

# 빅데이터·인공지능 산업 진흥을 위한 데이터 과학의 발전 전략 연구

Study on the strategies of big data & artificial intelligence  
for the promotion of industrial development





# 빅데이터·인공지능 산업 진흥을 위한 데이터 과학의 발전 전략 연구

Study on the strategies of big data & artificial intelligence  
for the promotion of industrial development



■ **집필위원장**

이 영 조 (한국과학기술한림원 정회원, 서울대학교 교수)

■ **집필위원**

박 성 현 (한국과학기술한림원 종신회원, 사회적책임경영품질원 회장)

장 중 순 (아주대학교 교수)

김 선 (서울대학교 교수)

박 태 성 (서울대학교 교수)

신 현 정 (아주대학교 교수)

이 재 길 (한국과학기술원 교수)

# 요약문

## 연구과제명

**국문** 빅데이터·인공지능 산업 진흥을 위한 데이터 과학의 발전 전략 연구

**영문** Study on the strategies of big data & artificial intelligence for the promotion of industrial development

연구책임자 | 이영조

한림원 소속부 | 이학부

- 갈릴레오, 뉴턴 등의 과학 혁명을 거치면서 데이터로부터 확실한 지식을 얻을 수 있다는 새로운 귀납적 방법론의 중요성이 대두되었고, 과학의 혁명은 기술의 발전과 더불어 산업혁명으로 이어져 인류 문화와 의식주 전반에 급격한 발전을 이뤘음. 물리학, 지구과학, 화학, 그리고 생물학의 비약적 발전으로 이어지던 과학 혁명은 이제 귀납적 방법론 자체에 대한, 즉 인간의 인지적 사고 과정들에 대한, 그리고 우리의 의사 결정에 도움을 줄 수 있는 데이터 수집, 저장, 요약, 분석 및 예측 등을 위한 수행 알고리즘에 대한 연구들로 초점이 맞춰지게 되었음
- 단 백년도 안 된 지난 세기에 컴퓨터와 통계학, 컴퓨터공학, 산업공학, 그리고 IT 기술의 급격한 발전은 컴퓨터, 스마트폰, 로봇, 인터넷 시대의 대두와 더불어 귀납적 방법에 대한 통계학 이론들이 컴퓨터, 데이터 처리 기술, IT, 로봇 기술들과 융합하여 새로운 빅데이터·인공지능 경제 시대를 열게 되었음. 즉 데이터를 원료로 인공지능 기술을 가미하여 로봇, 전기·전자, IT기술 또는 사물인터넷 등을 통해 새로운 제품과 서비스, 산업을 창출하는 4차 산업혁명 시대가 도래하게 됨

- 이제 자원이 데이터이며 빅데이터·인공지능 기술에 의해 새로운 산업과 서비스가 창출되는 4차 산업혁명 시대는 최고의 기술을 가진 기업이 이익을 독식하는 무한 경쟁 시대가 되어, 그 어느 때보다도 우리 자신의 기술이 우위가 되어야 하는 상황이 되었음
- 미·중 그리고 한·일 무역 갈등이 심화되는 이때, 다양한 분야의 학계·산업계 전공자들이 모여 다가올 시대를 준비하는 마음으로 지혜를 모아 빅데이터·인공지능의 개요와 역사 그리고 기술 수준을 살펴보고, 빅데이터·인공지능 관련 산업의 현황 및 변화를 살펴보고 최종적으로 데이터사이언스 발전 전략을 제안하는 것으로 “빅데이터·인공지능 산업 진흥을 위한 데이터사이언스 발전 전략 연구”에 관한 소고를 준비하였음
- 본 소고가 한국 데이터사이언스, 빅데이터·인공지능 기술발전에 도움이 되어 4차 산업 발전에 이바지하기를 기원하며, 다가올 미래를 위해 자원이 빈약한 한국은 빅데이터·인공지능 관련 고급 인력을 더욱 양성해야 한다는 10만 양병론을 제안함. 이제 한국은 차세기의 힘찬 도약을 위하여 빅데이터·인공지능 관련 산업들을 선도할 고급 인력을 양성하고, 이들의 국제적 활약으로 새로운 시대를 개척해 나아가야 한다고 전망함

# 목 차

빅데이터·인공지능 산업 진흥을 위한 데이터 과학의 발전 전략 연구

Study on the strategies of big data & artificial intelligence for the promotion of industrial development

I. 서론 .....	1
II. 빅데이터·인공지능 개요 .....	7
1. 데이터사이언스·빅데이터·인공지능의 정의 .....	8
2. 데이터사이언스 프로세스 및 체계도 .....	9
3. 데이터사이언스에서 다루는 문제 분류 및 모델링 .....	11
■ 참고문헌 .....	16
III. 빅데이터·인공지능 역사 .....	17
1. 데이터사이언스·빅데이터의 역사 .....	18
2. 인공지능·기계학습의 역사 .....	24
■ 참고문헌 .....	33
IV. 빅데이터·인공지능 기술 수준 .....	35
1. 데이터 관리 기술 수준 .....	36
2. 인공지능 기술 수준 .....	44
■ 참고문헌 .....	57
V. 빅데이터·인공지능 관련 산업 .....	61
1. 빅데이터·인공지능 산업 동향 .....	62
2. 제조업 .....	71
3. 바이오·의료 산업 .....	84
4. IT 산업 .....	93
5. 금융 산업 .....	98
6. 스타트업 .....	103
■ 참고문헌 .....	108

# 목 차

빅데이터·인공지능 산업 진흥을 위한 데이터 과학의 발전 전략 연구

Study on the strategies of big data & artificial intelligence for the promotion of industrial development

VI. 데이터사이언스 발전 전략 .....	111
1. 교육 .....	112
2. R&D .....	125
3. 산업 .....	132
■ 참고문헌 .....	140
VII. 결론 .....	141

---



## 표 목차

<표 4.1> 인공지능 기술 분류 및 주요 키워드 .....	45
<표 5.1> 영역별 국내 빅데이터 산업 규모 현황 .....	65
<표 5.2> 인공지능 분야 주요 글로벌 기업 M&A 동향 .....	67
<표 5.3> 유전 변이 데이터베이스 .....	89
<표 5.4> 인공지능 스타트업의 비즈니스 모델 .....	105
<표 5.5> 한국 인공지능 스타트업의 글로벌 시장 진출 사례 .....	105
<표 6.1> 미국 상위 10개 데이터사이언스 석사 과정 .....	113
<표 6.2> 석사 학위 프로그램의 핵심 과목들(2014년 8월 기준) .....	114
<표 6.3> 인공지능 대학원 교육 과정 상위 10개 대학 .....	115
<표 6.4> Carnegie Mellon 대학의 인공지능 석사 과정 커리큘럼 .....	117
<표 6.5> 2019년 국내 대표 데이터사이언스 학위 과정 .....	118
<표 6.6> 2019년 국내 인공지능 학위 과정 개설 현황 .....	120
<표 6.7> 데이터 전문 인력과 빅데이터 인력 수요 추정 .....	123

# 목 차

빅데이터·인공지능 산업 진흥을 위한 데이터 과학의 발전 전략 연구

Study on the strategies of big data & artificial intelligence for the promotion of industrial development

## 그림 목차

<그림 2.1> 데이터사이언스 프로세스 .....	9
<그림 2.2> 데이터사이언스 전체 체계도 .....	10
<그림 2.3> 분류예측 및 회귀예측 .....	12
<그림 2.4> 데이터 기술을 위한 군집화와 이상치 탐지 및 차원 축소 .....	13
<그림 2.5> 다양한 형태의 데이터 통합 .....	13
<그림 2.6> 기계학습 모델링 과정 .....	14
<그림 3.1> IFCS-96 학술대회 자료 .....	19
<그림 3.2> 과학의 패러다임 변화 .....	20
<그림 3.3> 정보의 디지털화 및 저장 용량 추세 .....	23
<그림 3.4> 인터넷에서 1분 동안 발생하는 데이터 규모 .....	24
<그림 3.5> 다트머스 콘퍼런스를 개최한 인공지능의 아버지들 .....	26
<그림 3.6> 시기별 인공지능에 대한 기대감과 기계학습의 대중성 .....	26
<그림 3.7> 홉필드넷 예시 .....	29
<그림 4.1> 리듀서가 하나일 때의 맵리듀스 데이터 흐름 .....	37
<그림 4.2> 스톱 토폴로지 예시 .....	40
<그림 4.3> New SQL 부류의 대표적인 제품들 .....	43
<그림 4.4> 대화형 추론 문제 .....	45
<그림 4.5> 시각 인식, 시각 이해, 이미지 생성 기술 구현 예시 .....	46
<그림 4.6> 온톨로지 방식과 워드임베딩 방식의 비교 .....	47
<그림 4.7> 컴퓨터 비전 기술 분류 .....	48
<그림 4.8> 음성 인식 모델 수행 과정 .....	49
<그림 4.9> 설명 가능한 인공지능 구현 예시 .....	55
<그림 4.10> 인공 신경망의 설명 가능한 노드에 대한 레이블 예시 .....	56
<그림 5.1> 세계 빅데이터 시장 전망 .....	62



<그림 5.2> 글로벌 및 중국 빅데이터 시장 비교 .....	63
<그림 5.3> 국내 빅데이터 시장 규모 추이 .....	64
<그림 5.4> 부문별 빅데이터 투자 비중 시장 규모 .....	64
<그림 5.5> 인공지능 분야 글로벌 투자 건수 및 액수 .....	66
<그림 5.6> 산업 분야별 인공지능 활용 변화 추세 .....	67
<그림 5.7> 인공지능 적용으로 인한 기업 업무별 경제적 효과 .....	68
<그림 5.8> 인공지능 적용이 산업계에 미치는 경제적 효과 및 매출 .....	69
<그림 5.9> 국내 인공지능 분야 시장 규모 .....	70
<그림 5.10> 제조업의 정보 시스템 .....	73
<그림 5.11> 제조업의 공정 관리 시스템 .....	76
<그림 5.12> 공정 데이터 분석 프로그램 .....	77
<그림 5.13> 인공지능 시각화 기술로 익은 토마토를 판별한 연구 사례 .....	85
<그림 5.14> 단일 염기 다형성 .....	88
<그림 5.15> 약물유전체학 연구 예시 .....	90
<그림 5.16> 유전체 정보 생성 및 분석 과정 .....	91
<그림 5.17> 인공지능 하드웨어 분야별 시장 규모 .....	94
<그림 5.18> 사물인터넷 디바이스 수 전망 .....	95
<그림 5.19> 사물인터넷 기술 발전 전망 .....	96
<그림 5.20> 인공지능 스타트업 투자 추이 및 구성 .....	103
<그림 5.21> 주요 기업들의 인공지능 스타트업 인수합병 현황 .....	104
<그림 6.1> 인공지능 발전 전략 .....	128
<그림 6.2> 서울대와 스탠포드대의 컴퓨터공학 및 인공지능 인제 수 비교 .....	132
<그림 6.3> 데이터 산업 시장 전망 .....	133
<그림 6.4> 데이터 산업 시장의 추진 전략 .....	134



# I



## 서론



## I

## 서론

- 그리스 철학자 플라톤은 그가 세운 대학 문 앞에 “기하학을 모르는 사람은 들어오지 마시오”라는 글을 써 넣었는데, 이는 기하학이 우주의 비밀을 밝힐 수 있는 열쇠라고 생각했기 때문임. 그러나 인류는 콜럼버스와 마젤란의 항해 이후, 평평한 세계에서 진리였던 그리스 로마의 유클리드 기하학이 더 이상 진실이 아닌 것을 알게 되었음. 즉 수학·논리학으로 대변되는 그리스 로마의 지식 획득 방법은 전제가 참이어야 하는데, 수학의 공리와 논리학의 기본 명제들이 더 이상 증명의 대상이 되지 않는다는 것을 알게 됨. 이로써 수학·논리학 등에 의해 획득한 지식이 참인지 아닌지 여부도 증명의 대상이 아닌 기본 명제들의 참인지 아닌지의 여부에 따름
- 따라서 확실한 지식을 얻기 위해서는 관측된 데이터와 데이터로부터의 새로운 지식 창출이 중요해지고, 갈릴레오·뉴턴 등의 과학 혁명으로 자연에서 관측하여 얻은 데이터들을 설명하는 새로운 귀납적 학문인 과학 시대가 도래하게 되었음. 과학의 시대는 기술의 발전과 더불어 산업혁명으로 이어져 인간 생활의 다양한 변화와 지식의 급격한 진보가 이뤄졌음. 한편 18세기 영국의 철학자 흄은 수학·논리학뿐만 아니라, 관측에 의한 귀납법으로도 인과 관계를 밝힐 수 없고 확실한 지식을 만들 수 없다는 불가지론을 설파하였음. 그러나 흄의 이러한 공격에도 굴하지 않고 일련의 학자들은 어떻게 데이터로부터 확실한 지식을 생성할 수 있느냐는 귀납적 방법을 부단히 탐구하여 왔음
- 17세기 파스칼은 추론의 확실성을 표현하기 위한 확률의 개념을 도입하여 간단한 계산기를 제안하였고, 18세기 베이즈는 데이터에 근거해 추론의 확실성을 개선하는 귀납 이론을, 19세기 배비지는 기계식 계산기를 설계하였으며, 20세기 현대 통계학의

아버지라고 불리는 피셔는 과학적 추론의 확실성에 대한 우도 이론을, 현대 컴퓨터 과학의 아버지라 불리는 튜링은 튜링 기계를 고안하여 알고리즘에 관한 이론적 기초를 정립하였음. 이러한 이론적 연구는 계산기가 컴퓨터로 대체되고 통계학의 귀납적 이론들이 컴퓨터 알고리즘으로 구체화되면서 급속한 발전을 이루게 됨

- 에니악은 1946년 미국서 제작된 컴퓨터로 세계 최초의 컴퓨터라고 알려져 있으나, 프로그래밍이 가능한 최초의 전자식 디지털 컴퓨터는 2차 대전 당시 베이즈 이론으로 독일의 암호를 해독한 영국의 콜로서스라고 생각됨. 그러나 이 컴퓨터의 존재는 1급 비밀이었으므로 1970년대까지 일반인들에게는 알려진 바가 없었음. 에니악은 오랫동안 최초의 프로그래밍이 가능한 컴퓨터로, 18,000개의 진공관을 이용하여 무게가 30톤에 이르렀고, 가격 또한 매우 비쌌음. 따라서 수소폭탄의 폭발 예측 등 주로 군사적 목적으로만 이용되었음. 아마 이 컴퓨터가 우리들이 사용하는 개인용 컴퓨터보다 성능이 열악하다는 것을 알면 그동안 기술 발전이 얼마나 급속도로 진행되어 왔나를 깨닫게 됨
- 한편 1945년 미국의 존 폰 노이만은 프로그램을 기억 장치에 내장하는 방식의 디지털 컴퓨터를 제안하였고, 1970년대 말부터 스티브 잡스에 의해 대형 컴퓨터가 소형화되면서 개인용 컴퓨터로 대중에 보편화됨. 여기에 인터넷과 스마트폰의 출현은 컴퓨터 기술을 폭발적으로 발전시키게 됨. 이제는 개별적으로 발전하던 기술들이 융합하면서 빅데이터, 인공지능, 데이터사이언스, 기계학습, 심층학습 등의 새로운 용어들과 더불어 구글, 페이스북, 아마존 등 새로운 형태의 기업들이 탄생하였으며, 마침내 인류의 4차 산업혁명 시대를 열게 됨
- 기존 과학 기술의 발전과 더불어 새로 개척된 빅데이터, 인공지능, 자율 주행 등의 혁신은 다른 과학, 의학, 그리고 공학 기술들과 융합을 통해 인류를 이제까지 생각지 못했던 새로운 시대로 이끌게 될 것임. 단 백년도 안 된 지난 세기에 통계학, 컴퓨터 과학 및 컴퓨터 공학, 산업 공학, 그리고 IT 기술 등은 급격한 발전을 이루었고 기술 융합을 통해 현재의 빅데이터·인공지능 경제 시대를 가능케 하였음. 자원이 데이터로 대체되고, 여기에 인공지능 또는 로봇 기술이 가미되어 새로운 산업과

서비스가 창출 되는 4차 산업혁명 시대가 되었음. 이에 다양한 분야의 학계·산업계 전공자들이 모여 다가올 시대를 위해 지혜를 모아 소고를 준비하였음

□ 다음의 II장에서는 데이터사이언스, 빅데이터, 인공지능, 기계학습, 심층학습 등의 용어들을 정의하고 데이터사이언스 과정 및 체계, 다루는 문제 분류 및 모델링에 대해 설명함. III장에서는 데이터사이언스, 빅데이터, 인공지능의 역사를 개괄함. 4차 산업혁명, 그리고 빅데이터·인공지능 경제시대에서는 대용량 데이터로부터 분석을 통해 새로운 가치를 지속적으로 창출해 내는 기술이 핵심임. 이를 수행할 수 있는, 즉 스스로 사고하고 문제를 해결하며 끊임없이 자가 발전이 가능한 학습 방법들의 개발이 중요함. 이에 IV장에서는 인간의 한계를 뛰어넘는 다양한 최근 기술들을 소개함. V장에서는 제조업, 바이오·의료, IT, 금융, 스타트업 등 각 산업 섹터별로, 데이터 솔루션, 컨설팅, 서비스, 인프라 등이 빅데이터·인공지능 기술을 통해서 얼마나 첨단화되고 있는지에 대한 진행 상황을 개괄함. VI장에서는 데이터 사이언스 발전을 통한 빅데이터·인공지능 기술 제고를 위해 교육, R&D 그리고 산업에서의 국내외 정책 및 투자 동향을 살펴보고 체계적이고 구체적인 발전 전략들을 제안해보았음

□ 최근 데이터사이언스, 빅데이터, 인공지능이 너무나도 다양한 분야에서 급격히 발전 하고 있어 소고를 읽을 때 자세히 알려고 하기보다는 대략적으로 살피면서 현재 어떤 것들이 개발되고 있는지 파악하는 것이 중요함. 지난 수세기 동안 분리된 전공들로 개발되어 왔기에 같은 개념을 다른 용어로 표현하는 경우도 있고 분리된 기술들이 일련의 과정으로 연결됨으로써 이제는 학문의 경계가 모호해졌기 때문임. 예시하면 최근 각광받는 심층학습의 경우 모형에 대한 이론적 연구는 통계학의 다양한 모형, 특히 최근 개발된 잠재 변수 모형들을 알면 더욱 깊이 이해할 수 있음. 학습을 위해 어떻게 빅데이터를 저장하고, 병렬로 빨리 계산해 신속히 예측하는 등의 알고리즘 연구는 컴퓨터 공학으로 연결됨. 그리고 이를 사용하여 어떠한 새로운 가치를 창출 하는가는 산업공학, 경영학, IT, 전기전자 산업의 영역으로 이어짐. 최종적으로는, 다양한 분야에서 개발되어 왔던 많은 방법들과 이론들에 대한 종합적·일반적인 이해를 갖추고, 실제 현실에서 새로운 제품, 서비스 그리고 새로운 산업을 개척하는 것은, 이 소고를 읽을

데이터사이언티스트들의 몫이 되리라 봄. 데이터사이언티스트가 되기 위한 발판으로서 소고를 준비했으니, 처음에 너무 구체적으로 탐구하기보다는 폭넓은 지식을 갖춘다는 마음으로 읽기를 제안함. 마지막으로, 본 소고가 한국 데이터사이언스, 빅데이터·인공지능 기술 발전에 큰 도움이 되기를 기원함



# II



## 빅데이터 · 인공지능 개요



## II

## 빅데이터 · 인공지능 개요

## 1. 데이터사이언스 · 빅데이터 · 인공지능의 정의

## □ 데이터사이언스 정의

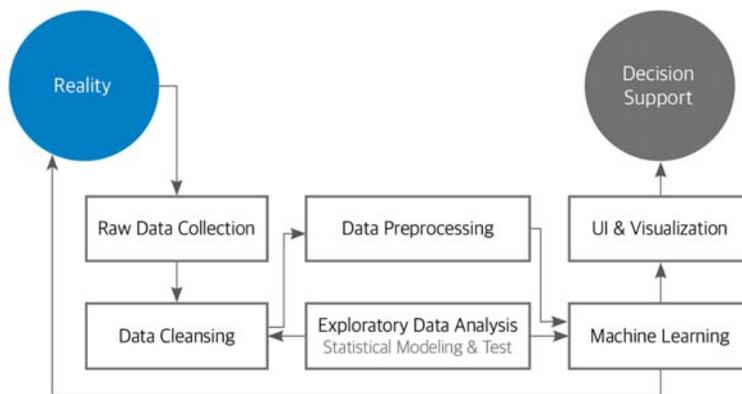
- 데이터사이언스(data science, DS)는 새로운 학문 분야로써 데이터의 수집과 저장에 필요한 데이터 프로세싱 기술과 데이터 분석에 관한 지식을 기반으로 다량의 데이터로부터 패턴을 찾아내고, 통계적 추정, 예측 모델링 등을 통하여 필요한 정보를 창출하고, 이를 실제로 활용하는 것을 연구하는 융합 과학임 (박성현 외, 2018)
- 빅데이터(big data, BD)는 기존 데이터베이스의 데이터 수집 · 저장 · 관리 · 분석의 역량을 넘어서는 구조적 및 비구조적 데이터를 포함하는 대용량의 데이터로, 의사 결정에 필요한 정보와 지식을 추출하는 관리 · 분석 기술로써, 대용량(volume), 다양성(variety), 빠른 속도(velocity)로 얻어지는 ‘3V’ 특징을 가지고 있음 (박성현 외, 2018)
- 인공지능(artificial intelligence, AI)은 인간의 인식, 판단, 추론, 문제 해결, 그 결과로서의 언어나 행동 지령, 더 나아가서는 학습 기능과 같은 인간의 두뇌 작용을 이해하는 것을 연구하는 학문 분야임 (TTA정보통신용어사전)
- 기계학습(machine learning, ML)은 인공지능의 한 분야로 컴퓨터가 데이터를 이용하여 학습한 내용을 토대로 하여 새로운 데이터에 대한 적절한 작업을 수행할 수 있게 하는 알고리즘 및 기술을 개발하는 분야를 뜻함 (TTA정보통신용어사전)
- 심층학습(deep learning, DL)은 컴퓨터가 많은 데이터를 이용하여 마치 사람 처럼 스스로 학습할 수 있게 하기 위해 인공 신경망(artificial neural network, ANN)을 기반으로 하는 기계학습의 한 기술임 (TTA정보통신용어사전)

## 2. 데이터사이언스 프로세스 및 체계도

### □ 데이터사이언스 프로세스

- 데이터사이언스는 현실의 데이터로부터 정보 분석을 통하여 의사 결정에 사용되는 실용적인 학문으로 데이터사이언스의 프로세스를 보면 <그림 2.1>과 같음

<그림 2.1> 데이터사이언스 프로세스



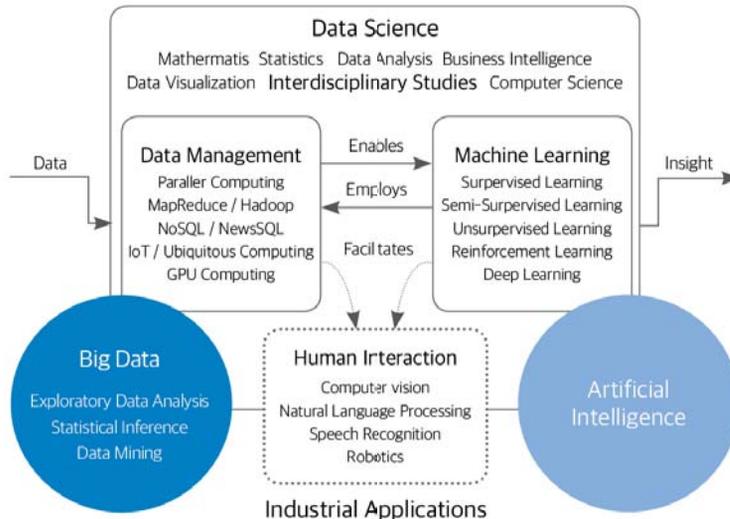
자료: O'Neil, C., & Schutt, R., 2013

- 데이터사이언스 프로세스의 시작은 현실(reality)로부터 원시 데이터(raw data)를 수집하고, 데이터를 정제(cleansing)하는 것으로 시작함
- 다음으로 데이터 전처리(preprocessing)를 통하여 초기의 데이터 셋을 구성하고, 탐색적 데이터 분석(exploratory data analysis)을 통하여 정보를 얻거나, 기계 학습을 통하여 필요한 알고리즘을 구성함
- 알고리즘의 결과물은 사용자 인터페이스(user interface, UI)로 시각화하여 사람이 보기 쉬운 보고서의 형태로 제공된 후, 의사 결정을 지원하는 데에 활용되고, 현실에 적용시키는 데에 도움을 줌
- 데이터사이언스는 <그림 2.1>의 프로세스 각 단계에 대하여 심도 있게 연구하는 학문으로, 앞으로 데이터 경제 시대에 가장 각광받는 학문으로 부상할 것임

## □ 데이터사이언스가 포함하는 넓은 의미의 전체 체계도

- 데이터사이언스는 데이터 입력으로부터 통찰력 있는 정보를 담은 결과물을 산출하기까지의 전 과정을 연구하는 학문으로서, <그림 2.2>의 상단에서 보는 바와 같이 수학(mathematics), 통계학(statistics), 데이터분석학(analytics), 컴퓨터과학(computer science), 데이터시각화(data visualization), 비즈니스 인텔리전스(business intelligence) 등을 아우르는 넓은 의미의 새로운 학문임

<그림 2.2> 데이터사이언스 전체 체계도



- 데이터사이언스는 빅데이터를 처리, 분석, 저장할 수 있도록 물리적 환경을 조성해 주는 데이터 관리(data management) 분야와 데이터로부터 모델을 만들어 인공지능으로 구현할 수 있도록 해주는 기계학습 분야로 나눌 수 있음
- 데이터 관리에 사용되는 기본적인 도구는 병렬 컴퓨팅과 맵리듀스(MapReduce)와 하둡(Hadoop), GPU 컴퓨팅 기술 등을 포함함. 한편 기계학습은 학습하는 방법과 분석 모델의 특성에 따라 지도학습(supervised learning), 비지도학습(unsupervised learning), 준지도학습(semi-supervised learning), 강화학습(reinforcement learning)이 있음. 최근에는 딥러닝(deep learning)을 따로 구분하기도 함

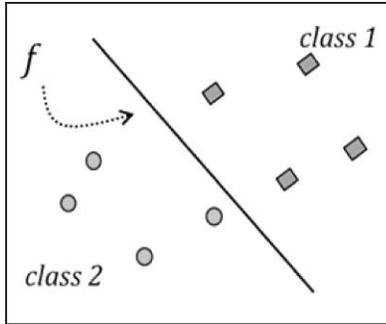
- 빅데이터 분석에는 탐색적 데이터 분석(exploratory data analysis), 통계적 추론(statistical inference), 데이터 마이닝(data mining) 등 통계학의 분석 기법들이 주를 이룸
- 통계학 및 기계학습 분석 기법들은 동일한 문제에 대하여 서로 다른 모델로 접근하는 경우도 있지만 공유하는 모델도 상당히 많음. 모델의 분석 결과물들은 다양한 응용을 통하여 인공지능 기술로 이어짐
- 데이터사이언스 프로세스의 모든 단계에는 사람이 개입하여 조절, 통제할 수 있음(human interaction). 이 때 사람과의 소통을 용이하게 할 수 있도록 해주는 인공지능 기술로 컴퓨터 비전(computer vision), 자연어 처리(natural language processing), 음성 인식(speech recognition), 로봇틱스(robotics) 등이 활용되기도 함
- 데이터사이언스를 편의상 여러 학문 분야 모듈들로 분해하였으나, 때로는 모듈화의 의미가 모호해질 만큼 학문 분야 간의 관계는 밀접함. 즉 데이터사이언스는 이러한 학문 및 기술 등의 유기적인 집합체라 할 수 있음. 이러한 다학제적 학문 분야(interdisciplinary studies)의 기술들은 다양한 산업에 적용(industrial applications)되어 새로운 제품 및 서비스를 창출하게 됨

### 3. 데이터사이언스에서 다루는 문제 분류 및 모델링

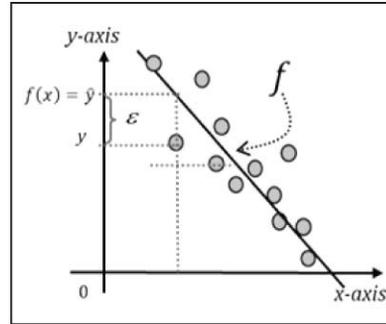
#### □ 문제 분류

- 데이터사이언스에서 다루는 문제들을 기술적으로 분류하면 대략 예측(prediction), 기술(description), 데이터 통합(data integration) 또는 데이터 퓨전(data fusion) 등으로 크게 나누어 볼 수 있음
  - 예측은 미래에 일어날 일을 가늠해 보는 것으로 분류예측(classification)과 회귀예측(regression)으로 구분할 수 있음. 예를 들어 주가가 내일 상승할 것인지 하락할 것인지를 알고자 한다면 분류를 수행하고자 한다는 것을 의미하고, 반면 주가가 내일 얼마가 될 것인지를 알고자 한다면 이는 예측을 수행하고자 한다는 것을 의미함

## &lt;그림 2.3&gt; 분류예측 및 회귀예측



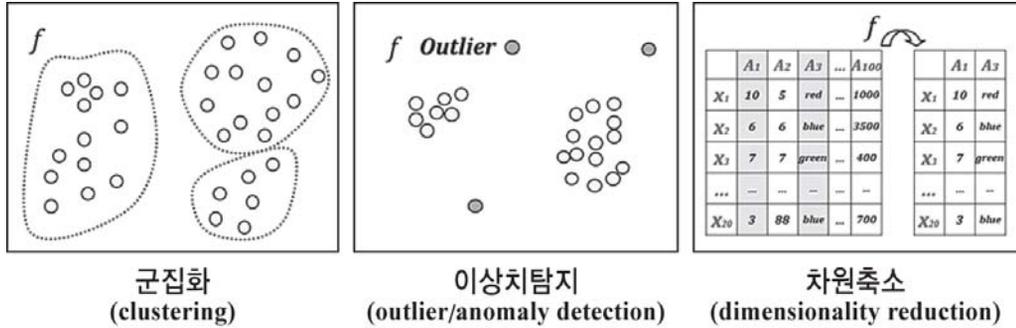
분류  
(classification)



회귀  
(regression)

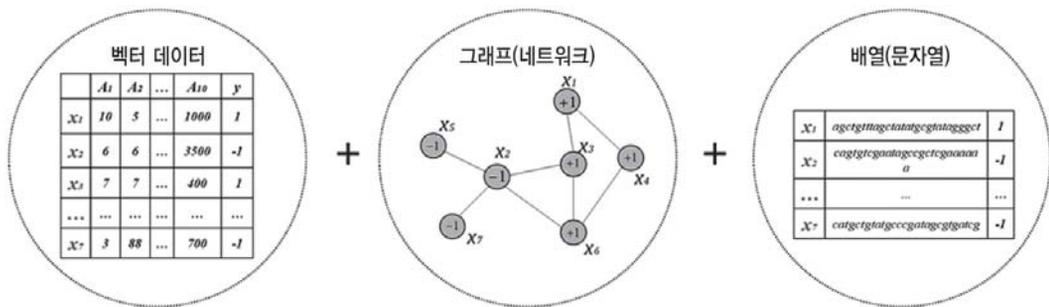
- 기술(description)은 미래를 예측한다기보다는 갖고 있는 데이터의 생김새를 파악하고 잘 설명할 수 있도록 서술하고 묘사하는 것이 목적임. 대표적으로 군집화(clustering), 이상치 탐지(outlier/anomaly detection), 차원 축소(dimensionality reduction)로 나뉘 볼 수 있음. 군집화는 데이터를 서로 비슷한 성격을 갖는 것들끼리 그룹화 하는 방법이며, 이상치 탐지는 일반적이라고 보기 어려운 의심스러운 데이터들을 찾아내는 방법임. 예를 들어 카드사나 은행에서 고객을 세분화 하여 맞춤 마케팅을 하고자 한다면 군집화를 수행할 것이고, 도용에 의한 거래나 금융 사기를 파악하고자 한다면 이상치 탐지를 수행하게 됨. 한편, 차원 축소에서 차원이란 입력 변수 또는 입력 속성의 수를 의미하는데 이를 줄이는 작업을 차원 축소라 함. 차원의 수가 줄어들면 데이터를 묘사하기가 수월해짐. 한편 기술적으로는 입력 변수의 수가 너무 많아지면 모델의 학습이 잘 이루어 지지 않는데, 이를 차원의 저주(curse of dimensionality)라 일컬을 정도로 통계 및 기계학습 분야에서는 비일비재한 문제임. 차원 축소의 방법으로는 많은 입력 변수들로부터 중요한 몇 개의 변수들을 선별해 내거나 혹은 입력 변수들이 갖고 있는 중요한 정보를 추출하여 몇 개의 변수들로 만들어 내는 방법이 있음

<그림 2.4> 데이터 기술을 위한 군집화와 이상치 탐지 및 차원 축소



- 데이터 통합(data integration fusion)이란 여러 이질적인 형식을 가진 데이터들을 함께 모델링에 사용할 수 있는 형태로 바꾸는 작업을 의미함. 빅데이터에는 비교적 분석이 용이한 정형 데이터 외에도 다양한 형태의 비정형 데이터가 혼재되어 있음. 그래프 혹은 네트워크(graph or network), 문자열(sequence), 이미지(image), 시계열(time series), 신호(signal), 웹로그 데이터(web-log data), 거래 데이터(transaction data) 등 다양한 형식의 데이터들을 분석 가능한 형태로 표현(representation)하여 통합 분석하면, 개별 데이터가 주는 정보보다 더 많은 정보를 얻을 수 있음

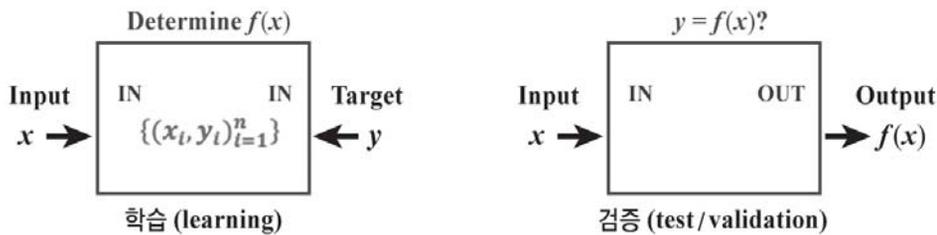
<그림 2.5> 다양한 형태의 데이터 통합



## □ 분석 모델링

- 수행하고자 하는 문제가 기술적으로 정해지면 적절한 모델을 선택하게 됨. 모델을 만드는 과정은, 데이터를 학습(learning) 또는 훈련(training)하는 단계와 만들어진 모델을 검증하기 위하여 학습에 사용되지 않은 데이터를 모델에 입력하고 결과값을 실제값과 비교해 보는 검증(test/validation)과정으로 이루어져 있음
  - <그림 2.6>에서 보면 학습 과정에는 입력값(input,  $x$ )과 목표값 또는 레이블(target or label,  $y$ )이 주어져서 모델  $f$ 를 결정하고 검증 과정에서는 모델 출력값(output,  $f(x)$ )과 목표값을 비교하여 모델의 성능을 평가함. 물론 모델 출력값과 목표값이 유사할수록 모델이 잘 학습된 것으로 판단함

<그림 2.6> 기계학습 모델링 과정



- 학습 방법에 따라 모델들을 크게 지도학습, 비지도학습, 준지도학습, 강화학습 및 심층학습(또는 딥러닝)으로 구분함
  - 지도학습(supervised learning): 입력값( $x$ )과 함께 레이블( $y$ )이 학습 데이터 셋  $\{(x_i, y_i)_{i=1}^n\}$ 으로 주어지며 모델  $f$ 는 입력과 레이블 간의 관계를 학습하여 분류 또는 예측을 수행함
  - 비지도학습(unsupervised learning): 데이터 셋에 레이블이 없음  $\{x_{i=1}^n\}$ . 데이터를 기술하는 용도로 주로 활용되는데 대표적으로 군집화, 이상치 탐지 및 차원 축소 등이 있음

- 준지도학습(semi-supervised learning): 레이블이 있는 데이터와 레이블이 없는 데이터를 동시에 활용하는 학습 방법임. 빅데이터는 레이블이 없는 데이터가 대부분임. 실패 없이 쏟아지는 입력 데이터에 레이블링을 한다는 것은 시간, 비용, 노력의 소요가 상당히 큼. 따라서 지도학습에서처럼 레이블이 있는 데이터를 학습하되, 레이블이 없는 데이터를 버리지 않고 비지도학습에서처럼 입력 데이터를 파악하고 기술하는 데에 활용하는 전략을 취함
- 강화학습(reinforcement learning): 입력이 환경(environment)으로 주어지며 모델이 취하는 행동(action)에 대한 보상(reward)으로 학습이 진행됨. 모델은 보상을 최대화 하는 방향으로 학습함. 주로 로봇 기술에 많이 활용됨
- 심층학습(deep learning): 딥러닝이라고 불리며 일종의 신경망 계열의 기계학습 모델임. 다만 방대한 학습 데이터의 규모 및 학습 시간, 비선형 조합에 의한 모델 복잡도 수준, 높은 추상화 기술, 뛰어난 성능 등이 강조되면서 따로 분류되어 다루어짐

## 참 고 문 헌

---

박성현·박태성·이영조(2018). 빅데이터와 데이터 과학, 자유아카데미.  
한국정보통신기술협회 TTA 정보통신용어사전. <http://terms.tta.or.kr>



# 빅데이터 · 인공지능 역사



## III

## 빅데이터·인공지능 역사

## 1. 데이터사이언스·빅데이터의 역사

## □ 데이터사이언스의 역사

- 데이터 기반으로 정보와 지식을 추출하는 기술은 통계학, 문헌정보학, 전산학 등의 학문에서 오래전부터 연구되어 왔으나, ‘데이터사이언스’라는 용어가 사용된 지는 불과 40년 정도밖에 되지 않았음
- Naur(1974)는 “Concise Survey of Computer Methods”라는 책에서 여러 응용에서 사용되고 있는 현대 데이터 처리 기법을 정리하면서 ‘데이터사이언스’라는 용어를 처음으로 사용하였음
- 1977년 International Association for Statistical Computing(IASC)이 International Statistical Institute의 한 부서로 설립되어, 전통적인 통계 기법, 최신 컴퓨터 기술, 도메인 전문가의 지식을 결합하여 데이터를 정보와 지식으로 변환하려는 비전을 수립하였음
- 1989년 Gregory Piatetsky-Shapiro는 제1회 Knowledge Discovery in Databases (KDD) 워크숍을 조직하여 개최하였고, 이는 1995년에 현재 최고 권위의 데이터 사이언스 국제학술대회인 ACM SIGKDD Conference and Knowledge Discovery and Data Mining(KDD)가 됨
- 1992년 제2회 일본-프랑스 통계 심포지엄 Montpellier 2 University에서 데이터 사이언스의 현대적인 정의가 논의되었으며, 다양한 원천, 형태, 구조의 데이터를 다루는 새로운 독립적인 분야의 등장을 폭넓게 인식하게 됨

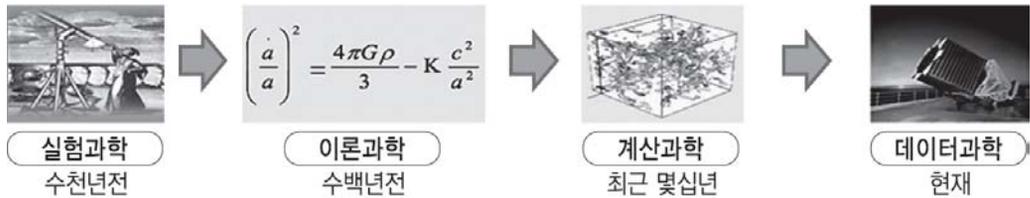
- 1996년 일본 고베에서 개최된 International Federation of Classification Societies(IFCS) 회의에서 “Data Science, Classification and Related Methods”를 발표했고, 데이터사이언스는 <그림 3.1>과 같이 학술대회의 제목으로 처음 사용됨

<그림 3.1> IFCS-96 학술대회 자료



- Cleveland(2001)는 자신의 논문 “Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics”에서 데이터사이언스는 통계 분야에서 데이터를 다루는 컴퓨팅 기술을 결합한 확장이라고 소개하며 독립적인 학문 분야로써 확고하게 언급하였음
- 2002년 4월 Data Science Journal이, 2003년 1월 Journal of Data Science가 설립되어 연구 결과들이 활발히 공유되기 시작하였음
- 2002년에는 튜링상(Turing Award) 수상자인 Jim Gray가 <그림 3.2>와 같이 “Data-Driven Science”를 과학의 4번째 패러다임이라고 제시함 (Gray & Szalay, 2002)
  - 즉 대량의 데이터를 컴퓨터로 분석하는 기법을 현대 과학의 주요 수단으로 보았음

### <그림 3.2> 과학의 패러다임 변화



자료: Gray, J. & Szalay, A., The World Wide Telescope: An Archetype for Online Science, 2002

- 데이터사이언티스트(data scientist)가 새로운 직업으로 등장함
  - 데이터사이언스가 새로운 학문 분야로 자리 잡음에 따라 이를 연구하는 학자와 각종 산업 분야에서 데이터사이언스를 구현하는 엔지니어들을 데이터사이언티스트로 칭하며 새로운 직업으로 인식하고 있었음
  - 1997년 11월 Jeff Wu 교수는 미시간 주립대학의 H. C. Carver 교수 취임 연설에서 “Statistics = Data Science?”라는 제목으로 강연을 하였으며, 통계의 핵심 요소는 데이터 수집, 데이터 모델링 및 분석, 의사 결정이라고 주장했고, 통계는 데이터사이언스로 통계학자는 ‘데이터사이언티스트(data scientist)’로 변경되어야 한다고 주장함
  - 2005년 National Science Board는 “Long-lived Digital Data Collections: Enabling Research and Education in the 21st Century”라는 원고에서 데이터사이언티스트의 정의를 더 구체적으로 제시함. 데이터사이언티스트는 디지털 데이터를 효과적으로 관리하는 데에 중요한 전산학자, 데이터베이스/소프트웨어 엔지니어/프로그래머, 도메인 전문가 등으로 정의함 (Simberloff *et al.*, 2005)
  - 하버드 비즈니스 리뷰(Harvard Business Review)는 데이터사이언티스트가 새로운 직업군으로 매우 유용한 직무를 수행할 수 있는 반면에, 그 수가 부족하여 여러 산업군에 걸쳐 인력난이 문제로 대두될 것이라고 주장함 (Davenport & Patil, 2012)
- 2013년 IEEE는 데이터사이언스 및 고급 분석에 관한 태스크 포스를 발족시킴. 그리고 2014년에 “International Conference on Data Science and Analytics(IEEE)” 국제학술대회를 시작함

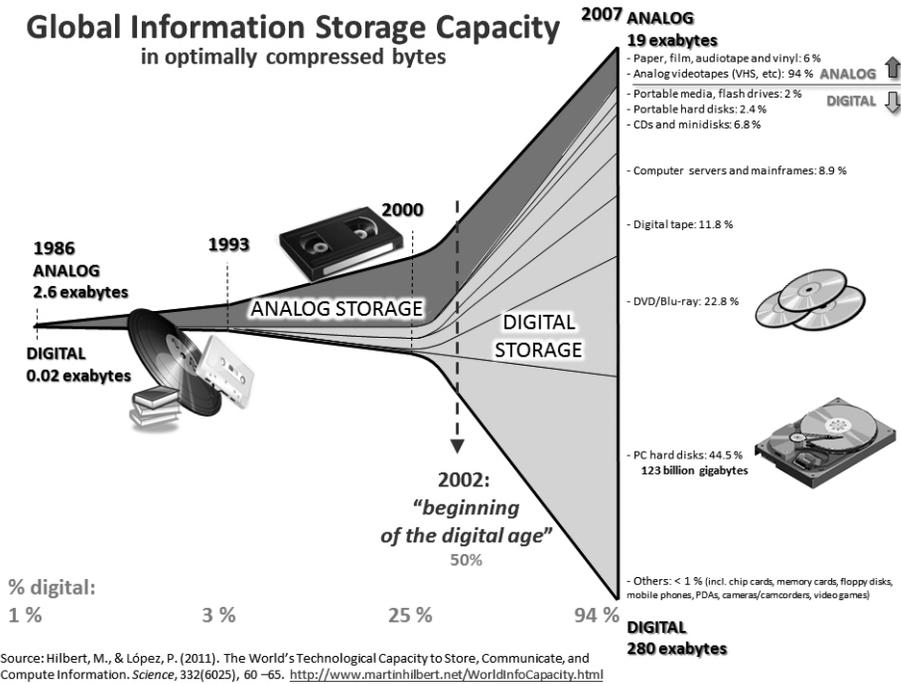
- 2014년 American Statistical Association은 Statistical Learning and Data Mining 부서를 Statistical Learning and Data Science로 개명함
- 2015년 Springer는 International Journal on Data Science and Data Analytics를 설립하여 데이터사이언스와 빅데이터 분석에 관한 연구 결과를 활발하게 출간하기 시작함

#### □ 빅데이터의 역사

- 1974년 대규모 데이터를 위한 데이터베이스 기술을 공유하기 위한 “Very Large Data Bases(VLDB)” 국제학술대회가 설립됨
- 1989년 베스트셀러 작가인 Erik Larson은 Harper’s Magazine에 기고한 글에서 자신이 받는 스팸 메일을 빅데이터 때문으로 보았으며, 이 글은 ‘빅데이터’라는 용어를 지금의 의미로 처음 사용한 것으로 인식되고 있음
  - 다음과 같은 두 문장을 적었음: “The keepers of big data say they do it for the consumer’s benefit. But data have a way of being used for purposes other than originally intended.”
- 1991년 8월 Sir Tim Berners-Lee는 최초의 웹 사이트를 CERN 내에 구축하고 이를 발표함. 추후 이 월드와이드웹(world wide web, WWW)은 빅데이터 생산과 공유의 주요 플랫폼이 됨
- 1996년에 이르러 디지털 저장 장치에 데이터를 저장하는 비용이 종이에 기록하는 비용보다 더 저렴해졌음 (Morris & Truskowski, 2003)
- Bryson, Kenwright, Cox, Ellsworth & Haimes(1999)가 Communications of the ACM(CACM)에 출간한 “Visually Exploring Gigabyte Data Sets in Real Time” 논문은 ‘빅데이터’라는 용어를 사용한 최초의 CACM 논문임
  - “Big Data for Scientific Visualization”을 장(章)제목으로 사용하였음
- 2004년 구글은 도서관을 디지털화하여 인터넷으로 열람할 수 있는 구글 북스 서비스를 발표함
  - 주요 학술 도서관에 소장된 2,000만 권 이상의 책을 디지털로 만들었음

- 2005년 웹 2.0의 시대 (O'Reilly, 2005)가 열려 대부분의 콘텐츠가 사용자에게 의해 제작, 공급되기 시작하면서 본격적인 빅데이터 시대가 열렸다고 볼 수 있음
  - 페이스북은 그 전해인 2004년에 서비스를 시작하여 2005년 당시 약 550만 명의 가입자를 보유하고 있었음
- 2006년 4월 대표적인 빅데이터 분산 병렬 처리 프레임워크인 아파치 하둡(Hadoop)의 0.1 버전이 발표되었음
- 2007년 3월 IDC의 Reinsel(2007)은 “The Expanding Digital Universe: A Forecast of Worldwide Information Growth through 2010”이라는 보고서에서 처음으로 디지털 데이터의 증가량을 구체적인 수치로 예측하였음
  - 이 보고서에서 2006년 전 세계 디지털 데이터의 양은 161EB이며, 2006년부터 2010년까지 6배가 증가하여 988EB가 될 것이라고 예측함
  - 2010년과 2012년에 출간된 같은 보고서에 따르면 디지털 데이터의 양은 예상보다 훨씬 더 증가하여 2010년에 1,227EB에 이르렀고 2012년에는 2,837EB까지 증가하였음
- 2008년 9월 네이처(Nature)는 빅데이터에 관한 특별호를 편찬하여 빅데이터가 현대 과학에서 갖는 의미를 살펴보았음
- 구글은 2009년부터 미국에서 자율 주행을 테스트하며 엄청난 양의 빅데이터를 축적하고 있음
  - 2016년 11월 기준 구글의 자율 주행 자동차 프로젝트 월간 보고서에 따르면 약 234만 마일에 대한 데이터가 축적됨
- Hilbert & Lopez(2011)은 사이언스(Science)에 실린 “The World’s Technological Capacity to Store, Communicate, and Compute Information”이라는 논문에서 <그림 3.3>과 같이 정보의 디지털화 및 저장 용량의 추세를 설명하였음

<그림 3.3> 정보의 디지털화 및 저장 용량 추세

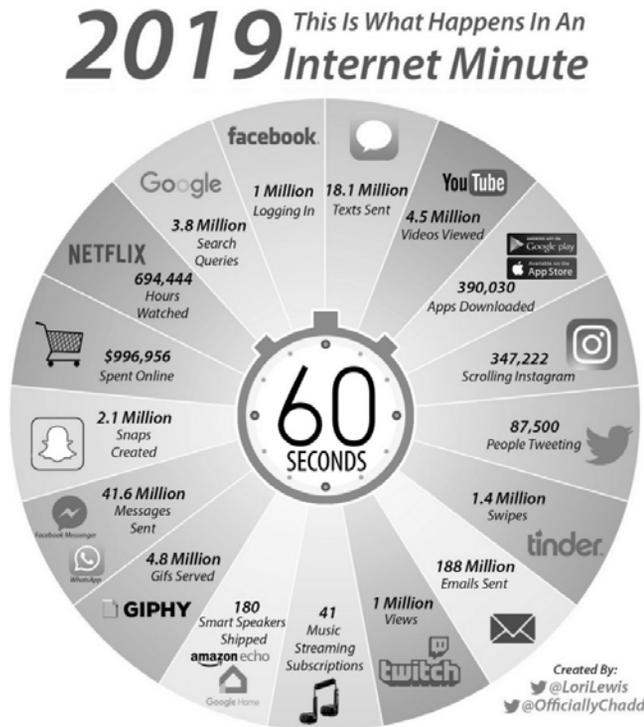


자료: Hilbert, M. & López, P., The World's Technological Capacity to Store, Communicate, and Compute Information, 2011

- 2002년부터 디지털로 저장된 정보의 양이 아날로그로 저장된 정보의 양을 추월하기 시작하였음
  - 2007년 당시 디지털로 저장된 정보의 양은 약 280EB로 전체 정보의 약 94%를 차지하는 것으로 추정되었음
  - 저장장치 중에서는 하드디스크가 44.5%로 가장 큰 비중을 차지하였음
- Cisco의 2016년 최근 보고서에 따르면 2030년까지 5천억 개의 장치(device)가 사물인터넷(Internet-of-Things)에 연결될 것으로 추산함 (Cisco, 2016)
- 각 장치는 센서를 내장하고 있어 데이터를 수집하고, 주변 환경과 상호 작용하고, 네트워크를 통해 서로 통신할 수 있음
- 인터넷은 빅데이터의 주요 원천 중의 하나이며 2019년 현재 인터넷에서 1분 동안 발생하는 데이터의 양은 <그림 3.4>와 같음

- 예를 들어 구글에서는 약 3천 8백만 건의 검색이 수행되며, 약 1.9억 건의 이메일이 전송되고 있음

<그림 3.4> 인터넷에서 1분 동안 발생하는 데이터 규모



자료: <https://www.visualcapitalist.com/what-happens-in-an-internet-minute-in-2019>

## 2. 인공지능·기계학습의 역사

### □ 인공지능의 탄생(1952~1956년)

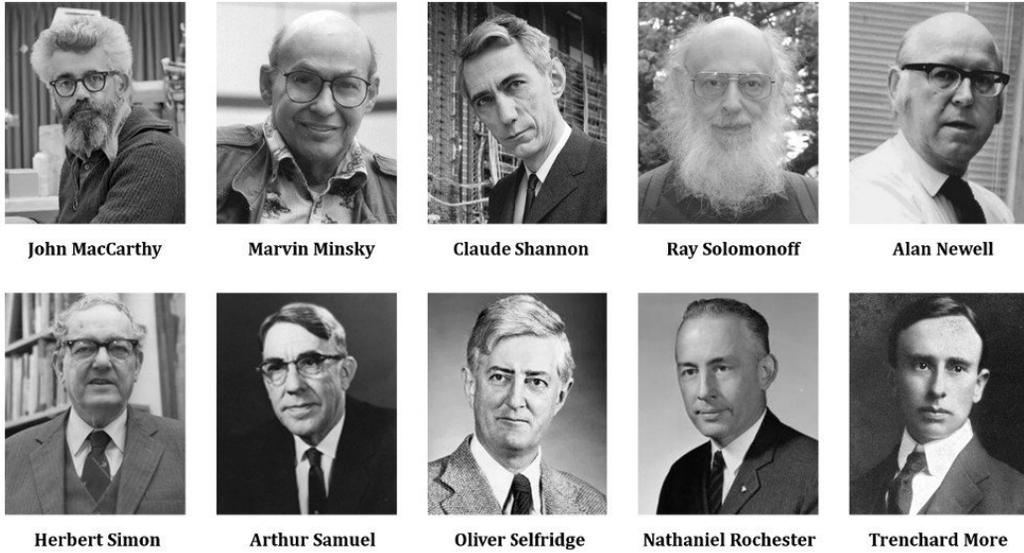
- 1940년대 후반과 1950년대 초반에 이르러서 수학, 철학, 공학, 경제학 등 다양한 영역에서 인공적인 두뇌의 가능성이 논의되었고, 1956년에 이르러 인공지능이 학문 분야로 들어섰음

- 튜링테스트(1950년): 앨런 튜링은 생각하는 기계의 구현 가능성에 대한 분석이 담긴 인공지능 역사의 시초가 되는 논문을 발표하면서 ‘생각’을 정의하기 어렵다는 점에 주목하여 튜링테스트를 고안함 (Turing, A., 1950)
  - 텔레프린터를 통한 대화에서 기계가 사람인지 기계인지 구별할 수 없을 정도로 대화를 잘 이끌어 간다면, 이것은 기계가 ‘생각’하고 있다는 충분한 근거가 된다고 제시하였음
- 다투머스 컨퍼런스(1956년): 존 매카시, 마빈 민스키, 앨런 뉴웰 그리고 허버트 사이먼 등 인공지능의 아버지라고 불리는 학자들이 모여 다투머스 대학에서 콘퍼런스를 개최함. “학습의 모든 면 또는 지능의 다른 모든 특성을 기계가 정밀하게 기술할 수 있고 이를 시뮬레이션 할 수 있다.”라는 주장을 제기함 (McCarthy, J. *et al.*, 1956)
  - 콘퍼런스에서 논리 이론을 소개하고, 이와 관련된 연구 분야를 인공지능으로 칭함으로써 명칭과 목표점, 첫 번째 성공과 이를 이룬 사람들, 그리고 넓은 의미의 인공지능 탄생을 알림

#### □ 제1차 인공지능 호황기(1956~1974년)

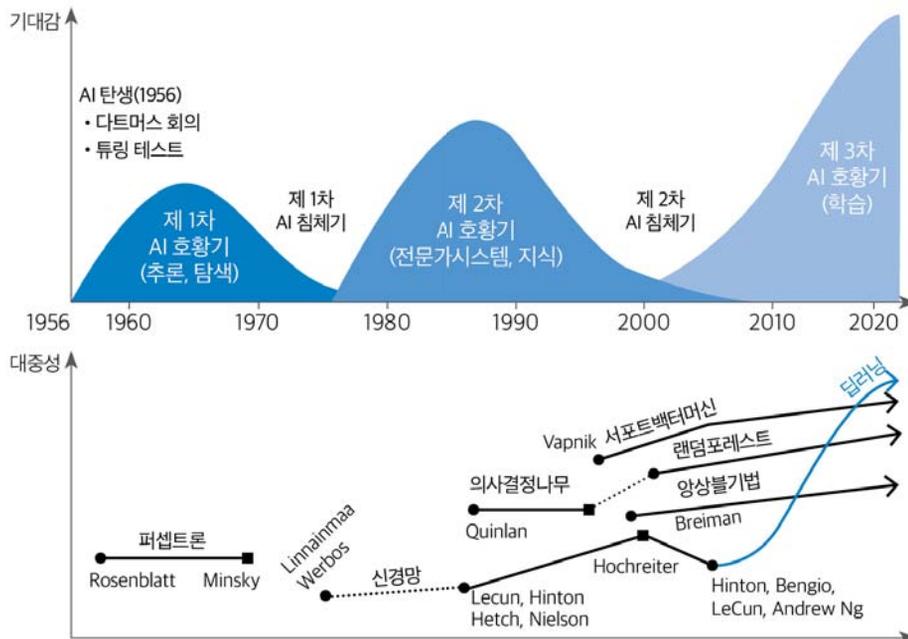
- 다투머스 콘퍼런스 이후, 인공지능이라는 새로운 영역에서 만들어진 프로그램은 대수학 문제를 풀었고, 기하학의 정리 증명 및 영어 학습을 이루어내며 많은 사람들을 놀라게 함
- 인공지능에 대한 기대감이 증폭되면서 일부 연구자들은 기계의 ‘지능적’ 행동으로 모든 것이 가능할 것이라 주장했고 연구자들은 낙관론을 펼치며 완전한 지능을 갖춘 기계가 20년 안에 탄생할 것이라고 예측함. 현재 미국 방위고등연구계획국(defense advanced research projects agency, DARPA)의 전신인 ARPA는 인공지능 및 관련 분야에 막대한 자금을 투입하게 됨

<그림 3.5> 다트머스 콘퍼런스를 개최한 인공지능의 아버지들



자료: <https://medium.com/rla-academy/dartmouth-workshop-the-birthplace-of-ai-34c533afe992>

<그림 3.6> 시기별 인공지능에 대한 기대감과 기계학습의 대중성



## ○ 탐색 추리

- 초기 인공지능 프로그램들은 기본적으로 동일한 알고리즘을 사용했고, 게임의 승리나 정리 증명 같은 목표 달성을 위해 한걸음씩 나아가는 step-by-step 방식을 사용함
- 한계점은 경우의 수가 천문학적으로 많았다는 것이었고 이를 해결하기 위해 추론이나 경험적으로 찾아낸 규칙으로 정답이 아닌 듯 보이는 경로를 지우는 방식을 사용함. 뉴엘과 사미언은 ‘범용 문제 해결기(general problem solver)’와 작동 알고리즘을 개발하고자 노력함

## ○ 자연어 처리

- 인공지능 연구의 중요한 목표는 인간의 언어로 컴퓨터와 의사소통 할 수 있게 하는 것이었으며, STUDENT라는 프로그램으로 고등학교 수준의 문제를 푸는 성공을 보였음 (Bobrow, D. G., 1964)

## □ 제1차 인공지능 침체기(1974~1980년)

- 1970년대에 이르자, 인공지능은 복잡한 문제를 해결하는 데에 거듭 실패했고 엄청난 낙관론에 비해 연구 성과가 미흡하여 자금 투자도 줄어들게 됨
- 동시에, 연결주의(connectionism) 또는 인공 신경망은 지난 10년 동안 마빈 민스키의 퍼셉트론(시각과 뇌의 기능을 모델화한 학습 기계, 지금의 신경망)에 대한 파괴적인 비판에 의해 완전히 중지되었음
- 1970년대 후반, 인공지능에 대한 긍정적이지 않은 대중의 인식에도 불구하고, 논리 프로그래밍, 상징 추론 등 여러 영역에서의 새로운 아이디어가 나타나게 됨
- 초기 인공지능의 한계점
  - 컴퓨터 능력의 한계: 메모리와 처리 속도가 충분치 않았음. 이 때문에 로스 켈리언의 자연어 처리 프로그램은 한번에 20개 단어만 처리할 수 있었음
  - 방법의 비효율성: 문제 해답을 찾는 데에 활용되었던 step-by-step 방법은 문제가 커질수록 지수적 시간이 요구됨. 간단한 문제에서도 최적의 해답을 찾는 데에는 막대한 시간이 소요되었음

- 데이터 부족: 영상 처리나 자연어 처리에서 학습을 위해서는 충분한 데이터가 필요하지만, 1970년대에는 방대한 양의 데이터베이스를 만들지 못했음
- 모라벡의 패러독스: 인공지능 프로그램에게 있어 이론 증명이나 기하학적 문제 해결은 비교적 쉬운 반면, 얼굴을 인식하거나 장애물을 피해 방을 가로지르는 것은 매우 어려움. 이는 인간의 직관을 컴퓨터가 갖출 수 없기 때문인데, 이로 인해 1970년대에는 영상 처리나 로봇에 대한 연구 성과가 미흡했음
- 프레임 문제: 존 맥카시와 같은 학자들은 규칙 기반 시스템에서 학습을 통한 규칙 변경이나 구조가 변화하지 않는다면 추론의 일반성을 도모할 수 없다는 것을 발견함

#### ○ 퍼셉트론과 커넥티즘의 암흑기

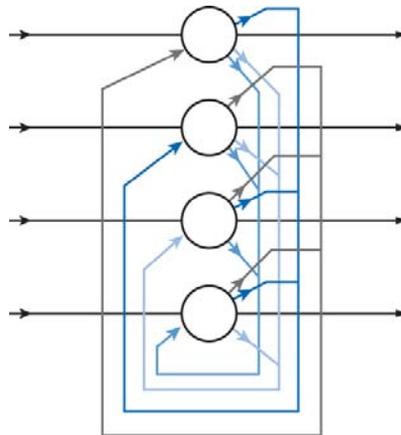
- 퍼셉트론은 1958년 마빈 민스키(Marvin Minsky)의 동료였던 프랭크 로젠블라트(Frank Rosenblatt)이 도입한 신경망의 한 형태로써, ‘퍼셉트론은 언어를 배우고 결정하며 번역하게 될 것이다’라고 예측하며 낙관론을 펼침
- 퍼셉트론 연구는 1960년대 활발히 진행되었지만 민스키와 페이퍼트(Papert)의 저서 ‘퍼셉트론(Perceptrons)’이 출간되며 갑자기 중단되었음
- 저서에는 퍼셉트론의 수행 영역에 심각한 한계가 있고 로젠블라트의 예측이 과장이었음을 암시하는 내용이 담겨있음. 이 책의 효과는 파급적이었고, 향후 10년간 커넥티즘에 대한 연구가 거의 이루어지지 않게 됨

### □ 제2차 인공지능 호황기(1980~1987년)

- 1980년대 인공지능 프로그램은 ‘전문가 시스템’이라고 일컬어지는 형태로 불려졌으며, 주로 검색에 초점이 맞춰졌음. 또한 존 홉필드와 데이비드 루멜하트가 신경망 이론을 부활시키며 두 번째 인공지능 호황기를 이끌어냄
- 전문가 시스템 활성화
  - 전문가 시스템은 특정 도메인에 대해 문제를 해결해주거나 질문에 대답해주는 프로그램이며 전문가의 지식에서 파생된 논리적 법칙을 사용함
    - 최초의 연구는 1965년 에드워드 파이겐바움이 분광계로부터 화합물을 식별하는 실험으로부터 시작되었음 (Lindsay, R. K. *et al.*, 1993)

- 가장 성공적인 전문가 시스템은 1972년에 개발된 MYCIN으로, 전염성 혈액 질환을 진단하는 데에 활용됨 (Shortliffe, E. H. & Buchanan, B. G., 1975)
  - 전문가 시스템은 일반성에 제한을 두어 특정 영역의 문제 해결에만 활용되었으며 단순한 디자인으로 프로그램 개발이 용이했음
  - 1980년 카네기 멜론 대학에서는 XCON이라 불리는 판매용 컴퓨터의 부품을 선정하고 시스템의 조립을 완성시키는 전문가 시스템을 개발하여 매년 4천만 달러를 절약시키는 매우 큰 성과를 달성했고, 전 세계 회사들은 1억 달러 이상을 인공지능 프로그램에 투자하여 전문가 시스템을 개발하면서 인공지능 하드웨어 및 소프트웨어 시장이 함께 성장함
- 신경망 이론의 부활
- 1982년 물리학자 존 홉필드는 홉필드넷을 개발하여 신경망의 학습과 정보 처리 방식에 대한 새로운 방법을 제시함 (Hopfield, J. J., 1982)

<그림 3.7> 홉필드넷 예시



자료: Hopfield, J. J., Neural networks and physical systems with emergent collective computational abilities, 1982

- 비슷한 시기에 신경망을 학습시키는 ‘역전파’ 방법 (Rumelhart, D. E. & Hinton, G. E & Williams, R. J., 1988; Werbos, P. J., 1990)이 개발되면서 1970년 이후 버려진 신경망 이론이 다시 부활함

## □ 제2차 인공지능 침체기(1987~1993년)

- 1987년 이후 인공지능에 대한 비즈니스 커뮤니티의 관심은 1980년대 버블 경제의 패턴과 같이 상승 후 하락하게 됨. 정부 기관과 투자자들의 인공지능에 대한 부푼 기대감이 꺼짐으로써 제2의 침체기가 도래하게 됨
- 인공지능 겨울(AI Winter)
  - ‘인공지능 겨울(AI Winter)’은 연구원들에 의해 만들어진 용어로, 전문가 시스템의 성과가 기대에 미치지 못하자 자금 지원이 삭감되는 상황을 은유함. 1980년대 후반과 1990년대 초반의 인공지능 연구는 재정적 어려움을 겪게 됨
  - 1987년 인공지능 하드웨어 시장은 붕괴를 겪게 됨. 애플과 IBM의 데스크탑 컴퓨터는 값싸고 좋은 성능을 보이면서 전문가 시스템을 구현하던 기존 인공지능 하드웨어를 넘어서게 됨
  - 상대적으로 XCON과 같은 초기의 성공적인 전문가 시스템은 과도한 유지 보수 비용이 소요되었고 업데이트와 학습이 어려워 범용성을 갖추지 못하면서 도태하게 됨
  - 1980년대 후반 Strategic Computing Initiative는 인공지능 자금을 대폭 삭감했고 DARPA는 인공지능이 ‘차세대 물결’이 아닐 수 있으며 투자에 대한 결과가 요원하다고 판단함. 따라서 연구 개발 자금은 즉각적인 결과를 낼 가능성이 있는 프로젝트로 투입됨
  - 300개가 넘는 인공지능 회사들은 1993년 말 폐쇄나 파산, 인수합병 되며 첫 번째 상업 물결을 스스로 종식하게 됨

## □ 제3차 인공지능 호황기(1993년~현재)

- 무어의 법칙(Moor's Law)
  - 1997년 5월 11일, 딥블루(Deep Blue)는 세계 체스 챔피언인 게리 카스파로프를 이긴 최초의 컴퓨터 체스 플레이 시스템이 되었음. IBM의 슈퍼 컴퓨터는 초기 버전보다 처리 속도를 두 배나 향상시키며 발전하게 됨

- 2005년 스탠포드 대학이 개발한 로봇이 예행 연습을 하지 않은 사막 트레일을 따라 131마일을 자율 주행하여 DARPA Grand Challenge에서 우승했고, 2년 후 카네기 멜론대학 팀은 자율 주행 차는 교통 위험 및 모든 교통 법규들을 준수 하면서 도시 환경에서 55마일을 자율 주행하여 DARPA Urban Challenge에서 우승함
  - 2011년 2월, IBM 왓슨은 제퍼디 퀴즈쇼에서 가장 위대한 챔피언 두 명과 상대 하여 우승을 차지함
  - 이와 같은 성공은 새로운 이론의 패러다임 발견 때문이 아니라 하드웨어 엔지니어링 기술 발전으로 인한 컴퓨터의 속도와 저장 용량이 대폭 증가했기 때문임
  - 이러한 급성장은 ‘컴퓨터의 속도와 메모리 용량은 2년마다 두 배가 될 것’이라는 무어의 법칙으로 일컬어짐
- 지능형 에이전트
- 1990년대에는 ‘지능형 에이전트’라고 불리는 새로운 패러다임이 다방면에서 활용되고 있었음. 인공지능 연구에 결정론과 경제성이라는 개념을 도입함으로써 경제학자들의 합리적 에이전트라는 정의와 컴퓨터 과학자들의 객체 혹은 모듈러 정의를 융합하여 지능형 에이전트의 패러다임을 완성함
- 인공지능 연구는 더욱 정교한 수학적 도구를 사용하게 되었고 수학, 경제학 등에서 많은 연구자들이 인공지능을 활용하게 됨. 수학적 방법을 널리 공유하고 기술 향상을 성취하게 됨으로써 인공지능은 더 엄격한 과학 학문으로 자리 잡게 됨
- 인공지능 알고리즘들은 거대한 시스템의 일부로 자리 잡으며 매우 어려운 문제 들을 해결하기 시작함. 데이터 마이닝, 산업공학에 의한 다양한 산업으로의 응용, 로봇공학, 논리학, 음성 인식, 은행 소프트웨어, 의학적 진단, 검색 엔진 등 다방면에서 유용성이 입증됨
- 21세기 초 대량의 데이터에 대한 접근, 고속 컴퓨팅과 정교한 기계학습 기법들이 경제 전반에 걸쳐 많은 문제에 성공적으로 적용됨. 맥킨지는 ‘미국 경제의 거의 모든 부문에서 평균 200TB의 저장된 데이터를 보유’라고 추정함 (Manyika, J., 2011)

- 2016년까지 인공지능 관련 제품, 하드웨어 및 소프트웨어 시장은 80억 달러 이상에 달했으며, 뉴욕타임즈는 인공지능에 대한 관심이 ‘광란’에 이르렀다고 보도함. 빅데이터의 응용 프로그램은 생태학의 훈련 모델 및 경제학의 다양한 응용 프로그램과 같은 다른 분야에도 적용되기 시작함
- 심층학습, 특히 컨볼루션 신경망 및 재귀 신경망의 발전은 이미지 및 비디오 처리, 텍스트 분석 및 음성 인식에 대한 연구를 촉진시킴

## 참 고 문 헌

---

- Bobrow, D. G.(1964). “Natural language input for a computer problem solving system”.
- Bryson, S., Kenwright, D., Cox, M., Ellsworth, D. & Haines, R.(1999). “Visually Exploring Gigabyte Data Sets in Real Time”, Communications of the ACM, Vol.42, No.8, pp. 82~90.
- Cisco(2016). Internet of Things at a Glance, White Paper, Cisco, <https://www.cisco.com/c/dam/en/us/products/collateral/se/internet-of-things/at-a-glance-c45-731471.pdf>.
- Cleveland, W. S.(2001). “Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics”, International Statistical Review, Vol.69, No.1, pp. 21~26.
- Davenport, T. H. & Patil, D. J.(2012). “Data Scientist: The Sexiest Job of the 21st Century,” Harvard Business Review, Vol.90, No.5, pp. 70~76.
- Gray, J. & Szalay, A.(2002). “The World Wide Telescope: An Archetype for Online Science”, Communications of the ACM, Vol.45, No.11, pp. 50~54.
- Hilbert, M. & López, P.(2011). “The World’s Technological Capacity to Store, Communicate, and Compute Information”, Science, Vol.332, No.6025, pp. 60~65.
- Hopfield, J. J.(1982). “Neural networks and physical systems with emergent collective computational abilities.” Proceedings of the national academy of sciences 79.8: 2554~2558.
- Lindsay, R. K. *et al.*(1993). “DENDRAL: a case study of the first expert system for scientific hypothesis formation.” Artificial intelligence 61.2: 209~261.
- Manyika, J.(2011). “Big data: The next frontier for innovation, competition, and productivity.” [http://www.mckinsey.com/Insights/MGI/Research/Technology\\_and\\_Innovation/Big\\_data\\_The\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation).

- McCarthy, J. *et al.*(1956). “Dartmouth conference.” Dartmouth Summer Research Conference on Artificial Intelligence.
- Morris, R. J. T. & Truskowski, B. J.(2003). “The Evolution of Storage Systems”, Vol.42, No.2, pp. 205~217.
- Naur, P.(1974). Concise Survey of Computer Methods. Petrocelli Books.
- O’Reilly, T.(2005). What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software, O’Reilly,  
<https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>.
- Reinsel, D.(2007). The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010, International Data Corporation.
- Rumelhart, D. E. & Hinton, G. E. & Williams, R. J.(1988). “Learning representations by back-propagating errors.” Cognitive modeling 5.3: 1.
- Shortliffe, E. H. & Buchanan, B. G.(1975). “A model of inexact reasoning in medicine.” Mathematical biosciences 23.3-4: pp. 351~379.
- Simberloff, D. *et al.*(2005). “Long-lived Digital Data Collections: Enabling Research and Education in the 21st Century”, National Science Foundation.
- Turing, A.(1950). “Computing machinery and intelligence. Mind LIX (236): 433-460.” Reprinted as 40-66.
- Werbos, P. J.(1990). “Backpropagation through time: what it does and how to do it.” Proceedings of the IEEE 78.10: pp. 1550~1560.

# IV

## 빅데이터 · 인공지능 기술 수준



## IV

## 빅데이터·인공지능 기술 수준

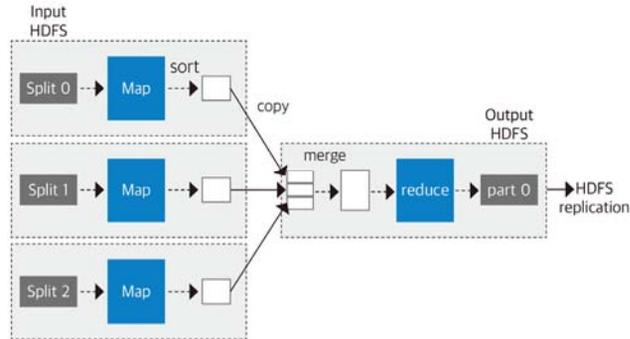
## 1. 데이터 관리 기술 수준

## □ 배치 분산 병렬 처리

## ○ 맵리듀스(MapReduce)와 하둡(Hadoop)

- 맵리듀스는 구글에서 대용량 데이터 처리를 위해 개발한 분산 병렬 컴퓨팅 프레임워크임 (Dean & Ghemawat, 2014)
  - 이 프레임워크는 일반 컴퓨터를 여러 대 연결하여 슈퍼 컴퓨터급의 성능을 낼 수 있도록 고안되었는데, 입력 데이터 셋을 여러 개의 독립적인 조각으로 나누고 각각을 따로 동시에 처리하는 방식을 취함
  - 맵리듀스에서 프로그래머는 맵(map) 함수와 리듀스(reduce) 함수를 구현함
    - 맵 함수에서는 나누어진 데이터의 각 조각(split)에 대해 주어진 작업을 수행하여 그 결과를 산출하고, 이러한 결과들은 정렬되어 리듀스 함수로 네트워크를 통해 전달됨
    - <그림 4.1>과 같이 리듀스 함수에서는 각 맵 함수에서 얻어진 결과를 취합하여 최종 결과를 얻음

<그림 4.1> 리듀스가 하나일 때의 맵리듀스 데이터 흐름



자료: White, T., Hadoop: The Definitive Guide, 4th Edition, 2015

- 이러한 맵리듀스 개념을 기존 알고리즘에 적용하여 기존 알고리즘의 확장성을 향상시키려는 시도가 활발히 진행되고 있음
  - 애초에 구글에서 맵리듀스를 개발한 목적이 정보 검색을 위한 역색인 (inverted index) 구조를 빅데이터에 대해 빠르게 구축하기 위함이었음
  - 데이터베이스, 데이터 마이닝의 많은 주요 연산들이 맵리듀스 버전으로 개발되었음
    - 데이터베이스 조인, k-means 알고리즘과 같은 군집화 알고리즘, k-최근접 이웃 알고리즘과 같은 분류 알고리즘 등을 비롯하여 삼각형 세기(counting) 등과 같은 그래프 알고리즘, LDA(latent dirichlet allocation)와 같은 텍스트 처리 알고리즘도 맵리듀스 형태로 개발되어 널리 활용되고 있음 (Shim, 2012)
- 맵리듀스가 간단하고 직관적인 개념으로 병렬 처리를 쉽게 구현할 수 있긴 하지만 이 개념이 널리 보급되는 데에는 맵리듀스를 구현한 오픈소스 소프트웨어인 아파치 하둡의 공이 컸음
  - 2011년 12월에 Hadoop 1 버전이 공식적으로 발표되었고, 2013년 10월에 Hadoop 2 버전이 처음으로 공식 출시되었으며, 2019년 7월 현재 Hadoop 3.1.2까지 출시된 상황임
- 하둡의 공헌을 좀 더 상세히 기술하자면, 분산 병렬 처리를 위해서는 알고리즘의 로직 이외에 여러 가지 일을 신경 써야 하는데, 이를 하둡이 자동적으로 처리해 줌으로써 개발자들은 알고리즘의 로직에만 집중할 수 있게 되었음

- 예를 들어 컴퓨터가 다운되었을 경우를 대비하여 각각의 데이터 블록을 여러 대의 컴퓨터에 복사해 놓아야 한다거나, 실제로 한 컴퓨터가 다운되었을 때 이를 빠르게 알아내어 복사본을 가져와 작업을 재개하거나, 또한 각 컴퓨터 간의 통신을 관장하는 등의 부가 작업을 하둡이 전적으로 처리해줌

### ○ 스파크(Spark)

- 하둡이 분산 병렬 처리를 통해 기존 알고리즘의 성능을 대폭 향상시킨 것은 자명한 사실이지만, 하둡 또한 몇 가지 약점이 드러나기 시작하였음
  - 실패 복원을 위해 하둡은 리듀스 함수의 결과를 HDFS에 저장하고, 일반적으로 HDFS는 디스크 상에서 동작하므로, 메모리에 비해 입출력 속도가 느릴 수밖에 없음
- 이 때문에 수행하려는 작업이 한 번에 끝나는 단순한 배치 작업이면 문제가 되지 않겠지만 이전 단계의 출력을 다시 입력으로 사용하여 반복적으로 수행해야 하는 작업이라면 여러 번 HDFS 상에서의 입출력이 발생하게 되어 성능이 저하되는 문제가 발견되었음 (Lee, 2016)
- 아파치 스파크는 방향성 비순환 그래프(directed acyclic graph, DAG) 형태의 데이터 흐름과 인 메모리 컴퓨팅을 지원하는 새로운 분산 병렬 처리 플랫폼이므로, 2014년 5월에 스파크 버전 1이 발표되어 2019년 7월 기준으로 스파크 2.4.3까지 이어져오고 있음
- 이러한 스파크의 특성은 하둡에 비해 대폭적인 성능 향상을 가져왔음
  - 예를 들어, 로지스틱 회귀분석(logistic regression)을 수행할 때, 스파크는 메모리를 사용할 때 하둡에 비해 약 100배 정도 빠르며, 디스크를 사용하더라도 약 10배 정도 빠른 것으로 밝혀졌음 (Aapche, 2019a)
- 스파크의 핵심은 RDD(resilient distributed datasets)라고 볼 수 있음 (Zaharia *et al.*, 2012)
  - RDD는 클러스터에 분산되어 있는 컴퓨터에 내장된 메모리의 집합체임
    - 즉, 한 대 혹은 여러 대의 컴퓨터에 있는 메모리를 모아 특정 데이터 셋을 위한 저장소를 구성한 것임
  - 스파크에서는 RDD가 일반적으로 메모리에 생성되므로 각 단계 간에 데이터를 전송하는 비용이 크게 감소함

- 데이터 설계자를 대상으로 한 설문 조사에 따르면, 이러한 장점으로 IT 관리자와 BI 분석가의 70%가 이미 사용 중인 맵리듀스보다 스파크를 선호하였음 (Yogurtcu, 2016)

○ 그래프 빅데이터 처리 기술(Giraph, GraphLab 등)

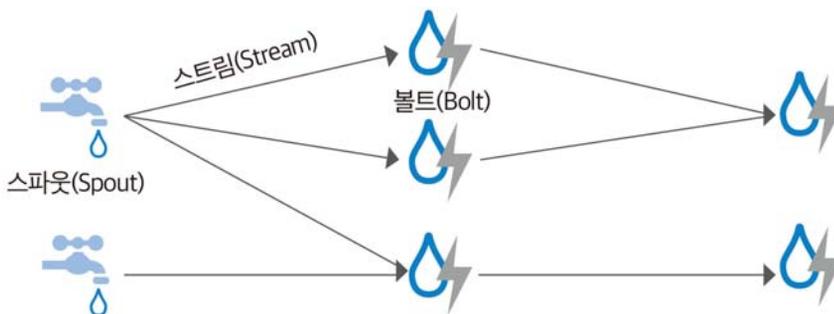
- 소셜 네트워크 서비스(social network service, SNS)의 보급에 따라 그래프 혹은 네트워크 형태의 빅데이터가 풍부해졌고, 이러한 그래프 데이터를 분석하고자 하는 시도가 생기기 시작하였음
- 그래프 빅데이터를 처리하기 위해 초창기에는 하둡을 적용하였으나 하둡이 적합하지 않음이 판명됨 (Malewicz *et al.*, 2010)
  - 그래프 데이터를 처리하기 위한 알고리즘들이 대부분 반복적임
    - 예를 들어 웹 페이지의 랭킹을 정하기 위한 알고리즘인 페이지랭크(page rank)는 각 웹 페이지의 랭크 값이 수렴할 때까지 반복함
  - 그래프 구조를 반복적으로 읽어 들여야 하는 비용이 큼
    - 예를 들어 페이지랭크 알고리즘에서 랭크 값을 계산하기 위해서는 그래프의 각 정점(vertex 혹은 node) 별로 이웃하고 있는 정점들을 알아야 하며, 이러한 그래프 구조 정보는 얻고자 하는 결과보다 그 크기가 훨씬 더 큼
- 그래프 형태의 빅데이터를 효율적으로 처리하기 위해 맵리듀스 대신에 새로운 정점 중심(vertex-centric) 패러다임이 도입됨
- 정점 중심 모델은 그래프의 구성 요소 중에서 정점을 중심으로 연산을 정의할 수 있도록 하였음 (Malewicz *et al.*, 2010)
  - 그래프 데이터는 가상적으로 그래프 형태를 그대로 유지하며 저장되고, 각 정점은 이웃 정점을 알아내 이웃 정점과는 자유롭게 통신할 수 있음
  - 맵리듀스와 달리 각 정점에 이웃 정점 정보가 이미 내재되어 있으므로 그래프 구조를 반복적으로 읽어 들일 필요가 없음
- 정점 중심 모델의 근간은 1990년에 발표된 BSP(bulk synchronous parallel) 모델임

## □ 실시간 분산 병렬 처리

### ○ 스톰(Storm)

- 앞서 설명한 기술들은 빅데이터가 수집되어 이미 저장된 정적인 데이터를 주로 처리하기 위해 개발되었으나, 사물인터넷의 발전으로 실시간으로 입력되는 동적인 스트림 빅데이터도 크게 늘어나게 되었음
  - 스트림 데이터는 끊임없이 빠른 속도로 계속 입력되는 형태의 데이터임
  - 예를 들면, 공장 장비에 있는 센서로부터 입력되는 측정값들이 스트림 데이터의 일종임
- 이러한 스트림 데이터를 위해 분산 병렬 스트림 엔진이 개발되었고 아파치 스톰(Apache, 2019b)이 가장 널리 활용되고 있음
- 아파치 스톰을 사용하기 위해서는 우선 토폴로지(topology)를 구성해야 하는데, 토폴로지란 컴퓨터 클러스터 내에서의 스트림 데이터의 흐름을 지정한 DAG임
  - 비유하자면 수도꼭지에서 흘러나온 물이 파이프를 통해 누구에게 어떻게 전달되어야 하는지를 지정하는 설계도라고 보면 되고, <그림 4.2>는 간단한 토폴로지의 예를 나타냄
  - 스파웃(spout)은 스트림 데이터의 소스를 뜻함
  - 연결선은 파이프를 보면 되고, 이를 따라 스트림 데이터가 전달됨
  - 볼트(bolt)는 스트림 처리의 핵심이 되는 모듈로서, 스트림 데이터에 적용해야 하는 연산(필터링, 그룹핑 등)이 프로그래밍 되어 있음

<그림 4.2> 스톰 토폴로지 예시



## □ 분산 NoSQL DBMS

- 전통적인 관계형 데이터베이스 관리 시스템(relational database management system, RDBMS)의 한계점이 대두됨
  - 전통적인 RDBMS는 입력 데이터가 구조화되어 테이블 형태로 깔끔하게 정리될 수 있을 때 적합함
  - 그러나 빅데이터 시대로 접어들면서 RDBMS의 몇 가지 문제점이 지적되었음 (Leavitt, 2010)
    - 확장성, 복잡도, SQL언어 적합성, 기능과다가 문제점으로 지적되었음
  - 이를 해결하기 위해 Web 2.0 회사들인 구글, 페이스북, 아마존을 중심으로 NoSQL DBMS가 개발되기 시작하였음
    - NoSQL은 애초에는 Non-SQL을 뜻하였으나 최근에는 일부 NoSQL DBMS가 SQL 유사 언어를 제공하기도 하므로 Not-Only SQL을 뜻하기도 함
- NoSQL DBMS의 등장
  - NoSQL DBMS는 기존의 RDBMS에 비해 더 간단한 디자인을 채택하였음
  - 우선 Horizontal Scale-Up이 간단한 데이터 저장 구조를 채택하여, 컴퓨터를 더 추가함으로써 쉽게 확장성을 얻도록 설계되었음
    - 테이블을 파티션함으로써 클러스터 내의 컴퓨터에 레코드들을 나누어 저장하고, 컴퓨터를 추가하면 기존의 저장된 데이터에 영향을 주지 않고서도 추가적으로 레코드들을 저장할 수 있어 쉽게 확장성을 향상시킴
    - 이 특징이 빅데이터 시대에 NoSQL DBMS가 주목을 받게 된 가장 큰 이유라고 볼 수 있음
  - 가장 대표적인 NoSQL의 저장 구조는 Key-Value 스토어임
    - 레코드 혹은 로우(row)에는 항상 키(key)가 할당되어 있고 이 키를 통해 레코드를 인출할 수 있으며, 레코드들은 키에 따라 정렬되어 있음
    - 값(value)에는 여러 컬럼들이 정의되어 저장될 수 있으며, 대부분 바이트 어레이로 저장됨
    - 일반적으로 NoSQL DBMS에서는 RDBMS와 달리 스키마를 엄밀하게 지정하지 않으므로 자유롭게 컬럼을 추가할 수 있음

- Key-Value 스토어의 일종인 Wide-Column Store에서는 수억 개의 컬럼을 갖는 것도 가능하며, 컬럼의 이름과 타입이 레코드 별로 서로 다른 것이 허용됨
- 구글에서 개발한 BigTable (Chang *et al.*, 2006)이 Wide-Column Store의 시조라고 볼 수 있으며, 가장 인기 있는 Wide-Column Store에는 HBase (Vora, 2011)와 Cassandra (Lakshman & Malik, 2010)가 있음
- 또한 Document Store도 Key-Value 스토어의 일종으로 구조화된 데이터가 아니라 문서를 저장할 수 있게 특화되어 문서의 인코딩 방법 (예: JSON (Bray, 2014))과 문서에 특화된 인덱스 구조를 지원함
  - 대표적인 Document Store 방식의 제품이 MongoDB (Kanade *et al.*, 2014)임

## □ New SQL 패러다임

### ○ No SQL DBMS의 한계점 대두

- No SQL이 높은 확장성으로 인해 빅데이터 시대에 점점 더 널리 보급되고 있지만, 역시 그에 대한 한계점들이 지적되기 시작하였음
- 많은 기업들이 재무 정보 등 오류에 상당히 민감한 데이터를 취급하고 있는데 ACID 특성이 보장되지 않는 DBMS에서는 서로 다른 값을 읽어 일관성에서 문제가 될 수 있음
- 따라서 ACID 특성을 보장하면서 빅데이터를 처리할 수 있는 시스템에 대한 필요성이 대두되었음
  - 이에 대한 기존의 해결책은 RDBMS를 성능이 매우 높은 한 대의 메인 프레임급 컴퓨터에서 운영하거나 혹은 RDBMS의 데이터를 분산하여 저장 하되 해당 응용에 특화된 미들웨어를 개발하여 운영하는 것임
  - 그러나 이러한 방식들은 모두 비용이나 효율성 면에서 뛰어나지 못함

### ○ <그림 4.3>의 제품들로 대변되는 New SQL 패러다임의 등장

&lt;그림 4.3&gt; New SQL 부류의 대표적인 제품들



- New SQL의 내부 구조는 개발사의 정책에 따라 매우 상이하나 크게 3가지 공통점이 있음
  - 관계형 데이터 모델을 택함
  - SQL 인터페이스를 제공함
    - SQL 인터페이스와 같은 High-Level 인터페이스가 존재하지 않으면 프로그래밍이 어려워져서 SQL 인터페이스가 다시 도입되었다고 사료됨
  - 무공유(shared nothing)구조에서 초병렬(massively parallel) 처리를 지원함
- New SQL의 개발 방식에는 크게 2가지가 있다고 볼 수 있음
  - 첫 번째는 완전히 새롭게 처음부터 개발하는 방식임
    - 무공유 구조에서 동작하도록 설계되었으며, 대표적인 예로는 구글에서 개발한 Spanner이며, 이를 기반으로 구글의 F1이 개발되어 구글의 자체 광고 서비스인 애드센스(AdSense)에 활용되고 있음
  - 다른 방식으로는 기존의 RDBMS를 무공유 구조에서 동작하도록 분산화하는 방식이 있음
    - 초병렬, 무공유 DBMS로서 인터넷 규모의 검색 엔진을 구현한 ODYS가 그 예가 될 수 있음 (Whang, 2013)

## 2. 인공지능 기술 수준

### □ 인공지능 개요

- 인공지능은 사람과 유사한 사고방식을 구현하는 인지적 모델링을 통해 컴퓨터의 학습 결과를 의사 결정에 반영하는 기술을 의미함. 최근 인공지능 기술은 현실에 적용될 수 있는 수준으로 비약적으로 발전함. 이는 수십 년에 걸쳐 축적된 연구 성과들이 그동안의 한계점을 넘으면서 가능해진 것임. 예를 들어 최근 딥러닝 기술은 이미지 분석, 음성 인식, 번역 등에 적용되어 탁월한 성과를 내고 있음. 이는 기계학습의 인공 신경망 기술이 빅데이터를 이용하여 학습할 수 있었으므로 가능했음. 빅데이터를 학습할 수 있는 기계학습 알고리즘들은 대용량 데이터 처리를 단시간에 처리할 수 있는 효율적인 알고리즘 및 강력한 컴퓨팅 인프라를 요구함. 최근의 소프트웨어 및 하드웨어 기술은 이러한 요구를 충족시킬 수 있는 기술 수준으로 발전함. 이러한 환경 조성이 지금의 인공지능 기술의 고도화를 이끌어냈음
- 인공지능의 일반화 수준에 따라 그 단계를 세 가지로 구분할 수 있음. 약인공지능(weak/narrow AI)은 특정 분야에 대해서는 지능을 갖추고 있지만, 스스로 사고하며 문제를 해결할 수 있는 일반화 능력이 없는 컴퓨터 기반 인공지능을 의미함. 강인공지능(strong AI)은 스스로 사고하며 문제 해결이 가능한 컴퓨터 기반의 인공지능으로, 자각 능력과 독립성을 갖추. 초인공지능(super/general AI)은 인간보다 훨씬 뛰어난 지능을 가진 인공지능으로, 효율성, 가치 보존의 의미, 창의성 등의 원초적 욕구를 기반으로 끊임없는 자가 발전이 가능함. 현재 인공지능 수준은 약인공지능에 해당함. 기계학습을 어느 분야에 적용하는 지에 따라 인공지능 분야는 크게 지식 추론, 시각 지능 및 언어 지능으로 구분됨
- 지식 추론
  - 지식 추론은 정보에 대한 가정과 전제로부터 지식을 도출해 내는 기술이며, 개별적 정보를 이해하는 단계를 넘어 정보 간 복잡한 관계를 파악하여 표현하는 기술임. 단순 응답·추천하는 기술에서 상대적인 관계를 유추하는 기술로 발전하였으며, 추론형 질의에 인간을 뛰어 넘는 기술이 등장함. 질문과 연관성이

있는 정보를 유추하거나, 데이터에 존재하지 않더라도 부분 정보들의 조합을 통해 답을 찾아가는 기술이 활발히 개발되고 있음

- <그림 4.4>는 대화형 추론의 예시임. 이 외에도 영상/이미지, 텍스트의 상대적인 관계를 추론하는 기술로는 구글 딥마인드(Google DeepMind)사의 관계형 네트워크 (relational network) 기술 등이 있음

<표 4.1> 인공지능 기술 분류 및 주요 키워드

기술 분류	주요 키워드
기계학습	지도학습, 비지도학습, 준지도학습, 강화학습, 딥러닝, 클러스터링, 베이지안 학습, 인공 신경망, 앙상블
지식 추론	지식 발견, 지식 큐레이션, 정보 추천, 대용량 지식 처리, 전문가 시스템, 질의 응답(Q/A), 대화 의미 분석, 의미 분석, 자동 추론, 추론 엔진, 정리 증명, 논리적 추론, 확률적 추론, 시/공간적 추론, 상식적 추론
시각 지능	객체 인식, 컴퓨터 비전, 행동 이해, 영상 지식 처리, 동영상 검색, 사물 이해, 장소/장면 이해, 비디오 분석 및 예측, 공간 영상 이해, 비디오 요약, 영상 기반 표정/감성 인식
언어 지능	자연어 처리, 텍스트마이닝, 온톨로지, 언어 분석, 대화 이해 및 생성, 자동 번역, 텍스트 요약, 음성 분석, 음성 인식, 화자 인식, 오디오 색인 및 검색, 잡음 처리 및 음원 분리, 음향 인식, 감성 분석

자료: 나영식·조재혁, 인공지능(SW), 2018

<그림 4.4> 대화형 추론 문제

<b>Information</b>	Mary walked to the bathroom. Sandra went to the garden. Daniel went back to the garden. Sandra took the milk there.	<b>Information</b>	It started boring, but then it got interesting.
		<b>Question</b>	What's the sentiment?
		<b>Answer</b>	positive
<b>Question</b>	Where is the milk?	<b>Question</b>	POS tags?
<b>Answer</b>	garden	<b>Answer</b>	PRP VBD JJ, CC RB PRP VBD JJ.

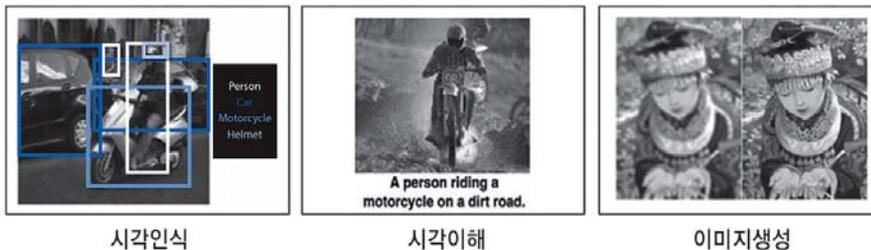
자료: Kumar, A. et al., Ask me anything: Dynamic memory networks for natural language processing, 2016

○ 시각 지능

- 시각 지능은 이미지/영상 등의 시각 정보로부터 객체(사물, 사람 등)를 인식하고 상황이나 감정 등을 이해하는 기술임. 이미지/영상 데이터 기반의 시각 지능은 이미 인간 수준을 넘어선 인식률을 달성했고, 상황을 이해하고 새로운 이미지를 생성하는 기술로 발전하고 있음. <그림 4.5>에서는 시각 지능의 세 분야인 시각 인식, 시각 이해, 이미지 생성 기술들을 묘사하였음

- 시각 지능의 초기 기술인 시각 인식은 이미지/영상으로부터 물체를 구분하는 경계를 인식하고 해당 물체가 무엇인지 판별하는 기술임. 더 나아가 시각 이해는 이미지/영상으로부터 요소 간 상관 관계를 분석하여 전체의 특징을 파악하는 기술로써 눈, 코, 입 모양 등의 상관 관계를 분석해 표정을 이해하거나 감정을 추측하고, 상황을 정확히 이해하여 언어로 표현하는 기술임. 이미지 생성 기술은 데이터가 부족한 환경에서 실제와 매우 유사한 이미지를 생성하는 기술로써, GAN(generative adversarial networks)이 대표적임

#### <그림 4.5> 시각 인식, 시각 이해, 이미지 생성 기술 구현 예시

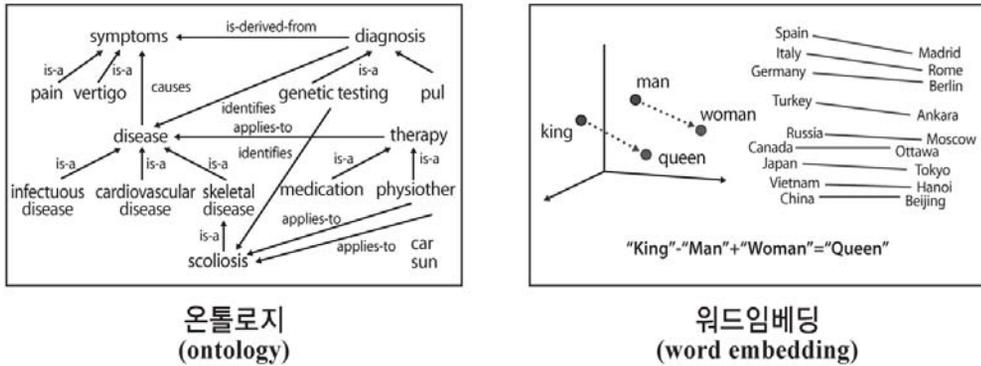


자료: Russakovsky, Ol. *et al.*, Imagenet large scale visual recognition challenge, 2015; Vinyals, O. *et al.*, Show and tell: A neural image caption generator, 2015; Goodfellow, I. *et al.*, Generative adversarial nets, 2014

#### ○ 언어 지능

- 언어 지능은 인간의 언어를 텍스트나 음성 형태로 컴퓨터가 인식하고 이해하며 지식을 도출하는 기술임. 최근 언어 지능 기술은 사람에 의존하지 않고 스스로 언어를 이해하며 인간의 억양과 유사한 음성을 생성하는 단계로 진입함. 크게 자연어 처리 기술, 음성 인식 및 생성 기술이 있음
- 자연어 처리 기술은 전문가가 각 단어 간 의미 관계를 일일이 설정하여 접근하는 온톨로지 기반 기술에서, 인간의 개입을 최소화하고 데이터 기반 학습을 통해 스스로 언어를 이해하게 하는 워드임베딩(word embedding) 방식으로 전환됨. 워드임베딩 방식은 단어를 벡터로 표현하여 유사성 및 관련성 등을 방식으로 확장성이 높고 도메인 의존성이 낮음. 구글은 'word2vec'이라는 워드임베딩 알고리즘을 활용하여 약 1,000억 개의 단어를 뉴스 서비스에 활용함

&lt;그림 4.6&gt; 온톨로지 방식과 워드임베딩 방식의 비교



자료: Mikolov, T. *et al.*, 2013

- 음성 인식 및 생성 기술은 사람의 목소리를 이해하고 생성하는 언어 지능 기술로써, 악센트뿐만 아니라 문장 단위에서의 억양까지 매우 정교한 수준으로 구현함. 마이크로소프트는 1990년 초반부터 2,400개의 전화 대화 녹음을 통해 음성 인식 성능을 5% 에러 수준까지 낮춤. 구글 딥마인드는 기존 음성 생성 기술인 TTS(text-to-speech)를 획기적으로 발전시켜, 방대한 데이터를 통해 사람의 목소리 패턴을 분석하여 언어를 생성하는 방식의 WaveNet 논문을 발표함 (Oord, A. *et al.*, 2016)

- 최근에는 세 가지의 기술 분야가 상세한 전문 기술 분야로 세분화되고 새로운 분야가 더해지기도 함. 분야로는 컴퓨터 시각 인식 기술, 음성 처리 기술, 자연어 처리 기술, 시계열 데이터 처리 기술, 컴퓨터 게임 기술, 네트워크 분석 및 지식 추론 기술, 자동 컴퓨터 프로그램 기술, 설명 가능한 인공지능 기술 등이 있음

#### □ 컴퓨터 시각 인식(Computer Vision) 기술

- 컴퓨터 비전 데이터를 이용한 인공지능 기술은 <그림 4.7> 예시처럼 이미지에 클래스 레이블(class label)을 부여하는 이미지 분류(image classification), 이미지에서 동시에 여러 객체를 검출하여 경계 박스(bounding box)로 표현하는 객체 검출(object detection), 이미지를 경계선까지 분리하기 위해 사용되는 이미지

분할(image segmentation) 등이 있음. 뿐만 아니라 인공지능으로 이미지를 만들어 내는 이미지 생성(image generation) 분야로까지 확장 연구가 활발함

#### <그림 4.7> 컴퓨터 비전 기술 분류



자료: Li, FF. & Karpathy, A. & Johnson, J., CS231n: Convolutional neural networks for visual recognition, 2015

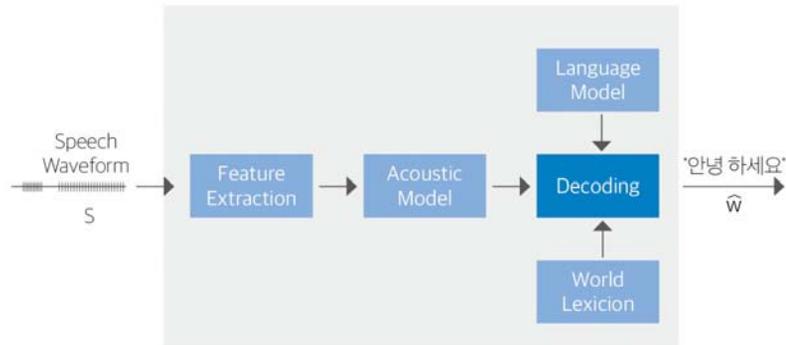
- 이미지 데이터 처리의 핵심 기술은 컨볼루션 신경망(convolutional neural network, CNN) 기반의 기술임. 이미지 분류 문제에서 AlexNet (Krizhevsky, A. & Sutskever, I. & Hinton, G. E., 2012)이 Dropout과 ReLU를 사용하여 ILSVRC 2012에서 1위를 차지한 이후로 CNN에 대해 활발한 연구가 진행되었음. 인공 신경망 모델을 더 깊게 만들어 성능을 향상한 GoogLeNet, VGGNet, ResNet 등이 개발되어 널리 사용되고 있음
- 객체 검출을 위한 기술로 R-CNN (Ren, S. *et al.*, 2015), Fast R-CNN, Faster R-CNN, YOLO 등이 개발되어 사용되고 있음. 핵심 기술은 다수의 경계 박스를 만들기 위한 지역 프로포잘(region proposal)을 CNN에 통과시켜 벡터를 만들어 내고 SVM과 같은 전통적인 분류 방법을 사용하는 것임
- 이외에도 생성망(generator)과 판별망(discriminator)를 경쟁적으로 학습시켜 신경망으로부터 이미지를 만들어내는 GAN (Goodfellow, I. *et al.*, 2014) 기술이 많이 사용되고 있음
- 또 다른 비전 응용 분야로는 비디오 스토리텔링이 있는데 이는 여러 장의 이미지가 순차적으로 들어왔을 때 이미지들의 순서에 맞게 이야기를 만들어 내는 것임. 이는 복잡한 문제이기 때문에 여러 기술이 사용됨. 예를 들면 GLAC Net

(Kim, T. H. *et al.*, 2018) 은 순차적 이미지를 ResNet0152를 이용하여 이미지 특성벡터(image feature vector)로 변환한 후 Bi-LSTM(bi-directional long short-term memory models)을 사용하여 인코딩 한 결과와 ResNet을 이용해 벡터로 만든 결과를 합쳐 디코딩하는 LSTM으로 스토리를 생성해 냄

## □ 음성 처리 기술

- 음성 데이터를 활용한 인공지능 기술은 애플 시리(Siri)나 구글 어시스턴트(Assistant)와 같은 인공지능 비서에 탑재된 음성 인식(speech recognition, <그림 4.8>)부터, TV 뉴스나 토론 프로그램에서 진행자 및 발언자를 식별하여 스크립트 작성 등에 도움을 줄 수 있는 화자 구분(seaker diarization), 악기의 소리를 감별하고 생성할 수 있는 인공지능 음악 작곡(AI music composition) 등 실생활과 밀접한 다양한 분야에 적용되고 있음

<그림 4.8> 음성 인식 모델 수행 과정



자료: 카카오 AI 리포트, 2017

- 음성 데이터는 음파로 이루어진 시계열 데이터이기 때문에 GMM(gaussian mixture model), HMM(hidden markov model) 결합 방식을 주로 사용하였음. 그러던 와중, 딥러닝 및 인공지능 기술이 발전함에 따라서 CNN, LSTM 등 딥러닝 모델로 교체하여 엔드투엔드(end-to-end) 방식으로 학습을 진행하여 획기적인 성능 향상을 가져왔음. 음성 인식의 대표적인 활용 사례로 카카오 AI

스피커, 네이버 클로버, 애플 시리, 구글 어시스턴트 등이 있지만, 가장 주목받는 사례는 구글 듀플렉스임 (Oord, A. *et al.*, 2016). 듀플렉스의 정확한 구조는 공개되지 않았지만 RNN과 WaVenet (Oord, A. *et al.*, 2016) 및 TTS (text-to-speech) 부분의 최신 모델인 Tacotron2 (Jonathan, S. *et al.*, 2018)이 포함된 것으로 보임. 듀플렉스는 구글의 시연 영상에 따르면 악센트나 발음이 거의 사람과 유사하면서, 의미 전달과 무관한 “음..”, “어..”와 같은 표현도 할 수 있는 모습을 보임. 또한 상대방의 질문 내용의 복잡도에 따라서 답변의 시간까지 조절하는 모습을 보여주기 때문에 음성 데이터를 활용한 인공지능 기술의 발전에 많은 기여할 것으로 생각됨

## □ 자연어 처리 기술

- 자연어 처리는 한국어, 영어, 프랑스어 등과 같이 인간이 사용하는 자연 언어를 컴퓨터로 처리할 수 있도록, 즉 분석, 이해, 생성할 수 있도록 하는 기술 분야임. 철자 검사기(spelling checker)와 같은 단순한 프로그램으로부터 보다 복잡한 자연어 정보 검색(information retrieval), 정보 추출(information extraction), 문서 분류(document classification), 문서 요약(text summarization), 감정 분석(sentiment analysis) 혹은 오피니언 마이닝(opinion mining), 자동 대화 시스템(question-answering system), 기계 번역(machine translation), 음성 인식(speech recognition, STT: speech-to-text) 및 음성 합성(speech synthesis, TTS: text-to-speech) 등과 같은 다양한 시스템들이 있음. 이 같은 시스템을 개발하기 위해서는 필수적으로 다음과 같은 언어 처리 기술이 요구됨
- 우선 어절들(띄어쓰기 단위의 연쇄)을 최소의 의미 단위인 ‘형태소’로 분석하는 형태소 분석이 자연어 처리의 첫 관문인데, 여기에는 반드시 중의성 해소(disambiguation)가 중요한 과제임. 형태소 분석이 완료되면 다음 단계는 문장의 구조를 파악하는 구문 분석(syntactic analysis 혹은 parsing)이 필요하고 나아가 의미 분석(semantic analysis)과 언어 외적 지식을 처리하는 지식 베이스(knowledge base) 혹은 온톨로지(ontology)의 개발도 필요함. 자연어 처리 기술의 적용 분야와 사용되는 인공지능 기술은 다음과 같음

### ○ 기계 번역(machine translation)

- 초기에는 규칙 기반(rule-based) 접근법을 적용하여 시스템을 개발했으나 정확도가 높지 않았으며, 이어 통계 기반 접근법을 적용하였으나 이 역시 한계에 봉착함. 이어 이 둘의 장점을 결합한 하이브리드(hybrid) 방식을 적용하다가 최근에는 인공 신경망 방식을 접목하여 큰 발전을 보임. 이는 일반적으로 입력 문장과 출력 문장을 하나의 쌍으로 두고, 가장 적합한 표현 및 번역 결과를 찾는 방식임. 대표적인 예가 구글 번역기임
- 구글은 2015년 인공 신경망 기반의 자연어 처리 플랫폼 ‘신택스넷(SyntaxNet)’을 개발했으며, 2016년에는 실시간 답변을 지원하는 ‘구글 어시스턴트’를 공개함. 한편 마이크로소프트는 2014년 가상비서 ‘코타나’를 개발하고, 애플은 ‘시리’를 공개함. 국내에서는 인터프리가 시스트란과 합작 법인을 설립하여 한국어 중심의 외국어 번역 앱인 지니톡을 개발하였는데, 이는 처음에는 하이브리드 방식과 학습 기반 방식을 적용하다 2016년에는 딥러닝 기술도 도입함. 네이버는 2012년 통계 기반 방식의 번역기를 개발하였다가 이후 2016년 인공 신경망 번역 방식의 파파고 앱을 출시함

### □ 시계열 데이터 처리 기술

- 시계열 데이터(time-series data)는 특정한 시간적 순서에 따라 관찰되고 정리된 데이터임. 시계열 데이터에서 관측치들은 시간적 흐름에 따른 관찰 대상의 변화량을 다루고 있기 때문에 시간적 순서에 종속적(dependent)이며, 이러한 종속적인 관계는 시계열 분석에서 주 연구 대상임. 시계열 분석의 세부 분야는 아래와 같음
  - 주식 시장, 기상 정보 등 과거에 측정된 데이터를 통해 미래의 관측치를 예측 (forecasting)하는 모델 (Rangapuram, S. S. *et al.*, 2018)
  - 주식 시장 등에서 정상적이지 않은 상황 등을 감지(anomaly detection)하는 것과 같이 주어진 시계열 데이터를 그 특징에 따라 분류(classification), 분할(segmentation)하는 모델 등이 있음

- 관측 과정에서 발생한 누락 데이터를 처리(time series imputation)하기 위한 알고리즘 개발도 활발하게 이루어지고 있음 (Che, Z. *et al.*, 2018). 최근에는 의료 정보 데이터 처리에서도 좋은 성과를 내고 있음
- 인공지능 기술은 의료 데이터 분석에 매우 중요한 역할을 하고 있음. 예를 들면 의료 기록 데이터, 의사, 간호사가 기록한 시계열 메디컬 노트 데이터는 매우 중요한 정보를 가지고 있지만 이러한 데이터의 분석은 매우 어려운 작업이어서 그동안 분석이 용이하지 않았음. 최근에는 RNN 기반 기술을 이용해 메디컬 노트 데이터 분석을 하여 심장 발작 시작 시점 예측 (Choi, E. *et al.*, 2017) 등 환자의 예후 및 예측이 가능성이 제시됨. 이 외에도 응급실에서 환자의 위급성을 인공지능 기술로 예측하는 연구도 이루어지고 있음

## □ 컴퓨터 게임 기술

- 컴퓨터 게임에는 기계학습 중에서도 강화학습(reinforcement learning)이 많이 이용됨. 강화학습은 미리 정의된 환경이 주어지며, 이 환경과 상호 작용하여 최적의 방법을 탐색하는 것을 목표로 하는데, 이는 플레이어가 주어진 규칙을 따르며 득점하는 게임에 적용하기에 적절함. 게임은 무수한 경우의 수를 다루어야 하는데 턴제(turn-based) 보드게임인 바둑, 체스의 경우 상대방 선수의 선택에 따라 경우의 수가 기하급수적으로 증가함. 따라서 컴퓨터 게임 연구는 현재 인공지능의 수준을 가늠하는 척도가 될 수 있음. 최근 강화학습은 전통적인 규칙 기반 알고리즘이나 다항식 함수로 근사 등의 방법이 아닌 딥러닝 인공 신경망을 이용하여 방대한 경우의 수를 효과적으로 탐색하여 바둑과 같은 고난도의 게임을 세계 최고 선수를 이기는 수준으로 비약적인 발전이 이루어지고 있음 (Silver, D. *et al.*, 2016)
- 컴퓨터 게임 연구의 기념비적인 성과에 대해 간략히 기술하고자 함. 첫째는 대용량 학습 데이터 없이도 학습이 가능해진 것임. AlphaZero (Silver, D. *et al.*, 2017)는 CNN을 사용해 어떤 사전 지식(prior knowledge)이나 기보 없이 오로지 셀프플레이(self-play)만으로 학습하며, 몬테카를로 트리 탐색(Monte Carlo tree search)을 이용해 세계 최고 수준으로 바둑을 학습할 수 있음을 보임. 두 번째는

한 가지 게임 기술이 여러 게임으로 보편적인 적용이 될 수 있다는 것임. 예를 들면 AlphaZero는 특정 게임에 대한 전문화 없이 바둑, 체스, 장기 등 2인 보드 게임에서 모두 최고 수준의 학습을 달성함. 세 번째는 클래식 턴제 보드 게임을 넘어서 온라인 전략 시뮬레이션 게임으로의 진화임. AlphaStar는 스타크래프트 2를 플레이하는 인공지능 기술임. 이 게임은 불완전한 정보를 토대로 상대의 전략보다 우위를 점하는 전략을 취해야 하고, 행동의 정의역의 크기가 크며, 게임이 실시간으로 진행되는 점에서 바둑보다 더 어려운 게임임. AlphaStar는 어텐션 트랜스포머(attention transformer)를 메인코어, 딥 LSTM 코어, 포인터 네트워크 등의 모델을 결합하며 프로게이머와 대등하거나 그 이상이 되는 수준의 플레이를 선보임. 네 번째는 인공지능이 2인 게임을 넘어서 다수 플레이어 게임을 전문가 수준으로 수행 가능하다는 점임

#### □ 네트워크 분석 및 지식 추론 기술

- 여러 개체 간의 복잡한 관련성을 모델링하기 위해 그래프 혹은 네트워크 구조가 자연스럽게 사용되는 추세임. 관련된 기술은 다음과 같음
  - 그래프 형태로 주어진 입력 데이터에 대해 네트워크 커뮤니티(network community)를 정의하는 네트워크 군집화 기술
  - 심층 생성 모델(deep generative model)을 그래프와 같은 복잡한 데이터에도 적용하는 그래프에 생성 기술. 이러한 기술은 인공지능 기술을 이용한 신약 합성 (Segler, M. H. S. *et al.*, 2019)등 여러 분야에 적용되고 있음
  - 네트워크에서 미래에 연결될 수 있는 잠재 링크나 혹은 결측 링크(missing link)를 추론하는 링크(link) 예측 기술
  - CNN을 그래프로 확장하여 그래프 데이터 셋으로부터 유의미한 패턴을 추출하여 데이터를 표현(feature representation)하는 그래프 컨볼루션 신경망(graph CNN, GCN) 기술 (Defferrard, M., Bresson, X. & Vandergheynst, P., 2016)
  - 그래프의 각 노드로부터 일정 길이의 숏 랜덤 워크(short random walk)을 수행하여 인접 노드들을 방문한 횟수 정보를 이용해 노드들을 저차원 공간에 매핑하는 그래프 임베딩(graph embedding) 기술

## □ 자동 컴퓨터 프로그램 기술

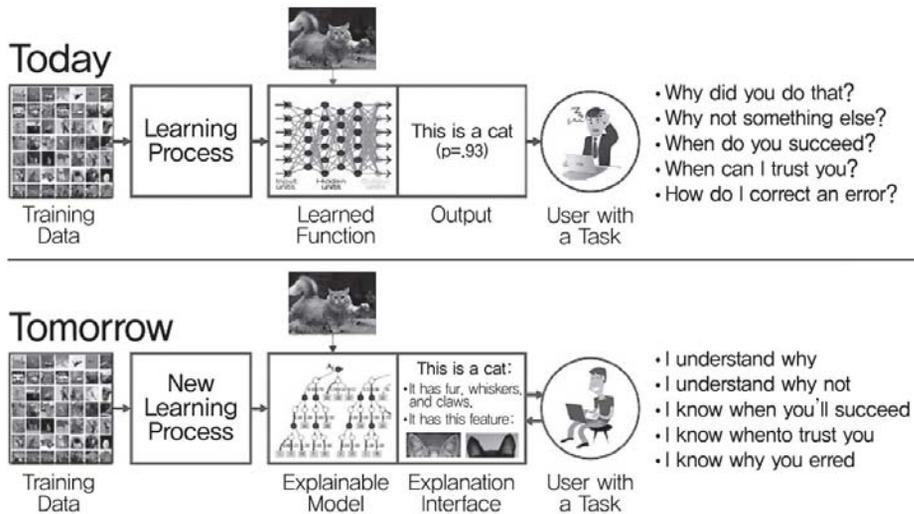
- 프로그램 유도 및 합성은 주어진 조건을 만족하는 컴퓨터 프로그램을 자동으로 구성하는 문제를 일컫음. 해당 문제는 기계학습 및 초기 인공지능 분야에서 오래된 연구 주제로써, 전통적으로는 논리적으로 구성된 다수의 규칙을 바탕으로 규칙 기반 프로그램 합성에 관한 연구가 진행되어옴. 그러나 잘 구성된 규칙 및 문법을 만드는 것은 프로그램을 직접 작성하는 것 이상의 노력이 필요하기 때문에, 최근에는 대부분의 경우 조건은 입력과 출력의 쌍으로 주어지며, 이를 만족하는 알고리즘은 인공지능 모델이 스스로 찾아가는 방식으로 학습이 진행됨. 기존의 방법들은 사람들로부터 검증된 규칙에 기반한 연역적 논리 추론(deductive reasoning) 또는 대용량의 데이터가 필요한 통계적 모델(statistical model)에 기반하였다면, 최근에는 딥러닝을 활용하여 이 문제를 해결하고자 하는 시도가 진행되고 있음. 많은 연구들 중에서 두각을 내고 있는 인공지능 기반 프로그램 개발 방법론은 구글에서 개발한 AutoML(auto machine learning) (Zoph, B. & Le, Q. V. 2017)과 마이크로소프트에서 제시한 DeepCoder (Balog, M. *et al.*, 2017)가 있음
- 유전 알고리즘 기법을 적용하여 딥러닝 구조와 모델을 자동으로 학습하여 이를 reinforcement learning 기법으로 복잡한 문제를 자동으로 해결하는 기법이 개발되고 있음. 예를 들면 2002년에 개발된 NEAT(Neuro-Evolution of Augumenting Topologies)를 구글이 확장한 PBT(Population Based Training) (Aderberg, M. *et al.*, 2017)을 이용하면 단일 사용자 게임들을 수행하는 인공지능 에이전트를 거의 자동으로 생성할 수 있음을 보임

## □ 설명 가능한 인공지능(eXplainable Artificial Intelligence, XAI) 기술

- XAI는 사용자가 인공지능 시스템의 동작과 최종 결과를 이해하고 올바르게 해석하여 결과물이 생성되는 과정을 설명 가능하도록 해주는 기술을 의미함. 예를 들어 고양이 이미지를 인공지능 시스템이 분류한다면, 기존 시스템은 입력된 이미지의 고양이의 존재 여부만을 도출하는 반면, XAI는 고양이의 존재 여부를 도출할 뿐만 아니라, 사용자에게 그 근거(털, 수염 등)까지 제공함

- 현재 인공지능·기계학습 기술은 빅데이터에 복잡한 알고리즘을 적용한 분석으로 사용자에게 의사 결정, 추천, 예측 등의 정보를 제공하지만 알고리즘의 복잡성으로 인해 도출한 최종 결과의 근거, 도출 과정의 타당성 등을 제공하지 못하는 한계가 있음
- 그러나 금융, 보험, 의료 등의 분야에서 고객의 신뢰를 기반으로 개인 정보와 자산 등을 다루는 인공지능의 경우 공정성, 신뢰성, 정확성 등을 보장하기 위해 인공지능 시스템으로부터 생성된 결과의 도출 근거와 과정의 타당성 등에 대한 확인이 필요하므로 XAI 기술의 발달을 요함

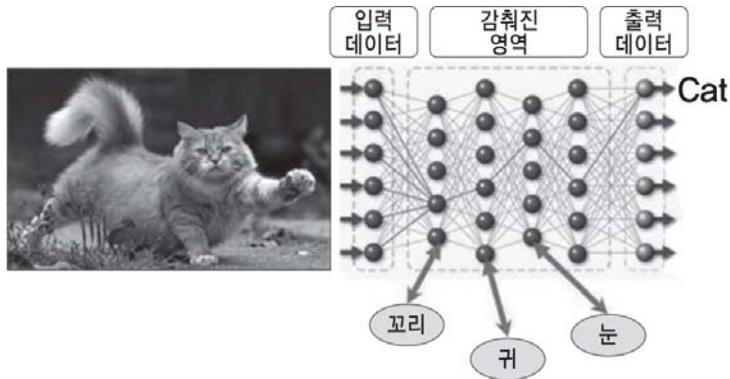
<그림 4.9> 설명 가능한 인공지능 구현 예시



자료: Gunning, D., Explainable artificial intelligence, 2017

- XAI 개발을 위한 기술적 접근
  - 딥러닝을 이용한 모델에서는 은닉 계층의 노드가 의미 있는 속성을 나타내도록 설명 가능한 특징들을 학습하도록 하는 기술임. 개와 고양이의 이미지에서 각 은닉 노드가 귀 모양, 꼬리 모양, 입의 위치 등을 나타내도록 학습해서 모델이 어떤 이미지를 개라고 판단하였을 때 활성화된 은닉 노드를 통하여 그 판단 근거를 알 수 있음 (Ribeiro, M. T., Singh, S. & Guestrin, C., 2016)

<그림 4.10> 인공 신경망의 설명 가능한 노드에 대한 레이블 예시



자료: Gunning, D., Explainable artificial intelligence, 2017

- 이외에도 인과 관계 모델 및 귀납 모델을 활용하여 XAI를 실현하려는 연구들이 있음. 해석 가능한 모델 중 인과 관계 모델로는 BPL(Bayesian Program Learning) 모델 (Edwards, L., & Veale, M., 2017) 및 확률적 AND-OR 그래프 기반 모델이 활용되고 있음 (Ribeiro, M. T. & Singh, S. & Guestrin, C., 2016) 귀납 모델로는 LIME(Local Interpretable Model-agnostic Explanations) 이라는 모델이 있음. 이는 이미지를 분류한 모델이 어떤 이미지를 개라고 판단했다면 이미 설명 가능한 다른 모델의 개에 대한 설명, 즉 개를 표현하는 픽셀들을 주어진 이미지와 대조하여 어느 부분이 개라고 판단한 근거인지 제시함

## 참 고 문 헌

---

- 김동현(2018). 미국방연구원 설명가능 AI(XAI), 스페셜리포트 2018-2. 한국정보화진흥원.
- 나영식·조재혁(2018). 인공지능(SW), KISTEP 기술동향브리프, 2018-16호, 한국과학기술기획평가원.
- 한지연·최재식(2017). “설명가능 인공지능”. 소음진동 통권 제150호, 한국소음진동공학회, pp. 8~13.
- Apache Software Foundation(2019a). Apache Spark Homepage, <https://spark.apache.org/>.
- \_\_\_\_\_ (2019b). Apache Storm Homepage, <http://storm.apache.org/>.
- Balog, M. *et al.*(2017). “Deepcoder: Learning to write programs.” arXiv preprint arXiv:1611.01989.
- Bray, T.(2014). The JavaScript Object Notation (JSON) Data Interchange Format, RFC 7159 Standard Document.
- Chang, F., Dean, J., Ghemawat, S., Hsieh, W. C., Wallach, D. A., Burrows, M., Chandra, T., Fikes, A. & Gruber, R. E.(2006). “Bigtable: A Distributed Storage System for Structured Data,” in Proc. Seventh Symposium on Operating System Design and Implementation (OSDI’06), pp. 205~218.
- Choi, E. *et al.*(2017). “Using recurrent neural network models for early detection of heart failure onset”, *Journal of the American Medical Informatics Association*, Volume 24, Issue 2.
- Dean, J. & Ghemawat, S.(2004). “MapReduce: Simplified Data Processing on Large Clusters”, in Proc. Sixth Symposium on Operating System Design and Implementation (OSDI’04).
- Defferrard, M. & Bresson, X. & Vandergheynst, P.(2016). Convolutional neural networks on graphs with fast localized spectral filtering, *Advances in neural information processing systems*.
- Edwards, L., & Veale, M.(2017). “Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for”. *Duke L. & Tech. Rev.*, 16, 18.

- Goodfellow, I. *et al.*(2014). Generative adversarial nets, Advances in neural information processing systems.
- Gunning, D.(2017). Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA), nd Web, 2.
- Jaderberg, M. *et al.*(2017). Population based training of neural networks, arXiv preprint arXiv:1711.09846.
- Jonathan, S. *et al.*(2018). Natural tts synthesis by conditioning wavenet on mel spectrogram predictions, 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP7). IEEE.
- Kanade, A., Gopal, A., & Kanade, S.(2014). “A Study of Normalization and Embedding in MongoDB”, in Proc. 2014 IEEE Int’l Advance Computing Conference (IACC’14), pp. 416~421.
- Kim, T. H. *et al.*(2018). “GLAC Net: GLocal Attention Cascading Networks for Multi-image Cued Story Generation.” arXiv preprint arXiv:1805.10973.
- Krizhevsky, A. & Sutskever, I. & Hinton, G. E.(2012). Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems.
- Kumar, A. *et al.*(2016). Ask me anything: Dynamic memory networks for natural language processing, International conference on machine learning.
- Lakshman, A. & Malik, P.(2010). “Cassandra: A Decentralized Structured Storage System”, ACM SIGOPS Operating Systems Review, Vol.44, No.2, pp. 35~40.
- Leavitt, N.(2010). “Will NoSQL Databases Live Up to Their Promise?”, IEEE Computer, Vol.43, No.2, pp. 12~14.
- Lee, H., Kang, M., Youn, S. B., Lee, J. G. & Kwon, Y. C.(2016). “An Experimental Comparison of Iterative MapReduce Frameworks”, in Proc. 25th ACM Int’l Conf. on Information and Knowledge Management (CIKM’16), pp. 2089~2094.
- Li, FF., Karpathy, A. & Johnson, J.(2015). CS231n: Convolutional neural networks for visual recognition, University Lecture.

- Malewicz, G., Matthew, H. A., Bik A. J. C., Dehnert, J. C., Horn, I., Leiser, N. & Czajkowski, G.(2010). "Pregel: A System for Large-Scale Graph Processing," in Proc. 2010 ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD'10), pp. 135~146.
- Mikolov, T. *et al.*(2013). Distributed representations of words and phrases and their compositionality, Advances in neural information processing systems.
- Oord, A. *et al.*(2016). Wavenet: A generative model for raw audio, arXiv preprint arXiv:1609.03499.
- Ren, S. *et al.*(2015). "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems.
- Ribeiro, M. T. & Singh, S. & Guestrin, C.(2016). "Why should I trust you?: Explaining the predictions of any classifier." Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM.
- Russakovsky, Ol. *et al.*(2015). "Imagenet large scale visual recognition challenge." International journal of computer vision 115.3 : pp. 211~252.
- Santoro, A. *et al.*(2017). A simple neural network module for relational reasoning, Advances in neural information processing systems.
- Segler, M. H. S. *et al.*(2019). "Planning chemical syntheses with deep neural networks and symbolic AI" Nature, pp. 604~610.
- Shim, K.(2012). MapReduce Algorithms for Big Data Analysis, VLDB 2012 Tutorial.
- Silver, D. *et al.*(2016). "Mastering the game of Go with deep neural networks and tree search" Nature 529, pp. 484~489.
- \_\_\_\_\_ (2017). "Mastering the game of Go without human knowledge" Nature 550, pp. 354~359.
- Vinyals, O. *et al.*(2015). Show and tell: A neural image caption generator, Proceedings of the IEEE conference on computer vision and pattern recognition.

- Vora, M. N.(2011). “Hadoop-HBase for Large-Scale Data”, in Proc. 2011 Int’l Conf. on Computer Science and Network Technology (ICCSNT’11), pp. 601~605.
- Whang, K. Y., Yun, T. S., Yeo, Y. M., Song, I. Y., Kwon, H. Y. & Kim, I. J.(2013). “ODYS: An Approach to Building a Massively-Parallel Search Engine Using a DB-IR Tightly-Integrated Parallel DBMS for Higher-Level Functionality”, in Proc. 2013 ACM SIGMOD Int’l Conf. on Management of Data (SIGMOD’13), pp. 313~324.
- White, T.(2015). Hadoop: The Definitive Guide. 4th Edition, O’Reilly.
- Yogurtcu, T.(2016). New Hadoop Survey Identifies Big Data Trends to Watch in 2016, Syncsoft News, <https://www.syncsort.com/en/about/news-center/press-release/new-hadoop-survey-identifies-big-data-trends>.
- Zaharia, M.(2012). Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing, in Proc. 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI’12)
- Zoph, B. & Le, Q. V.(2017). Neural architecture search with reinforcement learning, arXiv preprint arXiv:1611.01578.



# 빅데이터 · 인공지능 관련 산업



## V

## 빅데이터 · 인공지능 관련 산업

## 1. 빅데이터 · 인공지능 산업 동향

## □ 빅데이터 산업 동향

- 빅데이터 산업은 4차 산업혁명과 인공지능의 근간이 되는 분야로써 데이터 솔루션 구축 및 컨설팅, 서비스, 그리고 인프라 등의 세부 분야를 포함하고 있음
- 글로벌 빅데이터 산업 동향
  - 한국정보화진흥원(2019)에 따르면 세계 빅데이터 산업 규모는 2016년 280억 달러에서 2018년 1.5배 성장한 420억 달러로 2022년까지 5년간 연평균 11.1%의 성장세를 유지하며 710억 달러로 시장 규모가 확대될 전망이다

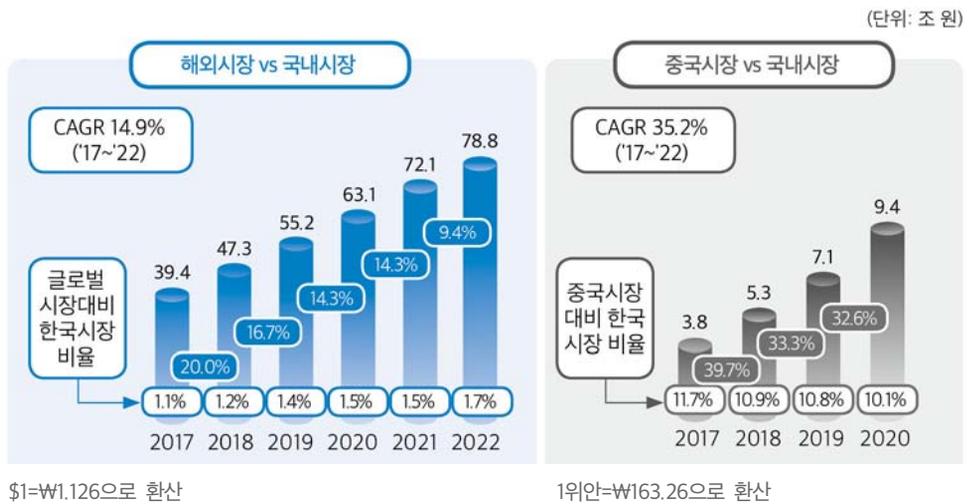
&lt;그림 5.1&gt; 세계 빅데이터 시장 전망



자료: 한국정보화진흥원, 2018 빅데이터 시장 현황 보고서, 2019

- 2022년까지 유망한 분야는 소프트웨어로 연평균 14.0%의 성장률을 보이며 270억 달러 규모로 성장할 것으로 예상됨. 빅데이터 기반 사업이 활성화될수록 하드웨어보다는 소프트웨어, 서비스 분야의 성장세가 높을 것으로 전망함
- 2018년 글로벌 빅데이터 산업 성장률은 20.0%이며, 가장 높은 국가는 중국으로 39.7%의 성장률을 나타냄. 한국은 28.5%로써 글로벌 수준보다는 높고, 중국 보다는 낮게 나타남. 규모는 한국이 글로벌 빅데이터 산업의 1.2%를 차지하는 반면, 중국은 10.9%를 차지함으로써 큰 격차를 보이고 있음

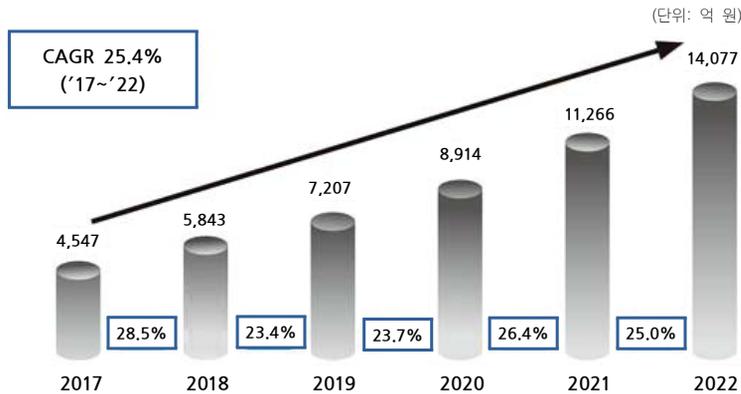
<그림 5.2> 글로벌 및 중국 빅데이터 시장 비교



자료: 한국정보화진흥원, 2018 빅데이터 시장 현황 보고서, 2019

## ○ 국내 빅데이터 산업 동향

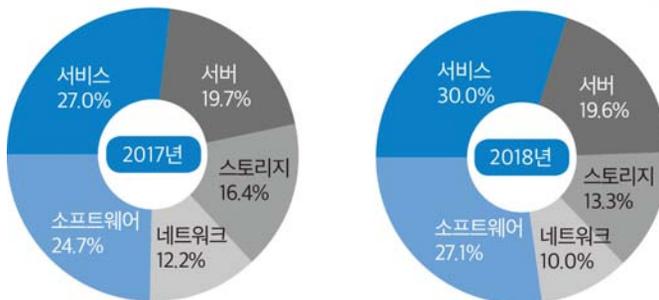
&lt;그림 5.3&gt; 국내 빅데이터 시장 규모 추이



자료: 한국정보화진흥원, 2018 빅데이터 시장 현황 보고서, 2019

- 2018년 국내 빅데이터 산업 규모는 5,843억 원으로, 전년대비 50% 이상 확증된 정부의 투자로 공공 부문과 민간 부문 모두에서 산업 활성화가 추진되고 있음
- 2019년 이후에도 정부 주도로 투자되는 빅데이터 플래그십 사업 및 플랫폼 사업 등의 활성화 정책이 효과를 내면서 20% 이상의 높은 성장세를 유지할 것으로 보임
- 2022년에는 국내 빅데이터 산업 규모가 약 1조 4천억 원에 달하여 2018년 대비 2.4배 이상 성장할 것으로 전망됨

&lt;그림 5.4&gt; 부문별 빅데이터 투자 비중 시장 규모



자료: 한국정보화진흥원, 2018 빅데이터 시장 현황 보고서, 2019

- 산업 세부 분야별 투자 규모에서는 서비스와 소프트웨어가 50% 이상을 차지하고 증가 추세에 있는 반면, 하드웨어로 대표되는 서버, 스토리지 및 네트워크의 비중은 감소 추세에 있음
- 이는 빅데이터 산업 추세가 과거 하드웨어에서 빅데이터를 활용한 서비스나 이를 가동하기 위한 소프트웨어로 중심 이동하고 있음을 나타냄
- 국내 빅데이터 산업은 정부가 선도적으로 활성화를 추진하며 공공 투자를 지속하고 있음. 이와 더불어 민간 부문에서도 빅데이터 산업 투자가 산업별로 확대되고 있으며 통신/미디어, 유통/서비스 등으로 확장될 전망이다. 또한, 대기업 뿐만 아니라 중견/중소기업에서도 빅데이터 플랫폼 도입이 시작됨
- 정부 투자의 경우 2018년 조달청에서 발주된 빅데이터 관련 사업은 266개로, 703억 원의 예산이 투입됨. 또한, 국가 정보화 시행 계획 중 빅데이터 관련 사업은 61개로 예산은 1,310억 원임. 총 규모는 2,014억 원 수준으로 크게 증가함

<표 5.1> 영역별 국내 빅데이터 산업 규모 현황

구분		2016년	2017년	2018년	CAGR
공공 부문	시장 규모	999억	1,338억	2,014억	42.0%
	성장률	43.1%	33.9%	50.5%	
민간 부문	시장 규모	2,441억	3,209억	3,829억	25.5%
	성장률	26.8%	31.5%	19.3%	
전체	시장 규모	3,440억	4,547억	5,843억	30.3%
	성장률	31.3%	32.2%	28.5%	

자료: 한국정보화진흥원, 2018 빅데이터 시장 현황 보고서, 2019

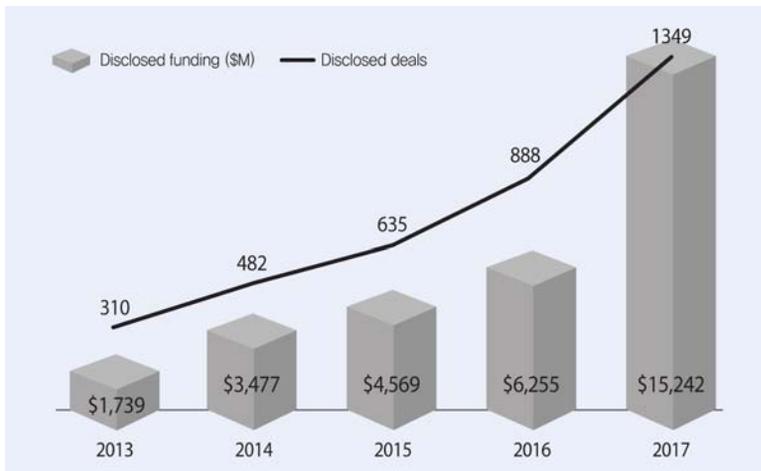
## □ 인공지능 산업 동향

### ○ 글로벌 산업 동향

- 인공지능 분야 글로벌 시장 동향은 조사 기관별 규모의 차이는 존재하나 2018년에는 전년 대비 30~50% 성장할 것으로 예측하였으며, 2025년까지 연평균 36~45%로 성장할 것으로 전망함 (나영식·조재혁, 2018)
  - 2017년 글로벌 인공지능 시장이 160.6억 달러에서 2018년에는 33.62% 성장한 214.6억 달러로 예측하였으며, 2025년까지 연평균 36.62% 증가한 1,906.1억 달러 규모로 성장할 것으로 전망함 (MarketNMarket, 2018)

- 인공지능 소프트웨어로 인한 글로벌 매출이 2017년 54억 달러에서 2018년에는 50.0% 성장한 81억 달러로 예측하였으며, 2025년까지 1,058억 달러로 연평균 45.05% 성장할 것으로 전망함 (TRACTICA, 2018)
- 인공지능 산업이 급성장함에 따라 글로벌 기업들은 적극적인 대규모 펀딩과 M&A를 확대하는 등 기술 경쟁력을 강화하기 위한 투자에 총력을 기울이고 있음
  - 주요 글로벌 기업들은 우수한 기술력을 보유하고 있는 스타트업을 적극 발굴하여 투자하고 있으며, 2013년 대비 2017년 펀딩 건수는 약 4.4배, 투자 규모는 8.7배 증가한 152.4억 달러로 특히, 미국과 중국의 기업 투자가 가장 활발함
  - 구글과 바이두 등 전 세계 선두 기업들은 2016년에만 200~300억 달러를 투자하였으며, 이중 90%는 우수한 기술력을 보유한 기업에 투자하고, 10%는 스타트업을 인수하는 데에 투자함 (Chui, M., 2017)

#### <그림 5.5> 인공지능 분야 글로벌 투자 건수 및 액수



Insights, C. B., Top AI Trends To Watch In 2018, 2018a

- 인공지능 분야 글로벌 스타트업의 수는 2018년 10월 기준 약 5,200개로 추정되며 2017년 기준 인공지능 관련 115개 스타트업이 구글, 애플 등 대기업에 인수되었고, 그 중 기계학습 기술을 보유하고 있는 스타트업의 비중이 49%로 가장 큰 것으로 나타남

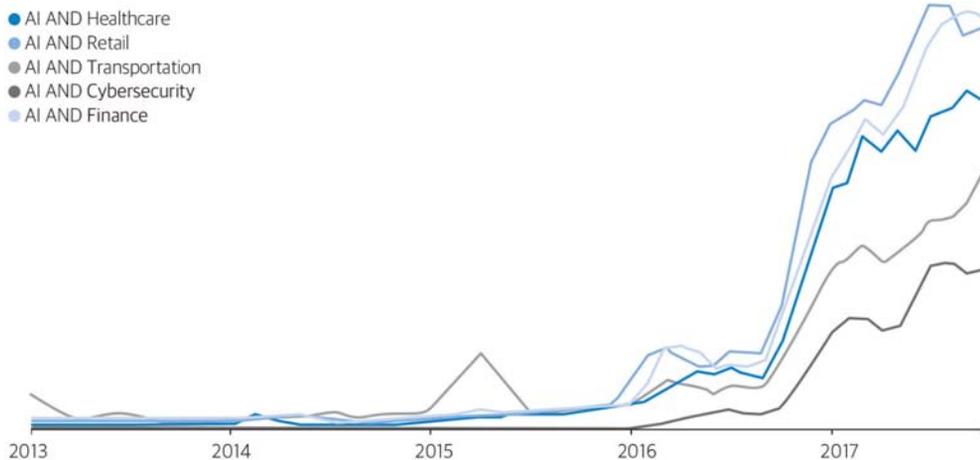
&lt;표 5.2&gt; 인공지능 분야 주요 글로벌 기업 M&amp;A 동향

대표 인수 기업	스타트업	보유 기술
구글(Google)	Api.ai	음성 인식, 언어 이해
	AlMatter	모바일 기반 컴퓨터 비전
	Halli Labs	딥러닝 및 기계학습 시스템 개발
아마존(Amazon)	Pop Up Archive	음악 검색 엔진
	Init.ai	대화 비서
	Regaind	컴퓨터 비전 소프트웨어
페이스북(Facebook)	Ozlo	통합 지식 플랫폼
	Zurich Eye	컴퓨터 비전 SW/HW
바이두(Baidu)	Raven Tech	인공지능 음성 비서
	xPerception	머신 비전

자료: <https://www.cbinsights.com/research/top-acquirers-ai-startups-ma-timeline>

- 글로벌 기업들은 인공지능 기술이 자사에 긍정적으로 기여한다고 인지하고 있으며, 다양한 산업으로의 확장 및 적용을 가속화하고 있음

&lt;그림 5.6&gt; 산업 분야별 인공지능 활용 변화 추세



자료: Insights, C. B., Game Changers 2018, 2018b

- 인공지능은 기업 업무의 반복적인 프로세스 및 작업을 자동화하고(60%), 현재 사용할 수 없는 새로운 전략적 통찰력(50%)을 제공하며, 80%의 글로벌 기업은 이미 인공지능을 제품에 적용하여 사용함 (Forbes, 2017)

- 인공지능을 활용한 주요 산업의 변화 추세는 2016년을 기점으로 급성장하였으며, 금융 및 소비, 헬스케어, 교통, 보안 순으로 활용이 증가하고 있음
- 이처럼 인공지능은 기업의 업무 기능과 다양한 산업에 적용되어 연간 3.5~5.8조 달러의 경제적 효과를 창출할 것이라고 전망함

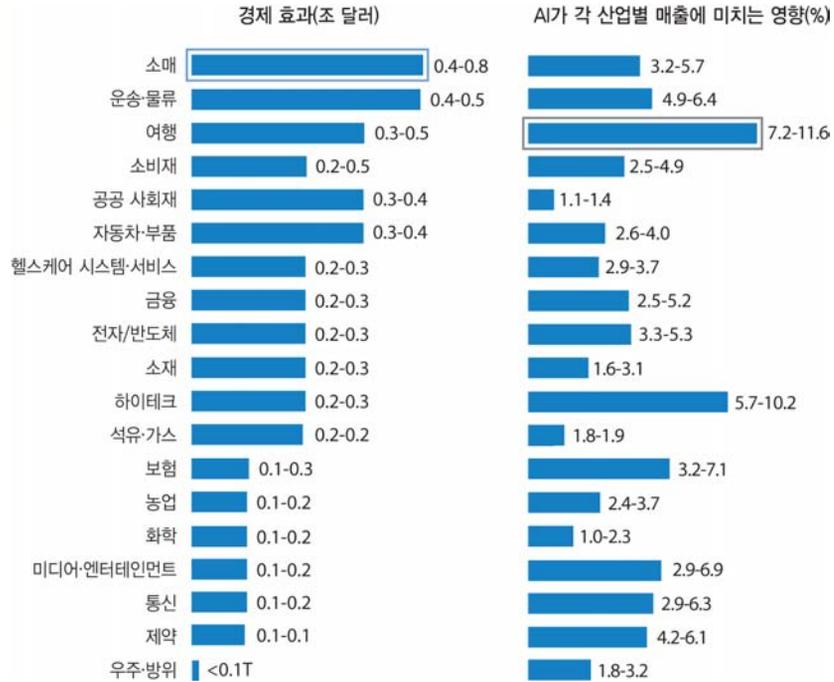
<그림 5.7> 인공지능 적용으로 인한 기업 업무별 경제적 효과



자료: Chui, M. *et al.*, Notes from the AI frontier: Insights from hundreds of use cases, 2018

- 산업 부문 중에는 소매 분야가 연간 최대 8,000억 달러, 운송·물류 분야는 연간 최대 5,000억 달러로 인공지능 도입으로 인한 가장 큰 경제적 효과를 창출할 것으로 전망함
- 인공지능이 매출에 가장 큰 영향을 미치는 부문은 여행 및 하이테크 부문으로 조사되었으며, 향후 데이터의 복잡성을 효과적으로 활용할 수 있는 영역에서 인공지능의 기업적 활용이 두각을 나타낼 것으로 전망함

&lt;그림 5.8&gt; 인공지능 적용이 산업계에 미치는 경제적 효과 및 매출



자료: 나영식·조재혁, 인공지능(SW), 2018

- 또한, 인공지능 전문가 확보를 위한 경쟁이 가열되고 있으며 경쟁사 전문가 영입, 학술대회 직접 채용 등 다각적 확보 전략을 추진하고 있음
  - 세계시장에 필요한 AI 인재는 100만 명 수준이나 실제 활동하는 전문가는 30만 명에 불과하고 세계 교육 기관으로부터 연간 2만 명을 배출하는 등의 수요 대비 공급 격차가 벌어지고 있음 (한국과학기술기획평가원, 2018)
  - 글로벌 기업들은 우수한 인재를 영입하기 위해 인공지능 관련 학술대회에 투자하거나, 경쟁사 최고 전문가 영입, 전문 인력풀이 우수한 국가에 현지 연구 센터를 구축하는 등 다양한 방법으로 인재를 확보하고 있음 (김주완, 2018)
  - 페이스북(Facebook), 아마존(Amazon) 등은 평균 4,500만 엔(약 4억 6,000만 원)에 달하는 대규모 연봉을 제시하며 인재 유치에 적극 나서고 있으며 최고 전문가들을 영입하고 있음 (과학기술정보통신부, 2018)
  - 구글은 중국의 AI 역량을 높게 평가하고 우수 인재를 미국으로 유치하기 위해 베이징에 'Google AI 중국 센터'를 설립함

## ○ 국내 산업 동향

- 국내 인공지능 시장은 2016년 5.4조 원에서 2020년 11.1조 원으로 연평균 19.7% 성장할 것으로 전망되며 음성 인식 및 통번역 기술의 규모가 가장 크게 나타남 (한국인터넷진흥원·KT경제경영연구소, 2017)
- 인공지능 세부 기술별 비중을 살펴보면, 음성 인식 및 통번역 기술의 규모는 4.2조 원으로 가장 큰 비중을 차지하며, 2020년에는 11.1조 원에 이를 것으로 전망함

&lt;그림 5.9&gt; 국내 인공지능 분야 시장 규모



자료: 한국인터넷진흥원·KT경제경영연구소, 2017년 인터넷 10대 이슈 전망, 2017

- 국내 주요 ICT 업체들은 기계학습, 시각 지능 및 언어 지능 등 범용적으로 사용 가능한 인공지능 플랫폼을 출시하며 본격적으로 시장을 공략하고 있음 (정보통신기획평가원, 2018)
- 삼성 SDS는 대용량 데이터 분석 플랫폼인 Bright AI를 개발하여 자연어로 대화하며 고객이 요청하는 업무를 지원하고 수행하는 언어 지능 플랫폼인 'Brity'를 출시함
- SK C&C는 IBM '왓슨' 한국어 버전 플랫폼인 '에이브릴'을 발표하고 대화, 자연어 이해, 언어 번역 및 이미지 인식 등의 기능을 구현함

- 솔트룩스는 IBM 왓슨과 유사한 ‘아담’을 공개하였으며, 언어 지능 분야 정부 R&D 프로젝트인 ‘엑소브레인’의 API를 개방함
- 국내 주요 기업들은 글로벌 선두 기업과의 기술 격차를 좁히기 위해 M&A 및 AI 연구소 설립, 인력 확보를 위한 투자를 통해 추격 중임 (오동현, 2018)
  - 네이버는 2017년부터 5년간 인공지능에 5,000억 원을 투자한다고 발표하였으며, 엔씨소프트는 2012년 AI Lab을 신설하여 100여명의 연구 인력을 유치함
  - 삼성전자는 AI 인재를 1,000명 이상 확보할 것이라는 계획을 밝히고 글로벌 연구 기관 및 기업들과 협업 및 스타트업 M&A에 적극 나서고 있음

## 2. 제조업

### □ 4차 산업혁명과 제조업

- 4차 산업혁명은 ‘유전자, 나노, 컴퓨팅 등 모든 기술이 융합하여 물리학, 디지털, 생물학 분야가 상호 교류하여 파괴적 혁신을 일으키는 혁명’으로 정의됨 (클라우드 슈밥, 4차 산업혁명). 또는 ‘인공지능, 사물인터넷, 빅데이터, 모바일 등 첨단 정보통신 기술이 경제·사회 전반에 융합되어 혁신적인 변화가 나타나는 차세대 산업혁명’으로 정의됨 (TTA정보통신용어사전)
- 4차 산업혁명은 ‘모바일 인터넷’, ‘클라우드 기술’, ‘빅데이터’, ‘사물인터넷’ 및 ‘인공지능’ 등의 기술로 특징짓는데, 공정 모니터링과 기록이 각종 스마트 센서와 소프트웨어를 통해 자동적으로 이루어지는 지능화된 공장인 스마트 팩토리를 출발점으로 하여 산업 사회 전반에 전개되어지고 있음
- 스마트 팩토리에서는 생산 원자재 입고 시점부터 완제품 출고까지 매 공정 진행 과정마다 생산 정보를 모니터링하여 빅데이터를 분석함으로써 최고의 품질을 위한 최적의 기계 가동 상태를 유지할 수 있도록 표준 플랫폼, 스마트 디바이스, 정보 수집 미들웨어 등을 활용함. 이를 통해 종래보다 생산 설비의 작동 중지로

인한 작업 중단이나 불량률의 예방, 설비 사후 유지 관리 비용이 큰 폭으로 절감되는 효과를 가져 오고 있음

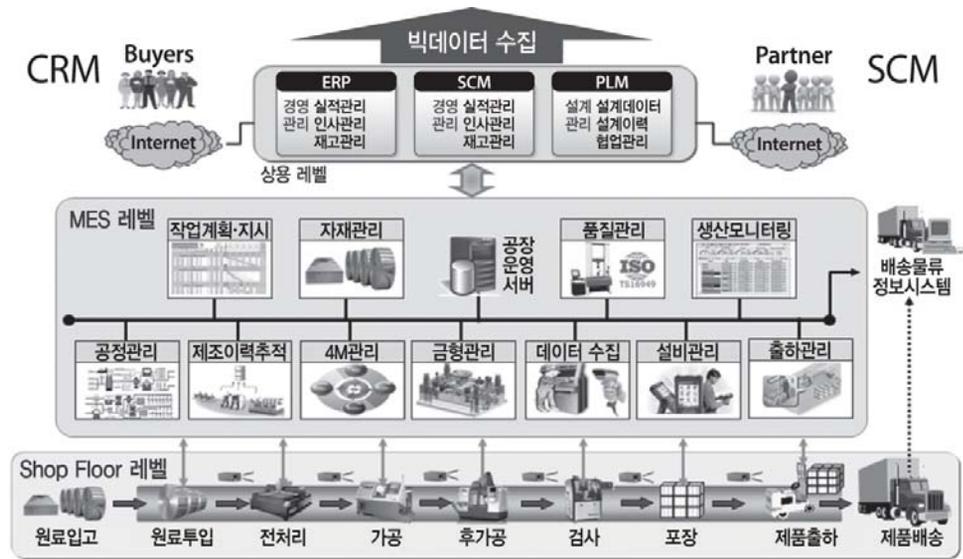
- 이러한 노력과 더불어 센서 기술과 IoT의 발전으로 이전에 수집하지 못하던 다양한 데이터를 측정할 수 있게 되었고, 네트워크 기술의 발전을 통해 측정된 데이터를 빠르게 저장소로 이동시킬 수 있게 되었음. 또한 대용량의 데이터를 낮은 비용으로 저장하거나, 분석의 속도가 빨라지게 되었음. 이러한 발전으로 인하여 4차 산업혁명은 ‘초연결성’과 ‘초지능화’를 그 특징으로 하게 되었음
- 특히 제조 분야에서는 딥러닝 등 기계학습과 빅데이터에 기반한 인공지능의 연계 및 융합으로 인해 제조의 ‘초지능화’를 추진하고 있음. 예를 들어 인공지능이 연계된 사이버 물리 시스템(cyber physical system, CPS)은 부품·제품이 주체가 되어 기계 설비의 서비스를 받아가며 스스로 생산 과정을 거치는 형태의 지능화된 산업 구조로 바뀌게 하고 있음
- 이와 같이 4차 산업혁명은 제조업을 근본부터 혁신시키고자 하는 움직임임. 4차 산업혁명은 정보화 혁명이 시작된 3차 산업혁명에서 얻어질 수 있었던 데이터의 양과 종류를 보다 풍부하게 함은 물론 3차 산업혁명에서는 크게 강조되지 않았던 실시간 빅데이터의 분석이나 활용을 가능케 함으로써 제조업에서는 생산성이나 품질 향상에 막대한 영향을 미치고 있음. 이런 의미에서 볼 때 4차 산업혁명은 제조 분야의 데이터 혁명이라고 할 수도 있음

## □ 제조 활동과 데이터

- 제조업은 인류나 사회에 유용한 제품을 공급하는 산업으로, 국가 경쟁력의 원천임. 제조업은 생산되는 제품의 품질과 비용, 생산 시간 등 경쟁력 요소는 물론 생산성 향상을 위하여 지속적인 의사 결정을 반복하고 있으며, 이 과정에서 많은 양의 데이터를 활용하며 또 만들어내기도 함
- 제조업에서 제품을 생산하기 위한 활동은 크게 두 가지로 분류할 수 있음. 먼저 제품을 생산하려면 고객의 주문을 접수하거나 수요 예측을 실시하여 생산 일정 계획을 수립하며, 지시된 계획에 맞추어 부품을 조달하고 생산을 실시하며,

만들어진 물건을 보관하거나 고객에게 인도하는 활동들과 이러한 제조 활동이 잘 이루어지고 있는지, 품질은 좋은지, 설비는 잘 작동하고 있는지 등을 모니터링하여 원활한 생산이 이루어지게 하는 활동들을 매일 반복적으로 수행하고 있음. 이를 온라인 생산 활동이라고 함

<그림 5.10> 제조업의 정보 시스템



자료: IT DAILY, <http://www.itdaily.kr>, 2018

- 한편 제품을 생산하기 위해서는 시장 조사나 설문 조사 등을 통하여 고객이 원하는 제품을 개발하고, 생산 방식이나 절차, 조건 등을 결정하며, 설비나 시설을 구매하고 설치하는 등 매일 반복이 되지는 않지만 제품 생산이 가능하게 만들어 주는 활동을 지속적으로 하고 있음. 이를 오프라인 생산 활동이라고 함
- 온라인 생산 활동이나 오프라인 생산 활동은 모두 많은 데이터와 정보를 필요로 하고, 많은 양의 데이터를 생성하는 원천이기도 함. 제조업에서는 이러한 데이터와 정보의 원활한 소통과 축적 및 활용을 위하여 전사적 자원 관리(enterprise resource planning, ERP), 제조 정보 시스템(manufacturing execution system, MES), 제품 수명 주기 관리(product lifecycle management, PLM), 공급 사슬

관리(supply chain management, SCM) 고객 정보 관리(customer relationship management, CRM) 등 5대 정보 시스템을 운용하여 왔음

- 제조업의 정보 시스템에는 ERP, MES 등의 정보 시스템 이외에도 업무 자동화 수준의 많은 시스템들이 구축되어 원료 입고에서 제품 배송까지의 라이프 사이클 동안 공정 관리, 제조 이력 관리, 설비 관리, 출하 관리 등 여러 가지 업무를 담당하고 있음
- 4차 산업혁명이 도래한 요즘에는 IoT와 다양한 센서 등의 활용으로 제조업의 데이터는 빅데이터화 되고 있으며, 향후 이러한 빅데이터의 활용 가능 여부가 그 기업의 경쟁력을 좌우하게 될 것임

#### □ 온라인 제조 활동

- 제조업의 온라인 생산 활동은 수요 예측이나 수주 내역으로부터 생산 계획을 수립하고, 소요 자재를 협력 업체에게 발주하거나 시장에서 구매한 다음 이를 바탕으로 생산을 지시하며, 공정 조건이나 자재 수급이 최적으로 유지되도록 생산량이나 품질, 설비나 공정의 상태 등을 모니터링하여 적기에 좋은 품질의 제품이 생산될 수 있도록 하고 있음. 또한 제품을 고객의 요구에 맞추어 출하하고, 고객의 사용 과정에서 발생하는 문제점 등을 모니터링하여 차후 제품 개발이나 생산에 피드백 하고 있음. 이러한 활동들을 원활하게 수행하기 위해서는 관련된 데이터나 정보를 효과적으로 수집하고 활용할 수 있어야 함
- 수요 예측
  - 수요 예측은 제품 개발은 물론 생산 계획 수립에 입력 데이터를 제공하는 활동으로 온라인 생산의 시작이라고 할 수 있음. 수요 예측은 과거에는 주로 과거의 수요나 판매 데이터를 분석하여 통계적 예측을 실시하여 왔음. 그러나 최근 들어서는 SNS 데이터나 인터넷 검색 순위에 있는 아이템 조사 분석, 온라인을 통한 고객의 수요 조사, 문화콘텐츠의 동향, 기술 동향 등 다양한 분야의 자료를 분석하여야 할 필요가 있게 되었음

- 현재 전 세계를 대상으로 제품을 판매하고 있는 우리의 제조 기업들은 과거와 같이 우리나라만의 수요를 예측하는 활동에서 벗어나 전 세계 모든 나라의 소비동향을 조사 분석하여야 하므로, 관련 데이터의 정확한 수집과 빠른 분석이 매우 중요해지고 있음. 더욱이 경쟁 기업의 동향이나 실적, 제품 특성 등도 아울러 수집하고 분석하여야 하는데, 그러한 데이터를 어떻게 효율적으로 확보하는가가 경쟁을 극복할 수 있는 실마리가 됨

#### ○ 생산 계획

- 생산 계획은 어떤 제품을 언제, 얼마만큼 생산할 것인가와 그에 필요한 자재나 설비, 작업 인력 등을 계획하는 활동임. 생산 계획은 원활한 생산 관리의 가장 중요한 활동으로 간주되어 그동안 많은 정보화가 이루어져 왔음. 예를 들어 현재 대부분의 기업이 활용하고 있는 ERP는 자재 소요 계획인 MRP를 바탕으로 발전된 것이며, 실제 작업을 지시하는 MES나 협력 업체와의 정보 소통을 위한 SCM도 생산 계획을 충실하게 수립하기 위한 정보 시스템의 요구에서 비롯되었음
- 생산 계획은 수요 예측은 물론 제조 라인과 협력 업체들의 상황 등을 입력으로 작성됨. 따라서 생산 계획을 수립하기 위해서는 많은 양의 데이터를 빠르게 처리할 수 있어야 하는데, 최적화가 어려운 경우가 많아 rule 기반 전문가 시스템이나 인공지능 등을 많이 활용하고 있음
- 스마트 팩토리에서 가장 중요하게 추진하는 제조 정보 시스템 MES도 유동적으로 변화하는 제조 현장의 상황을 실시간으로 수집하여 생산 계획, 특히 일정 계획에 반영하고자 하는 것임. 특히 과거에는 모든 제조 라인의 일정을 생산 계획에서 지시하였으나, 생산 라인의 유동성으로 부품과 설비의 대화를 가능하게 하여 부품 스스로 가공 일정으로 정할 수 있도록 하는 지능 자율형 생산 계획이 확대되고 있음. 이 기술은 IoT는 물론 최적화를 위한 인공지능 기술이 필요로 하며, 제조 라인 전체의 상황 데이터를 기반으로 의사 결정이 이루어져야 함

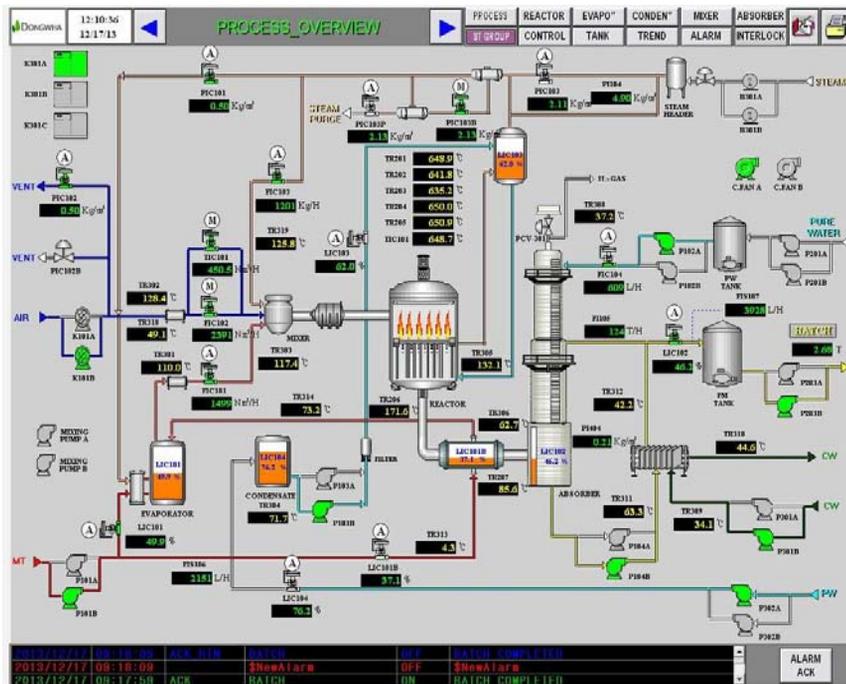
#### ○ 공정 관리

- 생산 계획이 수립되어 작업 지시가 이루어지고 나면 제조 라인의 실적을 파악하고, 상태를 모니터링하여 필요한 시정 조치를 취하는 공정 관리 활동을 수행하여야 함. 공정 관리는 과거에는 수작업에 의한 기록이나 점검에 의존하였으나,

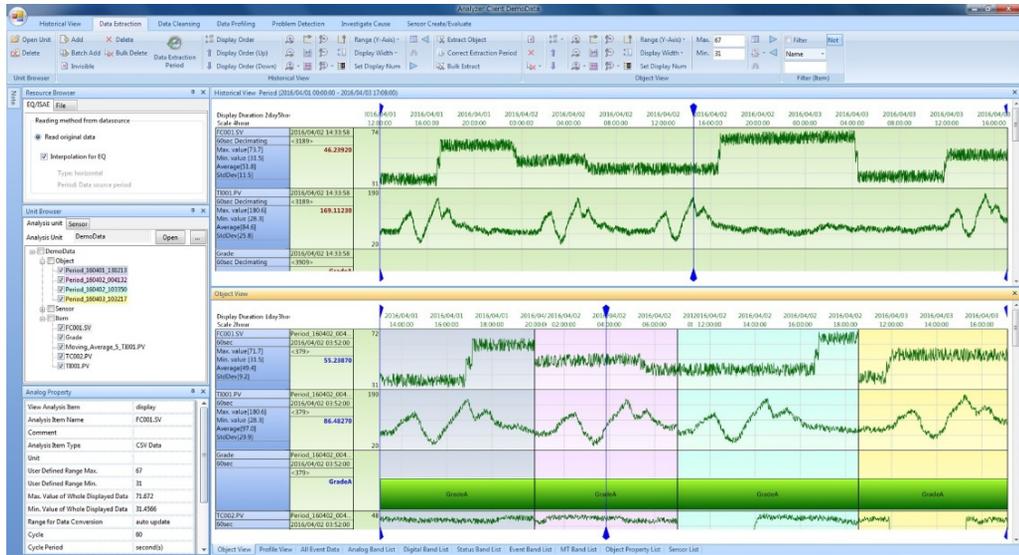
자동화가 진행되면서부터는 실시간 실적 집계와 공정 모니터링이 가능해지고 있으며, 4차 산업혁명에서는 다양한 센서와 측정 장비를 이용하게 됨에 따라 공정의 데이터는 빅데이터화되고 있음

- 대부분의 기업이 생산 실적에 대하여 관리 부서에서 집계하여 다수의 디스플레이 장치에 실시간으로 표시함으로써 생산 진도에 대하여 모든 구성원들이 인지할 수 있도록 하고 있음. 이러한 정보의 표시는 구성원들의 활동에 방향을 안내하는 역할을 하게 됨
- 한편 공정 관리에는 다음 그림과 같이 현재 공정의 상태나 조건을 단위 공정별로 시각화하여 표시하는 시스템들이 많이 활용되고 있음. 최근에는 제조 설비에 많은 센서와 논리 연산 제어 장치(programmable logic controller, PLC), IoT 기술 등이 응용되면서는 설비 내부의 모든 상황을 마치 의사가 환자 진단하듯 알 수 있게 되어 원활한 제조에 만전을 기할 수 있도록 하고 있음

<그림 5.11> 제조업의 공정 관리 시스템



&lt;그림 5.12&gt; 공정 데이터 분석 프로그램



자료: Business wire, <https://www.businesswire.com>, 2019

- 이와 같이 집계된 공정 데이터들은 <그림 5.12>의 좌측과 같이 데이터를 시계열적으로 나타내거나 우측과 같이 평균이나 표준편차, 또는 범위 등을 이용한 관리도로 표시되어 공정의 상태를 해석할 수 있도록 하고 있음. 과거에는 주로 단변량 데이터를 이용하는 수준이었으나, 공정 데이터의 빅데이터화로 다변량 데이터 분석이 더 중요해지고 있음. 더욱이 다변량 분석은 단지 한 공정이나 설비의 데이터를 다변량으로 수집하고 분석하는 것뿐만 아니라, 공정과 공정, 공정과 필드, 공정과 협력 업체 가치 연쇄 등간의 연관성을 분석하고 그에 따르는 적절한 조치를 취하는 것이 매우 중요해지고 있음. 예를 들어 어떤 공정이나 협력 업체에서 제조한 부품이 규격의 목표치에서 약간 아래나 위로 치우친 경우, 다음 공정에서는 이러한 정보를 전달받아 적절한 조치를 취할 수 있어야 지속적으로 좋은 제품을 생산할 수 있게 됨. 앞으로는 공정 데이터의 분석은 이와 같은 연계성 분석을 실시간으로 처리할 수 있는 방향으로 발전하게 될 것이며, 그에 필요한 많은 데이터를 어떻게 효과적으로, 효율적으로 수집하고 분석하는가가 중요하게 될 것임

- 최근 들어 제조 공정의 미세화가 지속적으로 진행되면서, 각 공정별 설비/센서에서 생산되는 데이터의 양과 종류는 폭발적으로 증가했으나, 이 데이터들을 분석하여 불량 원인과 시정 조치를 빠르게 찾아내는 분석 방법은 아직 한계가 존재함. 현재 선진 기업들에서는 스마트 공장 구축에서 가장 중점을 두는 것이 바로 이러한 분석 방법 및 체제의 구축임. 빅데이터 분석 기술이 발전하면서 대용량의 데이터 처리가 가능하기 시작했고, 보다 빠른 속도로 분석할 수 있는 방법을 시도하게 되었으나, 빅데이터를 단순히 분산/병렬 처리하는 기술만으로는 제조 현장의 요구를 충족할 수 없는 것이 현실임

#### ○ 설비 관리

- 우리나라의 제조업은 제품의 조립에서 출발하여 현재는 부품이나 소재를 생산하는 방향으로 발전하고 있으며, 이러한 제조에는 설비를 이용한 장치 산업이 많이 필요하게 되었음. 더욱이 많은 로봇의 사용 등으로 자동화가 이루어지면서 설비가 제조에 차지하는 중요성이 더욱 높아지고 있음
- 설비에 대한 모니터링은 앞서 살펴본 바와 같이 관리도 등 통계적인 방법을 이용하여 고장 발생 여부를 진단하는 방식이 주가 되어 왔음. 그러나 제조 설비가 복잡해지고 고가화됨에 따라 설비의 고장이 야기하는 손해와 비용이 막대해지고 있어, 고장이 발생하기 전에 미리 고장 발생 가능성을 예측하고자 하는 예지 관리(prognostics and health management, PHM) 기술이 보편화 되고 있음. PHM을 위해서는 설비의 고장 메커니즘을 나타낼 수 있는 특성치를 직간접적으로 모니터링하거나 아니면 고장을 유발할 수 있는 스트레스 요인들을 지속적으로 모니터링하여 얻어지는 데이터를 단변량/다변량 시계열 또는 확률 과정(stochastic process)으로 분석하여야 함. 이를 위한 센서나 계측의 개발은 물론 이들의 운용 과정에 필요한 에너지를 자연에서 얻을 수 있는 에너지 수확 기술, 인공 신경망이나 의사 결정 나무, SVM 등 다양한 다변량 분석 및 기계학습 방법들이 활용되고 있음
- 제조 설비에 PHM 기술을 적용하기 위해서는 기존의 설비를 개선하는 것도 필요하지만, 설비의 개발 단계부터 PHM 기술이 접목되도록 하여야 함. 최근에는 설비 개발에 이러한 PHM 기술을 많이 활용하고 있으며, 특히 IoT 기술이

이를 촉진시키고 있음. 제조 설비뿐만 아니라 자동차, 공작기계, 비행기 등 다양한 제품에 PHM이 적용되어 제조 기업이 원격으로 자사의 제품을 모니터링하여 현재 문제점이나 향후에 발생할 수 있는 고장을 미리 고객에게 통보하는 서비스를 실시하고 있음. 이러한 과정에서 얻어지는 데이터는 설비 상태뿐만 아니라 사용 환경 조건 등 다양해지고, 또 전 세계에서 수집되어짐으로써 빅데이터가 될 수밖에 없음. 향후에는 이러한 데이터를 얼마나 잘 수집하여 사전에 문제를 예방할 수 있는 서비스를 제공할 수 있는가가 경쟁력의 요소로 자리 잡게 될 것임

- 설비 관리의 최종 방향은 고장을 사전에 예지하여 미리 조치를 취하는 PHM 기술에서 설비가 생산하는 제품 품질에 문제가 발생하는 것을 사전에 예측하여 항상 목표에 일치된 제품만을 생산할 수 있도록 하는 품질 정비(quality maintenance)를 적용하는 것임. 이를 위해서는 생산되는 제품 품질 특성치를 다양하게 실시간으로 측정하여 수집하고 해석해야 함

#### ○ 품질 검사

- 제조 공정에 인공지능이나 빅데이터의 활용이 활발하고 또 중요해지고 있는 분야 중 하나는 품질 검사임. 과거에는 생산되는 제품에 대하여 어느 정도 불량률 허용하는 방향으로 관리를 할 수 있었으나, SNS 등의 보급으로 소비자의 힘이 매우 강력해진 요즘은 불량률 전혀 허용하지 않는 불량 제로의 방향으로 나아가고 있음
- 불량 제로를 위해서는 제품의 개발이나 설계, 공정 설계 단계에서 불량률 발생하지 않도록 하는 것이 최선이나, 현재의 기술로는 이를 100% 보증할 수는 없으므로, 생산 과정이나 후에 검사를 실시하는 것은 불가피함. 과거의 검사에는 단지 제품의 불량 여부만을 판정하는 계수치 데이터를 활용하였으나, 계측 장비의 발전으로 품질 특성치를 계량적으로 측정하여 분석하는 방향으로 나아가고 있음. 또한 검사 항목이나 검사 위치도 다양해지고 있으며, 카메라를 이용한 비전 검사의 활용은 이미지 데이터와 같은 비정형 데이터를 양산하고 있으며, 경우에 따라서는 소리나 진동, 전기적 특성, 기계적 특성 등 다양한 데이터가 검사에 활용되고 있음. 이러한 데이터들은 설비 모니터링 데이터와

더불어 잘 수집되고 분석되어 차후의 생산 계획이나 제품/공정 개발에 피드백 되어야 하며, 제조 기업들은 이러한 능력과 시설을 갖추고자 많은 노력을 하고 있음

#### ○ FRACAS(Failure Reporting, Analysis & Corrective Action System)

- FRACAS는 어떤 제조업이 생산한 제품에게서 사용 과정에서 발생하는 모든 문제점을 수집하고, 분석하여 필요한 시정조치를 취하는 시스템임. FRACAS는 소비자를 직접 접하는 B2C 업체들에게도 중요하지만 B2B 업체들도 자신의 제품의 현장에서의 수행도를 알아볼 수 있는 시스템으로, 품질 관리에 가장 중요한 정보 시스템임
- 과거에는 각 기업의 영업소에서 수동으로 기록되는 데이터를 수집하여 분석을 실시하였으나, 자동화가 진행되면서는 이를 컴퓨터로 수행할 수 있도록 시스템을 개발하였음. 그러나 앞서 살펴본 바와 같이 센서나 IoT의 발달로 각 제품의 상태를 원격으로 모니터링 할 수 있게 됨에 따라, FRACAS도 자동화, 지능화 하고 있음
- 그러나 어떤 제품이라 하더라도 점차 복잡도가 높아지고 있어 모든 부품에 이르기까지 모니터링을 실시하기에는 어려움이 있음. 이에 어떤 특성치를 모니터링 해야 최소 비용으로 최대 효과를 얻을 수 있을 것인가에 대한 연구가 중요 해지고 있음

#### □ 오프라인 제조 활동

- 비록 제조업의 온라인 제조 활동이 데이터를 많이 활용하고 또 생성시킨다고 하더라도, 이들이 정보화되어 오프라인 제조 활동에 활용되지 못한다면 그 효과성은 매우 제한적일 것임. 오프라인 활동은 제품을 설계하고 공정을 개발 하는 등 제조 활동이 가능케 하는 활동으로 실시간 데이터보다는 기술적 정보를 포함한 지식 데이터가 중요함

- 제품 설계에 사용되는 많은 데이터의 관리를 위해서 과거에는 제품 정보 관리 (product data management, PDM) 시스템이 개발되어 부품이나 제품의 개념, 설계, 개발, 도면, 마케팅 등의 모든 데이터를 단계별로 수집하여 관리하여 왔으나, 시장에서의 가용성, 비용 등 다양한 요소의 관리가 요구되면서 PLM으로 발전하여 제품의 개념, 설계부터 제품의 생산, 유통, 서비스 그리고 유지 보수, 폐기에 해당하는, 말 그대로 제품의 수명 주기 동안 발생하는 모든 데이터를 관리해주는 것을 말함
- PDM은 제품 개발에서 발생하는 방대한 양의 정보를 효과적으로 관리하여 양질의 제품을 보다 신속하고 경제적으로 개발하기 위한 것으로서 첫째, 신뢰도 높은 제품을 오류 없이 개발함으로써 제품의 품질을 향상시킴. 둘째, 신제품 개발 비용을 감소시킬 수 있음. 셋째, 고객이 요구하는 제품 출시가 빠른 시간에 가능할 수 있도록 하는 제품 개발 시간 단축 등의 특징을 가짐
- 한편 PLM은 제품의 수명 주기, 즉 제품의 기획 설계, 제조, 운영, 보수 유지, 폐기 등 모든 활동을 지원하고 그에 관련된 데이터, 정보, 지식 등을 관리하는 비즈니스 솔루션으로써, 제품 수명 주기의 모든 면에서 통합된 데이터/정보 위주의 접근 방법임. PLM을 활용함으로써 첫째, 고객 요구의 다양화로 인한 제품 수명 주기가 지속적으로 단축되며, 둘째, 다양한 제품 개발이 요구되는 시장 주도로의 경영 환경이 변화하고, 셋째, 협력업체 공급체인 및 프로세스상의 협업의 필요성이 강화되고, 넷째, 품질 요구 사항 증가로 그 필요성 또한 증대되고 있다고 할 수 있음
- 그러나 현대에 들어서는 고객의 목소리가 제품 개발에 가장 중요한 요소로 등장함에 따라 제품 정보 이외에도 고객의 요구를 적극적으로 수집하여야 할 필요성이 증대되었음. 특히 소비재의 경우에는 SNS나 인터넷 등을 통하여 소비 동향이나 자사의 제품은 물론 연관 제품에 대한 고객들의 의견을 적극적으로 파악할 필요가 매우 크게 대두되고 있어, 향후 이러한 활동을 인공지능을 활용하여 추진할 수 있는 방안의 모색이 필요함

- 오프라인 제조 활동에서 또 다른 중요한 문제는 개발 엔지니어들이 관련된 데이터를 손쉽게 접근하고 분석할 수 있도록 하는 것임. 6시그마와 같은 품질 혁신 활동의 결과를 우리나라의 대부분의 기업에서는 단지 경험에 의존하였던 과거와는 달리 실험을 실시하여 데이터를 분석하거나, 데이터베이스를 활용하여 설계를 진행하는 활동이 늘어나고 있음. 그럼에도 불구하고 이러한 데이터를 자유자재로 분석할 수 있는 패키지가 대부분 외국 소프트웨어로 활용되고 있으며, 이들을 이용하기 위해서는 통계 지식을 충분히 갖추고 있어야 함. 향후 인공지능을 이용하여 통계 지식을 모르더라도 필요한 정보를 분석해낼 수 있는 패키지의 개발이 필요함
- 제품 개발 과정에서 빅데이터나 인공지능을 활용하고자 하는 분야는 시험을 대체할 수 있는 가상 인증(virtual qualification)임. 가상 인증은 실제 시험을 하지 않고서 유한 요소법 등을 활용하여 시뮬레이션을 실시하고자 하는 것임. 과거에는 실제 시험 데이터가 축적이 되지 않아 top-down 방식으로 특정한 고장 메커니즘의 수명 방정식을 이용하는 방식으로 진행되어 정확성이 미흡한 경우가 있었으나, 최근 들어서는 그동안 축적된 시험 데이터를 직접 활용하여 시뮬레이션을 실행함으로써 해당 기업에 적합한 모의 평가가 이루어지고 있음. 이 기술은 제품의 기능이나 성능의 시험 이외에도 공정의 불량 가능성을 판단하는 데에 적극 활용될 것임
- 오프라인 제조 활동에서 가장 중요한 제품 개발 단계에서는 많은 데이터들이 생성되고 활용되어짐. 그러나 그동안 우리의 기업들은 중요한 데이터를 축적하여 활용하는 면에서 미흡함을 보여 온 것도 사실임. 외국의 경우에는 사소한 데이터 까지도 축적하고 분류하여 정보로 활용하고 있지만, 우리는 데이터의 중요성에 대한 인식이 확산되어 있지 않아 데이터를 수집하고 분류하여 축적하는 노력이 미흡하여 온 것도 사실임. 향후 데이터의 중요성을 인지하도록 하여 많은 데이터를 수집할 수 있도록 하여야 할 것임. 앞으로는 제품 개발 단계에서 필요한 의사소통과 브레인스토밍을 담당할 인공지능 기반 의사소통 또는 브레인스토밍 사이버 로봇의 활용이 필수적일 것으로 예상되는데, 이를 위하여서라도 관련 데이터의 축적이 중요한 상황임

## □ 제조와 데이터 분석

- 제조 공정에서 데이터 분석의 궁극적인 목표는 좋은 제품을 개발하고, 고생산성의 제조라인에서 고품질 제품을 지속적으로 생산할 수 있도록 하는 것임. 제조 공정에서의 데이터 분석은 시대별로 단변량 분석에서 다변량 분석으로, 다변량 분석에서 다시 빅데이터 분석으로 발전하여 왔음. 과거 제조 공정 데이터는 양적인 면에서 크지 않았고 측정치도 소수였기 때문에 각 변수를 개별적으로 관리할 수 있는 단변량 관리 기법이 널리 사용되었지만, 센서 등 계측 기기의 발달로 측정할 수 있는 변수들이 크게 늘었고, 이와 같은 다변량 공정에서는 개별적인 변수 정보와 더불어 변수 간 상호 관계를 분석할 수 있는 기법들이 필요 하게 되었음. 최근의 제조 공정은 공정 복잡도가 매우 증가하고 있으며, 많은 센서와 IoT의 활용으로 방대한 양의 데이터가 생성되고 있음. 따라서 이러한 데이터를 저장하고 효과적으로 분석하는 기술의 개발과 활용이 제조업의 경쟁력 증대에 큰 역할을 하게 될 것임
- 제조 공정에 생성되는 데이터는 기존 데이터베이스 형식에 맞게 저장할 수 있는 정형 데이터와 특정 저장 형식이 정해지지 않은 텍스트, 이미지 등 비정형 데이터가 있음. 정형 데이터는 주로 각종 센서와 계측 기기로부터 얻어지는 수치로 표현되는 데이터들로, 검사 공정이나 FRACAS에서 취급되는 데이터는 텍스트와 이미지 등이 있음. 이러한 데이터들은 양과 복잡도 면에서 기존 분석 기법만으로는 처리가 어려우며, 분산 처리 등의 빅데이터 관리 기법을 요구하고 있음
- 제조업은 방대한 양과 종류의 데이터의 보고임. 제조업에서는 현대의 경영 기법이 요구하는 ‘사실에 의한 관리’가 반드시 이루어져야 하는데, 이것은 바로 데이터에 기반한 의사 결정을 의미함. 이러한 면에서 볼 때, 제조업에서의 데이터 관리는 경쟁력 확보는 물론 기업 생존에 가장 중요한 요소라 할 수 있음

### 3. 바이오·의료 산업

#### □ 바이오 산업

##### ○ 바이오 실험 관련 인공지능 기술

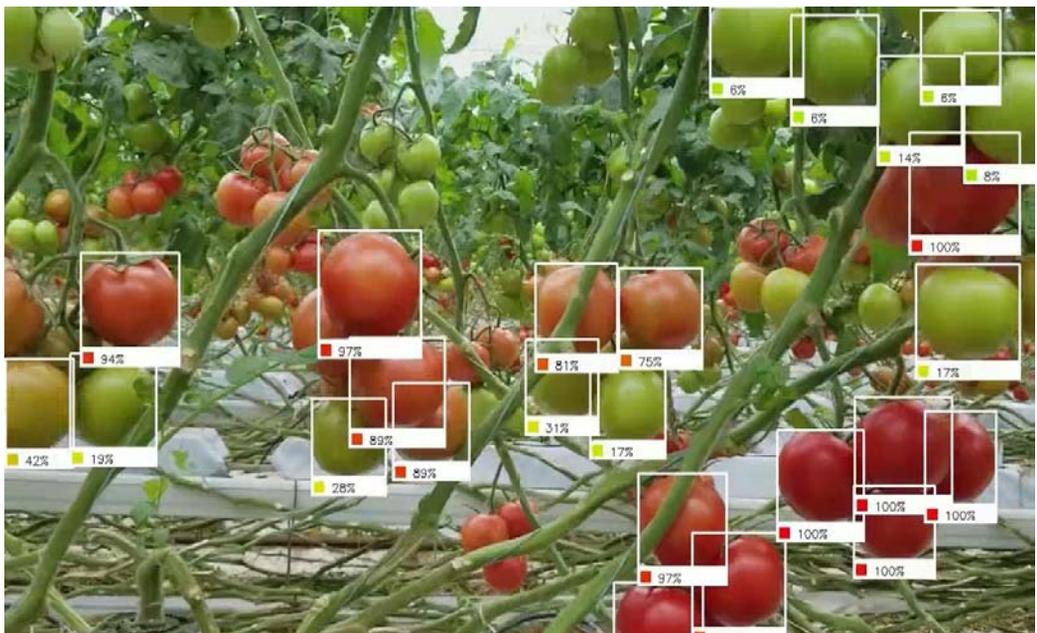
- 지금까지 바이오 실험은 기존 연구 경험에 기반하여 계획 및 수행되었는데, 이러한 기존 방법은 성공 확률도 낮고 비용도 많이 드는 문제가 있음. 특히 신기술에 적용하는 경우 실험자가 무엇을 해야 하는지 모르는 경우가 빈번함. 이러한 문제를 해결 방법으로 최근 인공지능 기술을 이용한 바이오 실험을 가이드하는 기술이 개발되고 있음
- 최근 각광받고 있는 유전자 편집 기술에 딥러닝 기술이 성공적으로 적용된 사례를 소개함. 최신 유전자 편집 기술인 CRISPR - cas9 시스템은 유전자에 결합해 특정 서열만을 편집할 수 있는 기술임. Cas9, Cpf1과 같은 핵산 내부 분해 효소 (endonuclease)가 가이드 RNA(guide RNA, gRNA)와 결합하면, gRNA에 상보적인 유전자를 잘라 교정할 수 있는데 DeepCpf1 (Kim, H. K. *et al.*, 2018)은 컨볼루션 신경망(convolutional neural network, CNN)을 바탕으로 임의의 gRNA 서열이 주어졌을 때 Cpf1의 활성을 예측하는 모델임
- 신약 개발 분야도 최근 AI 기술이 활발하게 적용되는 분야임. 최근 인공지능 기술은 신약 개발의 모든 과정에 적용되고 있음 (Vamathevan, J. *et al.*, 2019)
- 또한, 고해상도 생물 이미지 데이터를 처리하고 정보를 추출하는 데에 인공지능 기술이 성공적으로 사용되고 있음. CellProfiler는 2011년에 Broad Institute에서 출시된 open-source software package로 이미지 형태로 제공되는 생물학적 실험 결과(eg. 세포 염색 결과)에서 표현형(phenotype)을 확인하고 정량화하는 소프트웨어인데, 2018년에 공개된 CellProfiler 3.0 (McQuin, C. *et al.*, 2018)에서는 인공 신경망, 그 중에서도 CNN을 활용해 3D 이미지의 숨겨진 특성 (hidden feature)을 추출해 분석하는 기능이 추가되었음

##### ○ 기존 농업 관련 인공지능 로봇 기술

- 작물 재배 및 수확에도 로봇과 접목된 인공지능 기술이 사용되고 있음. 예를 들면 아래 그림에서 예시한 것처럼 인공지능 시각화 기술을 이용하여 익은

토마토만을 선별하는 기술 (Sa, I. *et al.*, 2016)을 로봇에 적용하여 수확하게 하는 기술이 널리 사용되고 있음. 대형 농장의 경우 항공기에서 촬영한 이미지를 분석하여 수확량을 예측하고, 수확 지역의 우선 순위를 정하는데 최근 AlexNet CNN (Krizhevsky, A. & Sutskever, I. & Hinton, G. E., 2012)을 이용하여 사용 가능성을 보여주는 연구 결과가 보고되었음. 이외에도 잡초 제거 (Li, W. *et al.*, 2019), 작물 및 토양 모니터링, 저장고 관리, 환경 분석(예: 날씨, 건조) 등 다양한 목적으로 인공지능 기술이 사용되고 있음

<그림 5.13> 인공지능 시각화 기술로 익은 토마토를 판별한 연구 사례



자료: <http://root-ai.com/>, rootAI

- 이러한 신기술은 상업화되어 여러 회사들이 AI를 Agriculture Robots에 접목하고 있는데 현재 상용화된 Agriculture Robots는 이미지 센서를 이용해 주로 작물들의 추수, 관리, 병해충 검출에 많이 사용되고 있음. Agriculture Robots는 주로 이미지 데이터 CNN을 통해 각종 임무를 수행하고 있으며, 빠른 이미지 분석을 위해서 faster R-CNN이 많이 사용되고 있음. 현재 상업 목적의 Robots는

주로 RGB 이미지만을 활용하고 있지만, 최근 연구는 근적외선 이미지 등을 융합해서 사용하려는 움직임이 보이고 있음

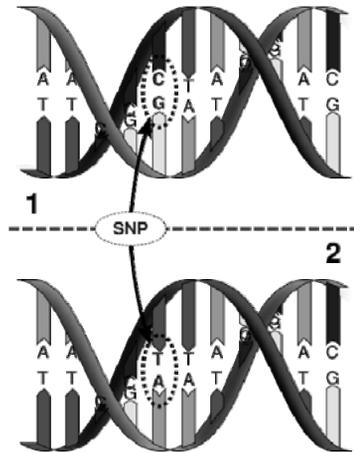
#### ○ 교배, 육종 및 유전자 조작 인공지능 기술

- 교배, 육종은 교배할 대상을 인위적으로 선별해서 특정 형질(예: 수확량, 가뭄 저항성 등)을 가진 식물이나 동물을 육종하는 것인데, 이를 전문가 지식에 의존한 수작업으로 하는 전통적인 방법은 시간이 많이 걸리고 성공 확률이 낮음. 최근 인공지능 기술을 육종에 접목하여 좋은 성과를 내고 있음
- 전통적인 육종을 넘어서 최근에 식물 육종 신기술(New Plant Breeding Techniques, NPBTs)이라고 불리는 유전자가위와 같은 첨단 유전자 교정 기술을 이용한 새로운 식물 육종 방법이 주목받고 있음. 이런 기술로 만들어진 신품종은 인위적인 조작이 가해진 위치의 유전자 염기서열 차이가 자연적 돌연변이 과정에서 발생하는 염기서열 차이 수준으로 줄어들게 되어 GMO(Genetically Modified Organism)에 관한 새로운 논의가 되고 있음. 예를 들면 경제협력개발기구(OECD) 생명공학규제조화작업반회의와 신규식품 사료작업반회의는 신기술 7가지를 그간에 상용화된 GM작물에 대한 안전성 논란을 피하거나 완화할 수 있는 식물 육종 신기술로 새롭게 명명하였음. 세계 각국에서도 식물 육종 신기술을 통해 만들어진 신품종에 대한 GMO 판단 및 규제 여부에 대해서 논의가 활발하게 진행되고 있음. 예를 들면 2016년 4월, 미국 펜실바니아 대학에서 크리스퍼-카스 (CRISPR-Cas9)로 불리는 유전자가위 기술로 개발한 갈변 억제된 GM버섯 (Waltz, E., 2016)은 원래 존재하는 6개의 폴리페놀 산화효소(polyphenoloxidase) 유전자 중 1개를 불활성화하여 버섯의 갈변 현상을 30% 정도 감소시킨 것인데, 미국 농무부(USDA-APHIS)는 외래 유전자가 존재하지 않는 이 버섯이 유전자 변형 생물체의 규제 적용 대상에서 제외됨을 공식적으로 밝혔음. 이외에 다른 유전자 가위 기반 작물들이 미국 농무부의 심사를 받고 있음
- 유전자 조작에 필요한 표적 유전자를 선택하는 데에 있어서 기존에 축적된 생물학적 지식에 기반한 예측이 필요하여 인공지능의 기술이 중요한 역할을 할 것으로 생각됨

## □ 의료 산업

- 의료 산업에서도 데이터 과학 기술이 활발히 적용되면서 중요성이 부각되고 있음.  
X-ray, CT, MRI, Ultrasound, ECG, 웨어러블 디바이스 및 정밀 의학 등 다양한 세부 기술에 대한 대기업과 스타트업의 활발한 연구가 진행 중임
- 정밀 의학
  - 1990년에 시작된 인간 유전체 사업에서 13년의 기간과 3조 원 가까이의 비용을 필요로 했던 유전체 분석은 이제 기술의 발달로 하루 이내의 시간과 100만 원의 비용으로 각 개인별 유전체 정보를 얻는 것이 가능하게 되었음
  - ‘정밀 의학(precision medicine)’은 이러한 인간 유전체 연구를 통해 얻어진 정보를 의학에 접목하여 질환을 예측, 예방, 그리고 진단하는 일련의 개인 맞춤형 의학 행위로서, 이러한 연구가 필요한 이유는 같은 환경, 동일한 질병 위험 인자에 노출되어도 개인에 따라 질환 발생 가능성에 차이가 있기 때문임
  - 따라서 정밀 의학은 개인별로 발생할 수 있는 질환의 가능성에 맞는 예방법과 치료법을 실시할 수 있게 하여 개인에게 가장 잘 맞는 처방 치료를 할 수 있는 방법이라 정의할 수 있음
  - 이러한 정밀 의학은 유전체 연구를 기반으로 하며, 2015년 1월 미국은 100만 명의 대규모 코호트를 구축하여 유전체 정보를 수집하고 특정 질병이나 형질과 연관이 있는 유전 변이를 찾는 연구를 진행하여 정밀 의학의 기반을 구축하려는 연구 계획을 발표하기도 하였음 (<https://obamawhitehouse.archives.gov/precision-medicine>)
- 유전체 정보
  - 각 개인별로 처방한 약물, 치료법, 질환에 따른 반응이 다른 이유는 인간의 유전적 다양성 때문임. 인간의 유전체를 이루는 30억 염기서열은 모든 사람에서 동일한 약 99%의 유전 정보와 약 1%의 다른 정보로 구성되어 있음. 이러한 1%의 유전 변이에 의해 피부색과 같은 외모부터 질병에 이르기까지 다양한 표현형의 차이가 생기게 됨
  - 가장 흔하게 나타나는 유전 변이의 하나인 단일 염기 다형성(single nucleotide polymorphism, SNP)은 <그림 5.14>와 같이 주어진 염기서열에서 한 개의 염기서열이 바뀐 것으로, 인구 집단에서 1% 이상의 빈도를 보이는 염기서열의 위치를 말함

&lt;그림 5.14&gt; 단일 염기 다형성



자료: <https://www.marshall.edu>

- 다양한 연구를 통해서 SNP가 눈 색깔, 맛을 느끼는 감각, 알코올 반응에 따른 피부색 변화부터 암이나 당뇨와 같은 복합 질환(complex disease)에 영향을 준다는 사실이 알려져 있음
- 전장 유전체 연관성 분석
  - 암이나 당뇨와 같이 특정 형질과 연관된 유전 변이를 찾기 위해서는 연관성 연구 (association study)를 수행해야 함. 연관성 연구는 주어진 인구 집단을 관심의 대상이 되는 형질을 가지는 그룹(case group)과 그렇지 않은 그룹(control group)으로 나눈 뒤, 각 그룹 간에 빈도에 차이를 보이는 유전 변이형을 찾아내는 방법임
  - 현재는 기술의 발전을 통해 약 100만 개 이상의 SNP에 대한 정보를 한 번에 알 수 있어 이러한 기술을 이용하여 대규모의 인구 집단에 대한 SNP와 인간의 다양한 형질 간의 연관성을 연구하는 전장 유전체 연관성 연구(genome-wide association study, GWAS)가 활발하게 진행되고 있음
  - 전장 유전체 연관성 분석은 다음과 같은 다양한 분야의 연구가 결합되어 진행되는 다분야 융합 연구로, 각 분야의 전문가가 참여하여 관심 형질에 연관된 유전 변이를 연구하게 됨

- 연구 대상의 선정 및 연구 대상의 형질을 조사, 연구하는 역학 연구
  - 연구 대상의 혈액과 같은 시료를 얻고 저장하고, 보관하는 유전 자원 연구
  - 얻은 연구 대상의 유전 변이에 대한 대용량의 데이터를 단시간에 효과적으로 분석할 수 있는 생물 정보학 및 생물 통계학적 연구
  - 분석된 결과에 의미를 부여하고 유전체를 종합적으로 분석하는 유전체 형질 연구 및 의학 유전체 연구
- 특정 형질과 관련된 SNP에 대한 기존의 연구 결과는 GWAS Catalog (<https://www.ebi.ac.uk/gwas/or> [www.genome.gov/gwastudies](http://www.genome.gov/gwastudies))에 모두 정리가 되어 있음. 2018년 5월까지 3,361개의 논문으로부터 3,375개의 형질(군)에 대해서 48,524개의 SNP가 보고가 되었고, 연구가 된 형질의 예로는 당뇨, 각종 암, 심장질환, 천식, 알레르기, 키, 몸무게, 비만도, 혈압 및 조울증이나 정신분열증 등과 같이 정신의학과 관련된 형질에 대한 연구도 있음
- 발굴된 유전 변이를 통한 효과적인 정밀 의학을 실시하기 위해서는 발굴한 정보의 축적과 정보의 효과적이고 빠른 해석을 가능하게 하는 기반의 구축이 필요함. 따라서 다양한 데이터베이스에 유전 변이 정보를 축적하여 연구 결과를 공유하고 이를 바탕으로 ‘정밀 의학’의 밑그림을 그리고 있는 추세임

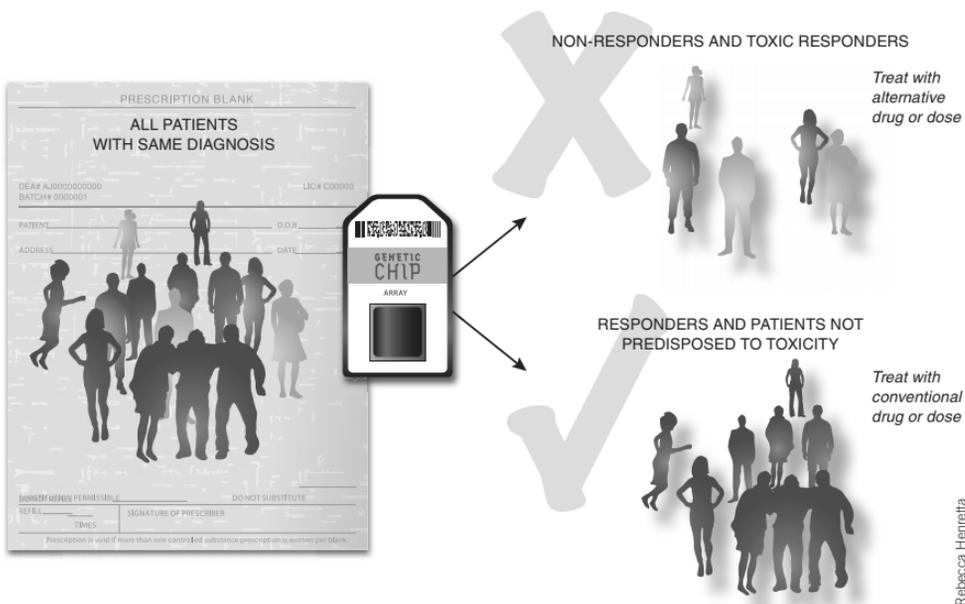
**<표 5.3> 유전 변이 데이터베이스**

데이터베이스	정보	출처
SNPedia	인간 형질에 영향을 주는 단일 염기 다형성을 정리	<a href="http://www.snpedia.com">www.snpedia.com</a>
DGV	유전체 구조 변이의 정보	<a href="http://projects.tcag.ca/variation">projects.tcag.ca/variation</a>
GEN2PHEN	유전자와 형질에 연관된 정보	<a href="http://www.gen2phen.org">www.gen2phen.org</a>
DECIPHER	유전체 구조 변이에 따른 형질 변화 정보	<a href="http://decipher.sanger.ac.uk">decipher.sanger.ac.uk</a>
HGVbase2G2P	전장 유전체 연관성 분석 정보	<a href="http://www.hgvbaseg2p.org">www.hgvbaseg2p.org</a>
VEGA	양질의 유전자 주석 정보	<a href="http://vega.sanger.ac.uk">vega.sanger.ac.uk</a>
NHGRI GWAS catalog	전장 유전체 연관성 분석 정보	<a href="http://www.genome.gov/gwastudies">www.genome.gov/gwastudies</a>

## ○ 약물유전체학(Pharmacogenomics)

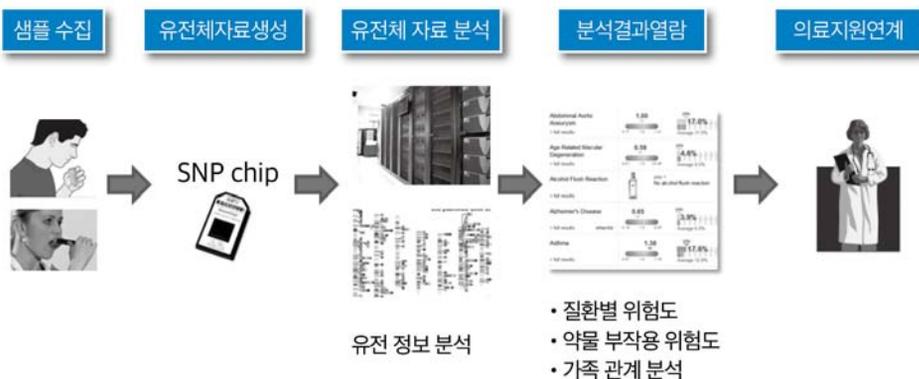
- 현재 대부분의 병원은 기본적인 검사와 진료 상담을 통해 환자의 의학적 상황이나 심리, 사회적 정보를 종합적으로 고려함으로써 환자에게 가장 적절하다고 판단되는 치료를 시도하고 있음
- 하지만, 현대 의학은 약물의 부작용과 치료법 선택 오류 등의 시행착오를 겪고 있으며, 이러한 시행착오는 환자에 따라 반응이 다르게 나타나게 됨. 예를 들어 처방받는 약물의 양에 따른 반응은 개인마다 다르기 때문에 환자에 따라 약물이 즉각적인 치료 효과를 가지기도, 효과가 전혀 없을 수도 있으며 때로는 심각한 부작용을 일으키기도 함
- 각 개인에 따른 약물 반응의 차이는 환경적, 심리적인 요인 뿐 아니라 유전적 요인에 의해서도 나타남. 따라서 개인의 유전 정보를 알게 되면 이러한 개인 간의 차이에 의해 발생하는 시행착오를 피함으로써 <그림 5.15>와 같이 각 개인에게 맞는 처방을 할 수 있을 것임

&lt;그림 5.15&gt; 약물유전체학 연구 예시



- 그 예로, 항응고제로 잘 알려진 ‘와파린’은 사용 양에 따라 중풍(뇌졸중)과 뇌출혈을 일으킬 수 있는데, 와파린의 적정 용량은 개인에 따라 최대 100배까지 차이가 발생함. 하지만 최근 개인의 유전 변이에 따라 와파린에 대한 약물 반응이 다르게 나타날 수 있다는 사실이 밝혀짐에 따라 개인의 유전 정보를 이용하여 와파린에 대한 적정 용량을 조절함으로써 맞춤 처방이 가능하게 되었음 (Johnson, J. A. *et al.*, 2017)
  - 정밀 의학 시대가 열리게 되면 유전체 검사를 통해 질병의 조기 진단 및 질병 위험률을 예측하여 미리 대비할 수 있고, 질병에 걸리게 되어도 개인에 맞는 처방을 통해 시행착오 없이 정확한 치료를 실시함으로써 의료비를 절감하고 질병으로부터 고통 받는 시간을 최소화할 수 있으리라 전망함
- 유전자 검사 서비스
- 유전자 검사 서비스는 개인의 유전체 자료를 분석하여 질환별 발생 위험도, 약물 부작용 위험도 및 간단한 가족 관계 등을 분석하며, 그 과정은 <그림 5.16>과 같음

<그림 5.16> 유전체 정보 생성 및 분석 과정



- 안젤리나 졸리는 2013년 유전자 검사를 통해 유방암과 난소암의 발병 위험을 높이는 BRCA1 변이를 발견하였고, 예방의 차원에서 정상 가슴을 절제하는 수술을 받았고, 그 결과 유방암 발병 위험도는 87%에서 5% 이하로 떨어짐

- 국내에서는 2017년 3월부터 고품암 10종(위암, 폐암, 대장암, 유방암, 난소암, 악성 뇌종양 등), 혈액암 6종(급성 골수성, 림프구성 백혈병 등) 및 유전성 질환 4개(난청, 망막 색소 변형 등)에 대한 유전자 검사를 의료 보험으로 지원하고 있음
- 해외 기업인 23&me와 FoundationOne이 제공하는 유전자 검사 서비스는 병원이나 의사의 도움 없이 바로 소비자가 회사에 연락해서 서비스를 받을 수 있는데, 이러한 서비스 형태를 direct-to-customer(DTC) 서비스라고 칭함
- 국내 DTC의 경우 한정된 형질에 대해서만 검사를 받을 수 있지만 검사 가능한 형질은 점점 더 늘어날 것으로 기대하고 있음. 또한 DTC 예측 서비스의 정확도는 모든 질환과 약물에 대해서 높은 것은 아니기 때문에 그 결과를 해석할 때에는 깊은 주의가 요구됨

#### ○ 정밀 의학을 위한 계속되는 노력

- 현재 정밀 의학 실현을 위한 많은 연구들이 진행되고 있으며, 특히 전장 유전체 연관성 분석 연구를 포함한 정밀 의학의 가장 중요한 기반이 되는 유전체 연구를 중심으로 정보의 축적을 통해 급속도로 발전하고 있음
- 기존의 전장 유전체 연관성 연구는 주로 유럽인들을 대상으로만 진행되어 아시아인, 특히 한국인에 대한 정보를 알 수가 없었으나, 최근 약 1만 명의 한국인 인구 집단을 대상으로 한 연구 결과가 발표되었으며, 이를 토대로 향후 더욱 다양한 형질에 연관된 유전 변이를 발굴할 수 있을 것으로 기대되고 있음 (Lee, J. Y. *et al.*, 2017)
- 약물 반응에 대한 유전 변이 연구인 약물 유전체 연구는 미국에서 가장 활발히 진행되고 있음. 미국은 국가일반의과학연구소(National Institute of General Medical Sciences, NIGMS)에서 약물유전체연구네트워크(Pharmacogenetics Research Network, PGRN)를 설립하여 약물 유전체 연구를 진행하고 있음. 특히 임상에서도 유전체 연구를 통한 약물 반응 연구가 활발히 진행되고 있으며, 최근에는 일부 지역에서 와파린 용량 조절을 위한 유전자 검사 비용을 국가 의료보험인 ‘메디케어(Medicare)’에서 보험 급여로 인정하고 있음

- 국내에서는 2003년부터 국가 지정 연구실 사업, 약물 유전체 연구 사업단을 통해 연구를 진행하고 있으며, 2009년에는 국제 공동 연구 컨소시엄을 통해 와파린 용량에 연관된 유전 변이를 밝혀냄. 또한 식약청에서는 약물과 유전 변이에 연관된 연구 결과를 토대로 다수의 특허를 출원 또는 진행하고 있음
- 이렇듯 ‘정밀 의학’은 많은 연구를 통해 점점 가속화되고 있으나 실질적인 시행을 위해서는 다음과 같은 해결해야 할 과제가 남아있음
  - 유전체 정보에 대한 접근성 증가에 따른 개인 정보 노출 위험이 발생할 수 있으므로, 유전체 정보의 개인정보보호법에 대한 연구와 입법을 통해 이를 보완해야 함
  - 유전체 분석과 해석을 위한 전문가 양성 및 유전체 연구 발전을 위한 지속적인 관심과 국가적인 지원이 필요함

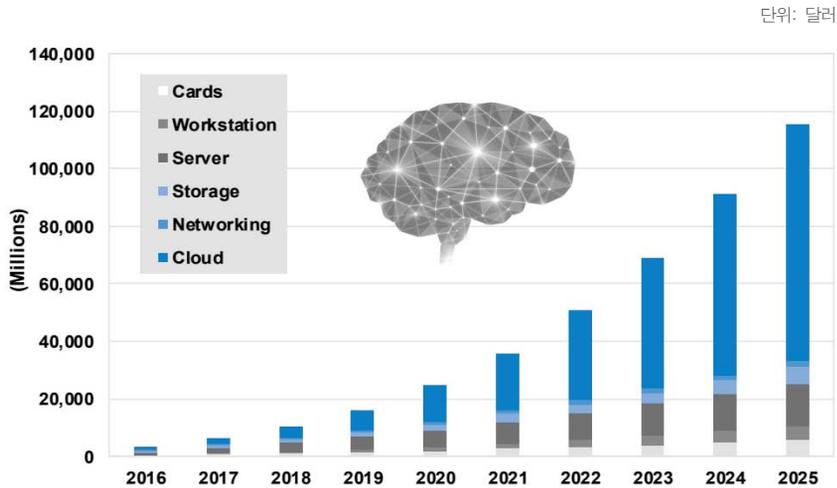
## 4. IT 산업

---

### □ 하드웨어

- 인공지능 시스템의 가속화를 목표로 관련 하드웨어 시장의 경쟁이 치열해지고 있으며, IBM, 아마존, 엔비디아, 인텔, 마이크로소프트, 레노버 등 유수의 글로벌 기업들이 인공지능 제품 및 솔루션에 중점을 두고 시장을 주도하고 있음
- 글로벌 시장 조사 기관인 트래티카의 조사 보고서 ‘인공지능 컴퓨팅, 네트워킹, 스토리지 및 클라우드 인프라 스트럭처: 글로벌 시장 분석 및 예측’에 따르면 지난 2년 동안 컴퓨팅, 스토리지 및 네트워킹과 같은 3가지 하드웨어 인프라가 크게 발전한 것으로 나타남 (정한영, 2018)

&lt;그림 5.17&gt; 인공지능 하드웨어 분야별 시장 규모



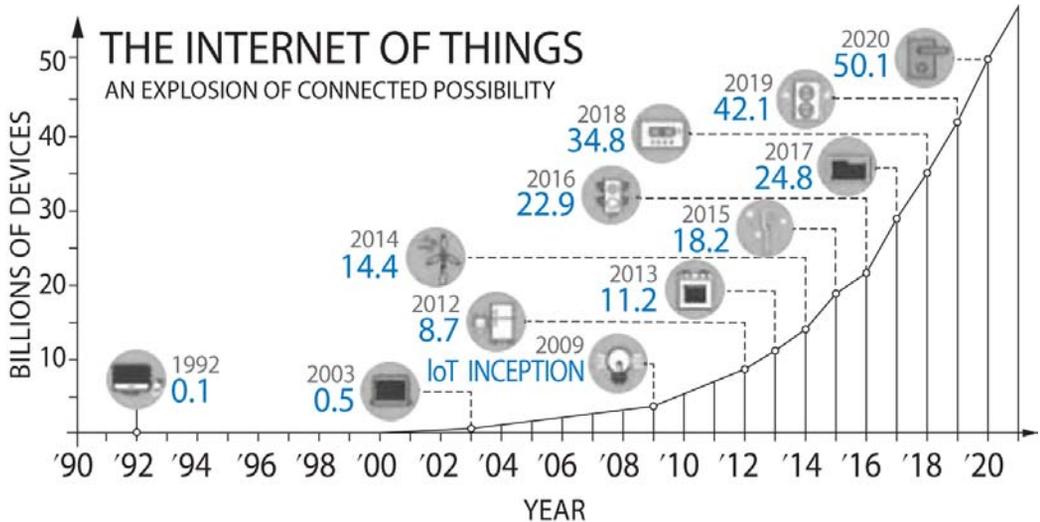
자료: Tractica, 2018

- 하지만 스토리지 및 네트워킹은 뒤쳐져 있으며 아직 인공지능 응용 프로그램과 관련된 주요 혁신을 보지 못했으며, 현재 업계에서는 가장 빠른 옵션을 사용하여 인공지능을 위한 솔루션으로 적극 홍보와 마케팅을 겸하고 있음
- 인공지능 기반 인프라에 대한 전 세계 시장은 2016년 35억 달러(약 3조 8천억 원)에서 2025년까지 1,154억 달러(약 124조 6천억 원)로 성장할 것으로 전망했으며, 클라우드 인프라, 서버, 스토리지, PC 카드, 워크스테이션, 네트워킹 등 6개의 핵심 세그먼트로 조사되었음

#### □ 사물인터넷(Internet of Things)

- 4차 산업혁명과 사물인터넷의 진화
  - 4차 산업혁명의 핵심적인 기술은 크게 사물인터넷, 클라우드, 빅데이터, 인공지능이라고 할 수 있으며, 데이터를 생성, 수집, 저장, 분석, 서비스하는 일련의 과정으로 구성됨 (한국정보화진흥원, 2019)
  - 사물인터넷은 현재 데이터의 생성과 수집에 주로 활용되고 있으며, 2020년에는 약 500억 개 이상의 사물들이 등장할 것으로 예상됨

&lt;그림 5.18&gt; 사물인터넷 디바이스 수 전망



자료: iNEMI Roadmap, 2017

- 대규모의 사물들(IoT 디바이스)을 서로 연결하고 이들이 생성하는 대용량의 데이터를 원활하게 수집하기 위해서는 사물인터넷 기술의 진화가 요구됨
- 2018년 정보통신기획평가원의 'ICT R&D 기술로드맵 2023'에 따르면 사물인터넷을 '인터넷을 기반으로 다양한 사물, 공간 및 사람을 유기적으로 연결하고, 상황을 분석·예측·판단하여 지능화된 서비스를 자율적으로 제공하는 제반 인프라 및 융복합 기술'로 정의하고 있음
- 사물인터넷은 연결형 사물인터넷의 제한된 기능인 사물의 상태 전달에서 더 발전하여 다른 사물들과 상호 소통하여 주변 상황을 인지하고 그에 대한 자율적 대응까지 수행하는 방향으로 진화할 것으로 보임
- 이와 같은 진화의 흐름에 따라 사물인터넷 기술 발전 단계를 1단계 연결형(connectivity) 사물인터넷, 2단계 지능형(intelligence) 사물인터넷, 3단계 자율형(autonomy) 사물인터넷으로 규정함

&lt;그림 5.19&gt; 사물인터넷 기술 발전 전망



자료: 정보통신기획평가원, ICT R&D 기술로드맵 2023, 2018

### ○ 사물인터넷과 빅데이터 · 인공지능

- 사물인터넷 기술이 지능형과 자율형으로 진화하는 데에는 빅데이터 · 인공지능 기술과 밀접한 관련이 있음
- 다양한 기기에서 나오는 데이터 양은 점차적으로 증가하기 때문에 그 데이터에서 의미 있는 정보를 얻고 주요 데이터만 효율적으로 처리 및 저장해야 하는데, 이를 위해서는 빅데이터의 실시간 분석 및 저장 기술이 필요함. 가까운 미래에 공장의 기계뿐만 아니라 가전제품, 대형마트의 상품, 자동차 등 대부분의 모든 사물에 센서가 부착될 것이며 그 센서에서 생성되는 데이터를 통하여 비즈니스 가치를 발굴하고자 할 것임

### ○ 기계 · 건물

- 일본 코마츠(Komatsu)의 경우 해당 회사에서 생산된 건설 기계 차량에 다양한 센서를 부착하고 센서 데이터를 수집 및 관리하고 있음
  - 굴삭기나 불도저의 동작 및 상태에 관한 모든 센서를 중앙 관리 센터로 보내고 있으며, 이를 활용하여 현재 작동 상태나 문제 발생 여부를 모니터링 하고 있음. 장비의 오류 및 연료 부족까지 실시간으로 데이터를 수집 및 분석하여 고객과 대리점에 대시보드 형태로 정보를 전송하고 있으며, 위치 추적을 통해 장비 도난을 사전에 방지할 수도 있음

- 미국 맥키니(Mckenney's)는 건물 자동화 및 제어 시스템 서비스를 제공하는 회사로, 미국 공군기지의 에너지 관리 프로젝트를 맡았으며 빌딩 내 에너지를 관리하는 시스템을 개발함
  - 빌딩 내 수만 개의 센서를 설치하고, 이를 통해 데이터를 수집 및 분석하여 실시간으로 모니터링할 수 있게 만들어, 온도, 습도, 빛 등 다양한 센서 데이터를 통해 비효율적으로 에너지를 사용하는 문제점을 발견하게 되었고, 해결책 또한 분석을 통해 찾아낼 수 있었음. 결과적으로 전체 에너지 사용량의 약 20%를 절감하는 효과를 얻을 수 있었음

#### ○ 병원

- 미국의 경우 블루투스 청진기 및 병원 내 자율 주행 카드 등 병원에서도 사물인터넷 기술을 활발히 적용하고 있음
  - 위치 기반 서비스를 통해 환자의 동선을 확인할 수 있으며 환자의 용무에 따라 걸리는 시간도 파악할 수 있음. 이런 데이터를 분석하여 환자에 따라 동선을 다르게 제공할 수 있고 예약 시스템에도 반영할 수 있으며, 분석 결과를 반영하여 대기 환자 및 대기 시간이 감소하는 효과를 얻을 수 있음

#### ○ 스마트 홈

- 외출 시 자동으로 집안 청소를 하고, 퇴근 전 보일러가 자동으로 켜지고, 주차한 자동차 위치를 파악하여 엘리베이터를 자동으로 불러주는 서비스는 이미 주위에서 쉽게 볼 수 있음
- 스마트 홈은 가정 내 가전 제품들이 인터넷을 통해 상호 연결되고 지능화되어 이를 통해 다양한 서비스가 제공되는 첨단 인텔리전트 서비스 시스템을 말함. 이를 위해 사물인터넷은 필수적인 요소임
- 해외 사례로 미국 애플이 스마트 홈 서비스를 선도적으로 시행하고 있으며, 2016년 스마트 홈 플랫폼 홈킷(Homekit)을 통해 아이폰이나 아이패드를 통해 잠금 장치, 조명, 카메라, 전원 장치 등을 제어할 수 있는 시스템을 출시함
- 국내 사례로는 SK텔레콤이 가정 내 홈 기기(냉장고, 청소기, 세탁기 등)를 원격 제어하여 편리하게 원하는 시간에 작동할 수 있도록 하고, 도어록 및 가스 장치를 제어할 수 있도록 시스템을 구현함

- 스마트 홈에서 쌓일 수 있는 데이터는 급속도로 증가하는 추세지만, 현재는 단순 기기들의 연결에만 집중하고 있고 데이터의 활용은 아직 미비한 상태임. 빅데이터 처리 기술 및 인공지능을 결합하여 개인 맞춤형 서비스 등의 새로운 비즈니스 모델을 발굴한다면 스마트 홈의 활용도는 더욱 높아질 것임

## 5. 금융 산업

### □ 빅데이터 활용 현황

- 국내 금융권의 데이터 활용을 통한 금융 혁신 등 산업 경쟁력 강화 노력은 해외 주요국에 비해 저조한 상황임 (금융위원회, 2018)
  - 해외 주요국 금융 시장의 경우 데이터가 다양한 영역에서 활용되고 있음
    - 미국과 유럽 연합의 경우, 신규 금융 상품 개발, 위험 관리 고도화, 마케팅 등 다양한 부문에서 데이터를 활용 중임. IBM 왓슨을 이용한 중·저신용자 대출 심사 정확도를 제고하거나(CitiBank), 자동차 운행 정보를 활용한 보험료 산정 시스템을 도입한(progressive) 사례가 대표적임
    - 핀테크 기업들도 다양한 데이터 분석·활용 등을 통해 금융 시장 진입을 가속화하고 있으며, 통신 기록이나 인성 검사 결과 등을 활용한 개인 신용 평가 모델을 개발하여 중·저신용자 고객 기반을 확대하고 있음
    - 후발주자였던 중국의 경우, 정부의 강력한 추진력과 풍부한 데이터를 바탕으로 금융강국으로의 부상을 도모하고 있으며, 통신·온라인쇼핑 정보 등을 활용, 거래 정보가 부족한 소비자를 대상으로 중금리 대출을 제공하고 있음
  - 반면, 국내 금융권의 데이터 활용은 아직까지 초기 단계에 불과함
    - 신용카드 보험업 등 일부에서 마케팅, 보험 사기 적발 등에 빅데이터 분석을 이용하는 등 제한적으로 활용 중이며, 데이터를 기반으로 금융 시장에 진입하려는 핀테크 기업 수도 많지 않은 상황임

- 국내 ‘금융 분야 데이터 산업’의 발달도 미흡하여 금융 데이터의 활용을 통한 부가 가치 창출도 미미함 (금융위원회, 2018)
  - 신용 정보 산업의 경우, 그간 빠른 성장에도 불구하고 독과점 구조 및 불합리한 규제 등으로 산업의 발전이 정체된 상황임
  - 개인 신용 평가 시장은 데이터를 보유하고 있음에도 불구하고 법령상 제한으로 다양한 데이터 산업 관련 업무 수행에 어려움이 있으며, 2개사(NICE, KCB)가 국내 시장을 양분하는 과점 구조가 고착화되어 경쟁 및 발달이 지연되고 있음
  - 미국의 경우, 대형 3개사(Experian, Equifax, Trans Union)외에도, 약 400여 개의 분야별 특화 CB사가 경쟁하고 있으며, 개인 신용 평가 업무 외에도 보유 데이터를 활용한 조사·분석·컨설팅 업무를 적극적으로 수행 중임
  - 기업 신용 평가 시장의 경우에도 과점 구조 등으로 발달이 지연되어 금융 산업에 대한 기업 신용 정보 인프라 기능이 미흡한 반면, 미국은 대표 기업 신용 평가사인 D&B가 약 95%의 기업체에 대한 여신정보·부도·소송·담보·세금 체납 정보 등을 데이터베이스로 제공하는 등 인프라가 잘 구축되어 있음
  - 평가 모형 개발 등 기본적인 데이터 업무와 신용 정보업과의 개념 구분이 현행법상 불명확한 점도 신규 진입의 제약 요인으로 작용하고 있음

## □ 인공지능 활용 현황

- 금융 산업에서 인공지능은 로보어드바이저, 시장 분석, 금융 보안, 신용 평가 등의 금융 서비스에 활용되고 있으며 (이은옥, 2017), 자체적인 연구팀뿐만 아니라 인공지능·기계학습 분야의 전문 기업과 연계하여 다양한 서비스를 개발하고 점차 범위를 확장하고 있음 (금융보안원, 2017a)
- 인공지능 기반 금융 서비스들은 업무 자동화를 통한 생산성 향상, 사이버 공격 및 이상 징후 탐지를 통한 기업의 리스크 최소화, 고객 친화적 등의 특성을 포함하고 있음 (Narrative Science, 2017)
- 업무 자동화
  - 자료 공유, 문서 분석, 고객 식별, 이상 징후 탐지 등의 업무들을 인공지능을 활용한 자동화 시스템으로 전환하여 생산성을 향상시키고 리스크를 최소화함

- 복잡한 현황 및 데이터의 통계 자료 산출, 자연어 분석을 통한 업무 문서의 주요 특징 추출, 임직원 및 고객이 입력한 데이터의 오류 탐지, 사고 손해액 예측 시스템 등에 인공지능을 활용하고 있음
- 해외에서는 자연어 분석을 통해 금융 기관들의 경제 지표, 실적, 주가 동향을 학습하고, 자동으로 문서를 생성하는 시스템을 구축하거나(골드만삭스), 자동차 보험 가입자들의 사고율 예측에 인공지능을 도입하여 기존 시스템보다 예측률을 향상시킨 사례(AXA다이렉트)가 대표적인
- 국내에서는 딥러닝 기반 이상 거래 탐지 시스템(fraud detection system)을 도입하여 금융 거래 정보를 학습한 인공지능이 이상 금융 거래를 탐지하도록 수행 중에 있으며(KB국민카드), 인공지능 기반 상권 분석 시스템을 도입해 데이터 분석, 시각 자료 작성 등을 자동화하는 시스템(BC카드) 및 기업과 관련된 뉴스, SNS 등에서 추출한 비정형 정보와 재무제표 정보 등을 분석하여 기업의 부실 위험을 예측하는 시스템(수출입은행)과 같은 사례가 있음

#### ○ 금융 서비스

- 사용자 중심의 대화형(텍스트, 음성) 인터페이스로부터 고객 질의를 입력받고 기계학습을 활용한 분석을 통해 고객이 원하는 다양한 금융 서비스(송금, 조회, 환전, 대출, 상품 추천 등)를 제공하는 인공지능 기반 서비스로써 고객 편의성과 만족도 향상에 기여하고 있음
- 대체로 챗봇을 활용한 헬프데스크, 콜센터와 같은 고객 대응 업무를 수행하면서, 이를 기반으로 송금, 조회, 이체, 금융 상품 추천, 마케팅 등의 금융 서비스를 제공하고 있음
- 해외에서는 IBM의 인공지능 왓슨(Watson)을 기반으로 한 콜센터, 스마트폰 앱, 또는 로봇으로 고객 대응 업무 수행하거나(미쓰이쓰미모토), 고객의 병원 기록, 의약 투여 기록 등의 문서를 분석하여 보험금 지급 업무의 생산성을 향상시킨 사례(후코쿠 생명)가 있음. 또한, 현금, 카드 또는 별도 앱의 도움 없이 쇼핑, 구입, 결제가 가능한 시스템(마스터카드)이나 고객의 자산 규모, 현금 흐름, 지출 금액 등을 기반으로 재정 상태를 파악하고 관리를 위한 피드백을 제공하는 시스템(뱅크 오브 아메리카) 등이 있음

- 국내에서는 ‘스마트 오픈링 시스템’을 통해 카드 승인 정보와 고객의 행위 정보를 실시간으로 모니터링하고, 고객의 행위 시점에서 적합한 혜택을 실시간으로 제공하는 시스템(KB국민카드), SNS나 온라인 커뮤니티 등에서 수집한 데이터로 고객 개인의 소비 성향을 분석하는 마케팅 프로파일링 시스템(BC카드) 등이 있음

#### ○ 신용 평가

- 고객의 금융 정보뿐만 아니라 비금융 정보(요금 납부 기록, 소셜네트워크 정보 등)를 인공지능 기반 분석으로 신용 등급을 세밀하게 평가함으로써 금융서비스 사각지대의 고객까지 서비스 범위를 확대하고 있음 (금융보안원, 2017a)
- 기존 시스템은 개별 심사자가 한정된 고객 정보에 의존하여 리스크를 평가하므로 세밀한 신용 평가가 어려웠지만, 인공지능 기술을 활용함으로써 사회 초년생, 노인 계층과 같이 금융 기록이 적은 사람들에게도 금융 서비스를 제공함
- 해외에서는 고객의 재무 정보, 거래 결제 정보와 약 10만 개의 회사들을 분석하고, 이를 이용하여 고객의 수익성을 예측하고 이를 바탕으로 대출 심사 및 금리를 결정하는 시스템(요코하마은행, 지바은행)이나 개인 대출을 위해 소셜 네트워크 서비스, 인터넷 사용 등을 분석하여 개인의 신용도를 측정하는 시스템(ZestFinance) 등의 활용 사례가 있음
- 국내에서는 고객 대출 심사 시 고객의 금융 정보뿐만 아니라 카드 사용 내역, 상담 기록 등 180여개의 비금융 정보를 함께 분석하여 고객 맞춤형 금리를 설정하는 인공지능 기반 신용 평가 시스템(신한카드), 글로벌 대안 신용 평가사 렌도와 협력 개발한 시스템(현대카드), 기업의 부도 가능성을 진단하는 빅데이터 기반의 시스템(우리은행) 등이 있음

#### ○ 트레이딩

- 로보어드바이저로 대표되는 시스템으로, 사람의 개입을 최소화한 자동화 시스템으로 고객의 투자 성향과 투자 금액에 따라 맞춤형 자산 운용 서비스를 제공함
- 전 세계 로보어드바이저 시장의 점유율은 미국이 약 57%로 1위이고, 독일이 약 9%로 2위이며, 한국의 점유율은 낮음 (Insights, C. B., 2017)

- 2015년 이전까지 미국 로보어드바이저 시장은 자산 운용 기술을 보유한 비금융 회사(Betterment, Wealthfront 등)가 주도했지만, 이후 금융 회사의 본격적인 투자 시작 이후부터는 금융 회사(Vanguard, Charles Schwab 등)가 시장을 주도하고 있음 (임병태, 2017)
- 국내 시장을 살펴보면, 금융 회사를 포함한 여러 기업들이 서비스를 개발하였으며, 현재는 금융 당국의 로보어드바이저 테스트베드를 통하여 알고리즘의 적합성을 심사하고 업무에 적용하고 있음

#### ○ 준법 감시

- 금융 회사가 다양한 규제들을 지키는 데에 많은 비용과 시간이 요구되면서 인공지능 기반의 준법 감시 시스템인 레그테크(RegTech)를 도입하여 효과적인 규제 이행으로 업무 생산성을 향상시키고 있음
- 레그테크는 2008년 국제 금융 위기 이후 변화하는 금융 규제 및 컴플라이언스 요구 사항을 효과적으로 해결하고자 인공지능 기술을 빅데이터에 적용하면서 시작되었고, 2025년까지 전 세계 기업의 30%가 인공지능 기반 준법 감시 시스템을 도입할 것으로 전망됨 (World Economic Forum, 2015)
- 금융 산업에서 레그테크 활용 분야로 데이터 관리, 위험 분석 및 예측, 금융거래 탐지, 고객 신원 관리, 기업 내부 통제, 컴플라이언스가 존재하며, 파생 상품의 적법성 여부를 실시간으로 판단하거나(Droit), 금융 시장의 주요 사건 및 가격 변동에 기반하여 위험을 분석하는 시스템(AlgoDynamix), 내부 직원의 행위 감시 및 위험 수준 등급화 시스템(Behavox), 가상화폐 비트코인의 블록체인에서 이상 금융 거래 탐지 및 식별 시스템(Ellipti) 등의 사례가 있음 (금융보안원, 2017b)
- 해외에서는 레그테크가 활성화되어 있지만 국내에서는 필요성 인식 단계에 그치고 있음

## 6. 스타트업

### □ 빅데이터·인공지능 스타트업 국외 동향

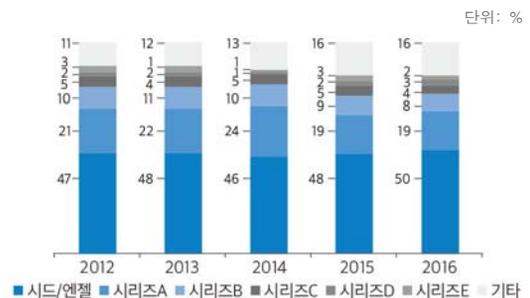
- 인공지능 스타트업은 인프라가 가장 먼저 갖춰진 미국을 중심으로 투자가 활발해지면서 시장이 확대되고 있음 (김보경, 2017)
  - 2016년 인공지능 스타트업 투자는 2012년과 비교하여 건수는 약 4.6배, 규모는 60% 증가함 (Insights, C. B., 2017)
  - 5년간 전체 투자의 67%가 시드·엔젤 단계의 초기 투자로 신규 스타트업이 계속 유입되고 있을 뿐 아니라 중기·후기 단계 투자도 꾸준히 유지되고 있어 인공지능 스타트업의 지속적인 성장세를 반영함
  - 2016년 기준 전체 투자의 62%가 미국 스타트업이었으며 영국(6.5%), 이스라엘(4.3%), 인도(3.5%) 등으로 투자가 다변화되고 있음
  - 2025년 인공지능 전체 시장은 약 600억 달러 규모로 2016년 기준 10억 달러 규모에서 연평균 52% 성장할 것으로 전망됨 (Tractica, 2017)

<그림 5.20> 인공지능 스타트업 투자 추이 및 구성

인공지능(AI) 스타트업 글로벌 투자 추이



인공지능(AI) 스타트업 글로벌 투자 구성

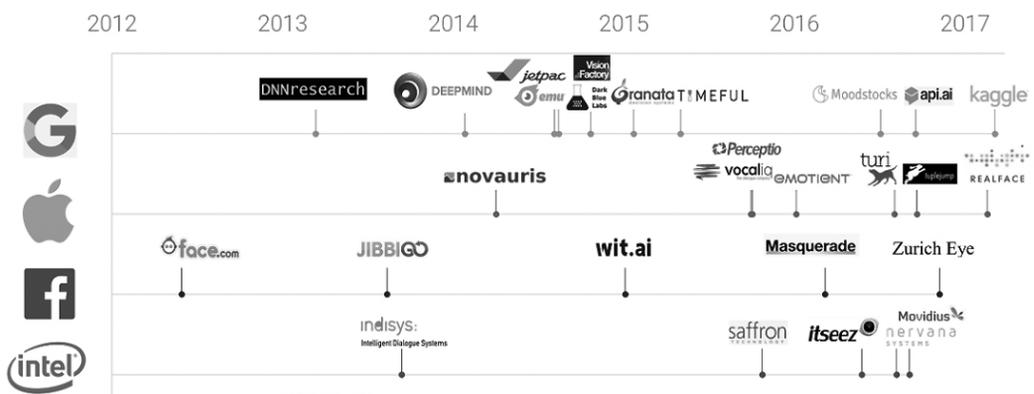


자료: Insights, C. B., 2017

- 구글, 애플, 페이스북, 인텔 등 글로벌 기업들은 '모바일 우선(Mobile First)'에서 '인공지능 우선(AI First)'으로 사업 전략을 전환하고 인공지능 스타트업 인수 합병 및 제휴를 통해 플랫폼을 선제적으로 구축하고 있음

- 인공지능 플랫폼은 스마트폰 플랫폼 이후 ICT 패러다임을 주도할 차세대 혁신 동력으로 주목받고 있으며, 글로벌 기업들은 자사 주도의 생태계 구축을 위해 공격적으로 관련 스타트업을 인수하고 있음
- 초기 시장 선점, 알고리즘 보완 등을 목적으로 2012년 이후 200여개 이상의 인공지능 스타트업의 인수합병이 있었으며 2017년도 1분기에만 30여건이 추진 되었음

<그림 5.21> 주요 기업들의 인공지능 스타트업 인수합병 현황



자료: Insights, C. B., 2017

- 유사하게 인공지능 플랫폼은 과거 스마트폰 플랫폼 구축 방식과 오픈소스 및 클라우드 형태로 개방하고 스타트업은 이를 기반으로 비즈니스 모델을 구축하는 개방형 혁신 생태계가 형성되고 있음
- 스타트업은 인공지능 기술을 내부적으로 보유하지 않고도 다양한 플랫폼을 활용해 각자의 비즈니스에 특화된 제품 및 서비스 개발이 가능함
- 구글은 기계학습 프레임워크 텐서플로우(TensorFlow)를 오픈소스로 제공하고 있으며 아마존의 대표적인 클라우드 서비스(amazon web service, AWS)에도 음성비서 에코를 비롯해 각종 기업용 인공지능 솔루션이 추가됨

## □ 빅데이터·인공지능 스타트업 국내 현황

- 국내에서도 인공지능은 새로운 비즈니스 인프라로 자리잡고 있으며 관련 스타트업이 활성화되고 있음. 비즈니스 모델은 크게 AI 플랫폼 스타트업, AI 솔루션 스타트업, AI 융합 스타트업으로 구분할 수 있음

<표 5.4> 인공지능 스타트업의 비즈니스 모델

비즈니스 모델 구분			정책
AI 인프라	AI 플랫폼 스타트업	B2B	AI 제품 및 서비스 개발을 돕는 PaaS(Platform as a Service) 형태의 비즈니스 운용
AI 적용	AI 솔루션 스타트업		기업의 니즈에 따라 다양한 AI 툴을 SaaS(Software as a Service) 방식으로 제공
		AI 융합 스타트업	B2C

자료: 김보경, 한국 인공지능(AI) 스타트업의 현황과 대응전략, 2017

- 국내 인공지능 스타트업 시장은 초기 단계임에도 불구하고 우수한 기계학습 기술을 중심으로 의료, 플랫폼 등의 분야에서 글로벌 시장 진출의 성과를 거두고 있음
  - 인공지능을 활용한 데이터 식별(자연어·음성 인식 및 분석) → 예측(수치 및 패턴 예측) → 대행(작업 자동화 및 최적화·추천) 과정 개발을 지원하거나 기업 및 소비자에게 관련 서비스를 제공하는 비즈니스로 구성

<표 5.5> 한국 인공지능 스타트업의 글로벌 시장 진출 사례

스타트업명	창업 연도	AI 비즈니스	해외 실적
플런티 (Fluently)	2015	챗봇 플랫폼 개발 지원 및 대화 앱 (텍스트 인식을 통한 자동 응답)	앱 다운로드 수 20만 돌파(미국, 캐나다 등), 해외 장애인용 기기 제조업체 기술 제휴
마인즈랩 (MINDsLab)	2014	음성 인식 등 다양한 인공지능 기반 솔루션(음성 인식을 통한 자동 응답)	미국 콜센터의 50만 달러 규모 음성 인식 솔루션 구축 사업 수주
네오팩트 (NEOFACT)	2010	재활 치료용 게임 앱 및 의료 기기 (개인별 재활 치료에 특화된 게임 추천)	미국 위스콘신주립대병원 및 재활 치료 병원 납품, CES 2017 혁신상 수상

자료: 김보경, 한국 인공지능(AI) 스타트업의 현황과 대응전략, 2017

- 플런티, 마인즈랩, 네오팩트는 각각 데이터 및 수요가 풍부한 글로벌 시장 조기 진출, 정부·연구소·기업의 연계를 통한 자사 중심의 인공지능 생태계 형성, 하드웨어 제품과의 융합을 통한 기존 수요 레버리지 등의 전략을 통해 성장하고 있음
  - 플런티는 데이터베이스 및 고객이 많은 영어권 텍스트 시장 조기 진출로 경쟁력을 확보하고 있음
    - 문자, 카카오톡, 페이스북 등 다양한 메신저에서 수신된 메시지에 적합한 답변을 추천해주는 대화 앱과 챗봇 형성 플랫폼 비즈니스를 운영하고 있음
  - 마인즈랩은 인공지능 생태계에 대하여 일종의 정보 공급망으로 인식하고 정부·연구소·기업 등 다양한 파트너십을 통한 지속적인 데이터 보완 체계를 마련하고 기술력을 개선시킴
    - 한국전자통신연구원(ETRI)에서 개발 중인 한국형 인공지능 엑소브레인(exobrain) 엔진을 활용하고 있으며 강원창조경제혁신센터와 협력하여 국내 최초 인공지능 콜센터를 추진하고 있음
  - 네오팩트는 재활 치료사 위주의 아날로그 재활 시장에서 인공지능 기반 하드웨어와 기능성 게임 간 융합을 통해 데이터 활용 범위를 넓히고 재활 서비스의 개인화라는 새로운 고객 가치를 창출하고 있음
    - 인공지능뿐만 아니라 게임 기획, 임상, 디자인 등 여러 분야의 박사급 전문가가 참여하는 스크럼(scrum) 방식을 통한 지속적인 게임 업데이트 및 피드백 반영으로 제품 완성도를 획기적으로 높이고 있음
- 향후 국내 인공지능 스타트업의 경쟁력 있는 비즈니스 모델 구축을 위해서는 인공지능 기술 확보 자체보다 데이터 수집, 분석, 활용 등의 데이터 중심의 접근법을 바탕으로 정밀하고 특화된 서비스 제공에 주력하는 것이 중요함
- 인공지능 기술을 목적이 아닌 비즈니스 개선을 위한 도구로 인식하고 작업 지능화(플런티-텍스트 자동 응답), 시스템 효율화(마인즈랩-콜센터 솔루션), 서비스 개인화(네오팩트-재활 훈련 추천)라는 AI만의 차별화된 고객 가치 창출이 필요함

- 초기 시장이 성숙되면 학습을 실시할 기반 데이터가 더 중요해짐에 따라 분야별로 전문적인 데이터 학습을 통해 핵심 가치를 제공할 수 있는 수직적 인공지능 스타트업이 부상할 것으로 보임
  - 인공지능 원천 기술은 구글과 같은 글로벌 기업들이 확고한 우위를 점하고 있으며 최근 오픈소스화 정책으로 사용 유인이 증대함
- 인공지능 기술에 대한 기초적인 진입 장벽은 낮아졌지만 실제 비즈니스 구현을 위해서는 해당 분야에 대한 도메인 지식(domain knowledge)의 지속적인 학습이 전제되어야 함

## 참 고 문 헌

- 과학기술정보통신부(2018). I-Korea 4.0 실현을 위한 인공지능(AI) R&D 전략.
- 금융보안원(2017a). 국내·외 금융권 머신러닝 도입 현황.
- \_\_\_\_\_ (2017b). 금융규제 이행을 위한 RegTech의 필요성 및 향후과제.
- 금융위원회(2018). 금융분야 데이터활용 및 정보보호 종합방안.
- 김보경(2017). 한국 인공지능(AI) 스타트업의 현황과 대응전략, 한국무역협회.
- 김주완(2018). 글로벌 기업 ‘AI 인재’ 쟁탈전…학술대회는 ‘스카우트의 場’, 한국경제.
- 나영식·조재혁(2018). “인공지능(SW)”, KISTEP 기술동향브리프, 2018-16호, 한국과학기술기획평가원.
- 오동현(2018). [AI시대] 기술·인재 잡아라... 韓 기업들 투자에 한창, 뉴시스.
- 이은옥(2017). “산업별 지능형 융합 서비스 미래상 전망”, 주간기술동향 제1786호, 정보통신 기술진흥센터, pp. 13~22.
- 임병태(2017). 최근 美 로보어드바이저 시장 동향, 금융투자협회.
- 정보통신기획평가원(2018). ICT R&D 기술로드맵 2023.
- \_\_\_\_\_ (2018). AI First, AI Everywhere로 전개되는 인공지능.
- 정한영(2018). 인공지능(AI) 하드웨어 시장 오는 2025년 16조원에 달할 것, 인공지능신문.
- 한국과학기술기획평가원(2018). “AI 인재 부족 우려... 국가별 대책 인터넷 시장 지각 변동 예상”, 과학기술&ICT 동향, 제111호. pp. 19~20.
- 한국데이터산업진흥원(2019). 2018 데이터산업 현황 조사.
- 한국인터넷진흥원·KT경제경영연구소(2017). 2017년 인터넷 10대 이슈 전망.
- 한국정보화진흥원(2019). 2018 빅데이터 시장 현황 보고서.
- \_\_\_\_\_ (2017). 우리나라 A.I. 기업 현황 조사 보고서.
- \_\_\_\_\_ (2019). 지능형 사물인터넷 네트워크로의 진화.
- Che, Z. *et al.*(2018). “Recurrent neural networks for multivariate time series with missing values.” Scientific reports 8.1 Article number:6085.
- Chui, M. *et al.*(2018). “Notes from the AI frontier: Insights from hundreds of use cases.” McKinsey Global Institute (Retrieved from McKinsey online database).

- Chui, M.(2017). “Artificial intelligence the next digital frontier?” McKinsey and Company Global Institute 47.
- Forbes(2017). 80% of Enterprises Are Investing In AI Today.
- Insights, C. B.(2017). A wealth tech world: mapping robo-advisors around the globe, viewpoint white paper.
- \_\_\_\_\_ (2017). The race for AI: Google, Twitter, Intel, Apple in a rush to grab artificial intelligence startups, Research Briefs.
- \_\_\_\_\_ (2018a). Top AI Trends To Watch In 2018.
- \_\_\_\_\_ (2018b). Game Changers 2018.
- Johnson, J. A., Caudle, K. E., Gong, L., Whirl-Carrillo, M., Stein, C. M., Scott, S. A., ... & Anderson, J. L. (2017). Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. *Clinical Pharmacology & Therapeutics*, 102(3), pp. 397~404.
- Kim, H. K. *et al.*(2018). “Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity” *Nature Biotechnology* 36, pp. 239~241.
- Krizhevsky, A. & Sutskever, I. & Hinton, G. E.(2012). “Imagenet classification with deep convolutional neural networks.” *Advances in neural information processing systems*.
- Lee, J. Y. *et al.*(2017). “Korean atrial fibrillation network genome-wide association study for early-onset atrial fibrillation identifies novel susceptibility loci.” *European heart journal* 38.34 (2017): 2586-2594.
- Li, FF. & Karpathy, A. & Johnson, J.(2015). CS231n: Convolutional neural networks for visual recognition, University Lecture.
- Li, W. *et al.*(2019). “Automatic Localization and Count of Agricultural Crop Pests Based on an Improved Deep Learning Pipeline” *Scientific Report* 9, Article number: 7024.
- MarketNMarket(2018). Artificial Intelligence Market by Offering (Hardware, Software, Services), Technology (Machine Learning, Natural Language

- Processing, Context-Aware Computing, Computer Vision), End-User Industry, and Geography - Global Forecast to 2025.
- McQuin, C. *et al.*(2018). “CellProfiler 3.0: Next-generation image processing for biology” *PLoS Biology* 16(7).
- Narrative science(2017). The Rise of AI in Financial Services.
- Piquette-Miller, M. & Grant, D. M.(2007). “The art and science of personalized medicine.” *Clinical Pharmacology & Therapeutics*, 81(3), pp. 311~315.
- Rangapuram, S. S. *et al.*(2018). “Deep state space models for time series forecasting.” *Advances in Neural Information Processing Systems*.
- Sa, I. *et al.*(2016). “DeepFruits: A Fruit Detection System Using Deep Neural Networks” 16(8).
- Tractica(2017). Artificial Intelligence Market Forecasts.
- \_\_\_\_\_ (2018). Deep Learning, Machine Learning, Natural Language Processing, Computer Vision, Machine Reasoning, and Strong AI: Global Market Analysis and Forecasts.
- \_\_\_\_\_ (2018). Artificial Intelligence Hardware and Software Infrastructure.
- Vamathevan, J. *et al.*(2019). “Applications of machine learning in drug discovery and development” *Nature Review Drug Discovery*. pages 463-477.
- Waltz, E.(2016). “Gene-edited CRISPR mushroom escapes US regulation.” *Nature News* 532.7599.
- World Economic Forum(2015). Deep Shift: Technology Tipping Points and Societal Input.
- <http://www.fluenty.co/ko/>
- <https://mindslab.ai/kr>
- <https://obamawhitehouse.archives.gov/precision-medicine>
- <https://root-ai.com>
- <https://www.cbinsights.com/research/top-acquirers-ai-startups-ma-timeline/>
- <https://www.marshall.edu>



# 데이터사이언스 발전 전략



## VI

## 데이터사이언스 발전 전략

## 1. 교육

## □ 미국 데이터사이언스·인공지능 교육 현황

## ○ 데이터사이언스 학위 과정

- 데이터사이언스 교육에서는 미국이 단연 앞서가고 있음. 미국에서의 데이터 사이언스 관련 교육은 네 가지 범주로 나누어지며, 이들은 학사 프로그램(bachelor's programmes), 석사 프로그램(master's programmes), 박사 프로그램(doctoral programmes)과 자격증 프로그램(certificate programmes)임
- 이들 중에 가장 인기가 높은 프로그램은 석사 프로그램임. 그 이유는 데이터 사이언스가 융합 학문이므로 학사 과정에서 다루기 쉽지 않고, 또한 실무 응용성이 뛰어난 학문으로 기업에서 석사 학위 전문가를 주로 찾고 있으므로, 박사 과정 교육이 크게 주목받지는 않고 있음
- 자격증 프로그램은 정규 학위 과정은 아니나 실무 능력을 배양해 주는 자격증을 주는 과정이므로 인기가 있음
- 데이터사이언스 석사 과정
  - 2019년 석사 과정을 가진 대학은 모두 141개이며, 석사 과정이 데이터사이언스 교육의 핵심적인 역할을 하고 있음. 이들 중에서 우수한 프로그램을 가진 상위 10개 대학은 <표 6.1>과 같음
  - 학생 대 교수 비율이 낮을수록, 등록금이 저렴할수록, 프로그램에서 집중 분야(concentration areas)가 있다면 좋은 평가를 하도록 기준을 설정함

&lt;표 6.1&gt; 미국 상위 10개 데이터사이언스 석사 과정

순위	학교	석사 과정
1	Purdue University - West Lafayette, Indiana	Master of Science in Business Analytics and Information Management
2	University of Rochester - Rochester, New York	Master of Science in Data Science
3	New York University - New York, New York	MS in Data Science
4	Carnegie Mellon University - Pittsburgh, Pennsylvania	Master of Computational Data Science
5	Columbia University in the City of New York - New York, New York	Master of Science in Data Science
6	Colorado State University - Fort Collins, Colorado	Master of Science in Applied Statistics (MSAS)
7	University of Iowa - Iowa City, Iowa	Full-time Master of Business Analytics
8	North Carolina State University - Raleigh, North Carolina	Master of Science in Analytics
9	Georgia Institute of Technology - Atlanta, Georgia	Master of Science in Analytics
10	DePaul University - Chicago, Illinois	MS in Data Science

자료: <https://www.datasciencedegreeprograms.net>

#### - 미국의 데이터사이언스 박사 과정

- 2019년 미국에서 박사 과정을 개설한 학교는 57개이며, 주로 큰 규모의 석사 과정을 개설한 대학들이 운영 중에 있음
- 데이터사이언스에서 실무적 감각이 좋은 현장형 전문가를 양성하는 것을 목적으로 하는 석사 과정이 가장 인기 있으나, 향후 데이터사이언스에서 많은 방법론들이 연구되기 시작하면서 박사 과정의 인기가 더 높아질 것으로 전망됨

#### ○ 데이터사이언스 교육 과정

- 미국 대학들의 데이터사이언스 커리큘럼은 대학 내에서 다양한 조직으로 운영 되므로, 조직 형태에 따라서 커리큘럼에 큰 차이를 보이고 있음
- 2014년 8월 기준으로 석사 과정을 개설한 대학 중에서 웹사이트에 과목을 공개한 대학 15개 대학의 현황을 조사하여 정리한 것이 <표 6.2>임

- <표 6.2>에서 석사 과정의 특징은 두 가지로 나타남. 하나는 통계 분석을 강조한 점임. Exploratory Data Analysis 과목은 Probability and Statistics의 심화 과정이라고 볼 수 있고, Data Mining과 Data Visualization 과목도 통계적 분석 방법에 기초를 두고 있음. 두 번째로, 데이터사이언스의 사회 적용 측면을 강조하여, 이와 연관된 과목들이 총 16과목 중에서 4과목(Information and Social Network Analysis, Research Methods, Social Aspects of Data Science, Healthcare Analytics)이나 된다는 것임

<표 6.2> 석사 학위 프로그램의 핵심 과목들(2014년 8월 기준)

Course	Number of universities offering the course among 15 universities
Exploratory Data Analysis	10
Database	10
Data Mining	9
Data Visualization	8
Statistical Modelling	8
Machine Learning	6
Information Retrieval	5
Information and Social Network Analysis	4
Data Warehouse	4
Introduction to Data Science	3
Research Methods	3
Social Aspects of Data Science	3
Algorithms	2
Data Cleaning	2
Text Mining	2
Healthcare Analytics	2

자료: Song, I. Y. & Zhu, Y., Big data and data science: what should we teach?, 2016

#### ○ 인공지능 학위 과정

- 미국은 인공지능(AI) 교육에서 단연 앞서가는 국가로, 대학에서의 학사·석사·박사 과정과 기업 및 연구소의 자체 과정으로 운영되고 있음. 주로 석사 과정을 중심으로 인공지능 교육이 이루어지고 있으나 최근 들어 학사 과정도 상당한 인기를 얻고 있음

- 미국 대학의 인공지능 교육은 주로 컴퓨터 과학(computer science), 컴퓨터 공학(computer engineering), 계산 과학(computing science) 혹은 전기 공학(electrical engineering) 등의 이름을 가진 학과 혹은 학부(department or school)에서 이루어지고 있으며, 이들 학과에서 하나의 전공 분야(track or specialization)로 이루어지기도 함. 그러나 대학에 따라서 컴퓨터 과학/ 공학과를 별도로 두지 않고 전기 공학과 안에서 컴퓨터 과학/공학을 하나의 전공 분야로 두면서 교육하기도 함
- 인공지능 전공을 포함한 학과를 가지고 있는 대학의 수는 219개 대학임. 데이터사이언스와 인공지능의 4차 산업혁명의 주요 학문으로 대두되고, 둘의 경계가 차츰 흐려지면서 인공지능 전공이나 관련 학과를 가진 대학의 수는 더욱 늘어날 것으로 보임
- 인공지능 대학원 학위 과정 현황
  - 미국은 인공지능 및 기계학습 분야를 주로 컴퓨터 과학 혹은 컴퓨터 공학과에서 다루고 있으며, 우수한 학사 과정을 가진 대학은 상당수가 대학원 과정도 가지고 있음. 대학의 성향에 따라서 중점을 두는 교육 과정에 차이가 있음
  - 인공지능 대학원 교육 과정 상위 10개는 <표 6.3>과 같음

**<표 6.3> 인공지능 대학원 교육 과정 상위 10개 대학**

순위	대학 이름	교육 단위 이름	학위
1	Carnegie Mellon University	School of CS <sup>1)</sup>	MS <sup>2)</sup> and Ph.D In AI and Innovation
2	Massachusetts Institute of Technology	Department of CS, CS and AI Laboratory	MS and Ph.D. with speciality in AI
3	Stanford University	Department of CS, AI Laboratory	MS and Ph.D. in CS, Graduate Certificate in AI
4	University of California, Berkeley	Department of EE <sup>3)</sup> and CS, Berkeley AI Research Laboratory	MS and Ph.D.
5	University of Washington	Paul Allen School of CS and Engineering	MS and Ph.D. in CS and Engineering

1) CS: Computer Science

2) MS: Master of Science

3) EE: Electrical Engineering

&lt;표 6.3&gt; 인공지능 대학원 교육 과정 상위 10개 대학(계속)

순위	대학 이름	교육 단위 이름	학위
6	Cornell University	Department of CS <sup>1)</sup>	MS <sup>2)</sup> and Ph.D. in CS
7	Georgia Institute of Technology	Georgia Tech College of Computing	MS and Ph.D. in CS with specialization of interactive intelligence or machine learning
8	University of Illinois	Department of CS	MS and Ph.D. in CS
8	University of Texas (tie)	Department of CS	MS and Ph.D. in CS
9	University of Michigan	Department of EE <sup>3)</sup> and CS	MS and Ph.D. in CS and Engineering or ECE <sup>4)</sup>
10	University of Massachusetts	Department of CS	MS and Ph.D. in CS

1) CS: Computer Science

2) MS: Master of Science

3) EE: Electrical Engineering

4) ECE: Electrical and Computer Engineering

자료: <https://observatory.tee.mx/edu-news/the-top-10-ai-graduate-programs-in-the-us>

### ○ 인공지능 교육 과정

- 미국에서의 인공지능 교육의 중심은 석사 과정이므로, 해당 커리큘럼을 살펴봄. 대학원 교육 및 인공지능 교육에서 1위를 차지한 Carnegie Mellon 대학의 석사 과정을 대표로 살펴봄
- 커리큘럼의 핵심은 기계학습(machine learning), 언어 기술(language technologies), 혁신적 AI 시스템(innovative AI systems)이며, 특히 현실의 실질적인 문제를 풀기 위한 실습 프로젝트인 캡스톤프로젝트(capstone project)를 중요시함. 커리큘럼은 핵심 과목, 지식 요구 과목, 선택 과목 등으로 구성되어 있으며 <표 6.4>에 정리하였음

**<표 6.4> Carnegie Mellon 대학의 인공지능 석사 과정 커리큘럼**

구분	개설 과목
핵심 과목	<ul style="list-style-type: none"> <li>- AI and Future Markets</li> <li>- Law of Computer Technology</li> <li>- AI Engineering</li> <li>- AI Innovation</li> <li>- Capstone Project</li> </ul>
지식 요구 과목	<ul style="list-style-type: none"> <li>- Coding Bootcamp</li> <li>- Machine Learning</li> <li>- Machine Learning with Large Datasets</li> <li>- Natural Language Processing</li> <li>- Applied Machine Learning</li> <li>- Deep Learning</li> </ul>
선택 과목	총 29개의 선택 과목 중 최소 3개 과목 선택 및 이수

자료: <https://www.cs.cmu.edu/overview-programs>

- 따라서 Carnegie Mellon 대학에서 ‘Master of Science in Artificial Intelligence and Innovation’으로 석사 학위를 받으려면 최소 14과목(5개 핵심 과목, 6개 지식 요구 과목, 3개 이상의 선택 과목)을 이수해야 함. 다른 대학들은 경우에 따라서 8~10개 과목을 요구하고 석사 논문을 작성하면 학위를 주기도 함

#### ○ 데이터사이언스 관련 자격증

- 미국에는 정규 학위 과정은 아니지만 실무에서 데이터사이언스를 실행하는 데이터사이언스티스트나 미래에 데이터사이언스티스트가 되기를 희망하는 사람들을 위한 ‘데이터사이언스 자격증 프로그램(data science certificate program)’을 운영하는 대학들이 많음. 이는 실무에서 빅데이터를 다루는 전문가들이 학위가 없어도 전문성을 발휘하도록 자격증을 주는 제도임
- 대부분이 온라인 프로그램으로 운영되며, 보통 12~18 학점을 이수하고, 일정 시험을 통과하면 자격증을 받을 수 있는 제도임. 이 프로그램을 운영하는 대학에 따라서 미국 데이터사이언스위원회(Data Science Council of America, DASCА)로부터 인정을 받은 대학도 있고, 그렇지 않은 대학도 있음. DASCА는 미국의 데이터사이언스에 대한 표준과 신뢰성 측정을 담당하는 기관임

## □ 국내 데이터사이언스·인공지능 교육 현황

### ○ 데이터사이언스 학위 과정

- 인공지능은 이미 오래 전에 우리나라 대학에 소개되어 컴퓨터 과학(공학)과에서 강의되어 왔으나, 빅데이터를 다루는 빅데이터 학과 혹은 데이터사이언스 학과(전공)는 비교적 최근에 우리나라 대학에 소개되었음. 2012년 3월에 전국 최초로 미래창조과학부 사업의 지원으로 충북대학교에 빅데이터 전문가 석사 과정으로 비즈니스데이터융합학과가 개설되었음
- 2019년 기준 국내 데이터사이언스 학위 과정을 운영하고 있는 대학은 모두 42개이며, 온라인 교육이나 사이버 교육 프로그램은 제외하였음. 이들 42개 대학 중에서 대표적인 10개 대학의 프로그램을 살펴보면 <표 6.5>와 같음
- 한 학교에서 두 개 이상의 프로그램을 운영하는 대학들도 있음. 예를 들면 국민대학교는 일반대학원에 데이터사이언스 학과, 경영대학 경영학부에 빅데이터 경영통계 전공, 경영대학원에 빅데이터 경영 MBA 과정을 두고 있음. 이처럼 한 대학에서 두 개 이상의 프로그램을 운영하는 대학도 국민대 외에 고려대, 부산대, 성균관대, 세종대, 이화여대, 충북대 등이 있음

<표 6.5> 2019년 국내 대표 데이터사이언스 학위 과정

대학	소속	학위명	학과명	설립연도
고려대학교	정보보호대학원	공학 석사	빅데이터 응용 및 정보보안학과	2015-03
	컴퓨터정보통신대학원	공학 석사	빅데이터융합학과	2016-03
	세종캠퍼스 공공정책대학 경제통계학부	인문 학사	빅데이터전공	2017-03
국민대학교	일반대학원(인문사회계열)	일반 석사 일반 박사	데이터사이언스학과	2014-03
	경영대학 경영학부	경영 학사	빅데이터경영통계전공	2013-03
	경영대학원	경영 석사	빅데이터경영MBA과정	2013-03
단국대학교	일반대학원(공학 계열)	공학 석사	데이터사이언스학과	2014-12
성균관대학교	일반대학원 지능정보융합원	공학 석사	데이터사이언스융합학과	2018-03
	정보통신대학원	공학 석사	빅데이터학과	2016-03

&lt;표 6.5&gt; 2019년 국내 대표 데이터사이언스 학위 과정(계속)

대학	소속	학위명	학과명	설립연도
세종대학교	소프트웨어융합대학	공학사	데이터사이언스학과	2017-03
	경영전문대학원	빅데이터 경영학 석사	빅데이터경영학과	2018-03
아주대학교	일반대학원(공과대학)	공학 석사	인공지능데이터사이언스학과	2017-09
연세대학교	정보대학원	정보시스템 석사	빅데이터석사과정 (기업연계 계약학과 형태)	2014-03
이화여자대학교	빅데이터분석학협동과정	데이터분석학 석사	빅데이터분석학	2014-09
	경영전문대학원	경영 석사	빅데이터MBA과정	2015-09
충북대학교	비즈니스데이터융합학과 (대학원, 계약학과)	빅데이터 전문가 석사	비즈니스데이터융합학과	2012-03
	경영정보학과	데이터 융합학사	빅데이터학과	2015-03
KAIST	경영대학 정보미디어 경영대학원	경영 석사	정보미디어MBA	2012-09

## ○ 데이터사이언스 교육 과정

- 데이터사이언스 교육 커리큘럼은 대학마다 다양하게 구성하고 있으나, 아래와 같은 세 가지 분야의 기초 지식에 중점을 두고 있음
- 첫 번째, 전산 과학 분야로 프로그래밍, 데이터베이스 및 알고리즘 등을 다룸.
- 두 번째, 통계학 및 응용 수학 분야로 확률과 통계, 이산 수학, 통계적 모델링, 기계학습, 최적화 이론 및 계산 프로그래밍 등을 다룸. 마지막으로, 실무 적용에 필요한 지식 분야로 경영정보시스템, 데이터 시각화, 시뮬레이션, 빅데이터 프로젝트 등을 다룸

## ○ 인공지능 학위 과정

- 국내 대학의 인공지능 교육은 미국과 동일하게 컴퓨터과학과, 컴퓨터공학과 등에서 하나의 전공 분야로 교육되어 왔음
- 최근 인공지능이 4차 산업혁명의 핵심으로 부각됨에 따라서 정부는 2018년 ‘인공지능 대학원 지원 사업’을 발표하고 우선적으로 3개 대학(고려대, 성균관대, KAIST)을 선정함. 고려대와 KAIST는 2019년 9월부터 AI 전문 대학원을 설치할

예정이며, 성균관대는 이미 2018년 3월에 인공지능·로봇학과를 설립하여 운영하고 있으며, 이를 전문 대학원으로 격상시킬 예정임

- 이들 3개 대학을 포함하여 인공지능 학과(전공)를 가진 대학들을 조사하여 보면 <표 6.6>과 같이 11개 대학이 있음. 이들 대학 외에도 많은 대학의 컴퓨터 과학과나 컴퓨터공학과 등에서 AI를 하나의 분야(혹은 전공)로 교육하고 있음

<표 6.6> 2019년 국내 인공지능 학위 과정 개설 현황

대학명	소속	학위명	학과(전공)명	설립연도
경희대학교	사이버대학	학사	AI사이버보안전공	2019-03
고려대학교	일반대학원(공학계)	석박사 통합 및 공학 박사	인공지능학과	2019-09
단국대학교	정보·지식재산대학원	공학 석사	인공지능공학과	2017-03
대구가톨릭 대학교	일반대학원	의학 석사	학과 간 협동과정 인공지능융합의학과	2018-03
서강대학교	연계전공	학사	인공지능연계전공	2019-03
성균관대학교	일반대학원 지능정보융합원	공학 석사 공학 박사	AI·로봇학과	2018-03
세종대학교	일반대학원	석사	인공지능언어공학과	2018-03
연세대학교	공학대학원	공학 석사	인공지능	2018-10
전남대학교	융합전공	공학사	IoT·인공지능융합전공	2019-03
청주대학교	소프트웨어융합부	공학사	인공지능소프트웨어전공	2017-03
KAIST	공과대학원	공학 석사 공학 박사	AI 대학원	2019-09

#### ○ 인공지능 교육 과정

- 우리나라에서는 인공지능을 주로 교육하는 학과(학부)는 서울대의 경우와 같이 기존의 컴퓨터과학과(학부)나 컴퓨터공학과(학부)임. 이 학과(학부)에 하나의 중요한 과목으로 AI가 있고, 대학원에서는 AI 전공을 선택하여 공부하거나 논문을 쓸 수 있음

- <표 6.6>에 있는 인공지능학과의 대표적인 경우로 단국대학교 정보·지식재산 대학원의 인공지능공학과(Department of AI Engineering)의 교과 과정을 살펴봄. 과목 수는 매우 많으나 매학기 모두 개설되는 것은 아니며, 필요에 따라서 일부만 개설됨
  - 뇌과학 개론, 뇌영상 과학, 인공지능 개론, 인공지능 프로그래밍 입문, 인공지능 알고리즘, 심층학습, 기계학습, IoT 응용시스템, 빅데이터 처리, 데이터 마이닝, 자율 주행 시스템, 지능형 로봇, 원격 헬스 케어, 드론 운용기술, 자연어 처리와 챗봇, 인공지능 설계
- 위 과목들을 통해 대학원의 인공지능 교육 과정을 대략적으로 이해할 수 있음. 서울대학교의 컴퓨터공학부의 커리큘럼과 큰 차이가 있음

#### ○ 데이터사이언스 관련 공인 자격 제도

- 빅데이터 공인 자격 제도
  - 빅데이터 전문 인력에 대한 기업의 인력 요구가 계속 증가하면서, 기업이 원하는 빅데이터 전문 인력 양성을 위한 자격 제도에 대한 관심이 높아지고 있음. 자격 제도는 국가 자격, 민간 자격으로 크게 나눌 수 있음
  - 빅데이터와 관련해서 한국데이터베이스진흥원에서 운영하는 데이터 전문가 자격 검정으로 ‘데이터 분석 전문가’와 ‘데이터 분석 준전문가’ 자격 제도가 있음. 한국데이터베이스진흥원은 2014년에 데이터 분석 자격 검정 민간 자격을 등록하여 제1회 데이터 분석 전문가 자격시험을 시행했음. 이 시험에 대하여 2015년에 국가 공인을 승인받았고, 이 시험은 필기시험과 실기시험 까지 합격해야 자격 취득이 가능하며, 평균 합격률이 5% 정도로 합격이 어려움
  - 그러나 데이터 분석 준전문가 자격시험은 필기시험만 치며, 합격률은 30% 정도임. 건강보험심사평가원은 2010년에 ‘DW 분석사’ 국가 공인 사내 자격 제도를 운영하고 있음. 여기서 DW는 데이터 웨어하우스(data warehouse)를 의미하는 용어로 기업 활동을 지원하는 대규모 데이터베이스를 의미함

#### - 인공지능 공인 자격 제도

- 대통령 직속 4차산업혁명위원회는 2018년 5월 15일 제6차 회의를 개최하고 ‘인공지능 R&D 전략’을 심의·의결했음. 이 전략의 목표는 향후 5년간 (2018~2022) 2.2조원의 R&D 자금을 투자하여 ‘세계 4위 AI 강국 도약’, ‘AI 우수 인재 5천 명 확보’, ‘AI 데이터 1.6억 건 구축’을 달성하겠다는 것임. 이 전략의 추진 방법으로는 ‘세계적 수준의 AI 기술력 확보’, ‘최고급 인재 양성’, ‘개방 협력형 연구 기반 조성’을 내세웠음
- 그러나 추진 방법 중에 들어 있는 ‘최고급 인재 양성’ 속에는 인공지능 전문가 양성을 위한 ‘인공지능 공인 자격 제도’는 들어 있지 않음. 대학에서만 인공지능 우수 인재 5천 명을 양성하기는 어려움이 있으므로 ‘인공지능 공인 자격 제도’의 도입을 권장함

### □ 국내 데이터사이언스 교육 발전 전략 제안

- 빅데이터와 인공지능 교육을 아우르는 데이터사이언스 교육 커리큘럼의 작성
  - 빅데이터나 인공지능이 모두 데이터 기반의 기술이므로 이 둘을 모두 아우르는 데이터사이언스 교육 커리큘럼이 바람직함. 현재는 빅데이터 중심의 데이터사이언스 교육에서 인공지능이 거의 포함되어 있지 않고, 인공지능 교육에서는 빅데이터 교육이 큰 비중을 차지하고 있지 않음
  - 따라서 현재의 데이터사이언스 교육 커리큘럼에서 기계학습을 중심으로 하는 인공지능 교육이 충분히 이루어져야 함. 향후에는 과학기술의 발전에서 보면 빅데이터와 인공지능의 영역이 매우 가까워질 것임
- 데이터사이언스 공인 자격 제도의 신설
  - 현재는 빅데이터 공인 자격 제도가 민간을 중심으로 활발하며, 인공지능 공인 자격 제도는 없음. 국가에서 권장하는 데이터사이언스 공인 자격 제도가 신설 되었으면 함. 대학에서 배출하는 인재 수는 사회에서 요구하는 수요를 충족시킬 수 없을 것임
  - 지난 2018년 6월 26일에 열린 대통령 직속 4차산업혁명위원회 제7차 회의에서 내놓은 ‘데이터 강국’을 위한 ‘데이터 산업 활성화 전략’의 청사진 중에는 2019년

부터 ‘데이터 분석 국가 기술’ 국가 자격 제도를 신설하여 빅데이터 분석사를 양성하고, 데이터 전문 인력을 추가로 5만 명 정도 양성하겠다는 계획이 있었음. 이 계획이 실현되면 데이터사이언스 공인 자격 제도가 될 것임

○ 빅데이터 인력 양성 문제의 조속한 해결

- 한국데이터진흥원의 조사에 의하면 2017년 말 우리나라의 데이터 전문 인력과 빅데이터 인력은 <표 6.7>에서 보는 바와 같이 각각 109,320명과 9,955명임. 여기서 데이터 전문 인력은 데이터를 다루는 모든 인력으로 데이터사이언티스트, 데이터 아키텍처, 데이터 개발자, 데이터 분석가, 데이터베이스 관리자, 데이터 컨설턴트, 데이터 기획/마케터를 포함함. 빅데이터 인력은 빅데이터를 다루는 전문가로 데이터사이언티스트, 빅데이터 개발자, 빅데이터 분석가 등으로, 데이터 전문 인력에 포함됨

<표 6.7> 데이터 전문 인력과 빅데이터 인력 수요 추정

분류	2017년	2020년	2022년(추정치)
데이터 전문 인력	109,320명	122,657명	150,000명 (4차산업혁명위원회 추정치)
빅데이터 인력	9,955명	15,963명	26,000명 (기업 수요 조사 결과 추정치)

자료: 한국데이터진흥원, 2018 데이터산업백서, 2018

- 데이터사이언티스트 등으로 이루어진 빅데이터 인력은 2017년에 약 1만 명이나 2020년에 약 1만 6천 명 정도 필요하고, 2022년에는 2만 6천 명(한국데이터진흥원의 기업 수요 조사 결과와 4차산업혁명위원회의 ‘데이터 산업활성화 전략’에서 기업 빅데이터 이용률이 2017년 7.5%에서 2022년 20%로 증가하기 때문에 필요한 인력)이 필요함
- 따라서 2017년에 비하여 2022년에는 데이터 산업의 급신장으로 인하여 추가로 약 1만 6천 명의 빅데이터 인력이 필요함. 이러한 인력을 공급하기 위해서 대학의 정규 교육, 국가 인증 자격 제도에 의한 인력 양성, 민간 인증 자격 제도의 확충 등이 필요함. 대학의 데이터사이언티스트 양성 계획이 시급한 실정임

### ○ 데이터사이언스 학사 과정 개설

- 국내 데이터사이언스 과정은 대부분 대학원에 개설되어 있으나, 데이터사이언스란 학문이 뿌리를 내리고, 우수 인력을 양성하기 위해서는 학사 과정부터 개설하여 운영할 필요가 있음. 이를 통해 데이터사이언스가 학사 과정 → 석사 과정 → 박사 과정으로 이어지는 신학문으로 성장하면 빅데이터·인공지능 및 관련 산업의 진흥에 크게 기여할 것임

### ○ 대학에서 데이터사이언스 교양 과목의 개설

- 데이터사이언스 전공 학생 이외의 학생들에게도 데이터사이언스에 관한 소양 교육이 필요하며, 대학의 교양 과목으로 개설하여 많은 대학생들이 수강할 수 있도록 유도하는 것이 바람직함. 21세기는 데이터 경제 시대이므로, 데이터 경제 시대를 살아가는 지식인들에게 데이터사이언스 교육은 필수적인 소양이 될 것임

### ○ ‘데이터 고속도로’를 뚫고, 빅데이터·인공지능 학습용 데이터를 많이 만들어야

- 문재인 대통령은 2018년 8월 31일 경기도 판교 스타트업 캠퍼스에서 열린 ‘데이터 경제 활성화 규제혁신’ 현장 방문 행사에 참석해 데이터 경제 육성 방안을 발표하면서 “산업화 시대의 경부고속도로처럼 이제는 데이터 고속도로를 구축 하겠다.”고 말했음
- 또한 공공 데이터를 원시 데이터 형태로 최대한 모으고 분야별 빅데이터 센터를 구축하고, 빅데이터·인공지능 학습용 데이터를 많이 모아 개방하겠다는 정책을 발표했음
- 정부의 추진 방향은 옳으며, ‘데이터 고속도로’를 건설하여 어디에서나 누구나 신속하고 자유롭게 분야별로 빅데이터·인공지능 학습용 데이터를 사용할 수 있어야 함. 이렇게 되면 자연스럽게 데이터사이언스는 발전할 것임
- ‘데이터 고속도로’를 효율적으로 깔기 위해서도 국내 클라우드 산업을 적극적으로 육성할 필요가 있음

## 2. R&D

### □ 해외 주요국 데이터사이언스 R&D 전략 현황

#### ○ 미국의 R&D 전략

- 인공지능 기술을 선도하고 있는 미국은 민간이 주도하는 연구 환경이 조성되어 있지만, 건강, 교육, 환경 등의 공공 영역에서는 민간 투자 활성화를 위하여 정부가 적극적인 중장기 R&D 전략을 발표하고, 또한 투자 주체로도 나서고 있음. 정부의 주요 R&D 전략 활동은 다음과 같음
  - AI와 관련이 깊은 뇌과학 분야의 기초 연구를 위해 2013년부터 10년간 총 30억 달러를 투자하는 ‘Brain Initiative’ 시작(2013.4.)
  - ‘AI, 자동화, 그리고 경제(Artificial Intelligence, Automation and the Economy)’ 보고서 발표함(2016.10.). AI 확산이 일자리와 기존 산업에 미치는 영향 및 대책에 대한 발표임
  - ‘AI R&D 전략 계획(National Artificial Intelligence Research and Developing Strategies Plan)’ 발표함(2016.10.). 정부의 장기 투자를 강화하기 위한 R&D 전략 방향을 제시함
  - ‘The 2018 White House Summit on Artificial Intelligence’ 개최함(2018.5.). 혁신을 위한 규제 장벽 제거와 ‘AI 위원회(Committee on AI)’ 설치 계획 등 AI 활성화를 위한 정책적 실행 안을 논의한 내용임
- 이러한 미국 정부의 꾸준한 AI에 대한 관심과 장기 투자는 오늘날 미국이 AI를 비롯한 뇌과학, 자율 주행 자동차 등에서 선도적인 기초 원천 기술을 확보하게 된 주요 원인임
- 미국은 기초 원천 기술 분야에서 명확한 정부의 투자 계획을 발표함과 동시에, 사회 문제와 직결되는 응용 서비스의 경우에는 민간 투자 영역으로 이양하여 시장 주도의 원동력을 만들어 낼 수 있도록 간접 지원 방식을 취하고 있는 것이 특징임
  - 미국 국방부 방위고등연구계획국(DARPA)은, 2009년부터 민·관 합동 시냅스(SyNAPS) 프로젝트를 통해 차세대 두뇌형 반도체인 뉴로모픽(Neuromorphic) 칩 개발을 지원하였으며, 이를 통해 IBM은 2014년 100만 개의 뉴런 및 2억

5,600만 개의 시냅스가 작동하는 뉴로모픽칩을 개발할 수 있었음 (조선비즈, 2012)

#### ○ 일본의 R&D 전략

- 일본은 인공지능 기술 경쟁력 확보를 위해 산업 경제 관점의 혁신 성장 정책으로 ‘일본재흥전략(2016.6.)’과 이를 발전시킨 ‘미래투자전략(2017.6.)’, 과학기술 관점의 기술 정책으로 ‘제5기 과학기술 기본계획(2016.1.)’ 등이 있음. 또한 혁신 산업 육성에 초점을 둔 ‘신산업구조비전(2017.5.)’과 인공지능에 관한 ‘AI 산업화 로드맵(2017.2.)’ 등을 발표하여, 단계적으로 치밀한 전략을 세우고 추진하고 있음
- 특히 ‘일본재흥전략’에서는 생산성 혁명 차원에서 AI 경쟁력 확보와 적극적인 투자를 강조하였음. 또한 일본이 보유한 강점인 로봇, 제조업 등과의 융합을 통한 차세대 인공지능 R&D를 추진하면서 자율 주행, 스마트 공장, 소형 로봇 등의 신기술 도입을 통한 고부가가치 창출을 위해 2020년까지 30조 엔 투자 계획을 제시하였음. 그리고 새로운 대학과 대학원의 창설과 실리콘밸리 - 일본의 연계 프로젝트 등 인재 양성에 관한 시책에서 인공지능 분야의 기술 및 법·제도에 대한 교육을 강조하였음

#### ○ 중국의 R&D 전략

- 중장기 전략
  - 중국 국무원은 ‘인터넷 플러스 3년 액션플랜(2016)’, ‘차세대 AI 발전 계획(2017.7.)’을 발표하면서, 2030년까지 인공지능 핵심 산업에 1조 위안(약 180조 원)을 투자하고 관련 산업 규모를 10조 위안까지 확대하겠다고 발표했음. 인공지능 발전 계획에서는 6대 중점 임무(개방·협력형 인공지능 과학 기술 혁신 체계 구축, 최첨단·고효율의 스마트 경제 육성, 차세대 인공지능 프로젝트의 선도적 추진 등)를 핵심적으로 제시하고 있음. 이와 같은 장기적인 정책 집행을 위한 컨트롤 타워로 2017년에 ‘차세대 인공지능 발전계획 위원회’를 설립했음
- 인재 양성 전략
  - 중국은 국가 차원의 대규모 선행 투자로 인공지능 R&D와 산업육성은 물론 인재 양성을 종합적으로 추진하고 있음. 중국은 3단계 국가 교육 전략을 통해

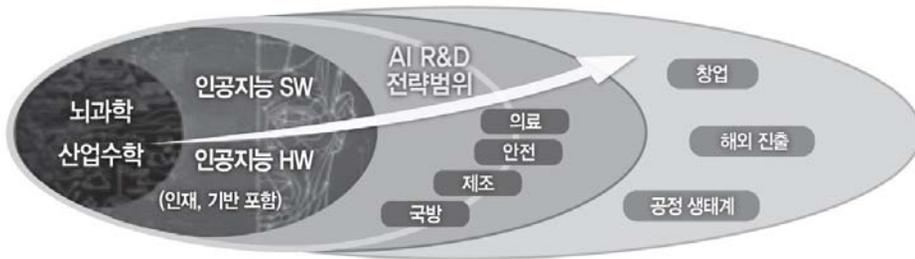
‘전 국민 스마트 교육 프로젝트’를 실시하고, 초·중·고등학교에 인공지능 관련 과목을 개설하여 AI 교육을 확산할 것이라고 밝혔다. 중국은 2022년까지 인공지능 교수 500명, 대학원생 5,000명을 육성할 계획이며, 중국 교육부, 베이징대, 시노베이션벤처스가 공동으로 출자하여 교육 기관을 만들어 인공지능 교육이 필요한 모든 조직에 무료 교육을 실시할 예정이다

## □ 국내 데이터사이언스 R&D 전략 현황

### ○ 빅데이터·인공지능 R&D 육성 전략

- 우리나라도 4차 산업혁명의 중심 기술인 빅데이터·인공지능의 발전에 큰 관심을 가지고 최근에 다음과 같은 여러 가지 정책과 전략을 발표해 왔음
- ‘I-Korea 4.0 실현을 위한 인공지능 R&D 전략’ (4차산업혁명위원회, 2018a)
- ‘데이터 산업 활성화 전략, I-Korea 4.0 데이터 분야 계획’ (4차산업혁명위원회, 2018b)
- ‘4차 산업혁명 대응 과학기술·ICT 인재 양성 지원 계획 - I-Korea 4.0 인재 Action Plan’ (관계부처합동, 2018. 11.)
- ‘데이터·AI 경제 활성화 계획[‘19~‘23년] - 혁신 성장 전략 투자’ (관계부처합동, 2019. 1. 16.)
- 위의 전략과 계획으로부터 정부의 주요 육성 전략의 기술 혁신 특징을 보면, 학습 가능한 양질의 데이터와 고성능 컴퓨터 및 차별화된 알고리즘 확보가 AI 서비스의 경쟁력을 결정하는 핵심 요소로 보고 있음
- 본 전략의 범위로는 <그림 6.1>에서 보는 바와 같이 뇌과학, 산업수학, 통계학 등의 기초 과학을 기반으로, 다음 단계의 기술 개발은 인지, 학습 등을 통한 인공지능 소프트웨어와 인공지능 성능 향상을 위한 하드웨어를 개발 및 연구하는 것임. 이 때 인공지능 개발 핵심 인재 양성이 필요하고, 적절한 데이터·컴퓨팅 파워가 제공되어야 함

### <그림 6.1> 인공지능 발전 전략



자료: 4차산업혁명위원회, I-Korea 4.0 실현을 위한 인공지능 R&D 전략, 2018a

- 다음 단계는 인공지능 R&D의 전략 범위로 의료, 안전, 제조, 국방 등을 주로 고려하고, 마지막 단계로 창업, 해외 진출, AI 플랫폼을 기반으로 타 산업과의 협력을 통한 개방형 생태계 구축 등의 분야에 관심을 갖는 것임
- 정부의 중점 추진 과제들은 다음과 같은 것임
  - 공공 분야(국방, 안전, 의료 등)에 대해 관련 부처와 대형 프로젝트를 공동 기획하여 5년간(2020~2024년) 2천억 원 규모로 예타 추진함
  - ‘AI R&D Challenge’ 대회를 ‘AI R&D Grand Challenge’로 확대 개편. Challenge 방식의 R&D는 정부는 문제만 제시 → 민간은 문제 해결을 위한 선행 연구 → 정부는 기술 개발 경진대회를 통해 선별된 우수 연구팀에 후속 연구비 지원을 하는 방식임
  - AI를 활용한 무인·지능형 스마트팜(데이터 팜) 구축 및 개인 건강 맞춤형(당뇨, 고혈압 등) 식물 생산을 위해 10년간(2020~2029년) 8천억 원 예타 추진
  - 스스로 학습할 수 있는 신경망 기반 인공지능 컴퓨팅(심층학습 알고리즘, 비정형 데이터 모델링 방법론, 기계학습을 위한 이론 등) 아키텍처 연구를 위해 ‘ICT 기초 원천 연구 사업’ 예타 추진(총 1조 원 규모)
  - 인공지능 고급 인재 양성(2022년까지 1,400명 규모)을 위한 R&D 사업
  - 인공지능 대학원 신설: 2022년까지 6개 대학원 신설(2019년에 KAIST, 고려대, 성균관대 3개 대학원 신설)

- 대학 연구 센터를 활용하여 고급 인재를 양성하기 위해 대학의 선도 연구 센터를 2022년까지 2개 추가 선발. 대학 ICT 연구 센터(ITRC) 중 인공지능 분야 확대
- 국제 공동 연구 지원: 해외 인공지능 우수 기관과의 긴밀한 협력 연구를 통해 2022년까지 50여명의 국내 고급 인재의 역량 향상
- 미국, 중국, 프랑스 등에 우리 대학들이 AI 관련 공공 연구 센터 또는 교육 프로그램을 마련할 경우 우선 지원
- 국내 인공지능 우수 인재를 대상으로 글로벌 기업 내부에서 우수 인력과 함께 리얼 데이터에 기반한 인공지능 연구를 수행할 경우에 집중 지원
- 해외 현황, 국내 경쟁력 등에 대한 전문적인 분석 기반으로 인공지능 정책을 총괄 지원하는 전담 조직을 신설하고, 인공지능 PM을 신설하여 총괄 조정 하도록 역할 부여

#### ○ 빅데이터·인공지능 시대 대응 전략

- 지금은 빅데이터·인공지능 경제 시대라고 볼 수 있음. 빅데이터·인공지능 경제란, 빅데이터를 기반으로 인공지능 기술을 활용하는 인프라를 구축하여 새로운 제품과 서비스를 창출하는 경제를 말함. 4차 산업혁명 기술 중에서 빅데이터와 인공지능이 가장 영향력 있는 기술이므로 지금은 빅데이터·인공지능 경제 시대라고 불러도 과언이 아님. 정부의 주요 계획은 다음과 같음
  - 데이터 가치 사슬 전주기 활성화
    - 공공·민간 분야(금융, 통신 등)별로 다양한 데이터를 수집하여 제공될 수 있는 빅데이터 센터를 100개 개설하고, 양질의 데이터가 결합·유통되고 새로운 서비스가 창출될 수 있는 빅데이터 플랫폼을 10개 구축
    - 중소·벤처기업이 데이터를 활용한 새로운 서비스를 개발할 수 있도록 데이터 구매 및 가공 비용을 지원하는 바우처 사업 추진(1,640개/년)
  - 세계적 수준의 인공지능 혁신 생태계 조성
    - 인공지능 관련 기초 연구(뇌과학, 산업수학, 통계학 등), HW(지능형 반도체, 양자 컴퓨팅 등) 및 소프트웨어 기술 개발에 집중하고 인공지능 분야 R&D를 고난이도 기술 개발에 적합한 Challenge 방식으로 추진

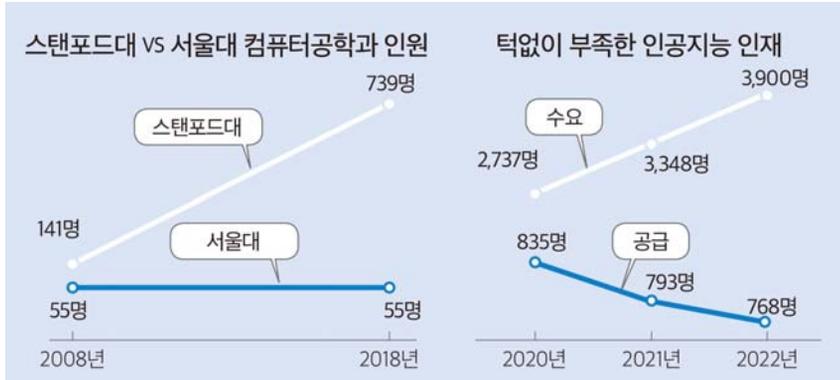
- 인공지능 전문 기업의 기술 개발을 집중적으로 지원하고 해외 및 국내 기업 간의 협력 프로젝트를 지원하여 기술 경쟁력 강화
- 데이터와 인공지능의 융합 촉진
  - 기업, 대학, 연구소가 한 곳에 모여 데이터와 인공지능 관련 연구 개발, 인력 양성, 창업 등을 지원할 수 있는 인공지능 융합 클러스터 조성 구축 (2020~2029년, 총 1조 원 규모 예타 추진)
  - 데이터와 인공지능의 활용을 촉진하는 제도(국가정보화기본법 개정) 마련과 안전한 활용을 보장하기 위한 제도(개인정보 관련 법령개정, 인공 지능 윤리 가이드라인 마련)를 강화
- 빅데이터·인공지능 경제 민관합동 태스크포스의 지속적 운영
  - 모든 산업의 혁신적인 변화를 이끌어 빅데이터·인공지능 경제 선도 국가로 도약하기 위하여 과학기술정보통신부 중심의 민관 합동 태스크 포스를 지속적으로 운영하고 주기적으로 점검

#### □ 국내 데이터사이언스 R&D 발전 전략 제안

- 위에서 미국, 일본, 중국의 R&D 전략을 간단히 살펴보고, 정부에서 추진하고 있는 R&D 전략을 간단히 기술하였음. 앞으로의 바람직한 R&D 전략으로 필자들의 의견을 다음에 몇 가지로 나누어 제안하고자 함
- 민간 기업 R&D가 더욱 활성화되도록 여건을 조성해 주어야 함
  - 앞에서 4차 산업혁명의 중심 기술인 빅데이터·인공지능의 발전을 위하여 정부 차원에서 작성한 전략과 계획은 전반적으로 우수하다고 보겠음. 그러나 정부 주도의 전략과 계획이 대부분임
  - 빅데이터·인공지능 연구가 활성화되려면 민간 기업 차원에서의 R&D가 활성화 되어야 함. 따라서 민간 기업들이 R&D에 더 투자하도록 여건을 조성해주는 것이 정부가 할 중요한 역할임. 이를 위하여 R&D 투자비에 대한 과감한 세금 삭감, R&D를 열심히 하는 기업에게는 법인세를 인하해 주거나 다른 세제 혜택을 주는 방안도 고려할 만함

- 우리나라는 국가 총 연구 개발비 중에서 민간이 차지하는 비율이 76.2%(2017년 기준) 정도로 정부에서 투자하는 연구 개발비만으로는 급변하는 빅데이터/ 인공지능 기술의 발전을 따라가기 어려움
- 과감한 ‘오픈 데이터’ 정책을 펴 빅데이터 산업을 키워야 함
- 정부의 각 부처에서 만드는 공공 데이터는 가장 양질의 빅데이터임. 이런 양질의 공공 데이터가 개인 정보 관련 3개의 법(개인정보보호법, 정보통신망법, 신용정보법)의 규제에 막혀 제대로 사용되지 않고 있음. 미국처럼 가명 정보만이라도 사용이 허락되면 빅데이터 산업이 커질 수 있을 것임
  - 공공 데이터는 특별한 문제가 없으면 기본적으로 일반에게 공개하는 ‘오픈 데이터’ 정책을 펴나가기를 권함. 이 정책이 받아들여지면 데이터를 활용하는 많은 벤처 기업이 탄생할 것이고, 대학에서 빅데이터·인공지능 교육에도 매우 유용하게 사용될 것임
- 빅데이터·인공지능 인력 양성을 위해 과감한 교육·연구 개혁이 따라야 함
- 빅데이터·인공지능에 필요한 인력이 급증하고 있으나 대학에서의 인력 양성은 이를 뒷받침하지 못하고 있음. 예를 들어 서울대학교의 경우 인공지능을 가르치는 공과대학의 컴퓨터공학부의 학부 입학 정원은 <그림 6.2>에서 보면 55명으로 10년 전(2008년)이나 지금(2018년)이나 동일함. 그러나 입학 정원을 수요에 따라서 순발력 있게 바꾸어주는 미국 대학의 경우는 필요한 경우 빠르게 변화에 대처하고 있음. 스탠포드 대학의 경우 컴퓨터공학부의 10년 전 입학자 수는 141명이었으나 지금은 739명으로 5배 이상 늘었음. 우리나라는 경직된 교육·연구 시스템을 가지고 있어 순발력 있게 수요에 대처하지 못하고 있으므로 개혁이 필요한 부문임. 인공지능 인재에서도 우리나라는 2022년에 수요는 3,900명이나 공급은 768명에 불과할 것으로 예측됨

<그림 6.2> 서울대와 스탠포드대의 컴퓨터공학 및 인공지능 인재 수 비교



자료: 소프트웨어정책연구소, 2018

- 빅데이터 · 인공지능 분야의 효율적인 연구 정책 수행을 위해 전문성을 가진 공무원을 배치해야 함
- 정부 차원에서 추진하는 빅데이터 · 인공지능 연구 전략과 계획은 크게 부족함이 없다고 봄. 그러나 이를 집행하고 점검하는 현장 인력(공무원)이 전문성이 없으며, 전문가들의 의견을 경청하지 않는 경우가 흔함. 따라서 전문성을 가진 인재를 과감하게 선발하여 현장 공무원으로 배치할 필요가 있음. 전문성이 있어야만 효율적인 연구 정책 실천이 가능할 것임

### 3. 산업

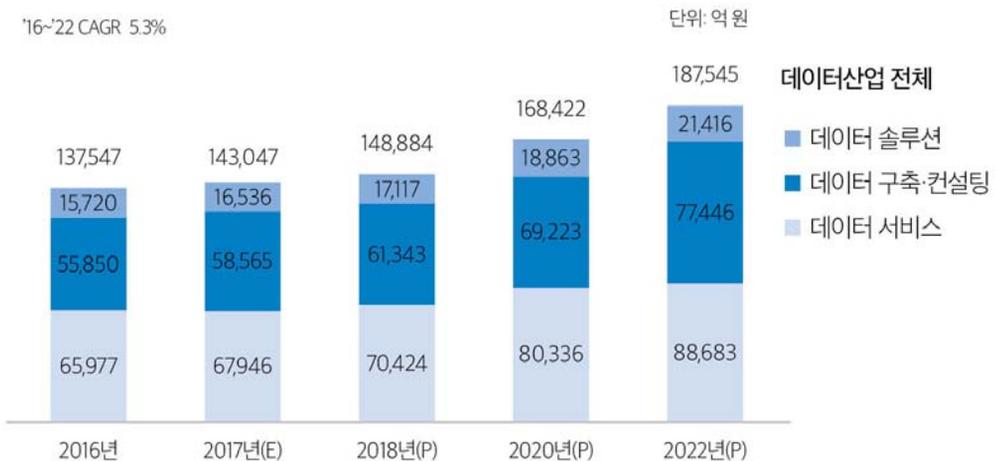
#### □ 국내 데이터사이언스 산업 전략 현황

- 4차 산업혁명의 핵심은 데이터에 있고, 데이터 산업은 4차 산업혁명의 기간 산업으로 간주되어야 한다는 것을 많은 전문가들이 주장하고 있음
- 데이터가 새로운 형태의 자산으로 주목 받으면서 데이터 유통에 기반한 ‘데이터 경제’ 개념이 비롯되었는데, 데이터 경제란 ‘데이터에 접근하고 활용할 수 있도록

협업하는 과정에서 데이터 생산, 인프라 제공, 연구 조사, 데이터 소비 등 서로 다른 역할을 담당하는 구성원으로 이뤄진 생태계'를 의미함 (European Commission, 2017)

- 한국데이터진흥원에서 발간한 '2018 데이터산업백서'에서는 우리나라의 데이터 산업이 2017년 14조 3,047억 원으로 형성되어 2010년 이후 연평균 7.5%의 성장세를 유지하고 있는 것으로 보고하였음. 동 백서에서는 세계 데이터 산업 시장의 규모는 솔루션의 시장이 2018년 999억 달러 규모로 2017년부터 2022년까지의 연평균성장률(CAGR)이 10.3%로 2022년에는 1,461억 달러로 성장할 것으로 예상하고, 온라인/오프라인 데이터 서비스 시장의 규모는 2016년 1조 6,000억 달러로, 2015년 1조 5,000억 달러 대비 6.4% 증가했으며, 2017년부터 2020년까지는 연평균 약 4% 내외의 성장세를 유지할 것으로 전망하고 있음. 이에 세계 데이터 산업 시장의 규모는 2018년 1조 9,000억 달러로 추산됨

<그림 6.3> 데이터 산업 시장 전망

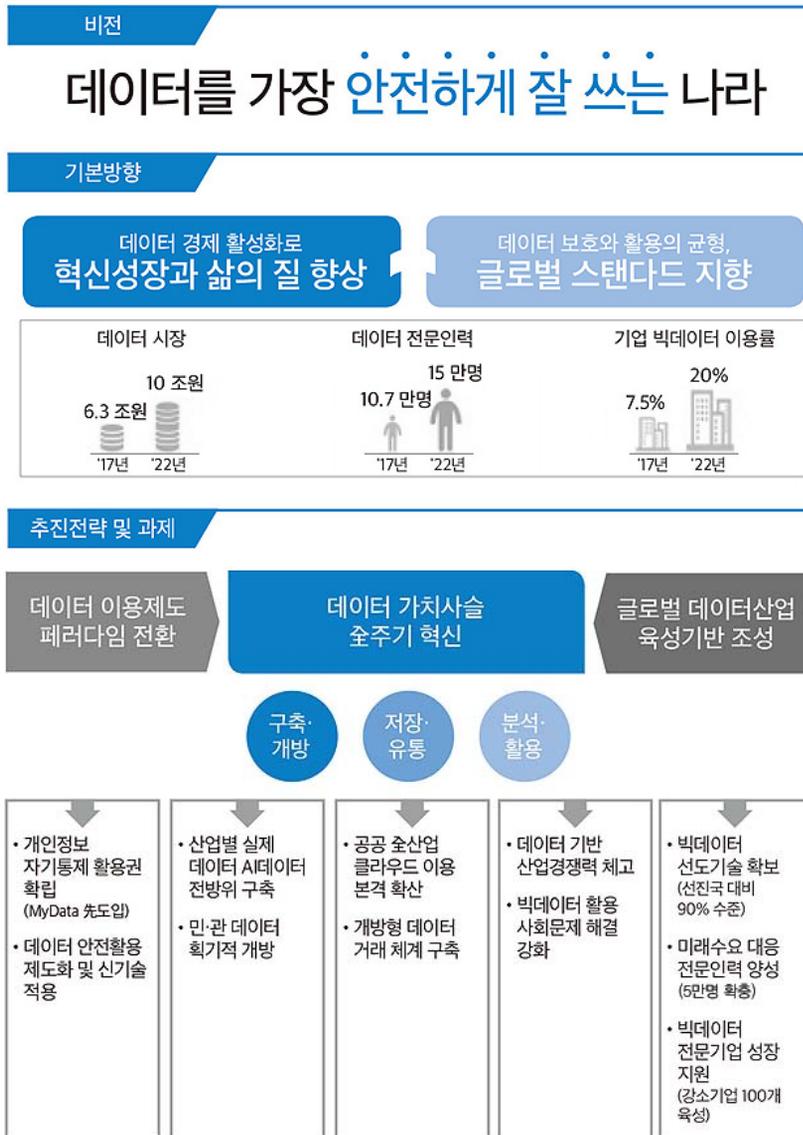


자료: 한국데이터진흥원, 2018 데이터산업백서, 2018

- 이와 같은 데이터 산업의 중요성을 인지한 정부는 2018년 6월 데이터 산업의 활성화를 위하여 '데이터를 가장 안전하게 잘 쓰는 나라'의 비전하에, '데이터를

통한 혁신 성장과 삶의 질 향상, ‘글로벌 수준의 데이터 보호와 활용’을 기본 방향으로, ‘데이터 이용 제도 패러다임 전환’, ‘데이터 가치 사슬 전주기 혁신’, ‘글로벌 데이터 산업 육성 기반 조성’의 3가지 전략을 수립함

<그림 6.4> 데이터 산업 시장의 추진 전략



자료: 4차산업혁명위원회, 데이터 산업 활성화 전략, I-Korea 4.0 데이터 분야 계획, I-Data, 2018b

## ○ 데이터 이용 제도 패러다임 전환

- 데이터 이동권 확립: 국민 데이터 주권 마이데이터(my-data) 찾기
  - 개인 데이터 관리를 기관(기업) 중심에서 정보 주체 중심으로 전환하는 마이데이터 제도 도입
  - 마이데이터는 정보 주체가 기관(기업)으로부터 자기 정보를 직접 내려 받아 이용하거나 제3자 제공을 허용하는 방식임
  - 국민들이 실생활에서 체감할 수 있는 의료·금융·통신 등의 분야에서 대규모 시범 사업 실시 예정. 이를 통해 건강 증진·재테크·통신비 절감 등 국민들이 체감할 수 있는 효과를 창출해 국민들의 인식을 전환할 예정임
- 개인 정보 안전 활용 촉진
  - 국민적 신뢰에 기반한 개인 정보의 안전 활용 제도 구축
  - 개인 정보 관련 ‘사회적 합의 결과’를 바탕으로 개인정보보호법 등 관련법 개정
  - 데이터 자체는 안 되지만, 분석 및 AI 개발 결과만 반출이 가능한 데이터 안심존 구축

## ○ 데이터 가치 사슬 전주기 혁신

- 양질의 데이터 구축 개방
  - 산업별 실제 데이터, 인공지능 학습 데이터를 전방위로 구축하고, 공공·민간 데이터의 획기적 개방을 추진
  - 산업별 빅데이터 전문 센터를 육성하고 각 빅데이터 센터 간 협력 네트워크를 확대
  - AI 산업의 발전을 위해 범용(이미지, 상식 등)·전문(법률, 특허, 의료 등) 분야 AI 데이터 세트를 수요 중심으로 단계적 구축·보급하며, 민간 수요가 높은 데이터는 국가 중점 데이터로 선정해 조기 개방을 확대
  - 데이터의 획기적인 개방을 위해서는 모든 공공기관이 보유한 공공 데이터를 전수 조사하고 데이터 현황을 보여주는 국가 데이터 맵을 구축해 공공 데이터 통합 관리 체계 마련

- 데이터 저장 유통 활성화
    - 데이터의 효율적인 저장 및 관리를 위한 민간 클라우드 이용을 확산시키고, 데이터 유통을 촉진시키기 위한 데이터 거래 기반을 본격적으로 추진
    - 양질의 데이터를 손쉽게 거래할 수 있는 데이터 유통 기반의 구축
  - 데이터 분석 활용 확산
    - 데이터를 활용한 산업적·사회적 가치를 본격적으로 창출하여 산업 전반에 걸쳐 성장 활력의 제고 및 국민 삶의 질 향상에 노력
    - 데이터를 전통적인 중소기업의 혁신 돌파구 마련을 위한 신제품 개발, 맞춤형 홍보 등에 적극 활용
- 글로벌 데이터 산업 육성 기반 조성
- 빅데이터 선도 기술 조기 확보
    - 2022년까지 선진국 기술 수준 대비 90% 이상 수준의 빅데이터 선도 기술을 확보
    - 4차 산업혁명 요소 기술(인공지능, 사물인터넷, 클라우드, 정보 보호)과의 융합 선도 기술 개발을 가속화
  - 미래 수요 대비 전문 인력 확충
    - 데이터 산업의 미래 수요에 대응해 향후 5년간 청년 고급 인재, 실무 인력 중심으로 5만 명의 전문 인력을 양성
    - 빅데이터 전문 교육 프로그램을 지원
    - 데이터 분석 대학원 전공, 연구 센터 운영을 확대
    - 데이터 분석 국가기술자격제도(빅데이터 분석 기사) 시행
  - 빅데이터 전문 기업 성장 지원
    - 빅데이터 전문 기업의 성장에 필요한 컴퓨팅 자원, 맞춤형 사업 지원
    - 정부에서는, 특히 데이터 거래 시장 활성화를 위하여 데이터 유통 플랫폼 운영 및 플랫폼 간 연계 지원, 데이터 바우처 지원, 데이터 안심 구역 구축 및 운영, 데이터 공정 거래 지원 등의 정책 추진과 더불어 데이터 거래를 지원하는 데이터 품질 관리, 데이터 거래 표준, 데이터 가치 평가(자산화), 데이터 소유권 등 법적 문제, 신기술 적용 방안 등의 관련 제도를 진행할 계획임

- 이와 같은 전략과 연관 과제들은 정부에서 조속히 추진하여야 하는 것들로 판단됨. 그러나 위의 전략들 이외에도 다음과 같은 항목들이 추진된다면 데이터 산업 발전에 더욱 큰 효과가 있을 것으로 기대됨

## □ 국내 데이터사이언스 산업 발전 전략 제안

### ○ 데이터 산업 선도 그룹 육성

- 데이터 산업은 데이터의 수집, 데이터베이스 구축, 분석, 해석 및 활용, 폐기에 이르는 라이프 사이클 동안 필요한 기술과 인프라를 확충하고, 자기 학습, 컨설팅 및 서비스 등을 통한 데이터 활용으로 데이터 기반 가치를 극대화하고자 하는 산업임. 그러나 데이터 관련 분야를 살펴보면 특히 데이터, 의료 데이터, 금융 데이터, 통신 데이터 등 이미 빅데이터가 갖추어져 해당 데이터베이스에 대한 효과적인 탐색과 분석이 필요한 분야도 있고, 경우에 따라서는 이제 데이터를 수집하여 DB를 구축하는데서 시작을 하여야 하는 분야도 있음. 또한 제조업과 같이 데이터가 수집되고 있다고 하더라도 공유하기 어려운 분야도 있고, 정부가 개인 정보를 제외하고 공유할 수 있는 공공 데이터베이스도 있음
- 우리나라에서는 데이터 수집이나 저장, 분석에 필요한 기본 기술을 연구·개발하기보다는 기 개발된 기술들을 빠르게 활용하여 응용 분야의 부가가치를 향상시키는 연계 연구에 주력하는 것이 바람직하다고 판단됨. 그러나 위와 같이 응용 분야의 특성이 서로 다르므로, 전 산업에 걸친 단일한 접근보다는 해당 분야에 적합한 맞춤형 접근을 실시하고, 그 결과 얻어지는 노하우를 기반으로 새로운 기술 개발로 나아가는 것이 바람직하다고 판단됨. 즉 데이터 산업 기초 기술보다 응용 분야에 특화된 활용 기술에 중점을 두자는 것임
- 이와 같은 점을 감안하면 우리나라의 경우는 데이터 기술자와 응용 분야 기술자의 협력이 크게 요구된다고 할 수 있음. 이러한 협력을 위해서는 분야별 선도 그룹을 선정하여 육성함으로써 해당 분야 데이터 산업의 신속한 발전을 기할 수 있을 것으로 판단되며, 이 접근법이 가장 효율적인 것으로 사료됨
- 선도 그룹은 해당 분야의 컨설팅, 솔루션 개발·제공업체, 대학이나 연구기관, 해당 분야의 업체 등으로 구성하여 데이터 활용 프로젝트를 수행하고, 해당

분야의 데이터 기술 적용 시의 문제점이나 필요한 기술과 절차 등에 대하여 제시하며, 수행 경험을 공유하고 해당 분야를 선도할 수 있도록 지원함

- 이러한 실무 경험이 있어야 데이터사이언스에 대한 교육이나 국가 자격 제도가 해당 분야를 크게 발전시킬 수 있는 인재를 양성할 수 있을 것임

#### ○ 데이터 페이지(data page) 구축 유도

- 데이터 산업을 육성하기 위해서는 인터넷의 발전을 참고할 필요가 있음. 인터넷은 관련 기술이나 인프라도 중요하였지만, 정보를 구축하고 공유하며 또 탐색하고자 하였던 사람들의 속성에 힘입은 바가 큼. 그 결과 개인과 기업, 기관들이 홈페이지를 구축하고 정보를 공개하기 시작하였으며, 인터넷의 급속한 발전을 가져오게 되었음
- 데이터 산업의 발전에도 이와 같은 전략을 구사하여 개인이나 기업, 기관들이 공개할 수 있는 데이터를 홈페이지에 별도로 구축할 수 있도록 적극적인 유도가 있었으면 함. 이를 ‘데이터 페이지’라고 부르기로 하며, 홈페이지와는 별도로 데이터를 공개할 수 있는 수단으로 인지할 수 있도록 유도함
- 이러한 데이터 페이지를 통하여 모든 사람들이 데이터에 대한 중요성을 인식하고, 데이터를 수집하고 분석하며 저장하는 활동을 생활화, 문화화 할 수 있도록 유도함. 이를 통해 데이터가 몇몇 전문가만의 소유물이 아닌 모든 사람들의 자산임을 인지토록 함
- 데이터 페이지를 구축하고 운영할 수 있는 기술은 솔루션 업체들이 개발함

#### ○ 데이터 분석 도구 오픈 소프트웨어 개발

- 데이터 페이지와 더불어 데이터 산업의 활성화에 기여할 것은 데이터 수집, 분석, 저장에 필요한 기본 분석 도구들을 파악하여 이를 소프트웨어로 개발, 언제 어디서나 용이하게 데이터를 처리·분석하도록 장려해야 함. 이와 더불어, 탐색 및 분석 가능한 데이터베이스가 여러 개 마련되었으면 함. 정부에서 공공 데이터 베이스를 공개할 예정인데, 전문가들만 접근할 것이 아니라 초등학생부터 모든 국민이 손쉽게 탐색하고 분석할 수 있도록 기본적인 데이터 분석 도구도 함께 제공하고, 탐색 및 분석 결과를 발표하거나 공유하여 데이터 처리 기술의 발전을 모색하고 데이터에 대한 인식을 전환할 수 있도록 하였으면 함. 공개 데이터 베이스 처리에 대한 경진대회를 개최하는 것도 데이터 처리 기술을 개발하고 공유할 수 있는 수단임

- 제조업의 경우 품질 관리를 위하여 이른바 ‘QC 7가지 도구’를 식별하고, 이를 자유롭게 활용할 수 있도록 적용 절차와 사례를 만들어 보급함으로써, 품질에 대한 인식을 갖게 하고 문제를 풀어낼 수 있는 능력을 배양하였음
- 따라서 데이터 산업의 경우에도 가장 기본적인 분석 도구를 오픈 소프트웨어로 개발 및 제공함으로써 모든 사람들이 늘 사용할 수 있도록 함. 또한 데이터를 이용하여 의사 결정을 하는 것이 합리적이라는 점을 강조하여야 함

#### ○ 데이터 처리 솔루션 국산화

- 현재 데이터 처리 솔루션은 외국 회사가 개발한 제품들이 대부분임. 우리나라에서는 앞서 논의한 바와 같이 이러한 솔루션들과 직접 경쟁하기 보다는 응용 분야에 활용하고 그 과정에서 획득되는 여러 지식을 활용하여 새로운 솔루션을 개발하는 것이 유리하다고 판단됨. 예를 들어 데이터 분석을 위해서는 전처리기가 선행되어야 하는데, 이를 특화된 분야에서 인공지능을 이용하여 자동화하거나, IBM 왓슨과 같은 시스템을 제조업에 응용하여 설계나 품질 등의 문제를 인공지능 로봇과 브레인스토밍 할 수 있는 시스템을 개발한다면 경쟁력 있는 솔루션을 개발할 수 있을 것임
- 그러나 빅데이터 처리에 필수적인 클라우드 컴퓨팅 등과 같은 시스템은 국내 업체의 경쟁력이 너무 낮으므로, 이의 개발을 지원하여 국산 솔루션을 가질 수 있도록 하여야 응용 분야의 적용 과정에서 시행착오를 줄일 수 있을 것임

#### ○ 데이터 보증 체계 개발

- 데이터의 상품화를 통한 데이터 거래는 데이터 산업의 궁극적인 목표임. 그러나 이러한 상품화를 가로 막는 가장 큰 요인은 데이터에 대한 신뢰성일 것임. 만일 상품화된 데이터를 믿지 못한다면 거래는 불가능할 뿐 아니라, 만일 거래가 되더라도 불량 데이터로 인해 발생하게 될 사회적 피해는 막대할 것임. 이를 해결하기 위해서는 상품화되는 데이터를 보증할 수 있는 보증 체계의 개발이 필요하다고 생각됨
- 외국에서는 데이터의 품질을 관련성, 시간 적절성, 정확성 등으로 구분하여 이를 제고할 수 있는 방법을 연구하고 있음. 이러한 결과를 바탕으로 데이터를 상품화 하는 데에 필요한 보증 체계의 개발과 인증이 필요하다고 판단됨

## 참 고 문 헌

- 4차산업혁명위원회(2018a). I-Korea 4.0 실현을 위한 인공지능 R&D 전략, 4차 산업혁명위원회 심의안건 1호, 2018.5.15.
- \_\_\_\_\_ (2018b). 데이터 산업 활성화 전략, I-Korea 4.0 데이터 분야 계획, I-Data, 4차산업혁명위원회 의결안건, 2018.6.26.
- 관계부처합동(2019). -혁신성장 전략투자- 데이터·AI경제 활성화 계획 ['19~'23년], 관계부처 합동 문건.
- 정원준·이나라(2018). “인공지능 활성화를 위한 주요국의 대응전략과 정책 제언”, 주간기술동향, 제1870호, 정보통신기술진흥센터, pp. 2~16.
- 김지섭(2012). 아이폰 '시리(Siri)'는 애초 미국방부 작품, 조선비즈, [https://biz.chosun.com/site/data/html\\_dir/2012/06/20/2012062000890.html](https://biz.chosun.com/site/data/html_dir/2012/06/20/2012062000890.html)
- 한국데이터진흥원(2018). 2018 데이터산업백서.
- De Veux, R. *et al.*(2017). Curriculum guidelines for undergraduate programs in data science, Annual Review of Statistics and Its Applications, p. 15~30.
- National Academies of Sciences, Engineering and Medicine(2018). Data Science for Undergraduates: Opportunities and Options, A Consensus Study Report of The National Academies of Sciences · Engineering · Medicine, Washington DC, The National Academies Press.
- Song, I. Y. & Zhu, Y.(2016). “Big data and data science: what should we teach? Expert Systems”, Vol. 33, Issue 4, pp. 364~373.
- White House(2016). Preparing for the Future of Artificial Intelligence, 2016. 10.
- \_\_\_\_\_ (2016). The National Artificial Intelligence Research and Development Strategic Plan, 2016. 10.
- <https://www.bachelorsportal.com>
- <https://www.datasciencedegreeprograms.net>
- <https://www.datasciencedegreeprograms.net/rankings/affordable-bachelors/>
- <https://www.mastersportal.com>

# VII



## 결론



## VII

## 결론

- 한국은 1967년 경제기획원이 IBM 1401 컴퓨터를 도입하여 인구 조사 결과를 처리한 것을 시작으로 컴퓨터가 사회 각층에 도입되기 시작하였고, 1975년 서울대학교가 관악으로 이전하면서 현재 본부 건물 2층 수백 평 공간에 대용량 컴퓨터를 설치 운용하였음. 그 큰 컴퓨터는 조그만 행렬의 역행렬을 계산하면 멈춰버리는, 지금 개인들이 가지고 있는 노트북보다도 성능이 훨씬 하잘 것 없는 것이었음. 미루어 보건대, 앞으로 수십 년 내 개인 핸드폰의 성능이 지금 한국 최고의 슈퍼 컴퓨터보다 성능이 훨씬 더 좋아져 우리가 상상할 수 없는 다양한 일들을 수행하게 될 것임
- 1975년에 신설된 계산통계학과는 컴퓨터학과와 통계학과가 한과에 있는 형태였으나 귀납적 방법론이 전공별로 세분화되는 시대적 흐름에 따라, 통계학과와 컴퓨터공학과로 분리되었고, 산업공학, 데이터 마이닝, 기계학습, 인공지능 등 다양한 전공들이 출현하였음. 그러나 이제는 통계학, 컴퓨터공학 및 과학, 산업공학, 경영학, 데이터 마이닝, 기계학습, 인공지능, 전기·전자, 로봇, IT 기술들이 융합 하는 빅데이터·인공지능 시대이며, 더불어 새로운 제품, 서비스, 신산업들을 창출하는 4차 산업혁명 시대가 도래하였음. 다가올 시대를 준비하고 새로운 미래를 개척하기 위하여, 빅데이터·인공지능 기술을 습득한 데이터사이언티스트들을 육성하는 것은 현 시대의 사명이라 할 수 있음
- 빅데이터·인공지능 기반 산업 활성화를 위해서는, 다양한 분야에서 개발된 데이터 사이언스 기술들을 폭 넓게 교육하는 정책이 필요하며 이어서 R&D 및 산업 분야에 대한 체계적이고 효율적인 정책·투자 등의 발전 전략이 요구됨

## □ 빅데이터·인공지능 기술을 위한 데이터사이언스 교육 전략

### ○ 정부

- 데이터사이언스 교육 커리큘럼 작성
- 데이터사이언스 공인 자격 제도 신설
- 데이터사이언스 인력 양성 문제에 대한 조속한 해결
- 데이터사이언스 교육 지원을 위한 공공 데이터 개방
- 빅데이터·인공지능 학습용 데이터를 많이 만들기 위한 ‘데이터 고속도로’ 구축

### ○ 대학

- 데이터사이언스 학사 과정 개설
- 데이터사이언스 교양 과목 개설

### ○ 중등 교육 과정

- 데이터사이언스 마인드 함양 및 교육

## □ 빅데이터·인공지능 관련 R&D 촉진 전략

- 민간 기업 R&D가 더욱 활성화될 수 있는 여건 조성
- 과감한 ‘오픈 데이터’ 정책으로 빅데이터 산업 성장
- 빅데이터·인공지능 인력 양성을 위한 과감한 교육·연구 개혁
- 빅데이터·인공지능 분야의 효율적인 연구 정책 수행을 위해 전문성을 가진 공무원 양성

## □ 데이터사이언스, 빅데이터, 인공지능 기술 산업 전략

- 데이터 산업 선도 그룹 육성
- 데이터 페이지(data page)의 구축 유도
- 데이터 분석 도구로 오픈 국산 소프트웨어 개발 유도
- 데이터 처리 솔루션의 국산화 개발
- 데이터 보증 체계 개발

- 국내 과학기술의 발전 방향이 추격형이 아닌 선도 그룹으로 자리 잡기 위해서는 더욱 심층적인 분석을 통해 선도형·융합형 인력 양성 방안을 제시하는 것이 필요함. 또한, 데이터사이언스 연구자들이 시장과 동떨어진 현학적인 연구에 빠지지 않도록 ‘수익을 창출하는 산업과의 연계성’이 강조되어야 함
  
- 4차 산업혁명 시대에서는, 다량의 양질 데이터를 확보하여 원료로 삼고, 최고의 빅데이터·인공지능 기술로 이를 가공하여, 새로운 산업과 서비스를 창출하는 기업이 거의 모든 이익을 독식하는 무한 경쟁 시대임. 미·중 그리고 한·일 무역 갈등이 심화 되는 지금의 국제 정세에서 우리는 그 어느 때보다 스스로 기술적 우위를 갖고 있어야 함. 다가올 시대를 준비하는 마음으로 다양한 분야의 학계 산업계 전공자들이 모여 지혜를 모아 제언한 이 소고가 한국 발전에 도움이 되기를 기원함

한림연구보고서 130

## 빅데이터인공지능 산업 진흥을 위한 데이터 과학의 발전 전략 연구

Study on the strategies of big data & artificial intelligence for the promotion of industrial development

발행일 2019년 12월  
발행처 한국과학기술한림원  
발행인 한민구  
전화 031) 726-7900  
팩스 031) 726-7909  
홈페이지 <http://www.kast.or.kr>  
E-mail [kast@kast.or.kr](mailto:kast@kast.or.kr)

편집/인쇄 (주)디자인여백플러스 02)2672-1535

I S B N 979-11-86795-48-4 94060

I S B N 979-11-86795-45-3 (세트)

- 이 책의 저작권은 한국과학기술한림원에 있습니다.
- 한국과학기술한림원의 동의 없이 내용의 일부를 인용하거나 발췌하는 것을 금합니다.





이 사업은 복권기금 및 과학기술진흥기금 지원을 통한 사업으로  
우리나라의 사회적 가치 증진에 기여하고 있습니다

