

시각 인식 기술을 중심으로 한 인공지능 기술의 역사와 발전 방향

□ 최중현 / 광주과학기술원

1. 서론

인간은 단순하고 반복적인 일을 싫어한다. 가능하면 이런 일을 피하고 싶어서 다양한 궁리를 한다. 작게는 식사 후 게임을 통해서 설거지 당변을 정하거나, 크게는 세탁기를 발명하여 세상을 바꾸기도 한다. 전기 세탁기가 발명된지 얼마 지나지 않은 20세기 중반, 컴퓨터 기술이 발달하면서 자연스럽게 인간의 지능도 컴퓨터로 모사하여 하기 싫은 일들을 맡길 수 있지 않을까라는 담론이 폭넓게 논의되면서 ‘인공지능’도 하나의 분야로 자리잡기 시작했다고 한다. 현대 컴퓨터의 아버지라 불리는 앨런 튜링이 1950년에 쓴 ‘Computing Machinery and Intelligence’라는 논문에서 제시한 ‘튜링 머신’이 인공지능 시스템의 프로토타입으로 여겨지며 [1], 이후 1956년에 존 매카시가 ‘인공지능’이라는 용어를 사용하면서 이 분야가 자리매김하게 되었다

고 한다. 이후 간단한 형태의 자연어 처리(Natural Language Processing, NLP) 시스템이 등장하면서 인공지능에 대한 기대는 더욱 커졌고, 금방이라도 인공지능을 완성할 수 있을 것 같은 분위기가 형성되었다. 1966년에는 MIT의 Seymour Papert 교수가 시각 정보를 처리하는 시스템을 여름 방학 동안 학부생 한 명과 마무리 하려는 야심찬(!) 계획을 세우기에 이른다 [2]. 하지만 기대와 달리 이 시스템은 원하는 만큼의 성능이 나오지 않았고, 다른 연구자들도 성능을 높이는데 어려움을 겪었다 (그리고 그 프로젝트는 ‘컴퓨터 비전’이라는 분야를 형성하였으며, 54년이 지난 지금도 연구가 진행중이다).

1970년대 초, 수 년을 투자한 뒤에도 이렇다 할 성과를 거두지 못한 미국의 많은 펀딩 기관들은 인공지능 연구에 대해 회의적인 시각을 갖게 되면서, 예산의 대부분을 삭감하기에 이른다. 이것이 이른바 ‘1차 인공지능 겨울(First AI Winter)’이다. 하지

만 사람들은 인공지능 연구에 대한 기대를 접지 않았다. 수많은 규칙을 사람이 일일이 정의해 놓은 ‘Expert system’이 컴퓨터 부품을 주문하는 과정을 자동화하는데 성공하면서 1980년대 초 일본에서 Expert System에 대한 대규모 투자가 이루어짐과 동시에 관련한 연구 프로젝트가 다양하게 진행되었다. 그러나 이 시스템은 모든 규칙을 하나하나 모두 입력해야 하기 때문에 다양한 일에 확장하기가 매우 어려웠다. 결국 1984년 Marvin Minsky가 ‘AAAI 연례 학회’에서 곧 AI산업계가 몰락할 것이라고 예측하였고, 3년 후인 1987년 당시 AI 시스템을 작성하는데 가장 널리 쓰이던 Lisp 프로그래밍 언어로 동작하는 소위 ‘Lisp 기계’를 만드는 수 조원 규모의 산업계가 몰락하면서, ‘두번째 인공지능 겨울(Second AI Winter)’이 찾아온다. 두 번의 겨울을 경험한 학계와 산업계는 인공지능 연구에 매우 조심스러워졌고, 발전 속도가 심각하게 둔화되었다.

한편, 두번째 겨울이 오기 직전인 1986년에 D. Rumelhart와 G. Hinton이 Artificial Neural Network (ANN 또는 NN, 인공 신경망 또는 뉴럴넷)을 학습하는 방법으로 Back propagation을 제안했다 [3]. 이 방법은 복잡한 데이터를 다층 구조의 대용량 뉴럴넷을 이용하여 학습할 수 있는 가능성을 열었지만, 당시 컴퓨터의 성능과 저장 공간의 한계로 작은 규모의 데이터를 가지고 단층 구조의 소형 뉴럴넷 밖에는 만들 수 없었다. 3년 후 Yann LeCun이 back propagation을 다층 구조의 뉴럴넷에 적용하여 손으로 쓴 숫자를 매우 높은 정확도로 인식하는 시스템을 만들어내고 나서야, 제한된 경우에 한해서 뉴럴넷이 좋은 성능을 내는 예를 찾긴 했지만, 다른 문제에 적용했을 때에는 좋은 성능을 보여주지 못했다. 게다가 뉴럴넷은 주어진 학습

데이터(training data)를 잘 학습하긴 하지만, 주어진 데이터 외의 입력에 대해서는 좋은 예측 값을 내놓지 못한다는 측면에서 머신 러닝(machine learning, 기계 학습) 커뮤니티의 외면을 받았다. 머신 러닝 연구는 주어진 데이터 외의 입력에 대해서 더 좋은 예측 값을 내는 (이를 ‘일반화(generalization)’라고 한다) 모델이나 시스템을 만드는 것을 목표로 한다. 뉴럴넷은 ‘일반화’ 성능 측면에서 좋지 않았으며, 이 때문에 주류 머신 러닝 연구에서 밀려났다. 그러는 사이 1990년대와 2000년대에 모델의 복잡도가 높지 않으면서 일반화 성능이 좋은 Support Vector Machine (SVM)이 머신 러닝 연구의 주류로 자리잡으며 뉴럴넷 연구자들을 연구 커뮤니티에서 아웃사이더로 만들었다. 하지만 SVM은 모델이 간단하기 때문에 일정 복잡도 이상의 데이터를 모델링 하는 데에는 상당히 취약했는데, 컴퓨터 비전이나 자연어 처리 같은 실제 태스크에서 그 한계를 뚜렷하게 드러내고 있었다 [4]. 이상적인 ‘일반화’는 주어진 학습용 데이터를 충분히 모델링 한 뒤, 그로부터 데이터의 핵심을 찾아내어 주어진 데이터는 물론 새로운 데이터도 잘 인식할 수 있게 하는 것인데, 기본적인 SVM인 선형 SVM은 모델이 간단하여 주어진 데이터가 복잡해지면 이를 모델링 하는 성능이 많이 떨어지기 때문에 일정 수준 이상의 인식 정확도를 내지 못했다.

또 한편, 1990년대와 2000년대에는 무어의 법칙으로 대표되는 컴퓨팅 하드웨어의 엄청난 발전이 있었다. 이에 힘입어 계산 속도가 엄청나게 빨라졌음은 물론이고, 저장공간의 크기도 놀라울 정도로 커졌다. 일례로 Yann LeCun이 back propagation이 동작하는 것을 보였던 1989년의 40MB 하드디스크는 \$5,000 이었지만, 2018년의 4TB 하드디스크는 \$100 미만의 가격으로 구할 수 있다. 즉, 용량

은 십만배가 늘었지만 가격은 50분의 1이 되었다. 이에 따라 모델을 학습할 수 있는 데이터의 크기가 천문학적으로 커졌고, 이를 처리할 수 있는 연산 장치의 속도도 그때에는 상상하기도 어려울 정도로 빨라졌다. 이런 하드웨어의 발전에 맞추어, 뉴럴넷을 믿고 있던 연구진들은 뉴럴넷의 약한 일반화 성능에도 불구하고 인식 성능의 기본이 되는 ‘충실한 데이터 모델링’에 집중하며, 적극적으로 새로운 하드웨어를 받아들여 꾸준히 성능을 개선하고 있었다.

II. 뉴럴넷의 귀환

2012년은 뉴럴넷 연구진들에게 잊지 못할 한 해였을 것이다. 인공지능 연구 커뮤니티, 특히 컴퓨터 비전과 자연어 처리 분야는 SVM의 한계에 부딪혀 이렇다 할 돌파구를 찾지 못하고 있었다 [4]. 그러던 중 뉴럴넷 커뮤니티가 초고속 연산 시스템과 대용량 데이터로 학습한 모델을 가지고 나와서, 컴퓨터 비전과 음성 인식 분야에서 난제로 여겨지던 물체 인식 [5]과 화자 독립형 음성 인식 분야 [6]에서 놀라운 성능 개선을 보였다. 특히 물체 인식 문제에 관해서 컴퓨터 비전 커뮤니티는 1,000개의 물체를 구분하는 벤치마크 문제인 ImageNet Challenge를 2009년부터 해마다 열고 있었는데, 매년 발표되는 새로운 이미지 변환 기술과 SVM의 변종들은 해마다 1%~3% 남짓한 인식 성능의 개선을 이루어 내는데 그치고 있었다. 그러다가 2012년 뉴럴넷 커뮤니티가 대용량 심층 뉴럴넷(Deep Neural Network)으로 전년 대비 10% 가깝게 개선하는 모델을 발표하면서 커뮤니티에 일대 돌풍을 일으켰다 [4]. 심지어 이 모델은 복잡한 이미지 변환도 필요없이 원본

이미지만을 가지고 이루어낸 성능 개선이었고, 모델의 많은 부분에서 튜닝을 하긴 했지만, 추가적인 개선의 여지가 많아서 큰 발전 잠재력을 가지고 있었다. 2012년을 기점으로 모든 인공지능의 하위 분야 - 컴퓨터 비전, 음성 인식, 자연어 처리 - 는 이 Deep Neural Network을 적극적으로 연구하기 시작했고, 다양한 분야에서 괄목할 만한 성능개선이 이루어졌다. 뉴럴넷 연구자들의 오랜 꿈과 믿음이 빛을 보게 된 것이다.

III. 기 대

2012년을 통해 인공지능 연구의 부활 가능성을 본 많은 연구자들은 이 오래되었지만 저평가된 기술을, 적용할 수 있는 거의 모든 분야에 적용하기 시작했다. Yann LeCun이 뉴럴넷의 가능성을 보여줬던 첫번째 응용 분야는 손으로 쓴 숫자 인식 테스트였고, 이는 시각 인식을 다루는 컴퓨터 비전 문제 중 하나이다. 컴퓨터 비전은 이미지를 다루는 분야 특성상 입력으로 주어지는 데이터의 크기가 매우 크기 때문에 3개의 인공지능 하위 분야 중 딥-뉴럴넷을 가장 적극적으로 적용하고 연구하고 있다. 여기에서 발견된 많은 기술 중 일부가 다른 분야에도 응용되어 사용되고 있는데, 그 대표적인 예가 바로 바둑에서 상대방에게 이기기 위한 수를 찾아내는 AlphaGo 알고리즘이다.

동양 문화권에서 바둑은 어려운 논리 문제의 대명사로 불릴 만큼 난해한 두뇌 게임이다. 많은 수의 바둑 세계 챔피언이 우리나라 사람이라는 사실에 필자 역시 은근한 자부심을 갖고 있었는데, 몇 년 전 영국의 한 start-up인 DeepMind Technologies (구글이 2014년 인수)가 만든 실험적인 알고리즘에

우리의 자부심이던 챔피언이 패배하는 모습을 보는 일은 충격 그 이상이었다. 이 이벤트를 통해서 사람들이 “바둑은 고차원적 인간 지능의 대명사 → 이를 알고리즘이 이겼음 → 이제 컴퓨터가 인간을 능가하는 지능을 가질 수도 있겠다”고 걱정을 하는 것도 크게 무리는 아니라 생각한다. 게다가 이런 생각에 응답이라도 하듯, 수많은 업체들은 다투어 ‘인간보다 더 정확한 얼굴 인식 시스템’, ‘인간보다 더 정확한 지문 인식 시스템’, ‘인간보다 사고를 덜 내는 자율 주행 시스템’ 등의 프로토타입을 선보이며 이 같은 우려가 사실일 수도 있겠다는 환상을 갖게 하고 있다.

또한 당 시대의 저명한 작가들도 정교한 상상력을 동원하여 AI의 미래상에 대해 다양한 담론을 내놓았는데, 대부분은 인간이 기계에 지배당할 것이라는 우울한 것인 경우가 많았다. 1991년에 개봉한 영화 터미네이터2, 1993년 Vernor Vinge의 The Coming Technology Singularity 등에서 인공지능으로 인해 우리의 미래가 매우 우울할 것이라고 예측한 것들이 그 좋은 예이다.

IV. 인공지능 연구의 미래

그렇다면 인간을 능가하는 인공지능이 나타나서 곧 인간을 지배하게 될까?

바둑을 비롯해 인공지능 연구가 다루고 있는 문제들을 자세히 보면 비교적 잘 정의된, 편협한 문제인 경우가 많다. 이를 테면, ‘정면 얼굴의 모습을 보고 이름과 연결(mapping)하는 얼굴 인식’, ‘조명을 일정하게 비추는 장치에서 일정한 각도로 촬영한 지문을 이름으로 연결하는 지문 인식’, ‘한정된 집합의 문장을 인식하는 자연어 기반 음성 인식’이 그

것이다. 이 잘 정의된 문제들은 일반적인 ‘얼굴 인식’, 일반적인 ‘지문 인식’보다 매우 제한적인 환경을 가정 하고 있으며, 이 가정하에 개발된 알고리즘들은 제한적인 환경의 특성을 최대한 이용하여 성능을 높이고 있다. 따라서 이 가정들이 없어질 경우 정확도가 눈에 띄게 떨어지는 경우가 많다.

컴퓨터는 계산기라는 그 이름에서 알 수 있듯이 주어진 계산을 빠르고 정확하게 수행해서 답을 내는 장치이다. 문제를 컴퓨터로 풀기 위해서는 적절한 가정을 통해 알고리즘으로 계산하기 용이한 정답 공간(solution space)을 정의하고, 고성능의 하드웨어를 이용하여 예측하는 해와 각 해마다의 불확실성(confidence interval)을 빠르게 계산해야 한다. 이 과정에서 컴퓨터는 계산을 할 뿐 ‘문제를 풀’지는 못한다. 예를 들어, 기계는 1부터 100까지 더하는 문제를 풀기 위해 논리적으로 $n(n+1)/2$, $n=100$ 이라는 추론을 하는 것이 아니라 빠르게 모두 더하거나, 인간이 준 $n(n+1)/2$ 라는 공식으로 계산을 한다. 바둑의 경우를 다시 보면, 현재 두는 수가 마지막에 이기는 수로 가는 길인지를 결정하기 위해서 현재로부터 마지막까지 가능한 수들의 이길 가능성을 계산하여 높은 가능성이 있는 몇 개의 수들로 경우의 수를 한정하여 가지치기를 하고, 각 경우마다의 게임 현황을 기억하여 끝까지 가상 경기를 해보아야 한다. 사람은 이 많은 정보를 손실없이 기억하는 것에 약하기 때문에, 고차원의 비유를 통해 각 수가 만드는 상황을 추상화하여 압축적으로 기록하면서 시뮬레이션을 하게 되리라 추측되는데, 기계는 저 모든 경우를 손실 없이 기억하며 빠르게 계산할 수 있기 때문에, 가지치기에 사용되는 ‘이길 가능성’을 정확히 계산한다면 바둑에서 기계가 인간을 이길 수 있는 가능성이 상당히 커진다. 인간의 암산 능력이 계산기보다 떨어진다고 해서 놀랄 사

람은 별로 없다. 이처럼 바둑의 본질을 계산 능력 자체로 보기 시작하면 AlphaGo의 승리가 그리 놀라운 일은 아니다. AlphaGo의 가치는 바둑을 이런 시각으로 해석하여 solution space를 정의하고, 이에 맞는 목적 함수(objective function)를 고안하여 이를 강화 학습으로 풀어낸 연구원의 통찰력과 실행력에 있다.

사람들이 걱정하는 ‘인간 수준의 인공지능(Human level AI)’, ‘일반적 인공지능(Artificial General Intelligence, AGI)’ 또는 ‘강 인공지능(Strong AI)’은 인간 수준의 고차원적 논리 전개를 요하는 시스템들인데, 이는 현재 인공지능 연구진들이 다루고 있는 연구 문제들과 거리가 멀다. 현재의 연구 주제들은 목적이 분명한, 잘 정의된 문제들(물체를 올바르게 인식 했다/안했다. 바둑을 이겼다/졌다)인 경우가 대부분이다. 인간에게도 목적이나 실체가 불분명한 문제들은 목적 함수를 정의하기가 어렵기 때문에 계산을 시작할 수조차 없다. 일례로, 인류가 시작된 이래 인간은 ‘사랑’의 본질을 찾아 수많은 각도(철학, 예술, 과학)에서 접근했지만, 우리는 아직도 이 문제를 풀기는 커녕 어떤 접근법이 옳은지조차 가늠하지 못하고 있다. 사랑을 이해하는 인공지능을 만들고 싶지만, 무엇을 계산해야 할지 모르는 문제를 푸는 데에 성능 좋은 계산기는 별 도움이 되지 않는다.

V. 결론

최근 인공지능의 괄목할 만한 발전을 보면 다양한 미래가 그려진다. 인류에게 도움이 되는 미래도 있지만, 그렇지 않은 것도 있다. 고성능의 하드웨어와 모듈화된 소프트웨어 기술들은 지루한 일들을 매우 효율적으로 할 수 있게 해주는 다양한 ‘인공지능’ 시스템을 만들어 내고 있다. 이는 산업혁명에 비유되어 소위 4차 산업혁명이라고 일컬어진다. 하지만 4차 산업 혁명의 실체는 기계가 인간을 지배하는 세상이 아니라, 혁신적인 생산성의 향상이다. 이런 비약적인 생산 효율성 증대의 과도기에서는 현재의 기술을 더 잘 이해함으로써 효율성 증대의 방점이 어디에 있는지를 명확히 파악하여 더 정확한 미래를 그려내고 준비하는데에 역량을 집중하는 것이 우리에게 다가올 미래를 올바르게 준비하는 자세가 아닐까 생각한다.

영화에서는 필자가 보기에도 불가능해 보일 만큼 놀라운 인공지능들이 등장한다. 하지만 영화는 아직 영화일 뿐이다. 필자는 용광로에 들어가기 직전 자신의 소멸을 슬퍼하는 존 코너의 눈물을 이해한다고 말하는 터미네이터 T-800의 ‘인공지능’이 이해한 눈물의 의미와 자신이 소멸해가는 순간에 존 코너에게 엄지 손가락을 세워 준 이유가 정말 궁금하다.

참고 문헌

- [1] A. M. Turing (1950) Computing Machinery and Intelligence. Mind 49: 433-460.
- [2] The summer vision project, <https://dspace.mit.edu/handle/1721.1/6125>
- [3] D. Rumelhart, G. Hinton & R. Williams (1986) Learning representations by back-propagating errors. Nature 323, 533-536 doi:10.1038/323533a0
- [4] Russakovsky, Olga & Deng, Jia & Su, Hao & Krause, Jonathan & Satheesh, Sanjeev & Ma, Sean & Huang, Zhiheng & Karpathy, Andrej & Khosla, Aditya & Bernstein, Michael & Berg, Alexander & Li, Fei Fei (2014) ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision. 115. 10.1007/s11263-015-0816-y.
- [5] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton (2012) ImageNet Classification with Deep Convolutional Neural Networks, Proceedings of NeurIPS 2012
- [6] Pieraccini, Roberto (2012) The Voice in the Machine. Building Computers That Understand Speech. The MIT Press. ISBN 978-0262016858.

필자 소개



최종현

- 2003년 : 서울대학교 전기공학 학사
- 2008년 : 서울대학교 전기컴퓨터공학 석사
- 2013년 : 어도비 연구소 연구인턴
- 2014년 : 디즈니 연구소 연구인턴
- 2014년 : 마이크로소프트 연구소 연구인턴
- 2015년 : 메릴랜드 주립대학교 전기컴퓨터공학 박사
- 2015년 : 컴캐스트 랩스(워싱턴, DC) 선임연구원
- 2016년 ~ 2018년 : 앨런 인공지능 연구소(시애틀, WA) 연구원
- 2018년 ~ 현재 : 광주과학기술원(GIST) 인공지능대학원/전기전자컴퓨터공학부 조교수