

인공지능 학습방법에 대한 분석

황호석*, 심재창*
 안동대학교 컴퓨터공학과
 e-mail : ghtjr991123@naver.com

Analysis about Artificial Intelligence Learning Method

Ho suk hwang*, Jaechang shim*
 Dept. of Computer Engineering, Andong National University

요 약

4 차 산업혁명의 대두로 인하여 인공지능이 전세계적으로 중요해지고 있는 상황이다. 이러한 상황에서 앞서 가기 위해 인공지능의 구현 방법중에 합리주의적 방법과 경험주의적 방법 그리고 인공지능 학습의 방법인 지도, 비지도, 강화 학습의 방법과 각 방법에 대한 특징 장점 단점 한계점 대하여 분석하여 작성한다.

1. 서론

현재 4 차 산업혁명이 중요한 이슈가 되며 2010 년대를 기반으로 클라우드, 빅데이터, 전반적인 컴퓨터의 파워향상 등으로 4 차 산업혁명 기술 중 하나인 인공지능이 암흑기를 벗어나 4 차 산업혁명의 핵심기술이 되었고 2016 년 알파고와 이세돌의 대국으로 인해 인공지능에 대한 관심이 매우 높아지며 인공지능에 대한 투자또한 높아져 개발이 되고있다[1]. 이러한, 상황에서 인공지능에 대한 지식은 매우 중요한 기술이 되었으며 이러한 상황에서 적응하기 위해 인공지능의 학습방법의 종류와 작동방식을 위주로 설명을 하였다.

2. 인공지능의 구현방법

인공지능의 구현방식은 합리주의자에 근거한 방식과 경험주의자에 근거한 방식이 존재한다. 합리주의자의 방식은 진리에 대한 지식은 이성으로만 얻을 수 있으며 감각이나 경험은 불필요한 것으로 판단하였다. 이러한 이유로 초기에는 합리주의자의 방식에 근거하여 전문가들이 1980 년대 전문가 시스템을 만들었으나 전문가 시스템은 결국 스스로 생각하는 것이 아닌 조건에 따른 결과를 도출해 내는 구조라서 많은 경우의 수가 필요했고 예외적인 일이 일어날 경우 가치가 떨어진다는 단점이 있어서 서서히 잊혀지게 되었고 인공지능의 겨울을 맞이 하게 되었다[2,3]. 그 후 컴퓨팅 파워의 향상으로 인해 모든 사건을 경험에 중점에 두는 경험주의자에 근거해 컴퓨터로부터 데이터를 직접 입력받아 학습하는 딥러닝이 발전하게 되었다.

3. 인공지능 학습의 방법

인공지능의 학습방식은 기본적으로 얼마나 데이터를 잘 분류하기 위해 데이터의 특징을 잘 정의하는게 본 연구는 과학기술정보통신부 및 정보통신기획평가원에서 지원하는 SW 중심대학사업(IITP-2019-0-01113)의 연구 결과로 수행되었음.

핵심이다. 데이터의 특징을 정의하기 위해 하는 학습의 종류는 크게 세가지로 나뉘어진다.

먼저, 지도학습은 어떠한 데이터를 입력하면 입력한 데이터에 대한 정답을 준다. 따라서 기계가 정답을 맞췄는지 아닌지 알기가 쉽다. 지도학습에는 크게 분류와 회귀가 존재한다. 분류란 주어진 데이터에 대해 어떤 카테고리에 따라 분류하는 문제를 말한다. 분류에는 다시 2 개의 방식으로 나뉘어지는데 어떠한 데이터에 대해 예/아니오로만 분류되는 문제를 이진 분류(Binary classification) 어떠한 데이터에 대해 여러 값중 하나로 분류할 수 있는 문제를 다중 분류(Multi-label classification)라고 한다. 우리가 통계를 사용하는 목적은 과거의 데이터를 학습하고 기억함으로써 최적의 예측값을 만드는 것이다. 이와같이, 회귀는 입력값과 결과값의 연관관계를 분석한 뒤 특정값에 대한 결과값을 예측하는 방법이다. 회귀 학습 방식에는 선형회귀가 존재한다[4]. 이 방법은 입력받은 데이터를 가장 잘 나타낼 수 있는 수식을 찾아내는 방법이다. 이러한 수식의 정확성을 알아보는 방법에는 최소 제곱 방법이 있다. 이 방법은 실제의 값과 예측의 값이 최소가 되어야 한다. 이러한 방식을 수식으로 나타내면 그림 1 과 같이 사용할 수 있다. 그림 2 와 같이 선형회귀 방식은 계산이 편리하다는 장점이 있지만 독립변수와 종속변수 사이의 비선형성으로 존재하는 모형의 설명은 신뢰도가 떨어진 다.

$$\sum_{i=1}^n x_i^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

그림 1.최소 제곱 공식

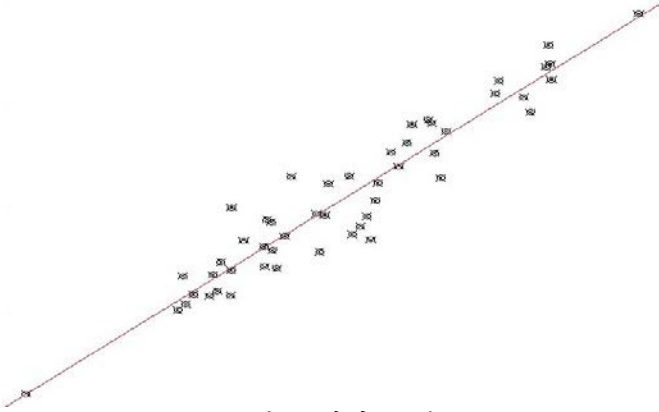


그림 2. 선형 모델

지도 학습 방식은 결과값도 같이 미리 주어지기 때문에 더욱 정확한 학습을 할 수 있으나 결과값과 출력값이 같이 들어가기 때문에 각 데이터들의 생성비용이 높아지고 들어갈 수 있는 데이터의 갯수 또한 작아진다는 단점이 있다.

다른 방법으로는 비지도학습이 있다. 이 방법은 입력한 데이터에 정답을 알려주지 않는다. 따라서, 기계가 스스로 판단하여 데이터 내부에 있는 비슷한 특징을 잡아 군집화(clustering)한다. 비지도학습에서 가장 중요한 것이 군집화 이지만 실제로 군집화의 과정에는 결과(label)이 없기 때문에 정답이 존재하지 않는다. 따라서, 군집이 얼마나 잘 만들어 졌는가를 기준으로 판단을 내리는데 이때 판단의 기준은 군집의 크기, 군집 간의 거리, 군집의 분산등을 사용한다. 군집 간의 거리는 멀수록 군집 내의 분산은 작을 수록 좋은 군집화의 결과이다. 군집화의 종류에는 K-평균 군집화와 계층적 군집화가 존재한다. K-평균 군집화는 각 군집에 중심을 기준으로 하나의 군집을 형성하게 된다. 이 군집화의 학습 과정은 그림 3 과 같이 기대(Exceptionation)를 사용하여 군집의 중심에 기반하여 군집을 나눈 다음 극대화(Maximization)를 사용하여 군집의 중심을 군집에 맞게 업데이트 해준다[5]. 이 방법을 반복 사용하여 더이상 변하지 않거나 사용자가 지정해준만큼 학습을 반복하면 학습이 끝난다. K-평균 군집화의 특징은 계산복잡성이 $O(n)$ 으로 높지 않아 가벼운 러닝을 할때나 여러번 러닝을 돌릴때 자주 쓰인다.

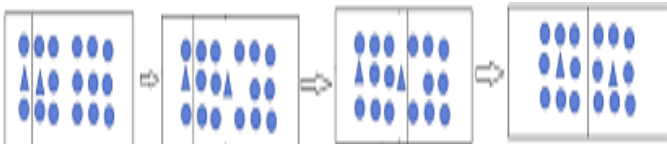


그림 3. 기대와 극대화

다른 방법으로는 계층적 군집화가 있다. 이 방법은 K-평균 군집화와 달리 그림 4 와 같이 계층적 트리모형을 사용하여 유사한 개체를 군집화하여 군집 수를 사전에 정하지 않아도 된다[6].

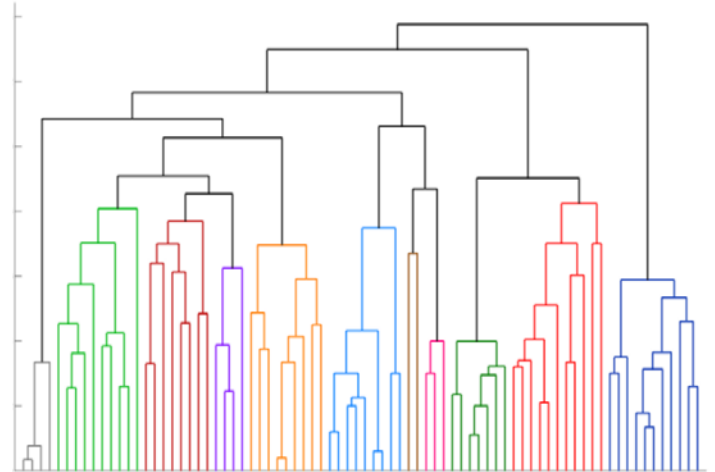


그림 4. 계층적 트리

계층적 군집화는 거리와 유사도가 미리 계산되어 있어야 작동이 가능하다. 계층적 군집화는 미리 군집수를 설정하지 않아도 된다는 장점이 있지만 계산복잡성이 $O(n^3)$ 으로 K-평균 군집화에 비해 복잡하다는 특징을 가지고 있다.

비지도방식은 지도방식에 비해 학습하기 어렵지만 클러스터링이 잘 이루어 진다면 매우 효과적인 방법이다.

마지막으로, 강화학습이 있다. 이 방법은 그림 5 와 같이 현재의 상태를 인식한 다음 어떠한 행동으로 인한 보상을 최대화하는 행동 또는 순서를 선택하는 방법이다[7]. 우리 일상또한 행동에 우선순위가 정해져 있어 강화학습은 매우 널리 사용되기 좋은 학습 방법이다. 강화학습은 어떠한 행동에 외부환경에 대한 보상이 지급되는데 상황에 따라 보상지급의 지연이 일어날 수도 있어 지도/비지도학습에 비해 난이도가 높다.

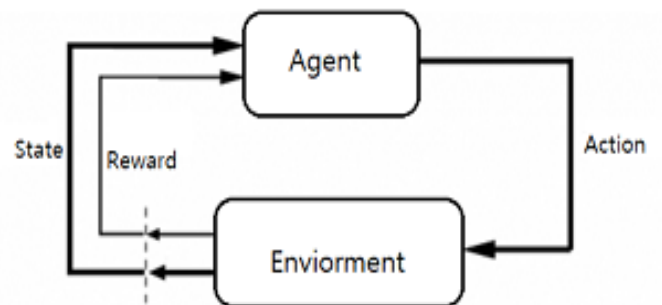


그림 5. 강화학습 학습도

4. 결론

현재 인공지능은 아직은 불완전하고 국내의 경우 미흡한 수준이다. 전 세계가 인공지능에 투자를 하고 있지만 인공지능은 아직 모든 분야에서 사용하지 못하고 부분적인 부분에 사용하는 추세이며 과학기술정보부의 " 주요국 인공지능 정책동향" 에서 제공한 표 1 에서 보이듯이 특히, 국내는 인공지능에 대한 투자또한 다른 강대국에 비해선 미미한 수준이다.

표 1. 국가별 경제 및 AI 투자 지표

국가	GDP/1인당 GDP	인구	정부예산		주요산업	ITU ICT
			전체	AI		
미국	20조 4천억 \$/6만 2천 \$	3억 2 천만	4400 조원	-	금융, 서 비스, 제 조, 교육, 정보통신	16 위
중국	14조 1천억 \$/1만\$	13억 8 천만	3300 조원	1135 조원 (30 년까 지)	농업, 제 조업, 정 보산업, 자동차	80 위
일본	5조 1천억 \$/4만 8천 \$	1억 2 천만	2200 조원	7865 억	서비스, 제조	10 위
독일	2조 9천억 \$/4만 5천 \$	8천 2 백만	1700 조원	-	자동차, 항공, 의 약, 기계, 선박	12 위
영국	2조 9천억 \$/4만 4천 \$	6천 6 백만	1200 조원	1조 5897 억원	기계장치, 자동차, 전자제품, 의약품	5위
프랑스	2조 9천억 \$/4만 4천 \$	6천 6 백만	1500 조원	1조 9887 억원	항공우주, 자동차, 의약	15 위
인도	2조 8천억 \$/1.8천\$	13억	322 조원	1조 1335 억원 (35 년까 지)	연료, 귀 금속, 전 자기기	134 위
캐나다	1조 7천억 \$/4만 4천 \$	3천 6 백만	720 조원	1조 1815 억원	농림수산 업, 광업, 정보통신, 항공	29 위
한국	1조 2천억 \$/3만 2천 \$	5천 1 백만	326 조원	-	정보통신, 자동차, 선박	2위
호주	1조 2천억 \$/5만\$	2천 4 백만	550 조원	250 억원	농업, 광 업, 서비 스업	14 위
UAE	4천억\$/4 만 \$	1천만	127 조원	2067 억원 (35 년까 지)	석유, 가 스, 서비 스, 관광	40 위
덴마크	3천억\$/5 만 \$	5백만	125 조원	132 억원	무역, 제 조, 건축	4위
핀란드	2.6천억\$/4 만\$	5백만	71 조원	2650 억원 (18- 22)	제조업, 도소매, 무역, 건 설	22 위
싱가폴	2.5천억\$	5백만	71 조원	125 억원 (AI Sing apor e)	제조, 서 비스, 금 융, 도소 매, 정보 통신	18 위

4차 산업혁명은 지금 현재진행형으로 계속 발전하고 있으며 그 중심에는 빅데이터를 기반으로한 AI가 있다. 4차 산업혁명에서 뒤처지지 않기 위해서는 체계적인 교육 및 지원, 매뉴얼이 필요하며 무엇보다, 빠르게 투입되어 이용가능한 전문인력의 양성이 중요하다.

참고문헌

- [1] 원동규, 이상필, "인공지능과 제 4차 산업혁명의 함의." 대한산업공학회, 제 23 권, 제 2 호, pp. 13-22, 2016.
- [2] 배정미, "해외 이민 한국인의 정신건강관리를 위한 웹기반 지능형 전문가시스템 개발 및 적용." 대한간호학회지, 제 43 권, 제 2 호 pp. 203-214, 2013.
- [3] 정동규, "인공지능 기술과 주요 적용 산업 동향." 한국정보기술학회지, 제 15 권, 제 2 호, pp. 21-18, 2017.
- [4] 임재현, 박윤기, 권진만, 서정옥. "사물인터넷 환경에서의 기계학습 기술." 한국통신학회지(정보와 통신), 제 33 권, 제 2 호 pp. 48-54, 2016.
- [5] 김연중, 김도연, 김윤, "K-평균 군집화를 이용한 객체 검출영역 보정 기법." 한국정보과학회, 한국정보과학회 학술발표논문집, pp. 2101-2103, 2017.
- [6] Lee, Jong-Sang, "Administrative Rezoning Using Hierarchical Cluster Analysis Based on Size and Connectivity of Administrative District." 한국지역개발학회지, 제 25 권, 제 3 호, pp. 91-107, 2013.
- [7] 민대홍, 정근우, 권기영, 박주영, "최신 강화학습 기법을 이용한 이동로봇 제어." 한국지능시스템학회, 제 21 권, 제 1 호, pp 67-70, 2011.