

차세대 인공지능의 연구 현황과 당면 과제

Research Status and Challenges of Next Generation
Artificial Intelligence

추형석

2019.1.

집필기관 및 참여인원 : 소프트웨어정책연구소

주형석

이 보고서는 2018년도 과학기술정보통신부 정보통신진흥기금을
지원받아 수행한 연구결과로 보고서 내용은 연구자의 견해이며,
과학기술정보통신부의 공식입장과 다를 수 있습니다.

목 차

제1장 서 론	1
제1절 연구배경 및 필요성	1
제2절 연구방법 및 범위	12
제2장 인공지능 연구 영역의 분류	13
제1절 개 요	13
제2절 인공지능 연구 영역의 분류 제안	18
제3장 차세대 인공지능의 연구 현황	22
제1절 개 요	22
제2절 인공지능 기초 연구 현황	23
제3절 범용 인공지능 연구 현황	39
제4절 소 결	48
제4장 요약 및 시사점	50
참고문헌	52
별첨 1. 인공지능 기술과 산업의 분류	54
별첨 2. 범용 지능에 대한 다양한 해석	56

Contents

Part 1. Introduction	1
1. Backgrounds	1
2. Methods	12
Part 2. Categorization of AI Research	13
1. Introduction	13
2. A Suggestion for categorization of AI Research	18
Part 3. Research Status of Next Generation AI	22
1. Introduction	22
2. Research Status of Basic AI	23
3. Research Status of AGI	39
4. Conclusion	48
Part 4. Summary and Implications	50
References	52
Attachment 1. Categorization of AI technology and industry	54
Attachment 2. Various interpretation of general intelligence	56

표 목 차

〈표 2-1〉 인공지능 기술의 주요 특징	14
〈표 2-2〉 인공지능 기술의 주요 특징과 시사점	17
〈표 2-3〉 인공지능 연구 영역의 구분과 정의	18
〈표 3-1〉 인공지능 최신 연구현황 분석 주제	22
〈표 3-2〉 신경망 구조 탐색의 성능	28
〈표 3-3〉 아실로마 23대 AI 원칙	43

그 립 목 차

[그림 2-1] 인공지능 연구 영역의 상관관계	20
[그림 2-2] 인공지능 연구 영역 구분의 도식화	20
[그림 3-1] 원샷 학습의 구조도	24
[그림 3-2] 메타 최적화 학습의 구조도	25
[그림 3-3] MAML의 개념도	26
[그림 3-4] 신경망 구조 탐색의 개념	27
[그림 3-5] 구조적 모수를 도출하기 위한 Controller	28
[그림 3-6] 스파이킹 신경망의 구조	31
[그림 3-7] 인공신경망과 스파이킹 신경망의 성능 비교	32
[그림 3-8] 전이학습의 예 : PathNet의 구조	35
[그림 3-9] 지속 학습의 기본 개념도	36
[그림 3-10] 유연한 가중치 강화(EWC) 손실함수의 의미	37
[그림 3-11] 심층 생성 모델을 활용한 지속 학습의 구조도	38
[그림 3-12] ACT-R의 개괄적인 흐름	45
[그림 3-13] Soar의 메모리 모듈 구조	46
[그림 3-14] CBMM의 시각지능 구조 가설	47

요 약 문

1. 제 목

차세대 인공지능의 연구 현황과 당면 과제

2. 보고서의 목적

이 보고서의 목적은 차세대 인공지능의 연구 현황을 분석하는 것에 가장 큰 목적을 둔다. 차세대 인공지능은 기계가 인간 수준의 지능을 달성하기 위한 연구 분야로 범용 인공지능, 인공생명 등의 분야가 있다. 이 보고서에서 차세대 인공지능을 다루는 이유는 우리의 미래 먹거리를 탐색하기 위한 현황을 파악함에 있다. 지난 5년간 인공지능 기술은 심층학습을 기점으로 급격한 발전을 했으나 여전히 인간 수준의 지능과는 거리가 멀다. 그럼에도 불구하고 인간 수준의 지능을 개발하기 위한 연구는 매우 활발히 이루어지고 있는 실정이다. 인류의 최후의 발명품일 가능성이 높은 차세대 인공지능에 대한 시의성 있는 분석은 성공적인 정책 설계를 위한 기초 자료가 될 것이다.

차세대 인공지능은 아직 먼 미래의 이야기일 수 있으나, 그것을 향해가는 연구들은 점차 증가하고 있다. 이 연구들은 알파고에서도 주목받았던 기술인 심층학습의 연장선상에 있다. 차세대 인공지능이 아직 개념적이며 실체가 없는 기술인 가능성이 높기 때문에, 이 보고서에서는 현재 인공지능이 겪고 있는 한계를 극복하기 위한 연구들을 분석했다.

3. 보고서의 구성 및 내용

심층학습으로 촉발된 현대 인공지능은 그 발전 속도가 빠르며 방향 역시 다양하다. 이에 따라 인공지능 활성화를 위한 정부의 정책 역시 급변하는 인공지능의 특성에 발맞춰가야 한다. 그 전제는 기술의 시의성 있는 분석으로부터 출발할 것이다.

인공지능은 다양한 문제 해결을 위한 도구적인 특성이 가장 짙다. 또한 인공지능은 대표적인 오픈 사이언스로 연구 결과가 빠르게 공유되고 검증할 수 있는 환경이 마련돼 있다. 이러한 환경은 인공지능 발전을 더욱 가속화하고 있다. 성공적인 인공지능 정책 제안을 위해서는 무엇보다 인공지능의 기술적·산업적 특성을 제대로 이해하기 위한 노력이 필요하다. 이를 바탕으로

실제 인공지능 기술을 이해하고 우리의 방향에 대한 이정표를 세워야 할 것이다. 이 보고서에서는 인공지능의 기술적 특성, 심층학습의 한계 등을 상세히 담아 인공지능의 속성을 이해하기 위한 자료로 활용하고자 한다. 또한 인공지능의 특성을 바탕으로 한 인공지능의 연구 영역의 구분을 제안한다. 이를 토대로 차세대 인공지능과 연관이 높은 인공지능 기초연구 사례를 분석했다. 마지막으로 우리의 당면 과제를 논의하는 것으로 결론 맺는다.

4. 활 용

인공지능 기초·원천 연구개발 기획의 기초자료로 활용

Summary

1. Title

Research Status and Challenges of Next Generation Artificial Intelligence

2. Purpose

The purpose of this report is to analyze the state of research on the next generation of artificial intelligence. Next-generation artificial intelligence is a field of research aimed at achieving human-level intelligence of machine such as artificial general intelligence or artificial life. The reason for dealing with the next generation of artificial intelligence in this report is to understand the current situation for exploring our future direction. Over the past five years, artificial intelligence technology has made rapid progress from deep learning, but it is still far from human intelligence. Nonetheless, there are many studies to develop human-level intelligence. A timely analysis of next-generation artificial intelligence, which is likely to be the last invention of mankind, will be the basis for successful policy design.

The next generation of artificial intelligence can still be a story of the future, but the research that goes toward it is increasing. These studies are an extension of deep learning, a technology that has also been noted at AlphaGo. Because the next generation of artificial intelligence is still a conceptual and immaterial technology, this report analyzes the research to overcome the limitations of current artificial intelligence.

3. Contents and Details

Modern artificial intelligence, triggered by deep learning, is fast-paced and diverse. Accordingly, the government policy for the activation of artificial intelligence must also keep pace with the characteristics of rapidly changing artificial intelligence. The premise will start from a timely analysis of technology.

Artificial intelligence is the most instrumental feature for solving various

problems. Artificial intelligence is a typical open science, and the results are quickly shared and verified. This environment is accelerating the development of artificial intelligence. In order to propose a successful artificial intelligence policy, it is necessary to make efforts to understand the technical and industrial characteristics of artificial intelligence. Based on this, we must understand real artificial intelligence technology and establish a milestone in our direction. I intend to utilize this report as understanding the properties of artificial intelligence in detail, including the technical characteristics of artificial intelligence and limitations of deep learning. I also propose a categorization of artificial intelligence R&D based on characteristics of artificial intelligence. Based on this, I analyzed the basic artificial intelligence research case which is related to the next generation artificial intelligence. Finally, I conclude what to be discussed about our immediate challenges.

4. Application

Utilize as basic data for planning basic and fundamental AI R&D

제1장 서론

제1절 연구배경 및 필요성

제4차산업혁명의 핵심 기술인 인공지능은 현대 산업과 사회를 획기적으로 변화시키는 동력이다. 특히 2016년 알파고 대국 이후 인공지능의 관심사는 폭발적으로 증가하고 있으며, 학계와 산업계는 ‘인공지능’이라는 키워드와 불가분의 관계가 됐다. 이처럼 유례를 찾기 힘든 전 세계적인 관심사는 인공지능 기술의 발전과 산업에의 확산을 가속화하고 있다. 또한 인공지능은 오픈 사이언스(Open Science)의 특성을 갖기 때문에, 빠른 속도로 연구 결과가 공유되고 공개SW를 통해 이를 검증하는 과정이 용이해졌다. 이렇게 끊임없이 발전하는 인공지능 기술은 그 발전상을 제대로 파악하는 것조차 어렵다.

따라서 성공적인 인공지능 정책을 제안하는 것 역시 인공지능 기술 발전상과 따로 생각할 수 없다. 또한 인공지능은 도구적인 성향이 강하기 때문에 인공지능의 기술적 특성에 입각한 접근이 필요할 것이다. 이 절에서는 인공지능의 기술적 특성을 크게 네 가지로 구분하여 살펴보고, 현대 인공지능의 핵심 기술인 심층학습(Deep Learning)의 한계에 대해서 논의함과 더불어 이 보고서의 주제인 차세대 인공지능에 대한 내용을 다루고자 한다. 이를 바탕으로 차세대 인공지능의 기초·원천기술 확보를 위한 우리의 당면과제에 대해 논의할 것이다.

1. 인공지능의 기술적 특성

인공지능 기술의 발전 속도는 매우 빠르며 그 적용 영역 역시 빠르게 확산되고 있다. 이러한 속도로 인해 인공지능의 실체를 잡고 구체화하는 노력 역시 그 속도에 발맞춰 가야 함에 이견이 없을 것이다. 가까운 미래에는 산업의 구석구석에 인공지능이 녹아 들어갈 것이다. 인공지능 기술에 대한 시의성 있는 이해는 우리가 무엇을 준비하고 실행해나가야 할 이정표를 보여준다. 인공지능의 기술 동향을 파악하는 것이 일견 진부해 보이는 접근방법으로 보일 수 있으나, 인공지능의 발전 속도와 활용 범위를 고려해 보면 제대로 기술 동향을 파악하는 것만으로도 충분한 의미가 있다.

현재 인공지능 산업의 키워드는 심층학습(Deep Learning)이다. 심층학습이 각광받는 이유는 바로 성능 때문이다. 컴퓨터 비전, 음성 인식, 자연어 처리, 데이터 마이닝 등 전통적인 컴퓨터 과학의 분야에서 심층학습이 차지하는 위상은 매우 높다. 이러한 인공지능의 핵심기술이 심층학습으로 인해 더욱 성숙해짐에 따라 산업과의 융합이 활발하게 이루어지고 있는 실정이다. 이러한 특성으로 인해 산업계에서는 인공지능을 문제해결의 도구로 바라보는 관점이 크다. 결국 산업에서 심층학습을 활용하는 용도는 문제해결의 도구이기 때문에, 문제를 해결하기 위한 시스템의 일부에 인공지능 기술이 녹아들어 있다고 볼 수 있다.

현대 인공지능 기술의 첫 번째 특성은 데이터에서 찾을 수 있다. 심층학습은 기본적으로 데이터를 학습해 패턴을 인식하는 기술이다. 양질의 데이터가 심층학습의 성능을 좌우한다는 점은 널리 알려진 사실이다. 그러나 양질의 데이터 구축은 그 반대급부로 많은 시간과 비용을 필요로 한다. 데이터의 절대량이 인공지능의 성능을 좌우하는 핵심 요소이라는 사실은 데이터로 인해 인공지능이 겪을 수밖에 없는 기술적 한계를 보여준다. 자율주행차를 예로 들어보자. 현재 사진에서 객체를 인식하는 기술은 발전을 거듭하여 약 3%의 오차를 달성했다.¹⁾ 그러나 이 성능은 학습 데이터와 유사한 데이터를 기준으로 측정된 것이므로, 실제 상황에서의 적용은 어려울 수 있다. 자율주행차가 운행하는 환경은 시점, 지형, 날씨에 따라 매우 상이할 것이다. 자율주행차가 완벽에 가깝게 주변을 인식한다는 사실은 곧 완벽에 가까운 데이터를 학습하는 것으로 귀결된다. 가능한 다양한 데이터를 구축하는 것이 자율주행차의 완성도와 직결된다는 것이다. 하지만 심층학습 기술은 아직까지 학습하지 않은 사실을 추론할 수 있는 단계에 도달하지 못했다. 강화학습으로 구현한 ATARI 2600의 벽돌깨기(breakout) 게임은 인간의 수준을 넘어설 정도로 우수한 성능을 갖췄으나, 게임 화면 중간에 벽돌을 삽입할 경우 전혀 작동하지 않는다. 결국 심층학습을 필두로 한 인공지능이 실제 산업에 적용되기 위해서는 무엇보다 데이터가 먼저다.

두 번째는 경험적인 결과에 의존한다는 점이다. 심층학습의 모태는 1970년대부터 회자되어 왔던 인공신경망(Artificial Neural Network) 기술이다. 따라서 인공신경망이 가지고 있던 한계가 고스란히 심층학습으로 이어진다.

1) 2017년 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 우승자의 예측 성능

인공신경망은 구조와 성능의 상관관계가 불분명하다. 다시 말하자면, 신경망을 구성하기 위한 이론적 배경이 얇고, 다양한 시도를 통해 최적의 결과를 도출하는 접근이 일반적이다. 이 사실은 인과관계가 분명해야하는 분야에서는 치명적인 단점이 될 수 있으나, 현재 심층학습은 이 단점을 무마시킬 정도로 성능이 좋기 때문에 다양한 분야에 적용되고 있다. 또한 이렇게 높은 성능 덕분에 심층학습의 이론적 연구가 강화되고 있는 추세다. 현재 심층학습의 이론적인 연구는 다양한 형태로 진행되고 있으나, 아직 구체적인 성과는 없는 상황이다. 아직 미래를 낙관할 수 있는 상황은 아니지만 심층학습에 대한 전 세계적인 붐은 그 돌파구를 열어 줄 것이다.

세 번째 인공지능의 특성은 산업에의 적용에 있어 사회적 합의가 중요하다는 점이다. 이것은 앞서 소개한 두 가지 특성과도 맞물려 있다. 바로 인공지능 시스템의 설명가능성, 안전성, 신뢰성의 불명확함에 따라 제기되는 문제다. 인공지능의 오작동에 관한 사례는 미디어를 통해 속속들이 보도되고 있다. 자율주행차의 사고, 지능형 챗봇의 인종차별 발언, 학습 정보의 왜곡 등 인적·물적 피해에서부터 사회적 문제까지 다양하게 분포하고 있다. 이 사실들은 인공지능 기술이 산업으로 전파됨에 있어 신중한 대처가 필요함을 시사한다. 단편적으로 본다면 인공지능의 오작동에 대한 책임소지의 명백한 규명과 합의가 필요하다. 인공지능의 오작동에 대한 책임을 개발자에게 물을 것인지, 시스템을 차용한 중간 사업자에 물을 것인지, 최종 소비자에게 물을 것인지에 대한 규명과 각 주체별로 책임의 범위를 합의해야한다. 그러나 이러한 논쟁은 소모적인 이권 다툼으로 변질 가능성이 있다. 인공지능 기술 자체가 보유하고 있는 한계가 산업에의 적용에 있어 닭과 달걀의 문제로 치부되기 때문이다. 인공지능 기술의 발전을 위해서는 무엇보다 데이터의 확보가 시급하다. 그러나 먼저 규제를 강조하고, 인공지능의 설명가능성과 합리성을 증명하는 것이 선행된다면 인공지능 산업의 발전은 그만큼 정체될 것이다.

인공지능 기술의 마지막 특성은 플랫폼이다. 현재 글로벌 IT기업은 인공지능 기술을 활용해 수익을 내는 구조로 플랫폼 서비스를 개발하고 있다. 인공지능 플랫폼의 가장 선두는 단연 구글이다. 구글은 자사의 심층학습 공개SW인 텐서플로우를 공개했을 당시, 심층학습 공개SW는 이미 카페(Caffe)나 토치(Torch)가 생태계를 선점하고 있었다. 그러나 구글은 텐서플로우의 저변확대를 위해 구글 클라우드와의 연계를 강조했다. 텐서플로우로 작성한

소스코드를 구글 클라우드에서 손쉽게 활용할 수 있게 환경을 조성하여 사용자의 편의를 도모했다. 또한 2017년 세계 최대 규모의 데이터 경진대회 플랫폼인 캐글을 인수하면서 텐서플로우의 확산을 꾀했다. 한편으로는 심층학습의 특성을 잘 반영한 AutoML 서비스를 출시하면서, 데이터의 특성을 자동으로 분석하여 최적의 인공신경망 구조를 제공하는 접근으로 심층학습에 대한 사용자의 진입장벽을 낮췄다. 그 밖에 페이스북은 메신저 플랫폼을 공개하여 챗봇을 쉽게 개발할 수 있는 환경을 제공하며, IBM은 왓슨을 API 형태로 제공하여 다양한 산업에의 접목을 시도하고 있다. 중국은 바이두, 알리바바, 텐센트 등 글로벌 IT 기업을 필두로 자율주행차, CCTV 안면 인식, 스마트 시티 등 인공지능을 직접 산업에 활용하는 교두보로써의 플랫폼을 적극 장려하고 있다.

인공지능 기술의 특성을 정리하자면 결국 첫 단추는 데이터로부터 시작한다. 데이터는 인공지능의 근간이 되는 원유로 그 원활한 공급이 무엇보다 선결되어야 하는 과제다. 기술적으로는 심층학습의 낮은 설명가능성으로 인해 인공지능을 도입함에 있어 사회적 합의가 필수적이다. 또한 글로벌 IT 기업의 인공지능 플랫폼으로 인해 인공지능을 활용하기 위한 진입장벽은 낮아지고 있으나, 승자독식의 원리에 의해 인공지능 산업의 구조가 격변될 가능성이 높다.

2. 심층학습의 한계와 차세대 인공지능의 부상²⁾

심층학습은 현대 인공지능의 핵심이다. 그러나 심층학습으로 모든 문제를 해결할 수 있는 것은 아니다. 심층학습은 단어 자체에서도 찾을 수 있듯이 데이터를 학습하여 패턴을 인식하는 기술이다. 따라서 데이터 수급과 관련된 문제는 곧 심층학습의 한계와도 맞물려 있을 수밖에 없다. 데이터가 심층학습의 외부적인 한계라면 기술자체 내부적인 한계도 존재한다. 심층학습의 한계는 크게 세 가지 관점에서 아래와 같이 논의할 수 있다.

첫 번째 한계는 데이터의 분포 문제다. 심층학습은 크게 두 단계로 나뉜다. 학습용 데이터(training data)는 심층신경망(Deep Neural Network)³⁾ 모델을 학습하는데 활용된다. 이후 학습이 종료되면 시험용 데이터(testing data)를 활용해 학습된 신경망의 성능을 측정한다. 문제는 학습 데이터와 시험 데이터의

2) 인공지능, 어디까지 왔나?, 소프트웨어정책연구소(2018.11.)에서 재정리

3) 심층학습의 심층(Deep)의 개념은 보통 층이 일정 수 이상의 인공신경망을 지칭하는 개념적인 용어로 사용되며, 심층신경망은 합성곱신경망, 순환신경망, 장단기메모리 등의 신경망 형태를 모두 아우르는 개념으로 활용함

분포다. 일반적으로 시험 데이터는 전체 데이터의 일부로 선택한다. 다시 말하자면, 전체 데이터가 100이라고 가정 할 때, 학습에는 80이 활용되고 나머지 20을 시험 데이터로 사용한다는 것이다. 바로 이 부분이 학습-시험 데이터가 동일한 분포에서 나왔다는 것을 의미한다. 이것이 의미하는 바는 무엇일까? 예를 들어 시뮬레이션 환경에서 로봇 팔이 물체를 집는 알고리즘을 심층학습을 활용해 성공적으로 구현했다고 보자. 여기서 시험 데이터가 여전히 시뮬레이션 환경에서 생산됐다면, 학습된 인공지능이 물건을 집는 행위를 잘 처리할 수 있을 것이다. 그러나 이 인공지능을 실제 로봇 팔에 적용한다면 필연적으로 실패할 가능성이 높다. 그 이유는 시뮬레이션 환경과 실제 환경이 매우 다르기 때문이다. 실제 환경에서는 중력, 공기와의 마찰, 관성, 기계의 갑작스런 오작동 등 고려해야 할 요소가 매우 많다. 따라서 실제 환경에 적용하기 위해서는 학습 데이터를 실제와 가깝게 구현하던가, 시뮬레이션 환경을 실제에 가깝게 구성해야 할 필요가 있다. 어느 쪽도 만만치 않다. 또 다른 예로는, 고전게임기 ATARI 2600의 벽돌깨기(breakout) 게임을 구현한 강화학습에서도 찾을 수 있다. 벽돌깨기 인공지능은 게임 인공지능의 혁신적인 결과 중 하나였지만, 벽돌깨기 게임의 명도를 20% 올렸을 때는 제대로 작동하지 않는다. 인간이 보기에는 큰 차이가 아니지만 인공지능이 바라보기에는 전혀 다른 문제라는 것이다.

심층학습의 두 번째 한계는 지속적인 학습이 불가능하다는 점이다. 우리는 흔히 인공지능을 실시간으로 학습할 수 있는 것처럼 오해할 수 있다. 그러나 현재 심층학습의 기술로는 지속적인 학습이 어렵다. 예를 들어 100만 개의 데이터로 학습한 인공신경망 모델이 있다고 가정하자. 여기에 10만 개의 데이터가 추가됐을 때, 이미 학습된 신경망에 추가적으로 10만 개를 학습하여 성능을 더 향상시킬 수 있을까? 이에 대한 결론은 “알 수 없다”가 답이다. 10만 개를 학습하여 성능이 올라갈지, 아니면 110만 개를 다시 학습하여 성능이 향상될지는 알 수가 없다. 효율의 관점에서 봤을 때는 10만 개를 추가적으로 학습하는 것이 타당해 보인다. 그러나 현재는 10만 개를 추가적으로 학습하던 110만 개를 다시 학습하던 일단 시도를 해 보고 선택하는 것이 일반적이다. 이러한 현상의 원인은 인공신경망의 구조적인 특성에서도 유추할 수 있다. 인공신경망을 학습하는 행위는 수학적으로 비용함수의 최소값을 찾는 과정이다. 그러나 비용함수의 차원은 인공신경망의 가중치 개수이며, 보통 수 백 만개를 넘어서기 때문에 결과적으로 수 백 만 차원의 비선형 함수에서 최소값을 찾는

문제로 귀결된다. 보통 5차원만 넘어서도 차원의 저주(curse of dimensionality)로 인해 해석적으로 최소값을 구하기 어렵고, 수치적인 근사로 찾는 방법이 일반적이다. 따라서 100만 개로 학습한 최적의 인공신경망을 10만 개로 추가 학습한다고 해도 의미 있는 결과를 보장할 수 없다.

심층학습의 세 번째 한계는 설명가능성이 낮다는 점이다. 현재 심층학습의 성능은 비록 설명이 불가능해도 학계에서 받아들여지고 산업계에 활용될 만큼 높다. 그만큼 높은 성능이 현대 인공지능의 붐을 이끈 원동력이다. 그러나 어떠한 방법론을 쓰던 성능이 좋기만 하면 되는 분야가 있는 반면, 그 인과관계를 철저히 규명해야하는 분야도 있다. 예를 들어 자율주행차가 사고를 냈는데, 그 원인을 특히 심층학습에서 찾으려 한다면 매우 어려울 것이다. 그 이유는 현재 심층학습 기술의 설명가능성이 매우 낮기 때문이다. 인공지능의 설명가능성은 보수적인 관점에서 인공지능이 우리의 삶과 공존할 수 있는 마지노선일 것이다. 한편으로는 설명가능성에 집중하기보다, 경험적으로 인공지능을 도입하는 것이 효율적이라면 응당 도입해야 한다는 관점도 있다. 현재 많은 과학기술들은 그 역효과를 철저히 규명하고 보장하기는 어렵다. 기술을 먼저 도입하고 나서 발생하는 문제를 단계적으로 대처해 나가자는 관점이다. 그러나 인공지능이 점점 고도화됨에 따라 인공지능이 인적·물적 피해의 직접적인 원인이 될 수도 있다. 이것이 설명 불가능하여 모든 책임을 단순히 인공지능을 사용한 주체에 부과한다면 매우 부조리 할 가능성이 높으며 사회적 문제를 야기할 것이다.

지금까지 심층학습의 한계에 대해 논의했다. 정리하자면 심층학습의 한계에 대한 원인은 데이터와 기술로 볼 수 있다. 기계학습 기반의 인공지능은 데이터를 학습한다는 속성으로 인해 데이터의 한계와도 이어진다. 특히 다양한 환경에서의 데이터 확보가 중요하다. 자율주행차가 지속적인 주행을 강조하는 것 역시 다양한 상황의 데이터 확보에 있는 것이다. 데이터는 현재 심층학습이 당면한 문제라면 장기적으로는 기술적인 해법이 필요할 것이다. 적은 데이터로도 성공적인 학습을 수행하며, 입력과 출력의 인과관계를 명확히 하는 설명 가능한 인공지능이 그 해법이다. 또한 심층학습의 틀에서 벗어나 다양한 지능적 행동을 수행하는 범용 인공지능의 접근도 심층학습의 한계를 뛰어 넘을 수 있을 것이다.

3. 국내외 인공지능 정책 현황

인공지능의 전 세계적인 관심으로 인해 인공지능 기술 확보를 위한 경쟁이 치열해지고 있다. 인공지능은 과거 각각 두 차례에 걸친 황금기와 암흑기를 거치면서 세 번째 황금기를 맞이했다. 많은 전문가들은 이번 인공지능의 황금기는 지속될 것이라는 긍정적인 반응을 보이고 있다. 이 긍정적인 반응은 이번 황금기를 계기로 인간 수준에 버금가는 인공지능(Human-Level AI)이 탄생할 것으로 예측한다.

인간 수준의 인공지능으로 가는 길은 단순히 기초·원천기술의 고도화가 응용 및 개발 연구에 이어지는 것은 아니다. 현재 인공지능을 주도하는 주체는 구글, 페이스북, 바이두와 같은 글로벌 IT 기업이다. 기업이 이윤을 창출해내기 위한 도구로 인공지능을 다양한 산업에 접목시키고 있으나, 앞서 소개했듯이 사회적·기술적 장벽이 존재한다. 이를 해결하기 위해 역으로 인공지능 기초 기술에 대한 연구개발 수요가 발생한다. 예를 들어, 자율주행차가 도입되어 우리의 삶과 공존하기 위해서는 인공지능이 어떻게 의사결정을 내리는지가 규명되어야 하며, 이것이 사회적으로 공감대를 이뤄야 한다. 그러나 현재와 같이 심층학습이 동작하는 방식이 여전히 블랙박스로 남겨져 있다면, 아무리 고도화된 자율주행차라 하더라도 우리 사회에 안착되기는 어려울 것이다.

따라서 인공지능 기술을 육성하고 산업을 진흥하기 위한 방향은 매우 복잡한 속성을 띄고 있다. 앞서 소개한 자율주행차의 인공지능이 설명 가능해진다고 해도 갈 길은 멀다. 자율주행차로 인해 인적·물적 피해가 발생하면 그 책임 소지를 분명하게 만들어야 하기 때문이다. 학계에서는 자율주행차에 개인처럼 보험의 대상으로 만들어 사고를 많이 발생시키는 자율주행차에 높은 보험료를 부과하는 의견도 나오고 있다. 인공지능은 비단 그 기술뿐만 아니라 사회적인 파급효과까지 고려해야 한다는 점에서 성공적인 인공지능 정책을 제안하기란 쉽지 않은 일이다.

이렇게 인공지능이 복잡함에도 불구하고 이렇게 전 세계적인 붐을 이끄는 동력은 무엇일까? 그 답은 계속 언급해 왔듯이 바로 성능이다. 특히 인공지능에 대한 학계와 산업계의 과도한 집중으로 인해 기술 발전 속도가 가속화되고 많은 기대를 걸고 있는 상황이다. 이러한 잠재력 때문에 세계 각국은 인공지능 기술의 조속한 확보를 위해 노력하고 있다.

미국은 인공지능 분야의 명실상부한 최고기술선도국이다. 구글, 페이스북, 아마존을 비롯한 인공지능 글로벌 IT 기업과 함께 인공지능 학습을 위한 연산처리장치 제조 기업인 Intel, NVIDIA, AMD 등이 포진하고 있다. 지난 2016년는 백악관에서 인공지능에 대한 정책 3건⁴⁾을 발표하여 미국이 인공지능 분야의 리더십을 유지하기 위해 비단 기술뿐만 아니라 사회·경제적 파급효과에 까지 대비를 한다고 밝혔다. 그러나 후술할 중국의 급격한 기술 추격으로 인해 미국은 기술육성 집중으로 노선을 변경했다. 2018년 2월부터 3차례 진행된 공청회⁵⁾에서는 미국이 인공지능 리더십을 주도적으로 유지해야한다는 결론을 내렸다. 이어 2018년 5월에는 미국 연방정부는 인공지능 특별위원회(select committee)를 구성하여 인공지능 육성 정책에 대한 의지를 표명했다. 또한 보고서 『기계의 부상』⁶⁾에서는 인공지능 분야에서의 중국의 추격을 가시화하고 미국이 명실상부한 인공지능 최고기술선도국으로 자리매김하기 위해 기술 육성에 집중한다고 밝혔다. 이 보고서에서는 심층학습을 비롯한 인공지능의 한계를 언급하면서 기술적으로 이를 보완하는 R&D가 중요하다고 언급했다. 미국은 트럼프 행정부가 집권하면서 슬로건으로 외친 “American First”의 철학이 미국의 인공지능 정책에도 담겨있다고 볼 수 있다. 따라서 현재 미국의 인공지능 정책 노선은 인공지능의 기술 육성에 치중해 있다고 볼 수 있다.

중국은 인공지능 분야에서 가장 가파른 성장을 하고 있다. 국내에서 조사된 2017년 ICT 기술수준조사⁷⁾에서는 중국과 미국과의 기술격차가 큰 폭으로 좁혀져 1.4년으로 조사됐다. 그러나 현장에서 체감하는 중국의 발전 속도는 더욱 빠르다. 인공지능 국제역량을 비교한 보고서⁸⁾에 따르면 중국이 양적인 측면(연구 건수, 피인용 횟수)에서 미국을 앞섰다. 질적인 측면에서는 여전히 미국과 주요국에 열세이지만, 인공지능의 변방이었던 중국이 이렇게 급성장한 이유는 중국 정부의 정책⁹⁾과도 맞닿아 있다. 중국은 2005년부터 중국어 자연어 처리를 위한 인공지능 기술 개발을 지속적으로 투자해 왔으며, 특히 2015년 『인터넷 플러스』 정책을 시작으로 본격적인 인공지능의 투자를 감행했다.

4) 국가 인공지능 R&D 전략(National AI R&D Strategic Plan), 인공지능, 자동화, 경제(AI, Automation and the Economy), a 미래 인공지능에 대한 준비(Preparing for the Future of AI)

5) Hearing: Game Changers – Artificial Intelligence Part I (2018.02.), Hearing: Game Changers – Artificial Intelligence Part II, Artificial Intelligence and the Federal Government (2018.03.), Hearing: Game Changers – Artificial Intelligence Part III, Artificial Intelligence and Public Policy (2018.04.)

6) Rise of Machines, Artificial Intelligence and its Growing Impact on U.S. Policy (2018.09.)

7) 2017 ICT 기술수준조사보고서, IITP (2018.05.)

8) 인공지능 연구역량 국제비교 및 시사점, 소프트웨어정책연구소 (2018.11.)

9) 인공지능 팔목상대, 중국, 소프트웨어정책연구소(2017.07.)

인공지능의 기념비적인 사건인 알파고 대국이 열린 2016년은 중국이 본격적으로 인공지능의 투자를 이어가는 계기가 됐다. 2017년 성능을 개선한 알파고와 바둑 세계챔피언 신성 커제와의 대결 성사도 중국이 인공지능에 대한 사회적 공감대 형성을 목표로 했다고 볼 수 있다. 당시 대국에서 알파고는 한 층 더 진보된 성능으로 커제를 2:0으로 완파하여 중국 사회에 인공지능의 중요성을 전파했다. 또한 중국 정부는 2017년 『차세대 인공지능 발전 계획』을 발표하여 2030년까지 인공지능 핵심 산업 규모 1조 위안, 인공지능 관련 산업 규모 10조 위안을 달성하기 위한 청사진을 발표했다. 인공지능 핵심 기술 육성을 통해 전 산업의 인공지능화를 목표로 한 이 전략은 이후 중국의 행보를 가속화하는 시발점이 됐다. 14억 중국 인구의 안면인식, 권역성별로 추진되는 인공지능 거점사업 등은 중국이 미래 먹거리 산업으로 인공지능을 선택했다는 사실에 이의가 없다. 특히 중국의 전략은 데이터에서 찾아볼 수 있다. 앞서 소개했듯이 현재 심층학습 기술은 불완전하다. 이 간극을 메워줄 수 있는 가장 빠른 조치는 다양한 데이터의 확보다. 이것은 개인의 권리와도 직결되는 부분이기 때문에 서구 사회에서는 상대적으로 느낄 수밖에 없는 사회적 합의를 통해 진행해야 한다. 그러나 중국은 정부의 지도 감독과 책임 하에 데이터 수집과 활용에 힘쓰고 있다. 바로 이 대목이 중국이 미국을 위협할 정도로 급성장 한 계기라고도 볼 수 있다. 중국은 이러한 데이터를 확보해 학습한 시스템을 플랫폼화하여 주변국으로의 영역을 확장시키고 있다. 이러한 관점에서 볼 때 중국은 현대 인공지능의 속성을 잘 이해하고 전략을 펼치고 있다고 본다. 또한 이러한 과감한 정책을 지탱해주는 인적·물적자원과 풍부한 내수시장이 결합된 결과로써, 중국은 단일 국가로 인공지능 분야에서 미국을 위협하는 유일한 나라로 성장하고 있음을 유추할 수 있다.

유럽연합(EU)은 2018년 5월 개인정보보호법(General Data Protection Regulation, GDPR)¹⁰⁾을 발효하여 인공지능 시대에서 개인의 정보 주권을 더욱 공고히 했다. 중국과는 반대로 데이터의 무분별한 활용을 방지한다는데 초점이 맞춰져 있다. 우리는 사실 우리가 모르는 사이 복잡한 이용 약관에 동의하고 서비스를 활용하고 있다. 예를 들면 인공지능 스피커의 가상비서 서비스를 활용하기 위해서는 개인의 음성이 클라우드 시스템으로 전송되는 것에 대해 동의를 해야 한다. GDPR은 이러한 과정에서 개인정보의 수집·저장·처리 등에

10) EU 개인정보보호법 발효 이후 각국의 대응 동향, 소프트웨어정책연구소 (2018.12.)

대한 과정을 더욱 투명하게 하여 정보 제공 주체의 권리를 더욱 높인다는 것이다. 데이터가 매우 중요한 인공지능 시대에 GDPR은 이를 역행하는 방향이라고도 판단할 수 있다. 그러나 EU는 인공지능의 가능성에 대해 매우 긍정적으로 판단하고 있으며, 그 진로를 데이터가 아닌 기초과학에서 찾는 경향이 짙다. 2013년부터 추진된 『휴먼 브레인 프로젝트』는 EU의 전통적인 강점인 기초과학 기술력을 결집하여 인공지능의 근간이 되는 뇌과학에 집중하고 있다. 심층학습의 많은 석학들이 심층학습을 개선하기 위해 우리의 뇌 구조에 더 집중해야 한다는 사실을 강조한 만큼 EU는 인공지능의 본질적인 성능향상을 위해 투자를 집중하고 있다. 데이터가 현대 심층학습의 구조적인 한계를 가장 쉽게 해결할 수 있는 대책이라면, 인공지능의 기초과학은 더 본질적인 성능향상을 추구한다는 데 방점이 있다. 실제로 휴먼 브레인 프로젝트의 일환으로 영국의 맨체스터 대학은 세계 최초의 뉴로모픽¹¹⁾ 슈퍼컴퓨터 스피네이커(SpiNNaker)를 개발하여 뇌모사 신경망 칩 분야에서 선도적인 역할을 수행하고 있다. 요약하자면 EU는 개인정보보호를 강화하여 정보제공 주체의 권리를 강화하는 한편, 인공지능 기초과학에 집중하여 본질적인 성능향상을 달성하는데 목표를 두고 있다.

일본은 크게 두 가지 특징을 바탕으로 그 정책방향을 설명할 수 있다. 일본 역시 전통적인 제조 강국으로 로봇 산업에 강점을 갖고 있다. 일본은 그간 축적해 왔던 로봇 기술을 인공지능을 활용해 더욱 강화하는 전략으로 첫 번째 정책 방향을 이해할 수 있다. 특히 지난 2017년 신설된 인공지능 국제표준¹²⁾에서는 워킹그룹 사용자 경험과 응용(Use cases and applications)에서 일본이 의장국을 맡으면서 로봇 산업의 표준 선도를 꾀하고 있다. 한편 일본은 미국이나 중국과 같이 인공지능을 선도하는 글로벌 IT 기업이 부재한 이유로 자국의 중소기업과 대학이 자유롭게 인공지능을 개발할 수 있는 대규모 클라우드 컴퓨팅 인프라 ABCI(AI Bridging Cloud Infra)를 구축했다. 약 195억 엔이 투자된 인공지능 전용 컴퓨팅 인프라 ABCI는 일본내 인공지능 연구·산업 생태계를 활성화시키는 일환이다. 또한 자국민이 안심하고 사용할 수 있는 신뢰성을 확보하여 일본의 인공지능 부흥에 마중물 역할을 할 것이다.

11) 뉴로모픽(Neuromorphic)은 동물의 신경계를 모사한 하드웨어로 실제 동작방식은 인공신경망과 유사함. 그러나 현대 컴퓨터와 구조(architecture)가 판이하게 다르기 때문에 현대 심층학습에서 활용되는 최적화 기법을 활용할 수 없고 스파이킹 신경망 구조를 학습해야 함. 뉴로모픽 연산처리장치는 통칭 뉴로모픽칩이라고 하며 저전력의 특징이 있음.

12) ISO JTC1 / SC42 Artificial Intelligence <https://www.iso.org/committee/6794475.html>

우리나라는 지난 2016년 알파고 쇼크를 직접적으로 경험한 나라로 인공지능에 대한 대중의 인식은 매우 높은 편이다. 또한 2018년 5월 공개된 『인공지능 R&D 전략』은 향후 5년 간 2.2조 원을 투입한다는 계획을 밝혀 우리나라의 인공지능 기술 육성과 산업 진흥을 위한 청사진을 보여줬다. 이 전략이 성공적으로 시행되어야 하는 점에는 이견이 없을 것이다. 또한 세계 최고 수준의 개인정보보호 규제를 보유하고 있는 우리나라¹³⁾는 데이터 문제와 관련된 개선안을 도출하기 위해 2018년 6월 『데이터 산업 활성화 전략』을 공개했고, 2019년 1월에는 『데이터·AI경제 활성화 계획』이 공개되어 인공지능과 데이터의 연계를 통한 발전 전략을 제시했다. 우리 정부는 인공지능을 둘러싼 문제 인식과 이를 해결하고자 하는 의지는 매우 강하다.

4. 우리의 당면 과제

인공지능 기술 선점을 위한 각국의 경쟁이 치열해지는 가운데, 국가별로 자국의 장단점을 고려한 전략을 펼치고 있다. 이러한 가운데 우리의 과제는 무엇일까? 데이터가 중요하다는 사실은 이미 많은 논의를 통해 중요성이 인식된 상황이다. 특히 4차산업혁명위원회에서는 네 차례에 걸친 해커톤에서 데이터와 관련해 개인정보, 위치정보가 각각 논의되었을 정도로 그 비중이 매우 높았다. 또한 『데이터 산업 활성화 전략』에서는 우리의 법·제도적인 규제 하에 데이터를 선순환시키는 체계를 마련하는 전략을 소개했다. 그러나 우리나라는 중국과 같이 정부의 적극적인 개입을 바탕으로 한 접근으로 인공지능 산업을 진흥할 수는 없을 것이다.

국내 인공지능 역량 강화를 위한 필자의 방향은 인공지능의 기초·원천기술에 대한 육성이라고 생각한다. 그렇다면 어떠한 기초·원천기술에 집중해야 할지에 대해 논의해야 한다. 그러나 그 선택지에 대한 도출은 쉽지 않다. 가장 근본적인 이유는 인공지능 기술이 빠르게 진화하며, 오픈 사이언스의 특징을 통해 매우 빠르게 공유되기 때문이다. 이 보고서는 인공지능의 기초·원천기술의 현황을 분석하고, 우리의 미래 먹거리 후보를 살펴보기 위한 차세대 인공지능의 현황을 논의하고자 한다. 용이한 분석을 위한 연구의 분류를 제시할 것이다.

13) 스타트업 코리아! 결과보고서, 아산나눔재단 (2017)

제2절 연구 방법 및 범위

이 보고서의 주된 연구 방법은 인공지능과 관련된 최신 연구의 분석에 있다. 효율적인 분석을 위해 외부 전문가의 도움을 얻었으며, 자체적으로도 최신 논문을 검토했다. 인공지능 연구 사례의 분석은 정책적 전략이나 제언을 도출하기 위해 가장 선행되어야 한다는 부분에서 초기 연구 단계에 지나지 않는다고 해석할 수 있다. 그러나 인공지능의 기술 자체가 발전하는 속도가 매우 빠르고 그 영역이 급속도로 확장됨에 따라 시의성 있는 분석 자체만으로도 의미가 있다고 판단된다. 이러한 필요성으로 인해 이 연구의 주된 목적은 인공지능 기술의 분석을 통해 정책 설계의 기반자료로 활용함에 있다.

이 보고서는 특히 인공지능의 기초·원천기술과 차세대 인공지능에 집중하여 그 현황을 살펴볼 것이다. 체계적인 분석을 위한 인공지능의 연구 영역을 제안하고, 이 틀을 활용하고자 한다. 인공지능의 연구 영역은 크게 네 가지로 구성되며 국가 연구개발의 단계로 널리 활용되는 기초·응용·개발 단계에 착안하여 인공지능을 기초·응용·개발 연구의 세 단계로 나눠 볼 것이다. 마지막으로 차세대 인공지능을 따로 분류하여 인간 수준의 인공지능을 향한 대표적인 연구를 소개할 것이다.

본격적인 논의로 들어가기에 앞서 인공지능과 심층학습의 관계에 대해 정의가 필요하다. 그 이유는 인공지능이라는 용어가 매우 포괄적인 개념인 반면, 인공지능이 부상한 계기가 바로 심층학습에 있기 때문이다. 심층학습은 인공지능의 방법론인 기계학습에서 인공신경망을 활용한 것으로 인공지능의 많은 방법론 중 하나에 불과하다. 그러나 심층학습이 기존의 방법론을 압도할 만큼 성능이 뛰어나기 때문에 심층학습을 인공지능 전체로 바라보는 시각이 있는 것이다. 실제 연구 현황에서도 심층학습의 비중이 매우 높기 때문에 인공지능과 심층학습을 동치로 볼 수 있으나 이 보고서에서는 본질적으로 심층학습이 인공지능의 한 방법론이라는 사실로 출발한다.

2장에서는 앞서 소개한 인공지능 연구 영역에 대한 분류를 소개하고 3장에서는 인공지능 기초 연구 현황과 차세대 인공지능의 한 분야인 범용 인공지능의 연구 현황을 살펴볼 것이다. 4장에서는 요약과 시사점으로 우리의 당면과제를 도출하여 마무리 지을 것이다.

제2장 인공지능 연구 영역의 분류¹⁴⁾

제1절 개 요

본격적으로 인공지능의 최신 연구 현황을 분석하기 앞서, 인공지능 연구 영역을 분류해 보겠다. 글로벌 시장조사기관은 인공지능 기술과 산업 분류체계¹⁵⁾를 제시했으나 기술과 산업이 혼재되어 있다. 인공지능의 특성상 기술과 산업을 분리하기는 어려우나, 이 보고서에서는 인공지능 산업을 인공지능 기술이 타 산업에 접목되는 형태로 구분하겠다. 이로써 이 연구에서 관심 있게 다룰 인공지능 기초·원천 연구와 차세대 인공지능 연구의 영역을 정의할 것이다.

인공지능 연구 영역의 분류는 인공지능 R&D와 산업의 특성을 바탕으로 가설적인 접근을 취할 것이다. 그 이유는 아직 인공지능에 대한 체계적인 분류가 매우 유동적임으로 인해 관점에 따라 쉽게 변화할 수 있기 때문이다. 또한 인공지능의 범위가 매우 넓기 때문에 데이터를 바탕으로 한 접근 역시 상이할 수 있다. 이 보고서에서는 가설이지만 인공지능의 굵직한 트렌드를 바탕으로 연구 영역의 구분을 제안할 것이다. 각 연구 영역은 인공지능의 특징에 따라 구분되고 각 연구 영역의 상호관계를 분석하여 연구 영역의 분류에 대한 타당성을 논의하겠다.

1. 인공지능 R&D와 산업의 특징

다음 <표 2-1> 인공지능 기술의 주요 특징을 기술한다. 이것은 전문가 회의, 국내외 문헌 자료 등에서 재구성한 것으로 세부적인 내용에 대해 후술할 것이다. 이 특징을 바탕으로 인공지능 연구 영역의 구분에 대한 가설을 제시할 것이다.

14) 인공지능, 어디까지 왔나?, 소프트웨어정책연구소(2018.11.)에서 재정리

15) 별첨 1. 인공지능 기술과 산업의 분류 참고

<표 2-1> 인공지능 기술의 주요 특징

1. 인공지능 기술의 발전 속도는 매우 빠르다.
2. 인공지능 기술이 적용되는 분야가 빠른 속도로 확장되고 있다.
3. 인공지능 기술을 주도하는 주체는 글로벌 IT기업이다.
4. 현재 인공지능 기술은 약 인공지능(weak AI)에 머물고 있다.

1. 인공지능 기술의 발전 속도는 매우 빠르다. 인공지능의 발전 속도가 매우 빠른 근본적인 이유 심층학습에 대한 연구의 과도한 집중 때문이다. 심층학습은 앞서 소개한대로 한계가 분명한 기술이지만, 매우 높은 성능을 보이고 있다. 이로 인해 인공지능이라는 키워드는 연구계와 산업계를 넘나들며 활용이 되고 있다. 구체적으로 심층학습 기술은 기존의 모델링이나 방법론을 대체하는 방향으로 진행되고 있다. 또한 오픈사이언스 생태계의 확대로 연구결과를 공유하는 arXiv, 개발한 코드를 공개하는 github 저장소, 수 십여 종에 이르는 심층학습 공개SW 등은 심층학습이라는 기술의 접근성 자체를 낮추는 데 기여했다. 또한 심층학습의 설명 가능성을 보장하기 위한 이론적 연구 역시 심층학습의 눈부신 성과로 인해 연구가 집중되고 있다. 예를 들면, 일반적인 인공신경망에서 은닉층(hidden layer)을 3층, 4층으로 구성했을 때의 차이점을 수학적으로 증명하는 내용 등을 다룬다. 혹은 각 은닉층이 갖는 수학적 의미를 도출하는 연구다. 이처럼 심층학습의 높은 성능과 낮아진 진입장벽으로 인해 유례를 찾기 힘들 정도로 전 세계적으로 연구가 활발히 이루어지고 있다. 이 사실은 결론적으로 심층학습의 빠른 발전속도에 기여하고 있다. 한 가지 특이한 점은 현재 인공지능 연구에서 심층학습이 차지하는 비중은 대략적으로 80% 이상으로 추정된다는 것이다. 따라서 현 시점의 인공지능은 대부분이 심층학습이라는 관점에는 이의가 없는 편이다. 결국 인공지능 기술의 빠른 발전 속도는 곧 심층학습의 방법론 자체의 성능향상에 기여하고 있다. 예를 들면, 적은 데이터로 학습하는 방법, 학습하는 방법을 학습하는 접근(Learning to learn), 설명 가능한 인공지능 등이 이에 속한다고 볼 수 있다. 이러한 접근은 기초·응용·개발의 R&D 단계로 볼 때 기초연구에 속한다.

2. 인공지능 기술이 적용되는 분야가 빠른 속도로 확장되고 있다. 앞서 언급한대로 현재 인공지능은 대부분 심층학습이 이끌어간다고 볼 수 있다. 심층학습은 매우 다양한 형태로 진화하고 있으나 본질적으로는 분류와 군집에 최적화된 방법론이다. 여기서 분류와 군집은 인식(recognition)과 탐지(detection)의 영역에서 필수적인 요소 볼 수 있다. 인식과 탐지는 컴퓨터 과학에서 대표적으로 컴퓨터 비전, 자연어 처리, 음성 인식, 데이터 마이닝 등의 영역이 속한다. 따라서 심층학습은 인공지능과 관련된 컴퓨터 과학의 영역을 재편했다고 볼 수 있으며, 실제로 언급한 분야에서 높은 성과를 달성했다. 이러한 성과는 곧 타 산업과 연구 영역으로 전파되고 있다. 컴퓨터 비전은 의료 영역에서 암을 진단하고, 자연어 처리는 기계 번역에 적용되며, 음성 인식은 스마트 스피커에 탑재됐다. 또한 많은 데이터가 누적돼있는 기초과학 분야에까지 심층학습이 전파되어 새로운 모델링 기법으로 활용되고 있다. 따라서 인공지능 기술의 지속적인 확대는 심층학습을 실제 응용 분야나 타 산업에 접목시키는 것으로 해석할 수 있다. R&D 단계에서 볼 때는 응용·개발연구의 영역이 혼합되어 있는 양상이다.

3. 인공지능 기술을 주도하는 주체는 글로벌 IT기업이다. 이 사실에 대해서는 큰 이견이 없을 것이다. 인공지능을 대표하는 글로벌 IT기업은 미국과 중국이 우세한 상황이며, 기술력과 생태계는 미국이 앞서고 있으나 중국이 무섭게 추격하고 있는 흐름이다. 글로벌 IT기업이 인공지능을 주도하고 있는 상황은 기술발전의 측면에서는 이례적이다. 보통 기초학문에서의 새로운 발견이 응용학문으로 이어지고, 이것이 산업계에 전파되는 과정이 일반적이다. 그러나 인공지능은 산업계에서 주도하고 있으며 이것이 역으로 기초학문에 영향을 미치고 있다. 이러한 이유로 이론적 접근보다는 실증적인 접근이 받아들여진다고 볼 수 있다. 이론적으로 탄탄한 배경보다는 실제 문제해결의 가능성에 집중하고 있는 인공지능의 활용은 곧 SW의 속성과도 일맥상통한다. SW는 승자독식의 원리가 적용되는 대표적인 분야로, 인공지능 역시 아직 승자는 없는 상황이나 승자를 차지하기 위한 경쟁이 본격화 되고 있다. 그 중심에는 플랫폼이 있다. 이 부분은 앞서 서론에서 소개한 인공지능의 기술적 특징과도 맞물려 있는 내용이다. 인공지능의 플랫폼화 전략에 단연 앞서있는 기업은 구글이다. 구글은 심층학습 공개SW, 클라우드 서비스에서도

후발주자였다. 심층학습에서는 텐서플로를 공개하기 이전 카페(Caffe)나 토치(Torch)와 같은 공개SW가 이미 선점을 하고 있었으며, 클라우드는 아마존 웹 서비스가(AWS)가 시장의 대부분을 차지하고 있었다. 이 상황에서 구글은 기계학습 클라우드 플랫폼(Machine Learning Cloud Platform)을 서비스하여 활로를 모색했다. 구글의 텐서플로로 작성된 소스코드는 구글의 대규모 클라우드에 활용할 수 있으며, 최근에는 데이터를 기반으로 최적의 학습 모델을 제공하는 AutoML을 출시했다. 비단 기계 학습 자체뿐만 아니라 고성능 OpenAPI¹⁶⁾도 지원하고 있다. 중국은 특히 인공지능 플랫폼 전략을 충실히 이행하고 있다. 앞서 정책 현황에서도 소개했듯이, 중국은 정부가 직접 데이터 활용을 장려하고 있다. 양질의 데이터로 학습된 시스템을 적극적으로 활용하기 위해서는 플랫폼 전략이 가장 적절하기 때문이다. 중국은 특히 동남아시아권 개발도상국의 인공지능 도입을 위해 플랫폼 형태의 서비스를 추진하고 있다. 이처럼 글로벌 IT 기업은 인공지능을 플랫폼화 하여 인공지능에 대한 진입장벽을 더욱 낮춤과 동시에 플랫폼을 통한 지속적인 성능향상으로 시장을 석권하기 위한 전략을 펼치고 있다.

4. 현재 인공지능 기술은 약 인공지능(weak AI)에 머물고 있다. 심층학습이 눈부신 성과를 거두고 있으나 기술적인 관점에서 현재 인공지능은 약 인공지능에 머물러 있다. 즉, 인간의 특정한 지능을 모사하는 것은 잘 하지만 여러 가지 지능을 단일 시스템으로 모사하지는 못한다는 것이다. 이를 극복하기 위한 연구는 현재 진행 중에 있다. 많은 연구자들이 심층학습이 그 돌파구를 열 것이라고 기대하고 있으나, 전통적으로 인간의 지능의 실체에 대해 연구해 온 뇌과학, 인지과학, 심리학 등의 분야에서는 여전히 범용 인공지능의 길은 어렵다고 추정한다. 감정(emotion)을 예로 들어보자. 현재 심층학습에서는 사람의 표정으로 감정을 예측하는 기술로 소위 감성 지능을 구현하기 위해 노력하고 있다. 그러나 심리학자, 뇌과학자의 관점에서의 감정은 다르다. 사람은 기뻐도 무표정 할 때가 있으며, 슬플 때 도리어 웃는 경우도 있기 때문이다. 이러한 것이 구현되려면 감정에 대해 탄탄한 이론적인 연구가 선행되지 않고서는 불가능하다는 입장이다. 그럼에도 불구하고 세 번째 황금기를 맞이한 인공지능의 관심은 여전히 현재 진행형이다. 인공지능 기술에서 단연 선두를

16) 시각 지능, 기계 번역, 음성 인식 등

달리고 있는 미국은 차세대 인공지능 AI Next Campaign을 추진하며 20억 달러를 투자한다고 밝혔다. 미국의 뒤를 바짝 추격하고 있는 중국 역시 2017년 차세대 인공지능 발전계획을 발표하며 범용 인공지능 개발에 대한 적극적인 의지를 보이고 있다. 현재 심층학습을 비롯한 인공지능 기술은 약 인공지능에 머물고 있으나 이를 극복하기 위한 차세대 인공지능 연구가 추진되고 있다고 볼 수 있다.

지금까지 인공지능 R&D와 산업의 특징으로 다섯 가지를 논의했다. 이 특징들을 바탕으로 도출한 시사점은 다음 <표 2-2>와 같다. <표 2-2>는 다음 절에서 제시할 인공지능 연구 영역 구분의 근거자료로 활용될 것이다.

<표 2-2> 인공지능 기술의 주요 특징과 시사점

1. 인공지능 기술의 발전 속도는 매우 빠르다.
→ 인공지능 자체의 성능 향상을 위한 인공지능 기초 연구
2. 인공지능 기술이 적용되는 분야가 빠른 속도로 확장되고 있다.
→ 인공지능을 도구로 활용하는 응용·개발 연구
3. 인공지능 기술을 주도하는 주체는 글로벌 IT기업이다.
→ 인공지능을 둘러싼 글로벌 IT 기업의 플랫폼 경쟁
4. 현재 인공지능 기술은 약 인공지능(weak AI)에 머물고 있다.
→ 범용 인공지능 등 차세대 인공지능의 대두

제2절 인공지능 연구 영역 분류 제안

인공지능 기술은 앞서 소개했듯이 범용 목적 기술에 가까운 도구적인 속성으로 인해 명확한 기준을 근거로 연구의 영역을 구분하기 어렵다. 또한 해외 기관들이 제시하는 기준들은 기술과 산업 분류가 혼재되어 있는 상태라 일관성이 부족하다. 이 보고서에서는 인공지능 R&D와 산업의 특징을 바탕으로 연구 영역의 구분을 위한 가설을 세워볼 것이다. 먼저 인공지능 연구 영역은 총 네 가지로 그 정의는 다음 <표 2-3>과 같다.

<표 2-3> 인공지능 연구 영역의 구분과 정의

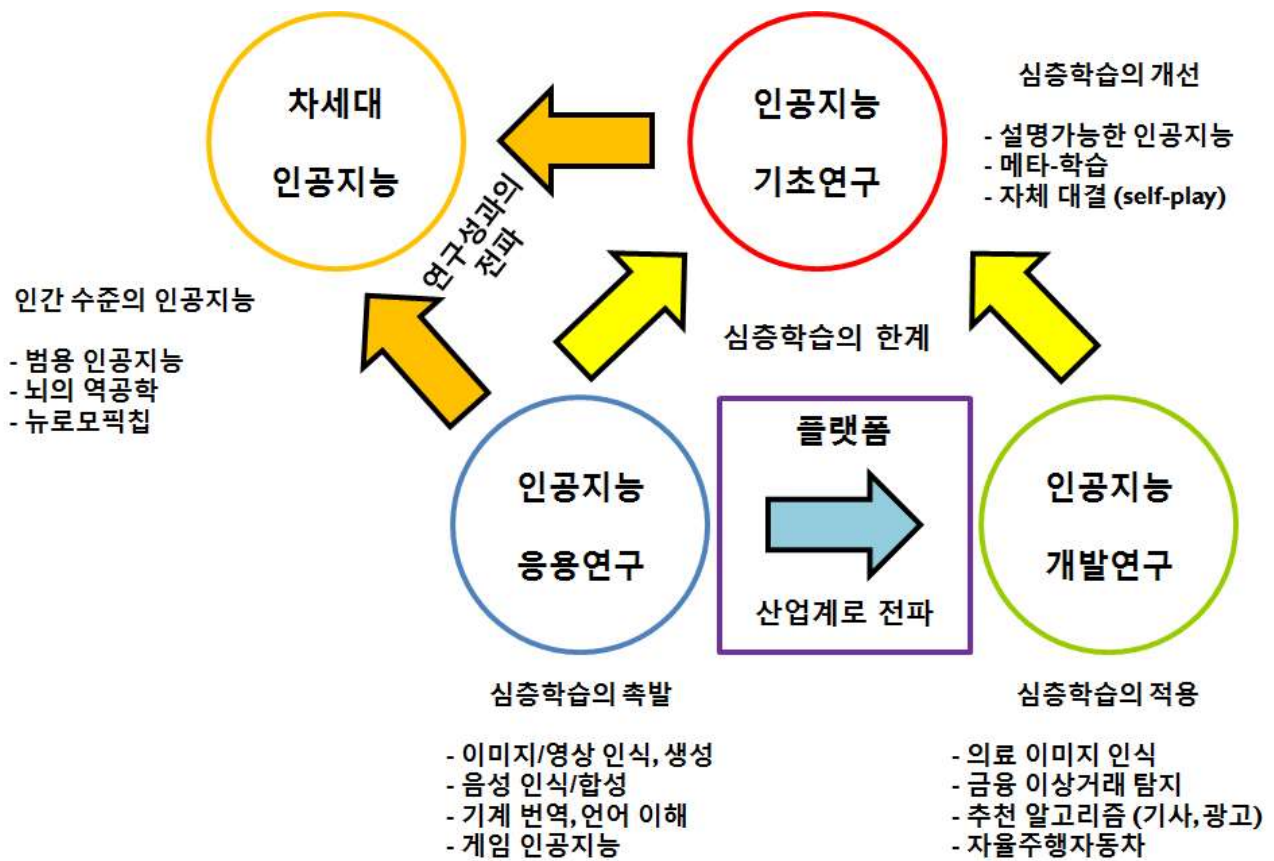
구 분	정 의
인공지능 기초연구	- 인공지능 성능 자체를 향상시키기 위한 연구 (향후 5년) - 예시 : 메타 학습, 설명 가능한 인공지능 등
인공지능 응용연구	- 컴퓨터 과학의 인식 및 탐지 분야 연구 - 예시 : 컴퓨터 비전, 자연어 처리, 음성 인식, 데이터 마이닝 등
인공지능 개발연구	- 인공지능 응용연구가 타 산업에서 활용되는 연구 - 예시 : 의료 이미지에서 암 진단, 금융 이상거래 예측 등
차세대 인공지능	- 사람 수준의 지능을 구현하기 위한 인공지능 연구 (향후 20년) - 예시 : 범용 인공지능, 인공 생명 등

<표 2-3>에서 인공지능의 연구 단계를 R&D 단계를 차용해서 썼으나, R&D에서 의미하는 기초·응용·개발 연구와는 의미가 일맥상통하지는 않으므로 해석의 유의가 필요하다. 또한 네 가지 영역 명확하게 구분되는 개념은 아니며 서로 교집합이 존재할 수 있다는 사실 역시 감안해야한다. 그것이 인공지능 R&D의 특징이기 때문이다.

〈표 2-3〉의 흐름은 다음과 같이 해석할 수 있다. 먼저 심층학습이 본격적으로 성과를 보여주기 시작한 것은 인공지능 응용연구였다. 특히 컴퓨터 비전의 분야에서 인공신경망 기술을 지속적으로 연구했던 결과가 심층학습의 모태가 된다. 심층학습의 4대 석학¹⁷⁾인 얀 르쿤 뉴욕대 교수는 필기 숫자 인식을 위해 합성곱신경망(Convolutional Neural Network)을 제안했으며, 심층학습의 창시자인 제프리 힌튼 토론토대 교수는 비지도학습 기반의 심층신뢰신경망(Deep Belief Network)의 개념을 통해 역시 이미지 인식의 성능을 획기적으로 향상시켰다. 따라서 인공지능 응용연구가 혁신을 일으킨 주체가 됐다. 특히 심층학습이 이미지, 음성, 언어 분야에서 괄목할만한 성능이 증명되자 산업계로의 전파가 활발히 이루어졌다. 과거 낮은 인식률로 인해 산업에의 적용이 어려웠던 분야가 점차 활력을 찾기 시작했다. 이는 인공지능 응용연구가 개발연구의 마중물 역할을 했다고 볼 수 있다. 인공지능 개발연구에서도 각 산업계에서 적용한 심층학습의 성과가 두드러지는 반면, 심층학습의 태생적인 한계로 인한 어려움이 발생했다. 바로 서론에서 소개한 심층학습의 한계다. 이를 극복하기 위한 노력이 바로 인공지능 기초연구라고 볼 수 있다. 플랫폼은 인공지능 응용연구가 개발연구로 빠르게 확산시키는 유헤유 역할을 하며, 차세대 인공지능은 인공지능의 개념이 태동했을 때부터 이어온 범용 인공지능 개발을 목표로 하는 것이다. 특히 차세대 인공지능은 심층학습의 성과에 힘입어 심층학습 기반의 범용 인공지능 개발에 집중하는 경향이 대두됐다. 이러한 관점에서 차세대 인공지능 연구는 인공지능 기초연구와도 맞물려 있다.

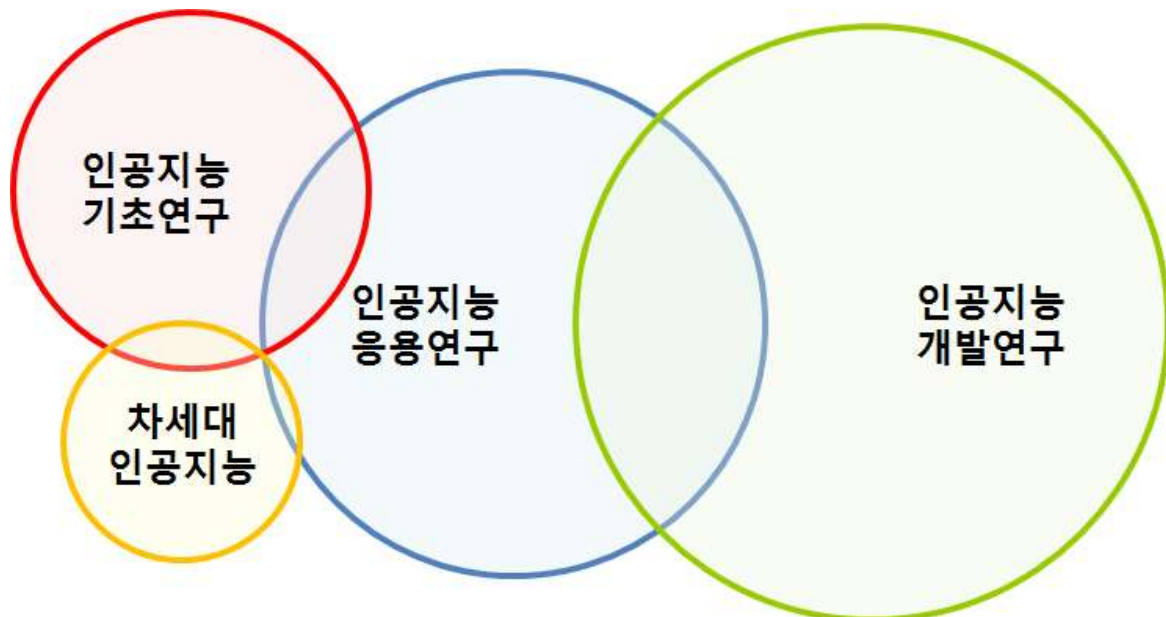
이 흐름을 도식화하면 다음 [그림 2-1]과 같다. 심층학습은 인공지능 응용연구에서 촉발되어 인공지능 개발연구에 빠르게 적용됐으며, 심층학습의 한계를 극복하기 위한 기초연구가 이루어지고 있다. 플랫폼은 인공지능 응용연구와 개발연구의 간극을 이어준다.

17) 2015년 Nature지에 게재된 Deep Learning의 저자 얀 르쿤(Yann Lecun), 요슈아 벤지오(Yoshua Bengio), 제프리 힌튼(Geoffrey Hinton) 교수와 스탠포드 대학 교수인 앤드류 응(Andrew Ng)



[그림 2-1] 인공지능 연구 영역의 상관관계

<표 2-3>을 도식화하면 다음 [그림 2-2]와 같다.



[그림 2-2] 인공지능 연구 영역 구분의 도식화

[그림 2-2]의 해석은 다음과 같다. 먼저 인공지능 연구 영역의 도식화는 가설 단계임을 고려해야 한다. 먼저 인공지능 응용연구와 개발연구는 그 경계가 모호할 수 있으나, 인공지능 개발연구가 대부분 인공지능 응용연구 결과를 차용한다는 가정으로 구분 지을 수 있다. 물론 개발연구의 결과가 응용연구에 직·간접적으로 영향을 미칠 수도 있다. 인공지능 기초연구는 인공지능 응용연구의 성과를 향상시킬 수 있다는 점에서 서로 교집합이 존재한다. 또한 반대로 응용연구에서 직면한 난제를 해결하기 위해 기초연구의 주제를 제시할 수 있다. 차세대 인공지능은 기초연구와 응용연구 모두와 관련이 있으나, 현 시점에서 도전적인 과제를 수행하기 때문에 상당부분 독립된 연구라고 볼 수 있다. 플랫폼은 인공지능 응용연구와 개발연구의 유기적인 전파를 위한 개념으로 본다면 보조적인 기능을 수행한다.

[그림 2-2]에서 연구영역을 원으로 표현했는데, 원의 크기는 현재 인공지능과 관련된 연구의 비중을 말한다. 이 도식화는 아직 가설을 세우는 단계로 원의 크기는 지속적인 연구를 통해 조정될 수 있다. 특히 어떠한 인공지능 연구가 어느 원에 속하는지 혹은 교집합에 포함 될 것인지를 통해 원의 배치가 변경되거나 새로운 영역(원)이 추가될 수 있을 것이다.

지금까지 인공지능의 R&D와 산업의 특성을 바탕으로 인공지능 연구 영역의 구분에 대한 가설을 제시했다. 이것을 고도화하기 위해서는 수많은 인공지능 관련 연구들을 대입하고 개선해야 하는 것이 숙제로 남아있다. 비록 가설로 출발하지만 제시한 인공지능 연구 영역의 구분이 심층학습의 여러 방법론과 산업에의 적용까지 모두 아우를 수 있다는 측면에서 의미가 있다고 본다.

제3장 차세대 인공지능의 연구 현황

제1절 개 요

이제 본격적으로 차세대 인공지능의 연구 현황을 살펴보고자 한다. 차세대 인공지능은 2장에서 소개한 연구 영역의 분류상으로 차세대 인공지능에 속하지만 상당부분은 인공지능 기초연구와 맞물려 있기 때문에 인공지능 기초연구에 대한 현황 역시 다룰 예정이다. 다음 <표 3-1>는 이 보고서에서 분석할 연구 대상을 나타낸다.

<표 3-1> 인공지능 최신 연구현황 분석 주제

구 분	주 제
인공지능 기초연구	① 메타 학습 : 학습하는 방법을 학습하는 심층학습 기법 ② 신경망 구조 탐색 : 최적의 인공신경망 구조를 도출하는 방법 ③ 스파이킹 신경망 : 뉴로모픽칩의 학습 방법 ④ 전이학습 : 학습된 신경망을 전이(transfer)하여 활용하는 방법 ⑤ 지속적인 학습 : 기존에 학습한 사실을 유지하는 기법
차세대 인공지능	- 범용 인공지능 ① 기호 인공지능(Symbolic AI)의 연구 현황 ② 확률적 프로그래밍 연구 현황

<표 3-1>에 나타나 있듯이 인공지능 기초연구에서 5가지 주제와 차세대 인공지능에서 범용 인공지능 연구 현황에 대해 분석해 볼 것이다. 인공지능 기초연구를 중점적으로 다루는 이유는 이 분야가 향후 응용연구와 개발연구에 접목되어 성능을 향상시킬 수 있는 기반이 되기 때문이다.

최신 연구 동향의 분석 범위는 다음과 같다. 먼저 기술이 등장하게 된 배경과 기존 기술과의 차별점, 성능과 적용 가능 분야, 마지막으로 향후 전망을 다룰 것이다.

제2절 인공지능 기초 연구 현황¹⁸⁾

1. 메타 학습

메타 학습(Meta learning)은 교육학에서 나온 메타 인지(Meta cognition)의 개념으로부터 출발한다. 메타 인지는 ‘내가 아는 것과 모르는 것을 즉각적으로 인지하는 것’을 의미한다. 다시 말하자면, 특정 사람이 특정 행위를 하는데 있어 이해하고 있는 부분과 그렇지 못한 부분을 얼마나 인지(cognition)하고 있는가를 판단하는 것이다. 따라서 메타 인지가 높은 집단은 일반적인 집단에 비해 동일한 낯선 환경에서 새로운 문제를 더 잘 해결한다는 실험적 결과가 존재한다.

메타 학습은 메타 인지로부터 파생된 개념이다. 메타 학습은 기존에 학습했던 정보를 바탕으로 새로운 문제를 해결하는 접근을 통칭하는 것으로 볼 수 있다. 더 범위를 좁혀 기계학습의 범주에서 메타 학습을 해석하자면, 이것은 기존에 학습한 시스템으로 새로운 임무(task)를 해결하기 위한 방법론을 연구하는 분야다. 여기서 새로운 임무의 범주는 아직 기초적인 수준에 머물러 있어, 이미지를 인식하는 시스템의 경우 새로운 임무 역시 이미지 인식을 의미한다. 아직 이미지 인식 시스템이 음성 인식까지 하기란 요원한 일이다.

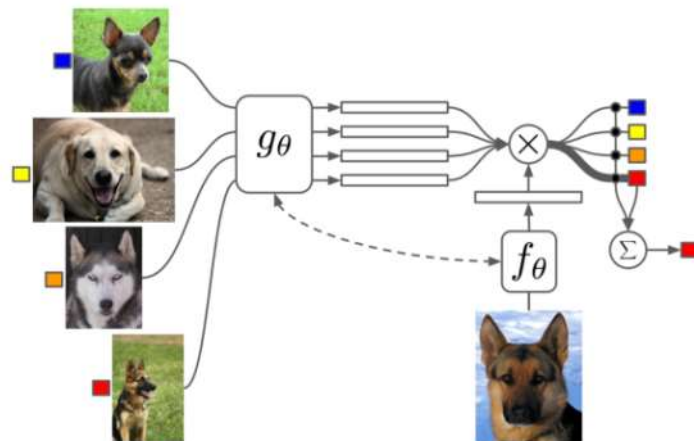
더 구체적으로 메타 학습의 배경을 살펴보자. 메타 학습은 일반적으로 학습하는 방법을 학습(Learning to learn)하는 분야를 포괄한다. 현재 연구계에서는 이미 학습된 시스템에 어떻게 새로운 것을 학습시킬 수 있는가를 고민하고 있다. 예를 들면, 100가지 객체를 인식하기 위한 심층학습 모델이 개발되었다고 가정했을 때, 100가지 객체 이외의 새로운 객체를 어떻게 “효율적”으로 학습시킬 수 있는지에 대한 것이다. 여기서 효율적이라는 점은 데이터와 학습 시간의 측면을 의미한다. 따라서 짧은 학습 시간에 적은 데이터를 활용해 효율적으로 학습하는 기법이 메타 학습 분야에서 현재 진행 중인 연구 과제다.

18) 인공지능, 어디까지 왔나?, 소프트웨어정책연구소(2018.11.)에서 재정리

메타 학습의 측정 방법 (Way, Shot)

- 메타 학습의 성능을 측정하는 기준으로 Way와 Shot이라는 단어를 활용하는데, 개념적으로 Way는 분류(Class)의 개수, Shot은 데이터의 수로 볼 수 있음
- 5 Way - 1 Shot의 경우 5개의 분류와 각 분류에 한 개의 데이터가 있는 것으로 이 데이터를 성공적으로 학습하느냐의 여부로 메타 러닝의 성능을 측정
- Shot은 낮을수록(데이터가 적음), Way는 높을수록(분류가 많음) 문제가 어려워짐

메타 학습 중 가장 대중적으로 알려진 원샷 학습(One-shot Learning)은 앞서 Shot의 개념을 살펴봤다시피 한 개의 데이터로 학습하는 방법이다. 다음 [그림 3-1]은 원샷 학습의 대략적인 구조를 표현한다.



[그림 3-1] 원샷 학습의 구조도

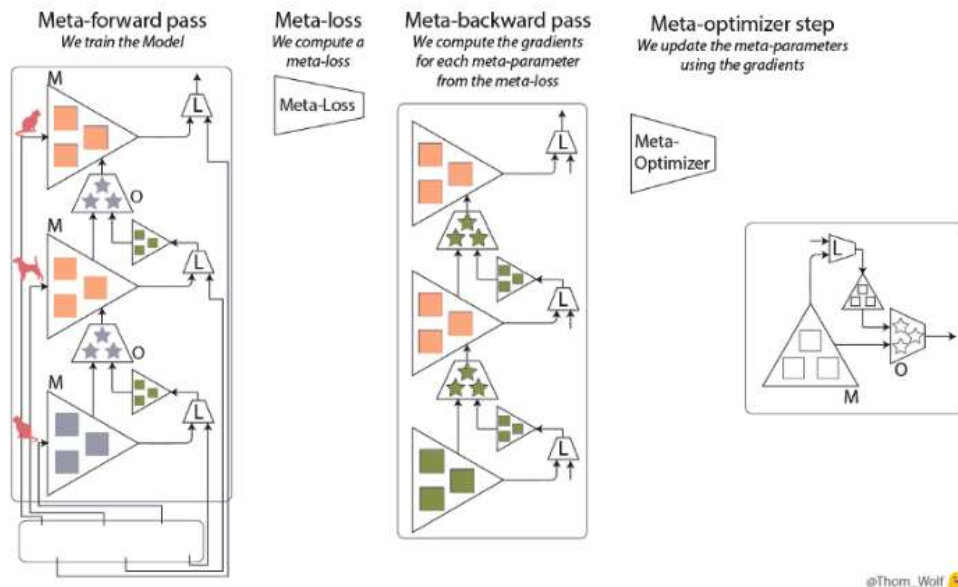
자료 : Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra, "Matching Networks for One Shot Learning," <https://arxiv.org/abs/1606.04080>

[그림 3-1]을 토대로 원샷 학습을 살펴보자. [그림 3-1]에서 원샷 학습의 목표는 네 장의 서로 다른 강아지 이미지를 분류하는 것이다. 그림 상으로는 4 Way - 1 Shot의 문제다. 이 목표를 달성하기 위해 원샷 러닝은 두 가지 함수를 활용한다. 먼저 g_{θ} 는 이미지에서 특성(feature)을 추출하는 역할을 수행하며, f_{θ} 는 추론(inference)하는데 활용된다. 일반적으로 g_{θ} 는 이미지 분류의 기능을 수행하기 위해 이미 학습된 인공신경망(pre-trained network)을 활용하여 특성19)을 추출한다. 이 특성을 재분류하는 것이 f_{θ} 의 역할이다. 이미 학습된

신경망이 Way(분류)별로 주어진 Shot(이미지)이 서로 구분될 수 있는 특성을 추출하고, 이것을 군집(clustering)하는 것이 f_θ 라는 것이다.

원샷 학습의 성능은 제한적이지만 이미지 인식 분야에서는 좋은 결과를 나타내고 있다. 그러나 한계점 역시 분명하다. 앞서 소개한 g_θ 는 이미 학습된 신경망이라는 점이다. 또한 g_θ 의 학습(training)에 활용된 데이터와 shot은 어느 정도 유사성이 보장되어야 한다는 것 역시 또 하나의 단점이라고 볼 수 있다.

또 다른 메타 학습으로는 메타 최적화 학습(Meta optimizer learning)이 있다. 기본적으로 인공신경망을 학습한다는 개념은 오차를 역전파(error back-propagation)하여 인공신경망을 구성하는 가중치를 최적화하는 것이다. 여기서 오차는 보통 손실함수(loss function)으로 정의한다. 손실함수의 가장 큰 예는 특정 학습 단계에 있는 인공신경망의 출력값과 실제값(label)의 차이의 합으로 본다. 메타 최적화 학습은 이 손실함수를 인공신경망으로 구현하는 접근이다. 이것은 메타 손실(meta-loss)라고 정의되며, 이 메타 손실 자체가 낮아지는 방향으로 학습한다면 적은 데이터로도 의미 있는 학습을 할 수 있다는 개념으로 [그림 3-2]와 같이 표현할 수 있다.



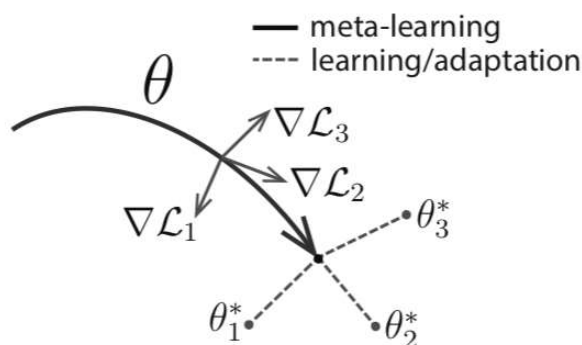
[그림 3-2] 메타 최적화 학습의 구조도

자료 : Thomas Wolf, From zero to research - An introduction to Meta-learning

<https://medium.com/huggingface/from-zero-to-research-an-introduction-to-meta-learning-8e16e677f78a>

19) 여기서 특징은 일반적으로 분류층(classifier) 이전에 있는 full-connected 층의 값을 활용함

다음 메타 학습 방법은 모델에 자유로운 메타 학습(Model-Agnostic Meta Learning, 이하 MAML)이다. MAML 방법은 학습 단계에서 다양한 임무(multi-Task)에 대해 정의하고 각 임무별로 손실(loss)도 설정한다. MAML이 학습하는 방법은 임무별 오차를 전역 손실(global loss)화 하여 이것을 최적화 하는 방향을 학습한다는 것이다. 여기서 최적화 하는 대상은 임무와 해당 임무를 구성하는 인공신경망의 가중치이기 때문에, 모델에 자유롭다는(Model-Agnostic) 개념이 쓰인 것이다. [그림 3-3]은 MAML의 가중치 갱신의 개념도다. 여기서 L 은 오차(Loss) θ 는 신경망의 가중치를 나타내며, 학습이 진행되는 방향은 화살표 방향이다.



[그림 3-3] MAML의 개념도

자료 : Chelsea Finn, Pieter Abbeel, and Sergey Levine, “Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks,” <https://arxiv.org/abs/1703.03400>

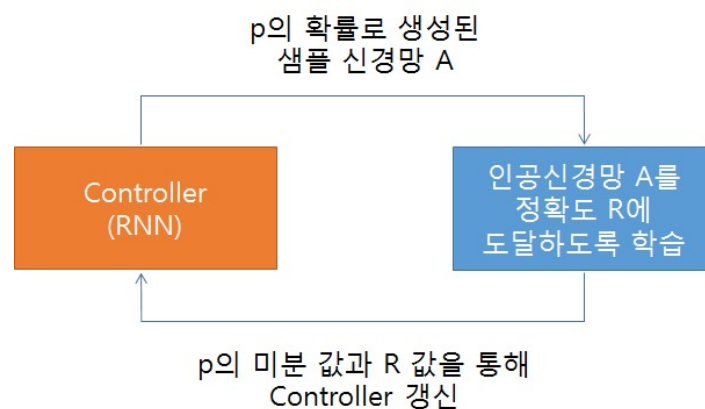
메타 학습은 학습하는 방법을 학습하는 방법론으로 부상하고 있으나, 현재 진행되고 있는 연구 성과로 봤을 때는 여전히 가야할 길이 멀다. 현재 메타 학습의 주요 과제는 학습의 “효율화”다. 심층학습 기술은 그 성능이 매우 뛰어나지만 그 성능을 달성하기 위해 수많은 데이터가 사용됐다는 것은 누구나 인지하고 있는 사실이다. 이에 반해 사람은 적은 양의 데이터로도 일반화시키는 능력이 매우 뛰어나기 때문에, 메타 학습은 이러한 사람의 학습방법을 모방하는 부분에 초점을 두고 있는 것이다. 사실 적은 데이터로 학습할 수 있는 기술이 개발된다면, 심층학습은 새로운 국면으로 진입할 것이다. 그만큼 데이터의 수급이 원활하지 않은 분야가 많으며, 기술 자체에 대한 수요가 높다는 것을 반증한다.

2. 신경망 구조 탐색

신경망 구조 탐색(Neural Architecture Search, NAS)은 기본적으로 메타 학습의 개념과 비슷하다. 바로 학습하는 방법을 학습하는 것인데, 신경망 구조 탐색이 다른 점은 인공신경망의 구조에 중점을 둔다는 것이다. 최적의 인공신경망 구조는 경험적인 결과를 바탕으로 한다. 구글 AlphaGo의 정책망은 13층의 합성곱신경망을, 페이스북의 딥페이스는 9층의 합성곱신경망을 활용했다. 이처럼 은닉층의 개수와 성능의 상관관계가 불분명한 이유로, 많은 연구자들은 경험적으로 많이 학습해보고 최적의 결과를 도출하는 접근이 일반적이다.

신경망 구조 탐색은 은닉층의 수, 합성곱의 크기와 모수 등 연구자가 변경해가며 학습해보는 방법을 자동화하는 것에 중점을 둔다. 물론 신경망 구조 탐색도 완벽하게 모든 경우의 수를 학습하는 접근은 아니기 때문에, 여전히 귀납적인 접근방법이긴 하나 용이하게 활용할 수 있다는 부분이 장점이다.

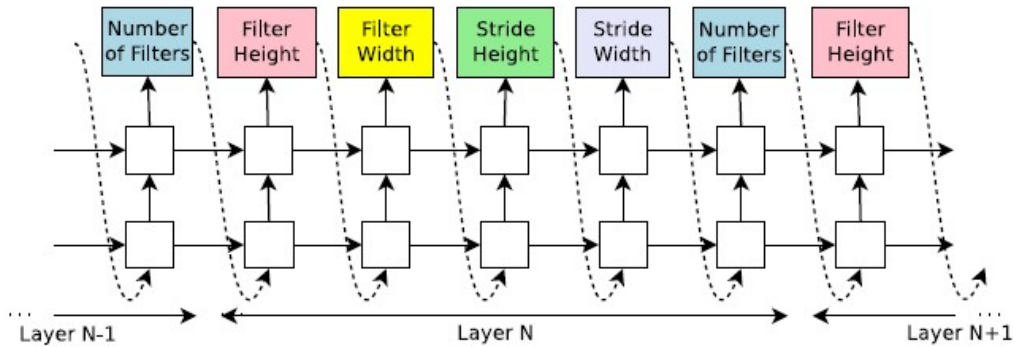
신경망 구조 탐색은 [그림 3-4]와 같이 순환신경망(Recurrent Neural Network)을 통해 파생되는 신경망을 학습하여 최적의 인공신경망 구조를 탐색하는 기술이다. 신경망 구조 탐색에서는 인공신경망을 일종의 다변 문자열(variable-length string)로 구체화한다. 문자열로 표현된 인공신경망(child network) A는 순환신경망의 구조를 갖는 Controller로부터 p의 확률로 생성된다. 인공신경망 A가 학습되고 난 뒤 검증 데이터로 정확도 R을 산출하고, R을 바탕으로 확률 p를 갱신한다. R이 높을수록 최적의 신경망 구조를 찾아간다는 개념이 신경망 구조 탐색의 원리다.



[그림 3-4] 신경망 구조 탐색의 개념

자료 : Neural architecture search with reinforcement learning에서 재구성

구체적으로 구조적인 모수(architectural hyperparameter)를 도출하기 위한 Controller는 다음 [그림 3-5]와 같이 순환신경망 구조로 표현할 수 있다. [그림 3-5]의 상단에 표현된 필터의 수(Number of Filters), 필터의 높이나 폭 등은 합성곱신경망에서 활용되는 구조적인 모수이다.



[그림 3-5] 구조적 모수를 도출하기 위한 Controller

자료 : Zoph, Barret, and Quoc V. Le. "Neural architecture search with reinforcement learning." arXiv preprint arXiv:1611.01578 (2016).

신경망 구조 탐색은 구글이 먼저 제안한 방법론으로 2016년에 처음 소개될 당시 수 백 장의 고성능 GPU를 활용한 결과였다. 말 그대로 귀납적인 방법을 풍부한 계산 자원으로 구현한 것으로 볼 수 있다. 그 결과 모델링을 근거로 한 방법론에 버금가는 성능을 달성했다. 구글은 2018년 효율적인 신경망 구조 탐색(Efficient NAS) 기술을 소개하여 고성능 GPU 한 장만으로도 좋은 성능을 달성함을 <표 3-2>와 같이 증명했다.

<표 3-2> 신경망 구조 탐색의 성능

CIFAR-10			Penn Treebank		
방법론	파라미터 수 (백 만)	오차율 (%)	방법론	파라미터 수 (백 만)	혼잡도
경험적인 방법 (DenseNet)	26.2	2.56	경험적인 방법 (LSTM)	22	56.0
NAS	3.3	2.65	NAS	54	62.4
ENAS	4.6	2.89	ENAS	24	55.8

자료 : Pham, Hieu, et al. "Efficient Neural Architecture Search via Parameter Sharing." arXiv preprint arXiv:1802.03268 (2018).

<표 3-2>에서 시험한 데이터 셋인 CIFAR(Canadian Institute For Advanced Research)-10은 6만 장의 사진을 10개의 종류로 구분한 이미지 데이터로, 이미지 분류의 성능을 측정하는 기준데이터로 활용된다. 파라미터 수는 인공신경망의 가중치 개수를 의미하며 파라미터 수가 작을수록 더 간소한 인공신경망을 의미한다. 신경망 구조 탐색은 이미지 분류 문제에서 기존의 경험적인 방법보다 더 적은 수의 파라미터로도 높은 성능을 달성했음을 확인할 수 있다. 반면 Penn Treebank 데이터셋은 언어 모델의 성능을 측정하는 데 활용된다. 언어 모델은 주어진 문장의 다음에 올 단어를 예측하는 것으로, 측정 기준은 적절함을 판단하는 혼잡도(Perplexity)를 활용하며 낮을수록 성능이 좋다. 마찬가지로 신경망 구조 탐색이 기존의 경험적인 방법에 버금갈 정도임을 증명했다.

구글이 신경망 구조 탐색을 개발한 이유는 최근 출시한 서비스인 AutoML에서 찾을 수 있다. 바로 기업이 심층학습을 적용하기 위한 장벽을 현저히 낮추는 서비스인 것이다. 과거 수 백 장의 GPU가 필요했던 신경망 구조 탐색을 한 장의 GPU만으로도 성공적인 학습이 가능했다는 점 역시 구글의 인프라 운영에도 효율을 높였다. 신경망 구조 탐색은 개인이나 기업이 심층학습을 적용하기 위한 간극을 낮췄다는 점에서 충분히 의미가 있다. 또한 구글이 신경망 구조 탐색을 비롯한 인공지능 기초연구에 몰두하고 있는 이유도 이러한 서비스의 성능을 향상시킬 수 있다고도 해석할 수 있다.

3. 스파이킹 신경망

스파이킹 신경망(Spiking Neural Network)은 인공신경망과 마찬가지로 신경계의 생물학적 동작 방식을 모사한 방법이다. 그러나 심층학습의 기반이 되는 인공신경망과 스파이킹 신경망은 상당부분 차이가 있다. 첫 번째 차이는 학습하는 방법이다. 인공신경망은 오차 역전파법(error back-propagation)을 활용해 신경망 전체의 가중치를 변경하여 학습한다. 반면 스파이킹 신경망은 세포 수준의 지역적인 학습을 수행한다. 두 번째 차이는 앞서 심층학습의 한계로 언급했던 지속적인 학습에 관한 것이다. 스파이킹 신경망은 구조적으로 인공신경망 보다 우리의 신경계와 가깝기 때문에 연속적인 학습이 가능하다. 세 번째 차이는 학습 시 전력의 소비다. 스파이킹 신경망은 뇌 구조를 모사한 뉴로모픽칩을 학습시키는데 활용되는 기술이다. 인공신경망은 수백 수천와트를 사용하는 GPU로 학습하는 반면 스파이킹 신경망은 밀리와트로도 학습이 가능하다.²⁰⁾

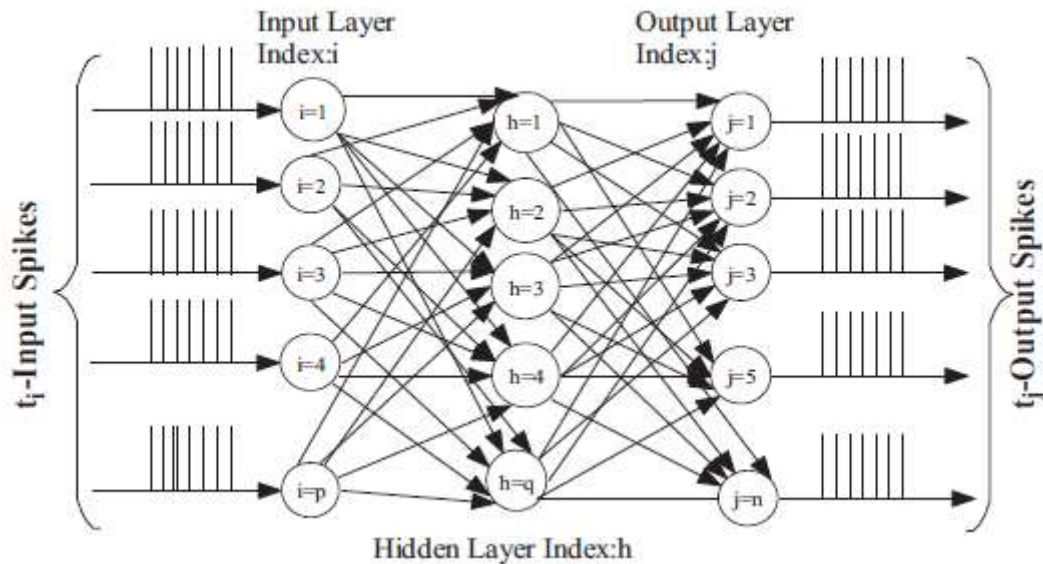
스파이킹 신경망에서 스파이크(Spike)의 의미는 생물학적인 뉴런이 정보를 교환할 때 발생하는 전기적 파동을 말한다. 스파이킹 신경망은 세 가지 생물학적 가정을 토대로 제안된 방법이다. 먼저 스파이킹 신경망을 구성하는 뉴런은 많은 입력을 받고 하나의 스파이크 신호를 출력한다. 두 번째 가정으로 스파이크가 생성될 확률은 입력이 클수록 증가하고 약할수록 감소한다. 마지막으로 오직 하나의 기준점으로 스파이크 생성 여부를 결정한다.

스파이크를 표현하는 수학적 모델은 디랙-델타 함수²¹⁾다. 시간에 따라 발생하는 간헐적인 스파이크는 각 스파이크에 해당하는 디랙-델타 함수의 합으로 표현되고, 이 합을 스파이크 트레인(Spike train)이라고 표현한다. 이 스파이크 트레인이 뉴런의 입력과 출력이 된다. 스파이킹 신경망의 학습 과정은 특정 두 인공신경세포를 연결하는 가중치를 조정하는 것이다. 이것은 생리학적 학습의 규칙에 따라 시냅스의 연결강도의 변화를 기술한 Hebb의 이론에 기반한다. 쉽게 표현하자면 두 개의 신경세포가 서로 반복적이고 지속적으로 점화(firing)한다면 상호간의 가중치가 강화된다는 것을 의미한다. 가중치의 변화량은 두 인공신경세포의 스파이크 트레인과 스파이크 트레인의

20) ESSER, S. K., et al. Convolutional networks for fast, energy-efficient neuromorphic computing. 2016. Preprint on ArXiv.

21) 한 점에서 값을 갖고 나머지에서 0의 값을 갖는 형태의 함수

저역통과필터(Low-pass filter)로 구성된다. 이 관계식을 통해 가중치가 변화하고 신경망이 학습된다. 스파이킹 신경망의 구조는 다음 [그림 3-6]과 같다.

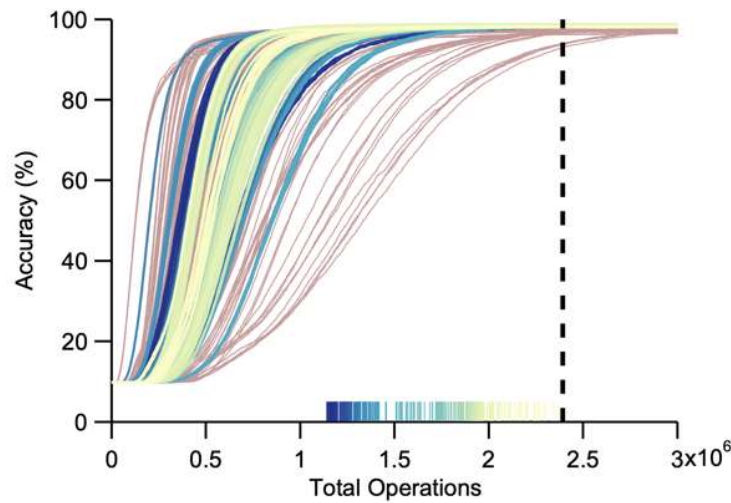


[그림 3-6] 스파이킹 신경망의 구조

자료 : Kulkarni, Santosh, Sishaj P. Simon, and Kinattingal Sundareswaran. "A spiking neural network (SNN) forecast engine for short-term electrical load forecasting." Applied Soft Computing 13.8 (2013): 3628-3635.

이제 스파이킹 신경망의 성능에 대해 살펴보자. 일반적으로 어떠한 시스템을 학습함에 있어 모델 미지수만큼 학습 데이터의 수가 보장되어야 한다. 쉽게 말하자면 연립방정식에서 미지수와 방정식의 관계라고 볼 수 있다. 그러나 그 모델이 인공신경망의 영역에 들어서면 관점이 달라진다. 사람의 뇌는 10^{14} 개만큼의 시냅스(미지수)가 존재하나, 실제로 사용하는 양은 10^9 개 정도이다. 즉, 모든 뉴런이 하나의 정보처리를 위해 사용되는 형태가 아닌 것이다. 이러한 관점에서 스파이킹 신경망은 스파이크 트레인 형태로 정보를 변형하여 학습하기 때문에 연산의 효율 측면에서 강점이 있다. 이것은 저전력의 특성과도 이어지는 부분이다. 아래 [그림 3-7]은 MNIST²²⁾ 데이터 셋에 대해 일반적인 인공신경망과 스파이킹 신경망의 성능을 비교한 그래프다.

22) Modified National Institute of Standards and Technology database로 손글씨 숫자 10개(0~9)에 대한 6만 장의 학습이 이미지와 1만 장의 시험이미지로 구성되어, 숫자 인식의 성능 측정을 하는데 기준점이 되는 데이터로 활용됨



[그림 3-7] 인공신경망과 스파이킹 신경망의 성능 비교

자료 : Tavanaei, Amirhossein, et al. "Deep Learning in Spiking Neural Networks." arXiv preprint arXiv:1804.08150 (2018).

[그림 3-7]의 x축은 연산의 수, y축은 예측의 정확도를 나타낸다. 검은색 세로 점선은 인공신경망을 활용해 98%의 정확도를 달성하기 위한 연산수이며, 다양한 색의 그래프로 나타낸 것은 스파이킹 신경망을 활용했을 때 연산수에 따른 성능향상을 나타낸 것이다. 분홍색 그래프는 98% 정확도를 달성하지 못한 경우를 의미하며, 나머지 녹색, 노란색, 파란색은 98%를 달성한 스파이킹 신경망을 의미한다. 종합적으로 볼 때 인공신경망의 연산수 대비해 스파이킹 신경망은 최대 1/4 수준의 연산수만으로도 98%의 정확도를 달성했다는 측면에서 연산의 효율이 높다.

스파이킹 신경망의 향후 전망은 크게 두 가지로 볼 수 있다. 먼저 성능 측면에서는 속제가 많다. 현재 동일한 임무에 대해서는 인공신경망의 성능이 더 좋다. 스파이킹 신경망은 기존의 인공신경망에 버금가는 성능은 어느 정도 증명했으나, 인공신경망의 성능을 초월하지는 못한다. 따라서 스파이킹 신경망의 특성을 살린 응용분야를 모색하고 성공사례를 만들어야 할 것이다. 두 번째 전망으로 저전력의 특성을 살리는 접근이 부각될 것이다. 현재 인공신경망을 학습하고 추론하기 위해서는 고성능 계산 자원이 필요하다. 스파이킹 신경망은 저전력으로 구현할 수 있다는 특성으로 학습에 필요한 자원을 획기적으로 낮출 수 있다. 최근 회자되고 있는 엣지(Edge) 컴퓨팅에서 그 영향력을 발휘할 수 있을 것이라 전망된다.

4. 전이 학습

전이 학습(Transfer Learning)은 이미 학습된 시스템의 지식을 전이 (transfer)시켜 새로운 임무를 해결하기 위한 학습 방법을 의미한다. 이것은 도메인 적응(domain adaptation), 일반적인 게임 플레이(general game playing), 다중 임무 학습(multi-task learning) 등 조금 범용적인 인공지능을 연구하는 기계학습 방법론이다. 그러나 2016년부터 심층학습에서는 전이 학습이라는 단어를 차용했었는데, 여기서 다루고자 하는 전이 학습과는 다른 것을 먼저 밝힌다.

심층학습의 전이 학습은 이미 학습된 인공신경망(pre-trained network)을 재사용하는 접근이다. 2017년을 끝으로 막을 내린 이미지 인식 경진대회 ILSVRC (ImageNet Large Scale Visual Recognition Challenge)에서는 이미지 분류를 성공적으로 수행한 인공신경망 구조가 공개됐다. 2012년에 우승한 AlexNet, 2015년에 우승한 ResNet이 대표적인 예다. 만약 한 개발자가 새로운 이미지를 분류하는 것을 심층학습으로 구현하고자 한다면, 이미 학습된 AlexNet을 활용하는 것이 심층학습에서의 전이 학습이다. 혹은 이미 학습된 신경망을 조금 변형한다는 측면에서 미세 조정(fine-tuning)이라고도 불린다. 이러한 접근법은 많은 분야에서 실험적인 성과를 거뒀다. AlexNet을 처음부터 학습하려면 길게는 일단위의 시간이 소요되는 반면, 전이 학습을 활용할 경우 몇 시간 안에 학습이 끝나기 때문이다.

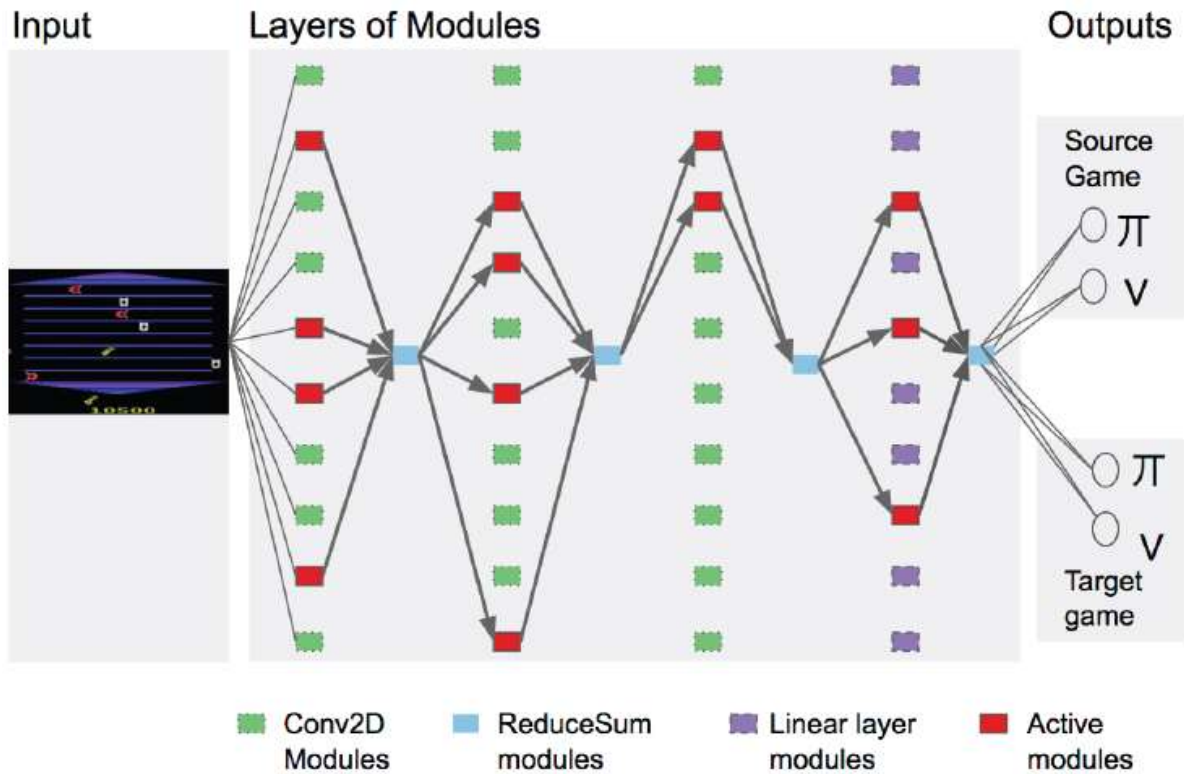
심층학습에서의 전이 학습과 미세 조정

- 심층학습에서의 전이 학습(transfer learning)과 미세 조정(fine-tuning)은 모두 이미 학습된 인공신경망을 활용하는 점에서 공통점이 있음
- 이 둘의 차이점은 ImageNet의 AlexNet이나 ResNet을 활용하는 경우 다음과 같이 구별될 수 있음
- 전이 학습의 경우 이미 학습된 신경망의 가중치 전체를 재학습시키는 접근으로 새로운 데이터가 이미지 분류의 임무이나 ImageNet 데이터와의 성격이 상이할 경우 활용됨
- 미세 조정은 ImageNet과 유사한 데이터를 재학습하는 경우에 주로 활용되며, 구조적으로 마지막 fully-connected 층과 classifier 층을 연결하는 가중치만 재학습하는 방법으로 비용이 저렴함

심층학습에서의 전이 학습은 큰 의미에서 본질적인 전이 학습과 일맥상통하는 부분이 있으나 매우 제한적인 접근이다. 심층학습의 전이 학습은 이미 학습한 인공신경망을 바탕으로 새로운 데이터의 학습을 통해 새로운 임무를 해결한다는 측면에서 일부 공통점이 있다고 본다. 그러나 본질적인 전이 학습에서는 이미지를 분류하는 인공신경망의 지식이 전이되어 음성 인식에 적용된다는 것을 의미한다. 현재 전이 학습이라는 용어가 인공지능 연구계에서 혼재되어 사용되는 측면이 있기 때문에, 용어를 활용하는 데 있어 주의가 필요하다.

전이 학습은 개념 자체가 범용 인공지능에 맞물려 있기 때문에 현 시점에서는 매우 도전적인 과제다. 전이 학습과 관련된 대표적인 연구는 구글 딥마인드의 PathNet이 있다. PathNet은 인공신경망이 학습을 진행하면서 신경망의 어떤 부분이 새로운 임무(task)에도 재사용될 수 있을지를 탐색하는 기법이다. 신경망의 재사용(path로 표현) 가능성을 측정하기 위해 유전 알고리즘(genetic algorithm)을 활용했으며, 이것의 적합도를 판단하기 위한 비용함수를 제안했다. 특히 실험적인 결과를 통해 신경망의 재사용 가능성이 높은 인공신경망으로 새 임무를 학습한 효율이 미세 조정(fine-tuning) 방법보다 우월함을 증명했다.

[그림 3-8]는 PathNet의 구조를 나타낸다. 이 그림에서는 먼저 A라는 게임을 학습하기 위해 4개의 은닉층과 각 층 당 10개의 노드를 활용했다. 여기서 재사용 가능성이 높은 노드가 붉은색으로 표현되며, 이것을 활용해 B라는 게임을 학습한다면 효율이 올라간다는 의미다.



[그림 3-8] 전이 학습의 예 : PathNet의 구조

자료 : Fernando, Chrisantha, et al. "Pathnet: Evolution channels gradient descent in super neural networks." arXiv preprint arXiv:1701.08734 (2017).

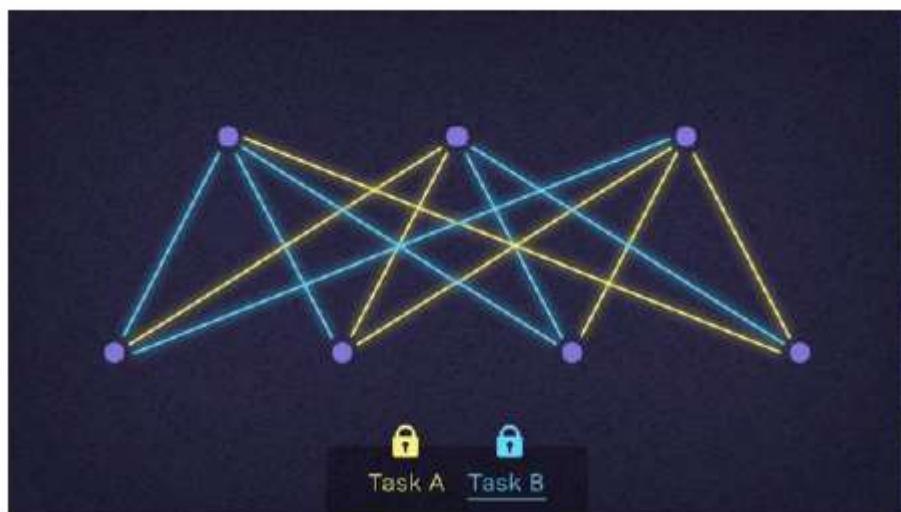
전이 학습은 앞서 소개한 스파이킹 신경망과도 관련이 있다. 스파이킹 신경망이 인공신경망의 전체를 활용한다기 보다 일부를 활용한다는 접근은 PathNet의 방법과 유사하다. 스파이킹 신경망이 전통적인 인공신경망 기법보다 뇌에 더 가까운 구조를 모사한다고 본다면, PathNet 역시 틀은 인공신경망이나 재사용의 관점에서 뇌의 구조적인 기능을 모사했다고 볼 수 있다. 또한 PathNet은 가장 먼저 소개한 메타 학습과도 맞물려 있다. 그 성능 역시 마찬가지다. 이 두 가지 연구의 궁극적인 목표는 범용 인공지능에 있기 때문이다.

현재 전이 학습의 연구는 여전히 같은 도메인 안에서의 전이(transfer)에 국한돼있다. 심지어는 데이터의 속성이 비슷해야 성공적인 전이가 이루어진다는 측면도 존재한다. 이러한 이유로 진정한 의미에서 도메인 간의 전이(transfer)가 이루어지기는 아직 요원한 일이라고 판단된다.

5. 지속 학습

지속 학습(Continual Learning)은 인간이 학습하는 방법인 순차적 학습(Sequential Learning)에서 작동 원리를 이해할 수 있다. 예를 들어 인간이 음식을 먹는 행위는 음식을 찾고, 음식이 있는 위치로 이동하며, 음식을 손으로 집어서, 입을 여는 행위로 세분화가 가능하다. 이 행위들은 독립적으로 학습된 것이 아니라 목표를 달성하기 위한 순차적인 세부 임무를 수행한다는 것이다.

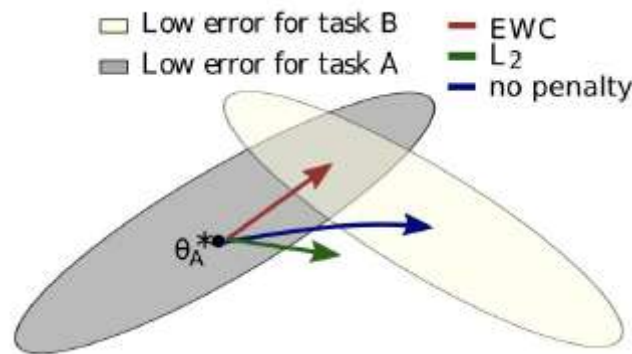
이제 심층학습으로 돌아와 보자. 심층학습은 기본적으로 독립적인 임무에 비약적인 성능 향상을 달성했다. 이 부분이 시사하는 바는 모델 자체가 일반화됐기 보다는 각 임무에 최적화됐다고 볼 수 있다. 특히 심층학습의 고질적인 문제인 치명적 망각(Catastrophic Forgetting)은 하나의 모델에서 다양한 임무를 수행하는데 한계를 보인다는 것이다. 다시 말하자면, 특정 임무 A에 특화되어 학습된 신경망을 임무 B로 학습시킬 경우 기존의 임무 A를 수행하기 위해 학습한 내용을 잊어버린다는 것이다. 이것을 보완하고자 나온 개념이 바로 지속 학습인 것이다. 지속 학습은 기존에 학습한 정보를 망각하지 않고 안정적으로 다른 임무를 학습하는 것이다. 그 개념은 다음 [그림 3-9]와 같이 표현할 수 있다. [그림 3-9]는 임무 A와 B가 서로 다른 가중치를 활용함으로써 하나의 모델로 두 가지 임무를 수행할 수 있다는 것이다.



[그림 3-9] 지속 학습의 기본 개념도

자료 : Google Deepmind Blog, “Enabling Continual Learning in Neural Networks”,
<https://deepmind.com/blog/enabling-continual-learning-in-neural-networks/>

지속 학습은 앞서 소개한 전이 학습과 마찬가지로 구글 딥마인드가 화두로 제시한 연구주제다. 특히 인공지능망에서 치명적 망각을 해결하는 방법론에 대해 논문을 발표했는데, 그 근거를 인간의 사고방식을 바탕으로 제시했다. 그 개념은 유연한 가중치 강화(Elastic Weight Consolidation, 이하 EWC)라는 손실함수를 활용하는 것이다. EWC는 새롭게 학습하는 정보가 기존에 학습한 정보에 영향을 미치는지에 대한 여부를 판단하여 가중치가 갱신되지 못하도록 고정하는 역할을 한다. [그림 3-10]은 EWC의 개념을 나타낸다.

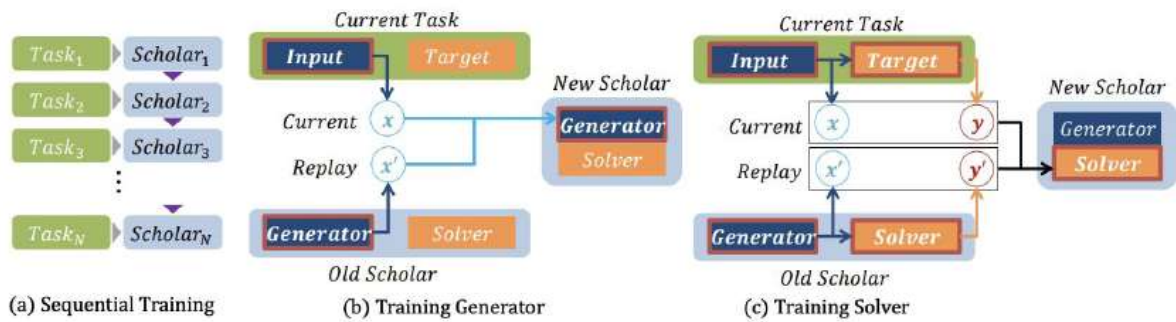


[그림 3-10] 유연한 가중치 강화(EWC) 손실함수의 의미

자료 : Vincenzo Lomonaco, “Why Continuous learning is the key towards machine intelligence”

[그림 3-10]에서 θ 는 인공지능망의 가중치를 의미하며 화살표는 각각의 손실함수를 통해 학습할 경우에 갱신되는 가중치의 위치를 의미한다. 회색 타원은 임무 A에서 낮은 오차를 갖는 가중치의 영역이며, 반투명한 노란색 타원은 임무 B에서 낮은 오차를 갖는 가중치다. 하나의 인공지능망이 임무 A와 B를 모두 잘 수행하는 영역, 즉 회색 타원과 노란색 타원이 겹치는 영역으로 가중치가 갱신된다면 두 가지 임무를 모두 성공적으로 학습할 수 있을 것이다. 이 방향이 바로 EWC다. 반면, 일반적인 손실함수의 경우는 임무 A를 학습한 정보를 잊어버리는 치명적 망각이 발생한다는 점에서 EWC가 의미 있는 것이다.

또 다른 지속학습의 연구 사례는 생성 모델을 활용했다. 이것은 인간의 뇌가 기존의 지식을 뇌에서 재생성 한다는 개념을 활용한 접근이다. 이것은 새로운 임무를 학습하기 위해 기존의 임무도 함께 학습한다는 개념인데, 기존의 임무는 대표적인 생성모델인 적대적 생성 신경망으로 만들어진 가상의 자료를 활용한다는 점이다. 그 개념은 다음 [그림 3-11]과 같다.



[그림 3-11] 심층 생성 모델을 활용한 지속 학습의 구조도

자료 : Hanul Shin, et. al., "Continual Learning with Deep Generative Replay,"
<https://arxiv.org/abs/1705.08690>

현재 단계에서의 지속 학습의 목표는 치명적인 망각을 해결하는 것이기 때문에 앞으로도 많은 연구가 지속될 것이다. 지속 학습은 궁극적인 메타 학습이나 전이 학습으로 가기 위해 반드시 필요한 분야다. 지속 학습 역시 서로 다른 임무의 범위가 응용 분야를 넘나드는 수준은 아니며 MNIST와 같이 실험실 수준의 데이터를 가지고 실험하는 것이기 때문에 그 한계는 분명하다. 그러나 소위 범용 인공지능을 개발하기 위한 노력이 다각도로 이루어진다는 점에서 그 미래는 긍정적이라고 볼 수 있다.

많은 인공지능 기초 연구사례를 살펴보면 그 연구 자체가 심층학습의 본질적인 한계를 극복하기 위한 노력을 이어가고 있다. 여기에 필요한 데이터는 사실 이미 공개된 데이터가 대부분이다. 우리가 인공지능 연구를 위해 다양한 데이터를 확보해야 한다는 커다란 숙제가 있는 반면에, 인공지능 기초연구는 이미 공개된 데이터를 활용하는 경향이 크다. 이 부분은 또한 성능측면에서 다른 연구와의 차별성을 찾기도 용이하기 때문에 인공지능 기초연구를 육성하는데 큰 무리가 없을 것이라고 판단된다.

제3절 범용 인공지능 연구 현황²³⁾

1. 개 요

많은 전문가들은 심층학습의 발전 속도가 유례없이 빠르다고 강조한다. 지난 2016년 이세돌 9단과 대결했던 AlphaGo는 경기 당시 이상한 실수를 하는 등 완벽하지 않았다는 해석이 많았다. 또한 인공지능 전문가들은 심층학습의 낮은 설명 가능성으로 인해 AlphaGo가 더 높은 수준으로 진화하기에는 수년의 세월이 필요할 것이라고 예측했다. 그러나 AlphaGo 개발진인 구글 딥마인드는 불과 1년 만에 인류의 바둑지식을 초월한 AlphaGo Zero를 공개했다. AlphaGo Zero가 더욱 충격적이었던 사실은 AlphaGo Zero가 어떠한 바둑 기사의 기보도 학습에 활용하지 않았다는 점이다.

이러한 심층학습의 놀라운 성과는 반대로 다양한 우려를 낳고 있다. 테슬라 모터스의 최고경영자인 엘론 머스크나 천재 물리학자 스티븐 호킹은 미래의 인공지능이 인류를 멸망시킬 것이라고 강조했다. 또한 사회적으로는 인공지능이 직업을 대체하여 인류의 일자리를 위협할 것이라는 걱정도 팽배해 있다. 이러한 부정적인 관점에 대한 원인은 현대 인공지능의 빠른 발전 속도로 인해 정말 사람 수준(Human-level)의 인공지능이 수 십 년 내에 만들어 질 것이라는 점이다. 사람 수준의 인공지능, 즉 범용 인공지능(Artificial General Intelligence)의 개발은 정말로 가능한 것일까? 현재 범용 인공지능과 관련된 연구는 무엇이 있을까? 이 절에서는 이 물음에 대한 답을 찾기 위해 범용 인공지능의 개념과 연구현황을 살펴보고자 한다.

본격적으로 범용 인공지능을 다루기에 앞서 심층학습에 대해 조금 더 살펴보자. 먼저 심층학습의 장점은 복잡한 데이터에서 패턴을 인식하고 예측하는 성능이 우수하다는 것이다. 예를 들어, 기계 번역은 end-to-end라는 심층학습 방법을 활용한다. 한국어를 영어로 번역하는 과정에서 한국어 문장과 그 번역에 해당하는 영어 문장을 대응시켜 방대한 데이터 셋을 만들고, 이를 학습시키는 과정이 end-to-end 방법이다. 이것은 전통적인 자연언어처리 방법과는 다르게 문장의 구조나 특정 문맥에서의 단어의 의미를 전혀 파악하지 않아도 된다. end-to-end 심층학습은 과거 전통적인 자연언어처리 기술을

23) 범용 인공지능의 개념과 연구 현황, 소프트웨어정책연구소 (2018.07.)에서 인용

활용한 번역보다 월등히 성능이 뛰어나기 때문에 현재 기계 번역 분야에서 많이 활용되고 있다.

심층학습은 이러한 장점과 대비하여 그 한계도 분명하다. 앞서 서론에서 심층학습에 대해 소개했지만 다시 한 번 그 내용을 상기시켜 보면, 먼저 심층신경망²⁴⁾의 성능은 특정 분야에 국한돼있다. 앞서 소개한 기계 번역의 end-to-end 심층신경망은 이미지나 음성을 인식할 수 없다. 인간은 뇌라는 하나의 신경망으로 사물을 인식하고 미래를 추론하며 바둑을 둘 수 있다. 심층학습은 이러한 관점에서 매우 제한적인 범용성을 가지고 있다. 흥미로운 것은 기술적인 진전이 조금씩 진행되고 있다는 것이다. AlphaGo Zero는 자체대국을 통한 자가 학습 과정을 일반화하여 AlphaZero라는 방법을 제안했다. AlphaZero는 바둑 이외에 체스, 쇼기 등의 보드게임에서도 최정상 실력을 입증하여 그 가능성을 시사했다. 그러나 AlphaZero는 여전히 보드게임이라는 틀에서 벗어나지 못한 것이 현실이다.

심층학습의 또 다른 한계는 설명가능성에 있다. 심층학습의 결과물은 매우 뛰어난 성능을 보여주나, 이 결과가 왜 산출되는가에 대해 명확히 설명 불가능하다는 것이다. 이것은 본질적으로 인공신경망의 구조에서 그 이유를 찾을 수 있다. 인공신경망은 다수의 인공신경세포를 연결해서 만든 구조다. 하지만 어떻게 연결해야 최적의 성능을 보장할 수 있는가에 대한 이론적인 근거가 없다. 이 사실로부터 심층학습 방법론은 경험적인 결과에 의존한다고 추론할 수 있다. 가능한 한 많은 경우의 수를 수행하여 최적의 결과를 도출하는 심층학습은 접근 방법 자체에서의 태생적인 한계가 존재한다는 것이다.

심층학습은 만능이 아니다. 심층학습은 일반적으로 학습을 위해 방대한 데이터를 필요로 한다. 또한 경험적으로 많은 시도를 해야 하기 때문에 필연적으로 많은 계산을 요구한다. 이렇게 심층학습을 객관적으로 보자면 심층학습이 발전하여 범용 인공지능으로 진화한다는 것은 시기상조일 수 있다. 그렇다면 심층학습과 범용 인공지능의 관계는 어떻게 될까? 심층학습이 범용 인공지능으로 가는 길에 가장 핵심적인 기술일까? 아니면 요소 기술 중 하나일까?

24) 심층신경망(Deep Neural Network)은 개념적으로 깊은 인공신경망(Artificial Neural Network)이다. 또는, 신경망 층(layer)을 2~3개 적층한 구조를 얕은 신경망(Shallow Network)에 대비되는 개념으로도 이해할 수 있다. 일반적으로 심층학습의 인공신경망은 심층신경망을 의미한다.

이 절에서는 범용 인공지능의 개념과 연구 현황에 대해서 논의할 것이다. 먼저 범용 인공지능의 개념과 접근방법을 소개하고, 분류별로 활발히 수행되고 있는 연구 내용을 살펴보겠다. 마지막으로 범용 인공지능의 미래에 대해 결론지을 것이다.

2. 범용 인공지능의 개념과 접근방법

범용 인공지능은 1956년 다트머스 회의에서 인공지능의 개념이 처음 정립됐을 때부터 출발한다. 여기서 논의된 궁극적인 인공지능의 형태가 바로 범용 인공지능이기 때문이다. 그러나 범용 인공지능으로 가기 위한 기술적 한계²⁵⁾는 극명했다. 이로 인해 인공지능은 지난 60여 년 간 부침의 역사를 겪었고, 결국 인공지능은 컴퓨터 과학의 한 분야로 인간의 지능적 행동의 일부를 모사하기 위한 접근을 취했다. 최근 범용 인공지능이라는 키워드가 다시 부상한 이유도 심층학습의 눈부신 성장 때문이라고도 볼 수 있다.

하지만 범용 인공지능은 아직 개발되지 않은 기술이기 때문에, 그 정의 역시 구체적이지 않은 것이 사실이다. 많은 학자들은 범용 인공지능을 인간 수준의 지능(Human-level intelligence)로 일컫는다. 이것은 인간처럼 정보를 학습하고, 의미를 부여하며, 다양한 문제를 인지하여 해결책을 도출하는 과정을 구현함에 있다. 물론 범용 인공지능을 단순히 사람 수준의 지능이라고 바라보는 것은 어폐가 있다. 범용 인공지능이 자의식(self-consciousness)이나 감정(emotion)이 필요한가? 혹은 범용 인공지능을 위해 물리적인 몸체(physical body)가 필요한가?에 대한 대답은 분야별 전문가에 따라 상당히 다르기 때문이다.

따라서 범용 인공지능에 대한 정의는 아직도 논의가 진행 중이다. 그 정의가 모호하기 때문에, 오히려 연구자들이 범용 인공지능에 대한 실체를 만들어 나가는 중이라고 볼 수 있다. 앞서 언급했듯이 범용 인공지능을 인간 수준의 지능을 구현한 기계장치로 바라본다면, 범용 인공지능은 다양한 목표를 달성하기 위해 능동적으로 해결책을 찾는 시스템이라고 볼 수 있다. 범용 인공지능 분야에서 활발하게 활동 중인 벤 괴르첼(Ben Goertzel) 박사는 이것을 핵심 범용 인공지능 가설(The Core AGI Hypothesis)이라고 제시했다. 이 가설이 맞다면 범용 인공지능은 좁은 인공지능(narrow AI)²⁶⁾의 반대개념으로 구분될 수 있다는 관점이다.

25) 인간의 뇌가 어떠한 방식으로 동작하는가에 대한 생리학적인 규명의 부재, 의식(consciousness)과 감정(emotion)에 대한 구현 방안, 인간이 학습하고 추론하는 방식에 대한 메카니즘 등

26) 좁은 인공지능(narrow AI)은 세계적인 미래학자인 레이 커즈와일(Ray Kurzweil)의 저서 『The Singularity Is Near』에서 소개된 개념으로 특정한 상황에서 특정한 지능적 행동을 하는 시스템으로 정의된다. 좁은 인공지능은 상황이나 행동을 약간

The Core AGI Hypothesis

- 충분히 넓고(인간 수준) 강한 범용성을 갖는 인공지능의 창조나 연구. 질적인 측면에서 상당히 좁고 약한 범용성과 반대되는 개념
- 원문 : the creation and study of synthetic intelligences with **sufficiently broad** (e.g. human-level) **scope and strong generalization** capability; is at bottom qualitatively **different from** the creation and study of synthetic intelligences with **significantly narrower scope and weaker generalization** capability

자료 : Artificial General Intelligence: Concept, State of the Art, and Future Prospects, Ben Goertzel (2017)

이러한 개념상의 범용 인공지능을 고려해보면, 심층학습은 아직 그 수준에 도달하지 못했다. 그러나 심층학습으로 촉발된 인공지능의 성능 발전 속도를 보면 범용 인공지능의 출현은 그리 먼 미래가 아닐 수도 있다. 실제로 비영리단체인 Future Life Institute(FLI)는 인공지능 전문가 1,634명을 대상으로 범용 인공지능의 출현 시기를 설문조사 했다. 그 결과 45년 뒤에 50%의 확률로 인공지능이 인간을 초월할 것이라고 전망했고, 인간의 모든 직업을 대체 하는 데는 120년이 걸릴 것으로 예측했다.²⁷⁾

만약 범용 인공지능이 실제로 출현하면 인류는 인공지능의 지배아래 놓이게 될까? 세계적인 미래학자인 레이 커즈와일은 그의 저서에서 밝혔듯이 범용 인공지능이 출현하는 시점을 특이점(Singularity)으로 정의했다. 특이점 이후의 인공지능은 자가 발전(self-improvement)하여 인간을 뛰어넘는 초인공지능(superintelligence)이 되는 것이다. MIT의 물리학과 교수인 맥스 테그마크 역시 범용 인공지능의 출현은 곧 초인공지능을 의미한다고 밝혔다.²⁸⁾ 이 관점은 인간과 범용 인공지능은 단순한 주종관계가 아니라, 서로 공존하는 관계로 정립돼야 함을 내포한다. 이에 인공지능 연구자들은 2017년 아실로마 학회에서 인공지능을 이롭게 활용하기 위한 23개 원칙 <표 3-3> 을 정립했다.²⁹⁾

변화시킨다면, 일정 수준의 프로그래밍과 재구성성이 필요하다. 심층학습은 좁은 인공지능의 대표적인 예로 볼 수 있다.

27) When Will AI Exceed Human Performance? Evidence from AI Experts (2018)

28) How to get empowered not overpowered by AI, Max Tegmark, TED talk (2018)

29) 아실로마 AI원칙, Future of Life Institute (2017), <https://futureoflife.org/ai-principles-korean/>

<표 3-3> 아실로마 23대 AI 원칙

용 어	내 용
연구목표	인공지능 연구의 목표는 방향성이 없는 지능을 개발하는 것이 아니라 인간에게 유용하고 이로온 혜택을 주는 지능을 개발하는 것이다.
연구비 지원	인공지능에 대한 투자에는 컴퓨터 과학, 경제, 법, 윤리 및 사회 연구 등의 어려운 질문을 포함해 유익한 이용을 보장하기 위한 연구비 지원이 수반되어야 한다.
과학정책 연결	인공지능 연구자와 정책 입안자 간에 건설적이고 건전한 교류가 있어야 한다.
연구문화	인공지능 연구자와 개발자 간에 협력, 신뢰, 투명성의 문화가 조성되어야 한다.
경쟁 피하기	인공지능 시스템 개발팀들은 안전기준에 대비해 부실한 개발을 피하고자 적극적으로 협력해야 한다.
안전	인공지능 시스템은 작동 수명 전반에 걸쳐 안전하고 또 안전해야 하며, 적용 가능하고 실현 가능할 경우 그 안전을 검증할 수 있어야 한다.
장애 투명성	인공지능 시스템이 손상을 일으킬 경우 그 이유를 확인할 수 있어야 한다.
사법적 투명성	사법제도 결정에 있어 자율시스템이 사용된다면, 권위 있는 인권기구가 감사 할 경우 만족스러운 설명을 제공할 수 있어야 한다.
책임	고급 인공지능 시스템의 디자이너와 설계자는 인공지능의 사용, 오용 및 행동의 도덕적 영향에 관한 이해관계자이며, 이에 따라 그 영향을 형성하는 책임과 기회를 가진다.
가치관 정렬	고도로 자율적인 인공지능 시스템은 작동하는 동안 그의 목표와 행동이 인간의 가치와 일치하도록 설계되어야 한다.
인간의 가치	인공지능 시스템은 인간의 존엄성, 권리, 자유 및 문화적 다양성의 이상에 적합하도록 설계되어 운용되어야 한다.
개인정보 보호	인공지능 시스템의 데이터를 분석 및 활용능력의 전제하에, 사람들은 그 자신들이 생산한 데이터를 액세스, 관리 및 통제할 수 있는 권리를 가져야 한다.
자유와 개인정보	개인정보에 관한 인공지능의 쓰임이 사람들의 실제 또는 인지된 자유를 부당하게 축소해서는 안 된다.
공동이익	인공지능 기술은 최대한 많은 사람에게 혜택을 주고 힘을 실어주어야 한다.
공동번영	AI에 의해 이루어진 경제적 번영은 인류의 모든 혜택을 위해 널리 공유되어야 한다.
인간의 통제력	인간이 선택한 목표를 달성하기 위해 인간은 의사결정을 인공지능 시스템에 위임하는 방법 및 여부를 선택해야 한다.
비파괴	고도화된 인공지능 시스템의 통제로 주어진 능력은 건강한 사회가 지향하는 사회적 및 시정 과정을 뒤엎는 것이 아니라 그 과정을 존중하고 개선해야 한다.
인공지능 무기 경쟁	치명적인 인공지능 무기의 군비 경쟁은 피해야 한다.
인공지능 능력에 관한 주의	합의가 없으므로 향후 인공지능 능력의 상한치에 관한 굳은 전제는 피해야 한다.
중요성	고급 AI는 지구 생명의 역사에 심각한 변화를 가져올 수 있으므로, 그에 상응 한 관심과 자원을 계획하고 관리해야 한다.
위험	인공지능 시스템이 초래하는 위험, 특히 치명적인 또는 실존적 위험에는, 예상된 영향에 맞는 계획 및 완화 노력이 뒷받침되어야 한다.
재귀적 자기 개선	인공지능 시스템이 재귀적 자기 복제나 자기 개선을 통하여 빠른 수적 또는 품질 증가를 초래한다면, 설계된 시스템은 엄격한 안전 및 통제 조치를 받아야 한다.
공동의 선	초지능은 널리 공유되는 윤리적 이상을 위해, 그리고 몇몇 국가나 조직이 아닌 모든 인류의 이익을 위해 개발되어야 한다.

자료 : 아실로마 AI원칙, Future of Life Institute (2017),

<https://futureoflife.org/ai-principles-korean/>

범용 인공지능의 접근 방법은 크게 두 가지가 있다.³⁰⁾ 첫 번째 접근 방법은 기호적 범용 인공지능(Symbolic AGI)이다. 기호적 범용 인공지능의 근간인 기호적 인공지능(Symbolic AI)은 과거 전문가 시스템에서 현대의 IBM 왓슨에 이르기까지 지속적으로 활용돼왔다. 사람은 물체를 인식할 때 기호를 부여한다. 그리고 그 기호 간의 관계를 정립함으로써 논리적인 구조를 만들고 비슷한 상황을 추론할 때 활용한다. 이렇게 기호적 인공지능은 지능이 기호를 다루는 행위로 정의하는 분야다. 기호적 범용 인공지능의 장점은 접근방법 자체가 일반화에 가장 적합한 도구라는 것이다. 따라서 심층학습과는 다르게 다양한 지능적 활동을 할 수 있다. IBM 왓슨이 질문응답 시스템(Q&A 시스템)으로 출발했지만, 현재 응용 분야는 의료, 금융 등 다양하게 그 적용 영역을 넓히고 있기 때문이다. 그러나 기호는 간단한 지각이나 동기로 인해 변화하거나 진화할 수 있기 때문에, 단순히 기호와 규칙만으로 범용 지능을 구현하기에는 한계가 존재한다.

범용 인공지능의 두 번째 접근방법은 창발적 범용 인공지능(Emergentist AGI)이다. 창발적 범용 인공지능 접근방법은 인간의 뇌의 작동 방식을 다양한 방식으로 모사하여 구현하는 인공지능을 의미한다. 특히, 뇌의 구조와 기능을 역공학(reverse engineering)으로 해석하여 기계로 구현하는 방법이 일반적이다. 심층학습 역시 큰 의미에서 보자면 창발적 접근방법이다. 심층학습에서 활용되는 인공신경망은 본질적으로 뇌 신경망을 모사하여 만들었기 때문이다. 흥미로운 것은 인공신경망의 구조가 조금 더 실제 뇌를 반영하도록 진화하고 있다는 것이다. 그러나 뇌의 생리학적 규명이 아직 완벽하지 않기 때문에, 더불어 창발적 범용 인공지능의 물리적 한계가 존재할 수 있다고 볼 수 있다.

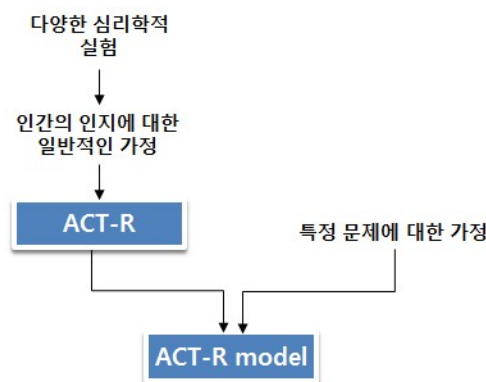
위에 소개한 두 가지 접근방법은 물과 기름처럼 정확히 나뉘는 것은 아니다. 실제로 두 가지를 혼합한(hybrid) 접근방법도 존재한다. 또한 위의 접근방법 이외에 다른 방식으로 범용 인공지능을 해석하는 관점도 존재한다. 이 보고서에서는 위에서 소개된 두 가지 접근 방법을 준용하여, 접근 방법별로 대표적인 범용 인공지능 연구 현황을 소개하겠다. 기호적 범용 인공지능으로 잘 알려진 ACT-R과 Soar를 소개하고, 심리학적 이론에 기반한 확률적 프로그래밍의 접근 방법에 대해 살펴보겠다.

30) 『Artificial General Intelligence: Concept, State of the Art, and Future Prospects』에서 소개된 내용을 정리

3. 범용 인공지능의 연구 현황

ACT-R은 대표적인 기호적 범용 인공지능 접근방법의 인지 아키텍처(Cognitive Architecture)³¹⁾로 1973년부터 시작된 유서 깊은 인공지능 연구다. ACT-R은 인간의 인지(cognition)가 어떻게 작동하는지에 대한 이론을 컴퓨터를 활용해 구현하여 다양한 임무를 해결하는데 접목됐다. ACT-R의 이론은 심리학 실험에서 파생된 수많은 사실과 인간의 인지의 과정에 대한 가정(assumption)을 통해 정립된다.

ACT-R은 일종의 SW 프레임워크로 정립된 이론을 통해 다양한 문제³²⁾를 해결하는 환경을 제공한다. 연구자들은 특정 임무를 해결하기 위해 ACT-R을 활용하여 모델(컴퓨터 프로그램)을 만들 수 있고, 여기에 연구자들은 문제에 대한 가정을 추가한다. 이 가정은 실제 사람이 동일한 임무를 수행한 것과 비교하여 검증하는 단계를 거친다. 비교의 측정 기준은 임무를 해결하는 데 소요된 시간과 정확도, 최근에는 사람의 기능적 자기 공명 영상(fMRI)도 활용한다.



[그림 3-12] ACT-R의 개괄적인 흐름

자료 : ACT-R homepage에서 재구성, <http://act-r.psy.cmu.edu/about/>

ACT-R을 활용한 연구결과는 약 700건 이상에 이르며, 대표적인 응용 분야는 인간의 기억, 자연언어처리, 인지 뇌과학, 교육 등 다양하게 활용된다. ACT-R은 개선을 거듭하여 2017년 7월 기준 7.5 버전이 공개돼 있다.³³⁾

31) 인지 아키텍처는 인간의 사고를 구조화하는 이론을 기계장치에 구현하는 분야

32) 하노이 탑 문제, 문서나 단어의 기억, 언어 이해, 의사 소통, 비행기 조종 시뮬레이션 등

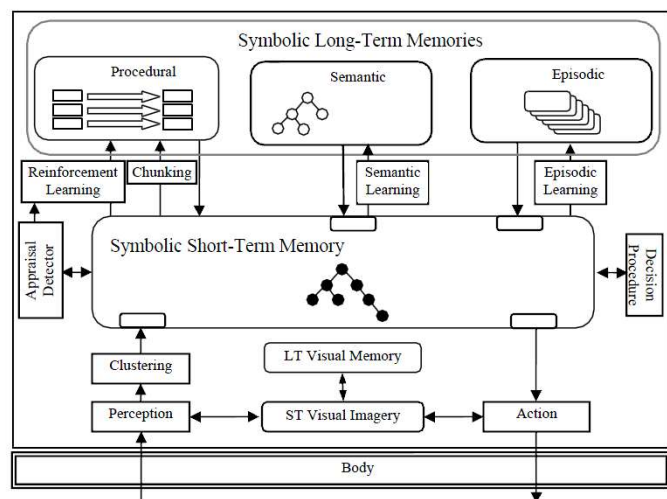
33) ACT-R software, <http://act-r.psy.cmu.edu/software/>

Soar는 1983년 처음 연구 프로젝트가 시작됐고 ACT-R과 함께 전통적인 인공지능 연구를 대표한다. Soar는 범용 지능을 갖는 에이전트를 만들기 위해 그 구성 요소기술(building blocks)을 개발하는 것을 목표로 한다. 이 에이전트는 인간의 거의 모든 지적활동을 대체할 수 있는 것으로 의사 결정, 문제 해결, 계획 수립, 자연언어처리 등 광범위한 일을 할 수 있는 범용 인공지능이다. Soar는 ACT-R과 마찬가지로 인간의 인지(cognition)에 대한 이론의 정립과 이를 컴퓨터 공학적으로 구현함에 있다. Soar는 범용 인공지능의 기능과 효율성에 집중을 한 반면, ACT-R은 인간의 인지에 대한 세부적인 모델링에 중점을 뒀다는 것으로 그 차별점을 찾을 수 있다.

Soar는 인간의 인지에 대한 가설 몇 가지를 제시한다. 첫 번째 가설인 문제 공간 가설(Problem Space Hypothesis)은 아무리 복잡한 임무도 세부적으로 분할될 수 있으며, 이것은 간단한 의사결정의 연속(sequence)로 볼 수 있다는 것이다. 두 번째 가설은 간단한 의사결정이 순차적이 아닌 병렬적으로 이루어진다는 점이다. 또한 Soar의 기본 구조는 모듈화 되어있고, 하나의 모듈이 언어나 계획 등 기능적인 지능 활동이라기보다는 장단기 기억, 시각적인 기억 등 기억에 관련된 모듈로 구성돼있다.

[그림 3-13]은 Soar의 메모리 모듈을 나타낸다. 기간별, 종류별로 기억을 세분화하여 다양한 지능적 행동을 수행하는 것이 가장 큰 특징이다.

Soar의 대표적인 응용 분야는 추론을 요구하는 퍼즐과 게임, 파일럿 시뮬레이션, 자연언어 이해, 로봇틱스 등이 있다. Soar 역시 ACT-R과 마찬가지로 인공지능 프로그램을 공개SW 형태로 제공한다.

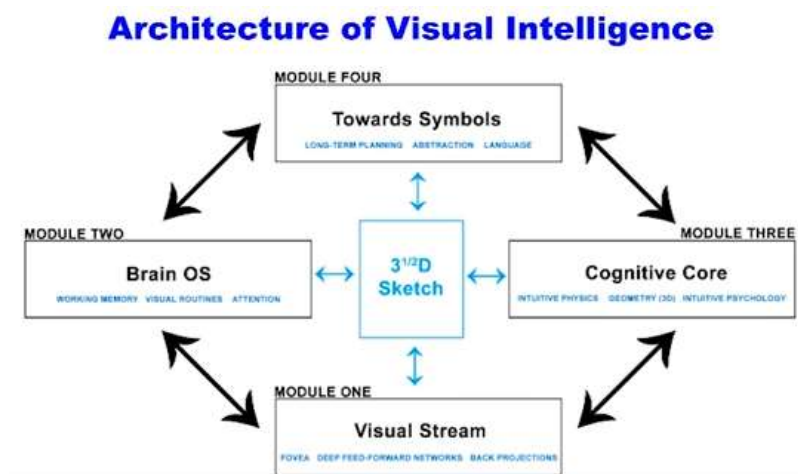


[그림 3-13] Soar의 메모리 모듈 구조

자료 : Extending the Soar Cognitive Architecture,
John E. Laird (2008)

CBMM(Center for Brains Minds+Machines)³⁴⁾은 MIT의 연구소로 인간의 뇌구조를 역공학으로 구현하는 연구를 수행한다. CBMM은 궁극적으로 범용 인공지능을 개발하는 것을 목표로 하고 있고, 현재는 시각 지능(Visual Intelligence)에 집중하고 있다. 현재 심층학습 기술은 이미지에서 사물을 인식하는데 매우 뛰어난 성능을 보유하고 있다. 그러나 사람이 가지고 있는 시각적인 지능은 복합적인 의사결정의 결과다. 예를 들어 500명이 앉아있는 강당의 첫 번째 줄에 있는 사람에게 “당신 뒤에 몇 명이나 있나요?” 혹은 “당신에서 가장 가까운 벽은 얼마나 떨어져 있나요?”에 대한 질문에 대해 뒤나 벽을 보지 않고도 대답할 수 있을 것이다. 이처럼 사람의 시각 지능은 비단 시각적인 정보뿐만 아니라 경험과 직관에 의해 구성된다는 점이다. 따라서 현대의 심층학습은 지능이라기보다 고성능의 패턴 인식 기술이라고 볼 수 있다.

사람처럼 행동할 수 있는 시각 지능을 구현하기 위해 CBMM은 다양한 분야의 전문가와 협업을 하고 있다. 특히 유아의 행동 발달과정을 분석하는 심리학자와 협업하여 시각지능을 구현하기 위한 가설을 세우고 실험을 통해 증명해 나가고 있다. 그 결과 유아의 행동을 역공학적으로 분석하여 아래 [그림 3-14]와 같은 가설을 세웠다. 이것을 구현하는 방법론으로는 확률적 프로그램(Probabilistic program)으로 소개했다. 확률적 프로그램은 기호적 지능 표현과 확률적 추론, 그리고 심층학습 구조를 융합하여 만든 개념이다.



[그림 3-14] CBMM의 시각지능 구조 가설

자료 : MIT AGI: Buliding machines that see, learn, and think like people (2018)

34) MIT의 AGI 강의 “MIT AGI: Buliding machines that see, learn, and think like people (Josh Tenenbaum)”에서 요약

제4절 소 결

지금까지 다섯 가지 인공지능 기초 연구 사례와 범용 인공지능을 살펴봤다.

먼저 기초연구의 공통점은 두 가지 특징으로 요약 할 수 있다. 첫 번째는 모두 인공신경망 기반의 접근 방법이다. 인공신경망은 동물의 뇌 신경계를 모사한 접근으로, 1980년대 주류를 이끌었던 방법론이었으나 데이터와 계산 자원의 한계로 인해 정체기에 진입했었다. 그러나 빅데이터의 출현과 저렴한 고성능 HW의 보급으로 인공신경망을 필두로 한 심층학습이 재조명됐다. 심층학습의 눈부신 성공으로 인해 현재 인공지능 기초연구 역시 심층학습 자체의 성능을 향상시키는데 초점이 맞춰져 있다고 볼 수 있다. 이 성능을 향상시키는 방점은 앞서 소개한 심층학습의 한계를 극복하기 위한 것이다.

두 번째 특징은 인공지능 기초연구가 아직 초기단계에 머물러 있다는 점이다. 분석한 인공지능 기초연구는 궁극적으로 심층학습의 효율성과 범용성을 높이는데 중점을 두고 있다. 그러나 효율성과 범용성 어느 측면에서도 아직 실험적인 단계이며, 많은 가정을 두고 도출한 결과다. 물론 심층학습 자체가 가지고 있는 기술적 한계를 고려하면 이러한 시도가 활발히 이루어진다는 면은 긍정적이라 볼 수 있다. 보통 R&D 단계에서 기초연구는 5년 이상의 장기과제가 대부분이다. 하지만 현재 심층학습이라는 키워드에 집중되고 있는 트렌드를 볼 때 2~3년 이내에는 기초연구에서 괄목할 만한 성과가 있을 것이라 기대된다.

인공지능 기초연구는 향후 인공지능의 방향을 좌우 할 수 있는 잠재력이 높다고 판단된다. 그만큼 달성하고자 하는 목표 역시 어렵다. 이 보고서에서 인공지능 기초연구를 중점적으로 다룬 것 역시 이러한 이유에서다. 현재 인공지능 연구의 영역에서 기초연구가 차지하는 비중은 그렇게 크지 않으나, 인공지능의 근원적인 성능을 높일 수 있다는 점에서 큰 의미가 있을 것이다.

차세대 인공지능이라고 볼 수 있는 범용 인공지능의 경우 관련된 문헌이나 활동을 종합해보면 범용 인공지능의 출현은 아직 먼 미래로 보인다. 다만, 심층학습을 필두로 한 인공지능의 성능은 눈부시게 발전하고 있다는 것은 사실이기 때문에 범용 인공지능에 대한 미래도 밝다고 볼 수 있다.

전통적인 기호적 범용 인공지능의 연구 사례를 살펴보면 심층학습과는 전혀 다른 접근 방법을 취하고 있다. 그렇기 때문에 과연 심층학습의 발전이 범용

인공지능과 상관관계가 있을까?라는 의문을 품게 된다. 전문가들은 기호적 범용 인공지능과 심층학습의 유기적인 결합을 강조한다. 각 분야에 특화된 임무가 서로 다를 수 있기 때문이다. 한편으로는 뇌를 역공학으로 분석하여 구현하는 창발적 범용 인공지능 연구가 진행되고 있다. 이 접근 방법은 유아의 행동 발달 과정이나 뇌의 네오코텍스(neo-cortex), 혹은 뇌 전체를 슈퍼컴퓨터로 구현하는 것을 목표로 한다.

많은 전문가들이 강조했듯이 범용 인공지능이 출현한다면 그것은 인간의 수준을 쉽게 뛰어넘는 초인공지능이 될 가능성이 높다는 것이다. 이 사실은 사람과 인공지능이 더 이상 주종관계가 아니고 공존하는 관계여야 함을 시사한다. 이에 많은 인공지능 연구자들은 미래 인공지능의 윤리와 정책에 대해서도 심도 있는 논의를 진행하고 있다. 또한, 인공지능은 반드시 사람에게 이롭게 활용돼야 한다는 대 원칙아래, 전 세계 인공지능 연구자는 인공지능을 이롭게 활용하기 위한 23가지 원칙을 정립했다. 이러한 움직임은 그만큼 범용 인공지능에 대한 파급력이 크다는 사실을 반증한다.

우리도 인공지능을 이롭게 활용해야 한다는 대 전제에 동참하여 앞으로의 인공지능 R&D 방향을 설정해야 할 것이다. 특히 인공지능에 대한 윤리는 본격적인 연구가 필요하다고 본다. 또한 현재 인공지능은 그 활용 측면에서 산업계의 확산이 매우 빠르게 일어나고 있다. 인공지능을 통한 산업 진흥도 중요한 분야지만, 더 막대한 파급력을 갖고 있는 범용 인공지능에 대한 투자도 아끼지 말아야 한다고 본다.

제4장 요약 및 시사점

인공지능이 세 번째 황금기를 맞이하며 성공가도를 달리고 있다. 많은 연구자들은 이번 황금기만큼은 지속될 것이라고 희망찬 예측을 하고 있다. 현대 인공지능은 과거 두 번의 황금기와는 배경이 다르다. 빅데이터의 출현과 값싼 계산 자원의 보급으로 인공지능 자체가 발전하기 위한 기반이 다져졌기 때문이다.

그렇다면 인공지능의 세 번째 황금기가 어떤 식으로 지속되고 있는지에 대한 이해가 중요하다. 인공지능 기술의 체계적인 이해를 돕기 위해 이 보고서에서는 가설단계이지만 인공지능 연구 영역의 가설을 제안했다. 그 근거는 현재 인공지능과 관련된 R&D와 산업의 특성에서 도출했다. 심층학습의 근원지인 인공지능 응용연구, 심층학습을 활용해 각 산업계에서 적용하는 인공지능 개발연구, 심층학습의 한계를 극복하기 위한 인공지능 기초연구가 그 주요 골자다.

특히 이 보고서에서는 파급력이 높은 인공지능 기초연구에 중점을 두고 다섯 가지 연구 사례를 분석했다. 물론 이것은 빙산의 일각에 불과하겠지만 인공지능 기초연구의 방향은 이해할 수 있을 것이다. 현재 인공지능 기초연구는 좁은 인공지능에 머물러 있는 심층학습의 범용성을 높이는 방향으로 진행되고 있다. 이것은 곧 차세대 인공지능 연구의 초석이자 응용·개발 연구에의 접목되어 성능을 향상시키는데 기여할 것이다. 인공지능 기초연구는 상당부분 범용 인공지능과 가까운 부분이 있으나, 인공신경망 기술에 중점을 두고 있다는 점에서 구별된다. 따라서 인공지능 기초연구가 어떠한 식으로 발전할지 그 귀추가 주목된다.

범용 인공지능의 경우 아직 그 출현 자체는 시기상조라 판단되나 인공지능에 대한 연구의 비상한 관심도는 다양한 임무를 수행할 수 있는 인공지능 개발을 가능케 할 것이다. 물론 이 범용 인공지능이 스스로 학습하는 초인공지능으로 진화할 것인가에 대해서는 논란의 여지가 있다. 그러나 인류가 만들어낼 최후의 작품의 초석은 범용 인공지능일 가능성이 높다. 향후 범용 인공지능 연구는 다양한 학문의 융합으로 발전될 것이다. 비단 컴퓨터 과학뿐만 아니라 뇌과학, 인지과학, 심리학 등과 다학제적 연구를 시작할 때라고 본다.

그렇다면 우리의 당면과제는 무엇일까? 인공지능은 제4차산업혁명의 핵심 키워드로 미래 사회를 대비하기 위해 선택지가 아닌 필수적인 기술로 자리매김하고 있다. 우리 정부역시 인공지능 기술 육성과 산업 진흥을 위한 노력을 이어가고 있다. 특히 이 보고서에서는 차세대 인공지능 개발을 준비하기 위한 인공지능 기초 연구의 중요성에 대해 논의했다.

차세대 인공지능은 우리나라뿐만 아니라 전 세계적인 도전과제다. 따라서 차세대 인공지능에 막대한 투자를 한다고 해도 유의미한 결과가 없을 수도 있다. 이러한 위험 부담을 안고 R&D에 투자한다는 것은 어느 국가나 쉽지 않은 선택지일 것이다.

그러나 인공지능 기초연구는 범용 인공지능으로 가는 길을 제시하고 있다. 현재 상태에서는 심층학습의 성능을 개선하는데 집중되어 있으나, 향후에는 기호적 인공지능과의 융합이나 새로운 방법론이 도출될 수 있을 것이다. 또한 인공지능 기초연구는 데이터에서 상대적으로 자유롭다. 예를 들어 사진 인식과 음성 인식을 동시에 수행할 수 있기 위해 단일 인공신경망을 개발하는 것은 현재 시점에서도 거의 해결 불가능한 과제다. 그렇다면 상대적으로 쉬운 데이터셋으로부터 출발할 수 있을 것이다. MNIST나 Tree-bank 데이터의 경우 연구용으로 공개된 데이터이고, 메타 학습을 비롯한 인공지능 기초연구는 이러한 기초 데이터만으로도 연구가 가능하기 때문이다.

인공지능 기초연구에 대한 과감한 투자는 우리나라 인공지능 기술의 기초체력을 육성한다는 점에도 큰 의미가 있다. 서론에서도 밝혔듯이 인공지능은 도구적인 속성이 강하다. 현재 인공지능 전문가라고 함은 사실 문제를 잘 해결하기 위한 데이터 과학자에 가깝다. 주어진 데이터로 문제를 해결하기 위해 심층학습을 비롯한 다양한 인공지능 기법을 활용하는 것이 바로 인공지능 전문가라고 볼 수 있기 때문이다. 문제 해결형 인재를 육성하는 것은 산업 진흥의 관점에서 큰 의미가 있겠지만, 인공지능 원천기술 확보를 위한 각국의 치열한 경쟁에서는 우위를 점할 수 없다.

인공지능 기초연구에의 투자는 인공지능 자체의 성능을 향상시킬 수 있는 고급 인재 육성과도 맞물려 있다는 점에서도 매우 중요하다고 본다. 우리 정부는 인공지능 기초연구 투자와 확산을 위한 정책적 선택지를 적극적으로 활용하여, 상대적으로 뒤처져 있는 기술의 추격과 전문 인재 양성을 통해 돌파구를 모색해야 한다고 본다.

참 고 문 헌

1. 2017 ICT 기술수준보고서, IITP (2018.05.)
2. 스타트업 코리아! 결과보고서, 아산나눔재단 (2017)
3. 공영일 외, 인공지능 팔목상대, 중국, 소프트웨어정책연구소 (2017.07.)
4. 심지섭, EU 개인정보보호법 발효 이후 각국의 대응 동향, 소프트웨어정책연구소 (2018.12.)
5. 이승환, 인공지능 연구역량 국제비교 및 시사점, 소프트웨어정책연구소 (2018.11.)
6. 추형석, 신경망 구조 탐색 연구 동향, 소프트웨어정책연구소 (2018.06.)
7. 추형석, 범용 인공지능의 개념과 연구 현황, 소프트웨어정책연구소 (2018.07.)
8. 추형석, 인공지능 어디까지 왔나?, 소프트웨어정책연구소 (2018.11.)
9. Esser, S. K., et al. "Convolutional networks for fast, energy-efficient neuromorphic computing. 2016." Preprint on ArXiv. <http://arxiv.org/abs/1603.08270>. Accessed 27 (2016).
10. Fernando, Chrisantha, et al. "Pathnet: Evolution channels gradient descent in super neural networks." arXiv preprint arXiv:1701.08734 (2017).
11. Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks." arXiv preprint arXiv:1703.03400 (2017).
12. Goertzel, Ben. "Artificial general intelligence: concept, state of the art, and future prospects." Journal of Artificial General Intelligence 5.1 (2014): 1-48.
13. Grace, Katja, et al. "When will AI exceed human performance? Evidence from AI experts." Journal of Artificial Intelligence Research 62 (2018): 729-754.
14. Kulkarni, Santosh, Sishaj P. Simon, and Kinattingal Sundareswaran. "A spiking neural network (SNN) forecast engine for short-term electrical load forecasting." Applied Soft Computing 13.8 (2013): 3628-3635.
15. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436.
16. Pham, Hieu, et al. "Efficient Neural Architecture Search via Parameter Sharing." arXiv preprint arXiv:1802.03268 (2018).
17. Shin, Hanul, et al. "Continual learning with deep generative replay." Advances in Neural Information Processing Systems. 2017.
18. Tavanaei, Amirhossein, et al. "Deep learning in spiking neural networks." Neural Networks (2018).
19. Vinyals, Oriol, et al. "Matching networks for one shot learning." Advances in neural information processing systems. 2016.
20. Zoph, Barret, and Quoc V. Le. "Neural architecture search with reinforcement

learning." arXiv preprint arXiv:1611.01578 (2016).

21. Hearing: Game Changers - Artificial Intelligence Part I, Information Technology Subcommittee, US Government (2018.02.)
22. Hearing: Game Changers - Artificial Intelligence Part II, Artificial Intelligence and the Federal Government, Information Technology Subcommittee, US Government (2018.03.)
23. Hearing: Game Changers - Artificial Intelligence Part III, Artificial Intelligence and Public Policy, Information Technology Subcommittee, US Government (2018.04.)
24. Rise of Machines, Artificial Intelligence and its Growing Impact on U.S. Policy, US Government (2018.09.)
25. 아실로마 AI원칙, Future of Life Institute (2017), <https://futureoflife.org/ai-principles-korean/>
26. AutoML, <https://cloud.google.com/automl/?hl=ko>
27. Google Deepmind Blog, "Enabling Continual Learning in Neural Networks" <https://deepmind.com/blog/enabling-continual-learning-in-neural-networks/>
28. ImageNet, <http://www.image-net.org/>
29. ISO JTC1 / SC42 Artificial Intelligence <https://www.iso.org/committee/6794475.html>
30. Thomas Wolf, From zero to research - An introduction to Meta-learning
31. Vincenzo Lomonaco, Why Continuous learning is the key towards machine intelligence

별 첨 1. 인공지능 기술과 산업의 분류

1. IDC

대 분 류	세부 분야
Hardware	Server, Storage
Software	Cognitive Application, Cognitive Platform
Service	IT Service, Business Service

자료 : Worldwide Semiannual Cognitive/Artificial Intelligence Systems Spending Guide, IDC(2017)

2. GVR (Grand View Research)

대 분 류	세부 분야
Hardware	The hardware segment includes chipsets such as the Central Processing Unit (CPU), Graphics Processing Unit (GPU), Application-Specific Integrated Circuits (ASIC), Field-Programmable Gate Array (FPGA)
Software	AI-based software solutions include libraries for designing and deploying artificial intelligence applications including those for, inference, primitives, video analytics, sparse matrices, linear algebra, and multi-hardware communication capabilities.
Service	The service segment includes installation, integration, and maintenance & support undertakings

자료 : Artificial Intelligence Market Analysis(By Solution By Technology By End-use, By Region, and Segment Forecasts 2014 - 2025, AI Market, GVR(2017)

3. Gartner

대 분 류	세부 분야
Core AI Technology	Machine Learning, Deep Learning and Neural Networks Natural-Language Processing, Speech Recognition and Text to Speech, Computer Vision, Machine Reasoning, Decision Making and Algorithms, Business Analytics and Data Science, Robots and Sensors
Leading AI Application Areas	Bots, Chatbots and Virtual Assistants, Conversational AI Platforms, Analytics and Predictive Analytics Models, Smart Objects, Sensors and Environments
AI in the Enterprise	General Enterprise AI Strategies, Customer Service and Support, Digital Business Initiatives, Digital Commerce and Digital Commerce Personalization, Governance and Information Management for AI, Human Capital Management/Recruiting, Internet of Things, IT Service Monitoring/Help Desk, Legal, Ethical and Social Issues, Sales, Security and Fraud, Workplace and Digital Workplace
AI in Vertical Industries	Automotive, Banking, Asset and Wealth Management, Capital Markets, Business Process Outsourcing, Customer Management, Finance and Accounting Procurement, and HR, Consumer Products With AI, Government, Healthcare, Life Science and Pharma Education (K-12 and Higher Education), Insurance Services — Property, Casualty and Life, Invest/Venture Capital and AI, Manufacturing Operations, Retail, Supply Chain, Telecom, Topics Specific to Regional AI Usage

자료 : A Framework for Applying AI in the Enterprise, Gartner(2017)

별 첨 2. 범용 지능에 대한 다양한 해석

앞서 범용 인공지능은 사람 수준의 지능이라고 소개했다. 그렇다면 사람 수준의 지능이라는 실체는 무엇일까? 사람의 일반적인 지능적 행동을 지칭하는 단어로는 범용 지능(General Intelligence)이 있다. 이 범용 지능을 구체화하기 위해 다양한 분야의 연구자들은 자신들의 해석을 밝혔다.³⁵⁾

먼저 실용적 관점(Pragmatic Approach)이 있다. 이 관점의 대표적인 예는 1950년 앨런 튜링이 제안한 ‘튜링 테스트’이다. 튜링 테스트는 사람과 인공지능의 대화에서 누가 사람인지 인식할 수 있는지의 여부를 판단하는 것이다. 이것을 확대해서 보면 사람의 지능과 인공지능의 기능적 측면을 비교하여 범용 인공지능을 판단하는 접근이다. 다시 말하자면, 사람의 지능적 행동에 방해될 만큼 인공지능 기술이 발전했다면 이것이 곧 범용 지능이라는 관점이다.

두 번째는 심리학적 관점이다. 범용 지능의 심리학적 관점은 실용적 관점보다는 더 원론적인 접근을 취한다. 심리학 관점에서는 범용 지능을 크게 두 가지 원칙으로 해석한다. 먼저 개인이 형성한 도메인 지식(Knowledge domains)은 서로 상관성이 있기 때문에, 해당 도메인별 지식의 수준(level)이 크게 다르지 않다. 이것을 개인 내 다양성(intra-individual variability)이라고 한다. 두 번째 원칙은 개인 간의 도메인 지식은 상당한 수준으로 상이할 수 있다는 것으로, 개인 간 다양성(inter-individual variability)라고 지칭한다. 심리학자 가드너(Gardner)는 이것을 바탕으로 인간의 지능에 대해 8가지³⁶⁾로 구분 지었다. 심리학적 관점에서 인간 수준의 범용 지능 논의하기 개최된 2009년 AGI Roadmap workshop에서는 14가지³⁷⁾의 분류체계를 제시했다.

다음은 인지 아키텍처(cognitive architecture) 관점의 범용 지능이다. 인지 아키텍처는 SW의 속성이 강하다. 따라서 인지 아키텍처 연구자들은 범용 지능을 구현하기 위해 SW적인 관점에서의 요구사항을 논의한다. 예를 들면, 다양한 작업을 수행함에 있어 SW 구조는 고정되어야 한다는 점이다. 새로운

35) 『Artificial General Intelligence: Concept, State of the Art, and Future Prospects』에서 소개된 내용을 정리

36) ① 언어적(linguistic), ② 논리적·수학적(logical-mathematical), ③ 음악적(musical), ④ 운동감각적(bodily kinesthetic), ⑤ 공간적(spatial), ⑥ 대인관계(interpersonal), ⑦ 자기 내부적(intrapersonal), ⑧ 자연주의적(naturalist) 지능, 다중 지능이론, Gardner(1999)

37) 인지(perception), 구동(actuation), 기억(memory), 학습(learning), 추론(reasoning), 계획(planning), 주목(attention), 동기(motivation), 감정(emotion), 모델링(modeling self and other), 사회적 상호작용(social interaction), 소통(communication), 정량적(quantitative), 창조(building/creation)

작업을 수행하기 위한 논리나 알고리즘이 SW상에 구현돼 있어 그 구조는 동일해야함을 의미하는 것이다.

네 번째는 수학적 방법이다. 수학적 방법의 주된 가정은 범용 지능을 구현하기 위해서 무한대의 계산이 필요하다는 점이다. 이것은 제한된 계산 자원으로는 범용 지능을 달성할 수 없으나, 특정 시스템은 다른 시스템보다 더 지능적일 가능성이 있다고 해석할 수 있다. 이 가정은 사람 역시 완벽한 범용 지능이라고 볼 수 없는 모순에 직면한다. 그러나 사람은 다른 시스템(곤충, 파충류 등)보다 더 지능적이라고 볼 수 있기 때문에 수학적 접근으로 지능을 해석하는 것이 일견 타당한 면이 있다.

마지막은 주변 환경과 밀접하게 관련된 적응주의적(Adaptationist) 관점에서 범용 지능을 접근하는 방법이다. 인공지능 연구자인 Pei Wang은 범용 지능을 ‘제한된 자원을 활용해 환경에 적응하는 것’으로 개념화했다. 지금까지 범용 지능을 특징화한 다섯 가지 접근에 대해서 살펴봤다. 이것을 정리하면 다음 <표 별첨-1>과 같다.

<표 별첨-1> 범용 지능을 특징화하기 위한 다양한 접근방법

접근 방법	내 용
실용적 관점	- 사람의 지능과 인공지능의 기능적 측면을 비교하는 접근 - 대표적인 예 : 튜링 테스트
심리학적 관점	- 지능의 해석 : 개인 내 다양성, 개인 간 다양성 - 인간의 지능 분류 : Gardner의 8가지 분류
인지 아키텍처 관점	- SW의 관점에서 범용 지능의 해석 - 다양한 작업을 수행함에 있어 SW구조는 고정돼야 함
수학적 관점	- 범용 지능을 구현하기 위해서는 무한대의 계산 자원이 필요 - 제한된 계산 자원으로는 상대적으로 우수한 범용 지능을 구현
적응주의적 관점	- 제한된 자원을 활용해 환경에 적응한다는 접근

자료 : Artificial General Intelligence: Concept, State of the Art, and Future Prospects, Ben Goertzel (2017)

주 의

1. 이 보고서는 소프트웨어정책연구소에서 수행한 연구보고서입니다.
2. 이 보고서의 내용을 발표할 때에는 반드시 소프트웨어정책연구소에서 수행한 연구결과임을 밝혀야 합니다.



[소프트웨어정책연구소]에 의해 작성된 [SPRI 보고서]는 공공저작물 자유이용허락 표시기준 제 4유형(출처표시-상업적이용금지-변경금지)에 따라 이용할 수 있습니다.
(출처를 밝히면 자유로운 이용이 가능하지만, 영리목적으로 이용할 수 없고, 변경 없이 그대로 이용해야 합니다.)