

# 신경회로망의 역사

2021254010 이지호

인공 신경망 모델에 관한 연구는 1943년 워런 맥컬럭(Warren McCulloch)과 월터 피트(Walter Pitts)로부터 시작되었다. 맥컬럭과 피츠의 모델의 논점은 네트워크 내의 단순한 요소들의 연결을 통하여 무한한 컴퓨팅 능력을 가진다는 점이다. 이들의 모델은 인간 두뇌에 관한 최초의 논리적 모델링이다.

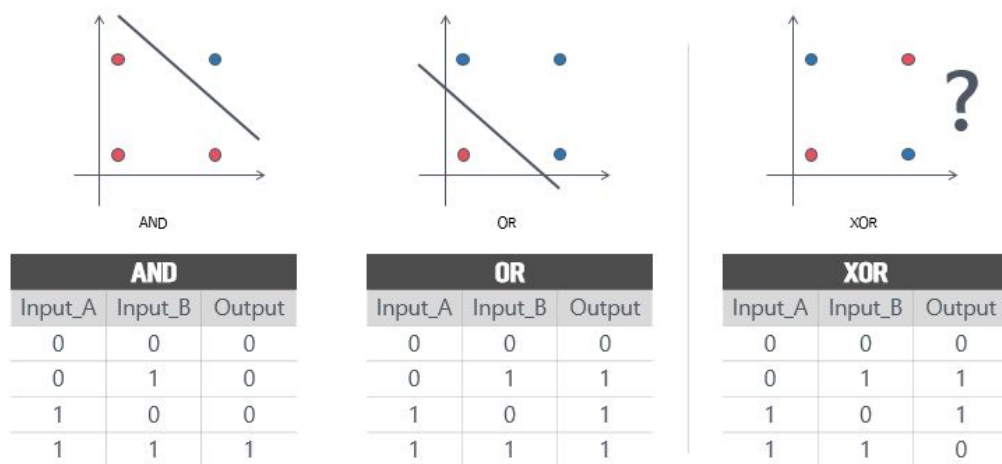
## 퍼셉트론 등장

그 이후 1949년에는 캐나다의 심리학자 도널드 헵(Donald Hebb)이 두 뉴런 사이의 연결 강도를 조정할 수 있는 학습규칙을 발표하였으며, 프랭크 로젠블랫(Frank Rosenblatt)은 1957년에 '퍼셉트론(Perceptron)'이라는 최초의 신경망 모델을 발표하였다. 이 퍼셉트론은 인간의 두뇌 움직임을 수학적으로 구성하여 그 당시 굉장히 많은 이슈가 되었다.

1958년 뉴욕 타임즈 기사 中. 걷고, 말하고, 보고, 쓰고, 스스로 만들어내고 심지어는 자아를 인식할 수 있을 것이라고 아직도 못 하고 있는 것을 할 것이라고 말하고 있다. 그 당시 인공 신경망에 대한 기대가 얼마나 컸는지를 알 수 있는 대목이다. 참 웃기기도 하고, 신기하기도 하다.

## 첫 빙하기 (XOR 문제)

그 당시 어떠한 것들을 해결했었는지를 살펴보면 간단한 and / or 연산을 수행했다. 이 연산을 기계가 스스로 풀 수 있으면 이러한 것들을 조합해서 생각할 수 있는 기계를 만들 수 있다고 생각했을 것이다.



AND / OR 연산 & XOR 문제

위 그림을 보자. AND와 OR의 경우 퍼셉트론을 통한 선형 분리(linearly separable)가 가능했다. 따라서 AND / OR 연산이 가능하게 훈련한 퍼셉트론을 보고 많은 사람들은 기계의 학습 가능성에 대해 큰 기대를 하게 되었다.

하지만 이러한 그들의 엄청난 기대에 찬물을 끼얹어버린 것이 XOR 연산에 대한 불가능이었다. 당시 하나의 인공 신경, 즉 퍼셉트론으로는 선형 분리가 불가능해 XOR 연산에 대한 학습이 불가능했다.

1969년 MIT AI Lab의 창립자였던 Minsky 그리고 Papert가 「Perceptrons」라는 저서를 통해 '현재의 퍼셉트론으로는 XOR가 절대 불가능하다.'라는 것을 수학적으로 증명을 했다. 그뿐만 아니라 하나의 퍼셉트론이 아닌 MLP (Multilayer Perceptron)을 통해서 XOR는 해결될 수 있지만, '각각의 weight와 bias를 학습시킬 방법이 없다.'라는 결론에 이르게 된다. 즉 이걸 절대 해결할 수 없다고 저술하였고, 많은 사람들이 이 책을 읽고 공감하게 된다.

따라서 인공신경망 연구에 대한 첫 붐은 이렇게 첫 번째 빙하기에 들어가게 됩니다.

## Backpropagation (역전파법) 등장

이후 Backpropagation 방법, 즉 앞의 진행방향에서 고쳐가는 것이 아니라 결과를 보고 뒤로 가면서 weight와 bias를 조정하는 역전파법을 고안하게 되었다.

역전파법이란, 샘플에 대한 신경망의 오차(error cost, 목표 출력과 실제 출력의 차이)를 다시 출력층에서부터 입력층으로 거꾸로 전파시켜 각 층(layer)의 가중치(weight)의 기울기 (gradient)를 계산하는 방법이다. 이를 통해 weight와 bias를 알맞게 학습할 수 있다는 것이다.

1974년 Paul Werbos가 박사과정 논문을 통해 이 Backpropagation을 제안했다. 하지만 그당시 인공 신경망에 꽤나 찬바람이 불던 시기였기 때문에 아무도 관심을 가지지 않았고, Minsky 교수에게 찾아가 말을 했지만 그 또한 냉랭했다고 한다. 1982년 다시 논문을 썼지만 역시 주목을 받지 못했다.

이와는 별개로 1986년 제프리 힌튼(Geoffrey Hinton)이 같은 방법을 독자적으로 고안해냈다. 이것이 주목을 받게 되었고, XOR 뿐 아니라 좀 더 복잡한 과정보도 해결할 수 있음을 보이며 다시 인공 신경망은 사람들의 관심을 끌기 시작했다.

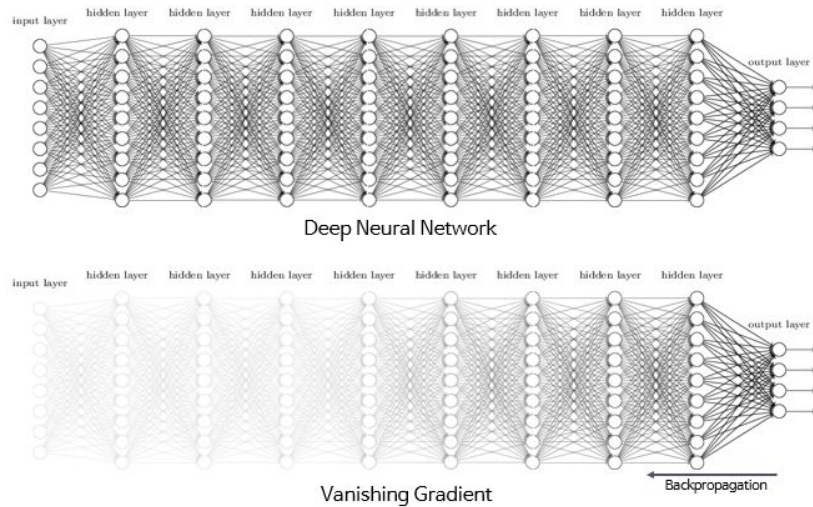
## 두 번째 빙하기

Backpropagation을 계기로 한 두 번째 붐은 90년대 초반까지 연구에 큰 진전을 이루었다. 하지만 이 시기의 연구의 붐 또한 90년대 후반에는 다시 끝나버리고 또 빙하기에 들어가게 된다.

이 두 번째 붐이 끝나게 된 원인으로는 크게 두 가지를 꼽을 수 있다.

1. 신경망의 깊이가 깊어질수록 원하는 결과를 얻을 수 없다.
2. 신경망 학습을 위한 파라미터 값의 최적화에 대한 이론적인 근거가 없었다.

## 1. 신경망의 깊이가 깊어질수록 원하는 결과를 얻을 수 없다.



깊은 신경망일수록 끝까지 전달이 되지 않는 문제

신경망이 깊어질수록 더욱 학습력이 좋아져야하는 것은 당연하다. 하지만 이상하게도 layer가 일정 개수 이상 늘어나면 기대하는 결과가 나오지 않는 일이 생기는 것이었다. Backpropagation을 수행할 때 입력층에서 멀리 떨어진 깊은 (deep) layer에 이르게 되면 기울기(gradient)가 급속히 작아지거나 너무 커져 발산해 버리는 Vanishing Gradient 문제가 발생했기 때문이다. 따라서 학습에 중요한 영향을 끼칠 입력층 부근에서 제대로 된 조정이 될 수 없었다.

## 2. 신경망 학습을 위한 파라미터 값의 최적화에 대한 이론적인 근거가 없었다.

신경망은 학습을 위한 여러 파라미터로 층 수나 유닛의 수를 갖는데, 당시에는 이 파라미터가 최종적으로 어떻게 성능으로 이어지는지를 알 수 없었다. 따라서 좋은 성능을 이끌어 내기 위한 파라미터들에 대한 노하우는 있었지만 이론적인 근거가 없었다는 것이다.

그러한 시기에 SVM(Support Vector Machine)이나 RandomForest 등 새로운 기계학습 알고리즘들이 등장했고, Deep Neural Net 처럼 복잡하지도 않으면서 더 좋은 성능을 나타낸다는 결과가 나오기 시작했다. 1995년도에는 그 당시 Neural net의 전문가였던 Yann Lecun 교수님 역시 neural network보다 새로운 기계학습 알고리즘의 성능이 더 좋음을 보였고, 이러한 이유로 90년대 후반부터 신경망 연구는 차가운 두 번째 빙하기를 맞이하게 되었다.

## Deep의 출현

길었던 neural net 연구에 대한 침체기에도 꾸준히 연구했던 분들이 있었다.

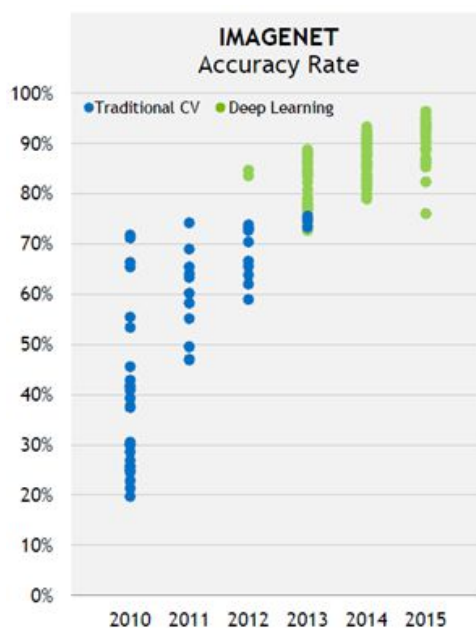
이후 2006년 Backpropagation을 고안했었던 제프리 힌튼(Geoffrey Hinton)이 "A fast learning algorithm for deep belief nets" 라는 논문을 통해 weight의 초기값을 제대로 설정하면 깊은 신경망학습이 가능하다! 라는 것을 보였다. 신경망을 학습시키기 전에 층 단위의 학습을 거쳐 더 나은 초기값을 얻는 사전훈련(pre-training) 의 최초 아이디어가 제안된 것이

었고, 적층 제약 볼츠만 머신(RBM : Restricted Boltzmann Machine)을 딥 빌리프 네트워크(DBN : Deep Belief Network)로 변환하는 형태로 사전훈련을 했다.

이어서 2007년에는 벤지오(Bengio) 팀이 "Greedy layer-wise training of deep networks"라는 논문을 통해 좀 더 간단한 자기부호화기(autoencoder)를 사용한 사전훈련 방법을 제안했다. 이 또한 깊은 신경망 학습이 가능하게 하기 위함이었고, 좀 더 어려운 문제들도 해결 가능함을 보였다.

눈치가 빠른 사람들은 알았을 수 있지만 위 두 논문의 이름을 보면 neural net이라는 단어를 찾을 수 없다. neural 대신 deep이라는 단어를 사용했다. 이 당시 neural network라는 말만 들어가면 논문에서 거절당한다는 말이 있을 정도로 신경망이 외면받던 시기였기 때문에 좀 더 사람들의 이목을 끌 수 있는 단어를 택했던 것 같다. 2006년 드디어 Deep Network, Deep Learning이라는 용어가 사용되기 시작했다.

## IMAGENET에서 딥러닝을 알리다



IMAGENET 대회에서 혜성처럼 등장한 딥러닝

컴퓨터 비전 분야에서 딥러닝의 저력을 크게 보여준 계기가 2012년 IMAGENET이라는 이미지 분류 대회에서 있었다. 캐나다 토론토 대학의 제프리 힌튼 교수의 제자 알렉스가 알렉스넷(AlexNet)이라는 딥러닝 기반 알고리즘으로 84.7%의 정확도를 보였다. 80% 이상의 인식률은 거의 불가능이라는 인식이 있을 당시였기 때문에 충격적인 사건이었고, 2012년 이후를 보면 대부분의 참가팀들이 알고리즘을 딥러닝으로 방향을 돌렸다. 현재는 인식률이 5% 이하의 정확도로 스탠포드 학생이 열심히 이미지를 공부한 후에 이미지 분류를 한 것보다 더 나은 인식률을 보였다고 한다.

## 딥러닝은 진행 中

2016년 3월에 있던 알파고 vs 이세돌

꽤 많은 일반인들(아마 특히 한국..)이 인공지능 혹은 딥러닝 이라는 단어를 알게 된 사건이 올해 있었다. 구글 딥마인드의 알파고(AlphaGo)가 이세돌에게 바둑으로 도전장을 내밀었기 때문이다. 나는 어릴적 바둑을 뒀었기 때문에 이세돌 바둑기사에 대해서는 알고 있었고, 기원에서 본적도 있었기 때문에 더욱 열렬하게 이세돌 선수를 응원했었다. 어쨌든 이세돌의 1:4 패배는 꽤 많은 사람들에게 인공지능의 무서움을 인식시켜주었고, 연구자와 개발자들에게는 꽤 큰 흥미를 주었다.

체스와 같은 게임은 오래전부터 인간을 넘어섰지만, 바둑은 경우의 수가 워낙 많기 때문에 이전의 기술력으로는 쉽지 않았을 것이다. 하지만 지금은 그 수많은 경우의 수를 다 따져볼 수 있을 정도의 속도를 지닌 HW 기술력과 새로운 딥러닝과 강화학습 능력을 갖추었기 때문에 이길 수 있었다. 다음 목표는 스타크래프트라고 들었는데 볼게 된다면 충분히 이길 것 같다.. 전략은 물론 모든 유닛에 대한 동시 컨트롤이 가능할테니.



최근에는 딥러닝이 사용되지 않는 분야를 찾기 힘들 정도로 다방면으로 사용되고 있다. 주가 예측은 물론 의료 분야에서도 큰 영향을 끼치고 있고, 자율주행에 있어서도 많은 방법들이 딥러닝을 활용하는 방식으로 대체되고 있다. 깊은 신경망을 통한 많은 계산량을 빠르게 수행가능한 GPU 컴퓨팅과 빅데이터 등 높은 기술력과 맞물려 크게 성장하고 있다. 물론 딥러닝이 만능일 것이라고는 생각하지 않는다. 이렇게 신나게 연구를 이어가다가 언제 또 한계에 부딪히게 될지 모르는 일이다.