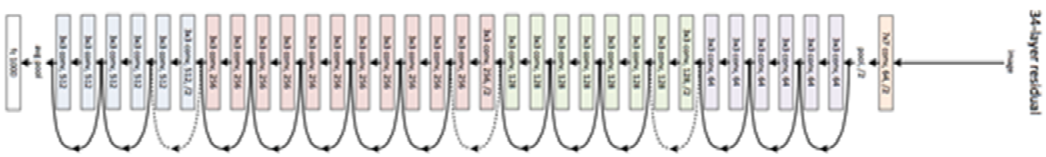
약 70%가 이 표지판을 ' 시속 45마일(72km) 제한'으로 잘못 인식하는 것을 밝혀냈습니다. 이는 ' 정지'와는 반대 개념으로서 자칫 대형사고를 유발할 수 있는 오류입니다.

*출처: https://m.news.naver.com/rankingRead.nhn?oid=092&aid=0002121075&sid1=105&date=20170807&ntype=RANKING

Deep Learning의 또 다른 문제점은 그 구조의 복잡성입니다. 2012년 딥러닝 시대의 서막을 알린 CNN(Convolutional Neural Network)의 은닉층은 겨우 8층이었지만 2015년에 발표된 CNN인 ResNet의 은닉층은 무려 152층이나 됩니다. 그래서 자율주



CNN의 구조



ResNet의 구조

행차의 주행 시 혹은 사고 발생 시 CNN 내부에서 어떤 노드들이 활성화되고 링크 가중치 값들이 어떻게 변화하는지, 그리고 그것이 추론과 판단에 어떤 영향을 미치

는지 인간 수준에서 이해하기란 불가능하다는 문제점이 발생합니다. 그리고 앞의 경우에서처럼 만약 영상인식이 잘못되어 길을 건너는 행인을 제대로 인식하지 못해 사망사고가 발생한 것이 확실해졌다면 누구에게 얼마만큼의 책임을 지워야 하는지도 현재로선 애매합니다.

인공지능의 이런 문제점과 그에 대한 대책을 다루는 용어가 바로 ' 투명성(transparency)'과 ' 책임성(accountability)'입니다. 투명성은 설명가능성(explainability), 해석가능성(interpretability), 역추적가능성(back-traceability)으로도 불리는데요, 쉽게 말해 인공지능 내부에서 사용 중인 입출력 데이터, 실행 또는 준비 중인 프로세스, 학습이나 판단에 사용되는 알고리즘 등을 사용자나 관리자가 요청할 때 일반인도 이해할 수 있는 수준의 언어나 그림, 상징을 이용해서 설명할 수 있어야 한다는 것입니다. 여기에서 중요한 것은 인공지능 개발 회사의 영업비밀이나 노하우가 담긴 소스코드를 공개하는 것을 요구하지는 않는다는 것입니다.

책임성은 투명성과도 연결된 개념으로서 인공지능이 상해나 사고를 일으킬 때를 대비하여 기술적, 법적, 사회보장적 대책을 수립하는 것을 의미합니다. 자율주행차 사고의 예에서처럼 사고가 날 경우 역추적가능성 기능을 이용해서 사고 원인을 최대한 빨리 찾을 수 있어야 하고, 관련 기술에 대한 주체에 대해 법적으로 명시된 책임을 물을 수 있어야 하며, 보험체계에 의해 적절한 보상을 할 수 있어야 합니다. 아울러 스마트시티 같은 사회 인프라 차원에서 유사한 사고가 재발하지 않도록 기술을 보완해 가야 합니다. 예를 들어 자율주행차가 제어불능 상태에서 어떤 선택을 하더라도 피해자가 발생할 수밖에 없는 트롤리 딜레마와 같은 상황이 발생하지 않도록, IoT, 빅데이터, 5G 기술을 이용해서 보행자를 보호하는 기술을 개발할 필요성이 있는 것입니다.

이외에도 인공지능과 로봇의 특성상 지속적으로 수집하는 데이터로부터 개인의 정당한 권리와 프라이버시를 보호할 수 있어야 하고, 악의적 의도를 가진 자에 의한 해킹이나 오남용으로부터 사용자를 보호하는 보안성, 소수의 데이터를 생성할 수밖에 없는 취약 계층이 차별받지 않게 하는 공평성(fairness), 오작동으로 인한 위험 최소화를 위해 킬스위치 기능을 갖추는 제어가능성(controllability) 등도 대책으로 수립해 가야 합니다. 이는 한두 주체의 노력이 아니라 설계자, 개발자, 공급자, 사용자, 관리자 모두 함께 협력하여야만 가능할 것입니다. 또한 정부, 기업뿐만 아니라 경실련같은 시민단체에서도 관심을 가지고 능동적으로 모니터링하고 목소리를 내어야 할 것입니다.

이러한 시대적 요구를 예측하여 세계적으로 EU 집행위원회, IEEE(세계전기전자공학인협회), UNESCO, OECD, 미국, 일본, 중국 등에서 인공지능의 윤리원칙과 가이드라인을 경쟁적으로 발표하고 있습니다. 또한 MS, 구글, 카카오 같은 기업에서도 자체적인 윤리 가이드라인을 구축하여 자율 규제를 실시하고 있습니다.

한국에서는 2007년에 초안이 작성된 로봇윤리헌장(가이드라인)의 개정판을 2018년 8월에 로봇윤리포럼에서 발표했으며, 그 내용은 다음과 같이 세 가지 기본가치와 다섯 가지 실천원칙으로 구성되어 있습니다.

기본가치

1. 인간의 존엄성 보호
2. 공공선 추구
3. 인간의 행복 추구

실천원칙

1. 투명성
2. 제어가능성
3. 책무성
4. 안전성
5. 정보보호

결론적으로, AI· 로봇 기술은 시대의 요구와 각국의 기술경쟁에 의해 자율주행차, 서비스로봇, 전투로봇 등 앞으로 다양한 형태로 등장할 것이며, AI· 로봇의 역기능을 최소화하고 인류가 함께 지속가능한 발전과 행복한 삶을 누리기 위해서는 각 이해당사자들이 긴밀히 협력해야만 한다고 생각합니다.

감사합니다.