

딥 러닝

기계 학습과 데이터 마이닝



문제

[분류](#) · [클러스터 분석](#) · [회귀 분석](#) · [이상 탐지](#) · [연관 규칙](#) · [강화 학습](#) · [구조 기반 예측](#) · [특징 공학](#) · [특징 학습](#) · [온라인 기계 학습](#) · [준지도 학습](#) · [비지도 학습](#) · [문법 유도](#)

지도 학습 (통계적 분류 · 회귀 분석)

결정 트리 학습법 · 앙상블 학습법 (배깅, Boosting, 랜덤 포레스트) · 최근접 이웃 탐색 · k -NN · 선형 회귀 · 나이프 베이즈 · 인공신경망 · 로지스틱 회귀 · 퍼셉트론 · 상관 벡터 머신(RVM) · 서포트 벡터 머신(SVM)

클러스터 분석

BIRCH · 계층적 군집화 · k-평균 알고리즘 · 기댓값 최대화 알고리즘 ·
DBSCAN · OPTICS · Mean-shift

차원 축소

인자 분석 · CCA · 독립 성분 분석 · LDA · 음수 미포함 행렬 분해 · 주성분 분석 · t-SNE

그래프 모형

베이즈 네트워크 · 조건부 무작위장 · 은닉 마르코프 모형 · 잠재 디리클레 할당

이상 탐지

k -최근접 이웃 알고리즘 · 국소 특이점 요인

인공 신경망

오토인코더 · 딥 러닝 · 생성적 적대 신경망 · 다층 퍼셉트론 · 순환 신경망 (LSTM · GRU) · 제한된 볼츠만 머신 · 변환기 (비전) · 자기조직화지도 · 합성곱 신경망

강화 학습

Q 러닝 · SARSA · 시간차 학습

이론

편향-분산 트레이드오프 · 계산학습이론 · 경험적 위험 최소화 · PAC 러닝 · 통계적 학습이론 · VC 이론

관련 문서

기계 학습 알고리즘 목록

v · t · e (https://ko.wikipedia.org/w/index.php?title=%ED%8B%80:%EA%B8%B0EA%B3%84_%ED%95%99%EC%8A%B5&action=edit)

시리즈
인공지능

주요 목표

지식 추론 · 계획 · 기계 학습 · 자연어 처리 · 컴퓨터 비전 · 로봇공학 · 인공 일반 지능

접근

심볼릭 · 딥 러닝 · 베이지스 네트워크 · 진화 알고리즘

철학

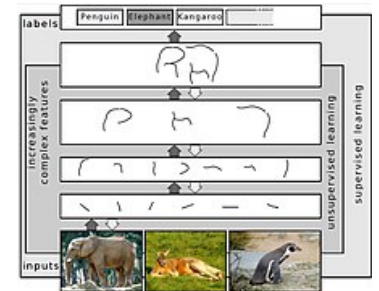
윤리 · 존재 문제 · 튜링 테스트 · 중국어 방 · 통제 문제 · 친절한 AI

역사

연표 · 진행상황 · AI 겨울

심층 학습(深層學習) 또는 **딥 러닝**(영어: deep structured learning, deep learning 또는 hierarchical learning)은 여러 비선형 변환기법의 조합을 통해 높은 수준의 추상화(abstractions, 다량의 데이터나 복잡한 자료들 속에서 핵심적인 내용 또는 기능을 요약하는 작업)를 시도하는 기계 학습 알고리즘의 집합^[1]으로 정의되며, 큰 틀에서 사람의 사고방식을 컴퓨터에게 가르치는 기계학습의 한 분야라고 이야기할 수 있다.

어떠한 데이터가 있을 때 이를 컴퓨터가 알아 들을 수 있는 형태(예를 들어 이미지의 경우는 픽셀정보를 열벡터로 표현하는 등)로 표현(representation)하고 이를 학습에 적용하기 위해 많은 연구(어떻게 하면 더 좋은 표현기법을 만들고 또 어떻게 이것들을 학습할 모델을 만들지에 대한)가 진행되고 있으며, 이러한 노력의 결과로 deep neural networks, convolutional deep neural networks, deep belief networks와 같은 다양한 딥 러닝 기법들이 컴퓨터 비전, 음성인식, 자연어 처리, 음성/신호처리 등의 분야에 적용되어 최첨단의 결과들을 보여주고 있다.



2012년 스탠포드대학의 앤드류 응과 구글이 함께한 딥 러닝 프로젝트에서는 16,000개의 컴퓨터 프로세서와 10억 개 이상의 neural networks 그리고 DNN(deep neural networks)을 이용하여 유튜브에 업로드 되어 있는 천만 개 넘는 비디오 중 고양이 인식에 성공하였다.^[2] 이 소프트웨어 프레임워크를 논문에서는 DistBelief로 언급하고 있다^[3]. 이뿐만 아니라 마이크로소프트, 페이스북^[4] 등도 연구팀을 인수하거나 자체 개발팀을 운영하면서 인상적인 업적들을 만들어 내고 있다.

딥 러닝의 역사

MIT가 2013년을 빛낼 10대 혁신기술 중 하나로 선정^[5] 하고 가트너(Gartner, Inc.)가 2014 세계 IT 시장 10대 주요 예측^[6]에 포함시키는 등 최근들어 딥 러닝에 대한 관심이 높아지고 있지만 사실 딥 러닝 구조는 인공지능경망(ANN, artificial neural networks)에 기반하여 설계된 개념으로 역사를 따지자면 최소 1980년 Kuniyiko Fukushima에 의해 소개된 Neocognitron^[7]까지 거슬러 올라가야 한다.

1989년에 안 르쿤과 그의 동료들은 오류역전파 알고리즘(backpropagation algorithm)^[8]에 기반하여 우편물에 손으로 쓰여진 우편번호를 인식하는 deep neural networks를 소개^[9] 했다. 알고리즘이 성공적으로 동작했음에도 불구하고, 신경망 학습에 소요되는 시간(10 개의 숫자를 인식하기 위해 학습하는 시간)이 거의 3일이 걸렸고 이것은 다른분야에 일반적으로 적용되기에는 비현실적인 것으로 여겨졌다.

많은 요소들이 느린 속도에 원인을 제공했는데, 그 중 하나는 1991년 Jürgen Schmidhuber의 제자였던 Sepp Hochreiter에 의해 분석된 vanishing gradient problem(지역최솟값에 머무르게 되는 원인)이었다^{[10][11]}. 또한 불연속 시뮬레이션에서 초기 상태를 어떻게 선택하느냐에 따라 수렴이 안되고 진동 또는 발산하는 문제, 트레이닝셋에 너무 가깝게 맞추어 학습되는 과적합 (Overfitting) 문제, 원론적으로 생물학적 신경망과는 다르다는 이슈들이 끊임 없이 제기되면서 인공지능경망은 관심에서 멀어졌고 90년대와 2000년대에는 서포트 벡터 머신 같은 기법들이 각광받게 된다.

본격적으로 딥 러닝이란 용어를 사용한 것은 2000년대 딥 러닝의 중흥기를 이끌어간다고 평가할 수 있는 제프리 힌튼^[12]과 Ruslan Salakhutdinov에 의해서이며, 기존 신경망의 과적합 문제를 해결하기 위해 이들은 unsupervised RBM(restricted Boltzmann machine)을 통해 학습시킬 앞막임 신경망(Feedforward Neural Network)의 각 층을 효과적으로 사전훈련(pre-training)하여 과적합을 방지할 수 있는 수준의 initialize point를 잡았고, 이를 다시 supervised backpropagation를 사용^[13]하는 형태로 학습을 진행한다.

또한 2013년에는 신호처리학회인 ICASSP에서 RBM을 대체하여 과적합을 방지할 수 있는 Drop-out^[14]이라는 개념이 소개되면서 사전훈련보다 훨씬 더 간단하고 강력한 형태로 과적합을 방지할 수 있게 되었다.

왜 다시 딥 러닝인가?

딥 러닝이 부활하게 된 이유는 크게 세 가지로 꼽힌다. 첫 번째는 앞서 딥 러닝의 역사에서 언급한 바 있는 기존 인공지능경망 모델의 단점이 극복되었다는 점이다. 그러나 과적합 문제만 해결되었다고 해서 느린 학습시간이 줄어드는 것은 아니다. 두 번째 이유로, 여기에는 하드웨어의 발전이라는 또다른 요인이 존재 한다. 특히 강력한 GPU는 딥 러닝에서 복잡한 행렬 연산에 소요되는 시간을 크게 단축시켰다. 마지막으로 언급하지만 가장 중요한 세 번째 이유로 빅 데이터를 들 수 있다. 대량으로 쏟아져 나오는 데이터들, 그리고 그것들을 수집하기 위한 노력 특히 SNS 사용자들에 의해 생산되는 다량의 자료와 태그정보들 모두가 종합되고 분석 되어 학습에 이용될 수 있다. - 인공지능경망의 학습에 사용되는 트레이닝벡터는 이름이 붙어 있는(labeled) 데이터여야 하는데

(supervised learning의 경우) 대량의 트레이닝셋 모두에 label을 달아주는 일은 불가능한 일이다. 이런 이유로 초기 학습에 사용되는 일부 데이터에 대해서만 지도학습(supervised learning)을 수행하고 나머지 트레이닝셋에 대해서는 비지도학습(unsupervised learning)을 진행하며, 학습된 결과는 기존 학습의 결과와 앞서 분석된 메타태그 정보들을 종합하여 인식이 완성 된다.

딥 러닝의 부활 이후 다양한 분야, 특히 자동 음성 인식(ASR, automatic speech recognition)과 컴퓨터비전 분야에서 최고수준의 성능을 보여주고 있으며 이들은 보통 딥 러닝의 새로운 응용들의 지속적인 성능 향상을 위해 만들어진 TIMIT(Texas Instruments와 MIT가 제작한 음성 Database), MNIST(이미지 클러스터링을 위한 hand-written 숫자 이미지 데이터베이스로 National Institute of Standards and Technology가 제작) 등의 데이터베이스를 사용했다. 최근에는 Convolution Neural Networks 기반의 딥 러닝 알고리즘이 뛰어난 성능을 발휘하고 있으며, 컴퓨터비전과 음성인식등의 분야에서 특히 탁월한 성능을 보이고 있다.

알고리즘

다양한 종류의 심층 신경망 구조가 존재하지만, 대부분의 경우 대표적인 몇 가지 구조들에서 파생된 것이다. 그렇지만 여러 종류의 구조들의 성능을 동시에 비교하는 것이 항상 가능한 것은 아닌데, 그 이유는 특정 구조들의 경우 주어진 데이터 집합에 적합하도록 구현되지 않은 경우도 있기 때문이다.

심층 신경망(Deep Neural Network, DNN)

심층 신경망(Deep Neural Network, DNN)은 입력층(input layer)과 출력층(output layer) 사이에 여러 개의 은닉층(hidden layer)들로 이뤄진 인공신경망(Artificial Neural Network, ANN)이다.^{[15][16]} 심층 신경망은 일반적인 인공신경망과 마찬가지로 복잡한 비선형 관계(non-linear relationship)들을 모델링할 수 있다. 예를 들어, 물체 식별 모델을 위한 심층 신경망 구조에서는 각 물체가 영상의 기본적 요소들의 계층적 구성으로 표현될 수 있다.^[17] 이때, 추가 계층들은 점진적으로 모여진 하위 계층들의 특징들을 규합시킬 수 있다. 심층 신경망의 이러한 특징은, 비슷하게 수행된 인공신경망에 비해 더 적은 수의 유닛(unit, node)들 만으로도 복잡한 데이터를 모델링할 수 있게 해준다.^[15]

이전의 심층 신경망들은 보통 앞먹임 신경망으로 설계되어 왔지만, 최근의 연구들은 심층 학습 구조들을 순환 신경망(Recurrent Neural Network, RNN)에 성공적으로 적용했다. 일례로 언어 모델링(language modeling) 분야에 심층 신경망 구조를 적용한 사례 등이 있다.^[18] 합성곱 신경망(Convolutional Neural Network, CNN)의 경우에는 컴퓨터 비전(computer vision) 분야에서 잘 적용되었을 뿐만 아니라, 각각의 성공적인 적용 사례에 대한 문서화 또한 잘 되어 있다.^[19] 더욱 최근에는 합성곱 신경망이 자동음성인식(Automatic Speech Recognition, ASR)을 위한 음향 모델링(acoustic modeling) 분야에 적용되었으며, 기존의 모델들보다 더욱 성공적으로 적용되었다는 평가를 받고 있다.^[20]

심층 신경망은 표준 오류역전파 알고리즘으로 학습될 수 있다. 이때, 가중치(weight)들은 아래의 등식을 이용한 확률적 경사 하강법(stochastic gradient descent)을 통하여 갱신될 수 있다.

$$\Delta w_{ij}(t+1) = \Delta w_{ij}(t) + \eta \frac{\partial C}{\partial w_{ij}}$$

여기서, η 는 학습률(learning rate)을 의미하며, C 는 비용함수(cost function)를 의미한다. 비용함수의 선택은 학습의 형태(지도 학습, 자율 학습 (기계 학습), 강화 학습 등)와 활성화함수(activation function)같은 요인들에 의해서 결정된다. 예를 들면, 다중 분류 문제(multiclass classification problem)에 지도 학습을 수행할 때, 일반적으로 활성화함수와 비용함수는 각각 softmax 함수와 교차

엔트로피 함수(cross entropy function)로 결정된다. softmax 함수는 $p_j = \frac{\exp(x_j)}{\sum_k \exp(x_k)}$ 로 정의된다, 이때, p_j 는 클래스 확률(class

probability)을 나타내며, x_j 와 x_k 는 각각 유닛 j 로의 전체 입력(total input)과 유닛 k 로의 전체 입력을 나타낸다. 교차 엔트로피는 $C = -\sum_j d_j \log(p_j)$ 로 정의된다, 이때, d_j 는 출력 유닛 j 에 대한 목표 확률(target probability)을 나타내며, p_j 는 해당 활성화

함수를 적용한 이후의 j 에 대한 확률 출력(probability output)이다.^[21]


심층 신경망의 문제점

기존의 인공신경망과 같이, 심층 신경망 또한 나이브(naive)한 방식으로 학습될 경우 많은 문제들이 발생할 수 있다. 그 중 과적합과 높은 시간 복잡도가 흔히 발생하는 문제들이다.

심층 신경망이 과적합에 취약한 이유는 추가된 계층들이 학습 데이터의 rare dependency의 모형화가 가능하도록 해주기 때문이다. 과적합을 극복하기 위해서 weight decay (ℓ_2 -regularization) 또는 sparsity (ℓ_1 -regularization) 와 같은 정칙화(regularization) 방법들이 사용될 수 있다.^[22] 그리고 최근에 들어서는 심층 신경망에 적용되고 있는 정칙화 방법 중 하나로 dropout 정칙화가 등장했다. dropout 정칙화에서는 학습 도중 은닉 계층들의 몇몇 유닛들이 임의로 생략된다. 이러한 방법은 학습 데이터(training data)에서 발생할 수 있는 rare dependency를 해결하는데 도움을 준다.^[23]


오차역전파법과 경사 하강법은 구현의 용이함과 국지적 최적화(local optima)에 잘 도달한다는 특성으로 인해 다른 방법들에 비해 선호되어온 방법들이다. 그러나 이 방법들은 심층 신경망을 학습 시킬 때 시간 복잡도가 매우 높다. 심층 신경망을 학습시킬 때에는 크기(계층의 수와 계층 당 유닛 수), 학습률, 초기 가중치 등 많은 매개변수(parameter)들이 고려되어야 한다. 하지만 최적의 매개변수들을 찾기 위해 매개변수 공간 전부를 확인하는 것은 계산에 필요한 시간과 자원의 제약으로 인해 불가능하다. 시간 복잡도를 해결하기 위해, 미니 배치(mini batch, 여러 학습 예제들의 경사를 동시에 계산), 드롭 아웃(drop out)과 같은 다양한 '요책'들이 등장하였다.^[24] 또한, 행렬 및 벡터 계산에 특화된 GPU는 많은 처리량을 바탕으로 두드러지는 학습 속도 향상을 보여주었다.^[16]

합성곱 신경망(Convolutional Neural Network, CNN)

 합성곱 신경망 문서를 참고하십시오.

합성곱 신경망(Convolutional Neural Network, CNN)은 최소한의 전처리(preprocess)를 사용하도록 설계된 다계층 퍼셉트론(multilayer perceptrons)의 한 종류이다. CNN은 하나 또는 여러개의 합성곱 계층과 그 위에 올려진 일반적인 인공 신경망 계층들로 이루어져 있으며, 가중치와 통합 계층(pooling layer)들을 추가로 활용한다. 이러한 구조 덕분에 CNN은 2차원 구조의 입력 데이터를 충분히 활용할 수 있다. 다른 딥 러닝 구조들과 비교해서, CNN은 영상, 음성 분야 모두에서 좋은 성능을 보여준다. CNN은 또한 표준 역전달을 통해 훈련될 수 있다. CNN은 다른 피드포워드 인공신경망 기법들보다 쉽게 훈련되는 편이고 적은 수의 매개변수를 사용한다는 이점이 있다. 최근 딥 러닝에서는 합성곱 심층 신뢰 신경망(Convolutional Deep Belief Network, CDBN)가 개발되었는데, 기존 CNN과 구조적으로 매우 비슷해서, 그림의 2차원 구조를 잘 이용할 수 있으며 그와 동시에 심층 신뢰 신경망(Deep Belief Network, DBN)에서의 선훈련에 의한 장점도 취할 수 있다. CDBN은 다양한 영상과 신호 처리 기법에 사용될 수 있는 일반적인 구조를 제공하며 CIFAR^[25]와 같은 표준 이미지 데이터에 대한 여러 벤치마크 결과에 사용되고 있다.

순환 신경망(Recurrent Neural Network, RNN)

 순환 인공 신경망 문서를 참고하십시오.

순환 신경망은 인공신경망을 구성하는 유닛 사이의 연결이 Directed cycle을 구성하는 신경망을 말한다. 순환 신경망은 앞먹임 신경망과 달리, 임의의 입력을 처리하기 위해 신경망 내부의 메모리를 활용할 수 있다. 이러한 특성에 의해 순환 신경망은 필기체 인식(Handwriting recognition)과 같은 분야에 활용되고 있고, 높은 인식률을 나타낸다.^[26] 순환 신경망을 구성할 수 있는 구조에는 여러 가지 방식이 사용되고 있다. 완전 순환망(Fully Recurrent Network), Hopfield Network, Elman Network, Echo state network(ESN), Long short term memory network(LSTM), Bi-directional RNN, Continuous-time RNN(CTRNN), Hierarchical RNN, Second Order RNN 등이 대표적인 예이다. 순환 신경망을 훈련(Training)시키기 위해 대표적으로 경사 하강법, Hessian Free Optimization, Global Optimization Methods 방식이 쓰이고 있다. 하지만 순환 신경망은 많은 수의 뉴런 유닛이나 많은 수의 입력 유닛이 있는 경우에 훈련이 쉽지 않은 스케일링 이슈를 가지고있다.

제한 볼츠만 머신(Restricted Boltzmann Machine, RBM)

볼츠만 머신에서, 층간 연결을 없앤 형태의 모델이다. 층간 연결을 없애면, 머신은 가시 유닛(Visible Unit)과 은닉 유닛(Hidden Unit)으로 이루어진 무방향 이분 그래프 형태의 모양이 된다. 결론적으로 모델의 층간 연결을 없앴으로써, 얻는 이점으로 뉴럴 네트워크는 깊어질 수 있었다. 가장 큰 이점은 가시 유닛이 관찰되고 고정(Clamped)되었을 때 은닉 유닛을 추론(Inference)하는 MCMC 과정이 단 한 번에 끝난다는 것이다. RBM은 확률모델의 계산이 불가능하기 때문에, 학습시 근사법인 MCMC 나 또는 제프리 힌튼 교수가 발견한 CD(Contrastive Divergence)를 사용하는데, RBM의 모델의 간단함에서 오는 추론에 대한 이점은 샘플링 근간의 학습법이 실용적인 문제에 적용되는데 기여했다. 자세한 학습 방법은 아래에 설명되어 있다.

RBM은 DBN의 기본 뼈대가 된다. RBM을 쌓아올리면서(Stacking), Greedy하게 학습함으로써, DBN을 완성한다. DBN을 아는 사람은 기본적으로 RBM은 무방향이기때, RBM이 방향성 있는 모델인 DBN이 되는지 의아할 것이다. 이는 제프리 힌튼의 제자인 Teh, Y. W의 우연 발견적인 연구 성과인데, 원래 힌튼 교수는 Graphical Model 이 방향성을 가질 때 생기는 Explaining Away 효과 때문에 학습이 매우 어려워 연구를 무방향성 모델로 선회했다가, RBM을 쌓으면, 다층 RBM이 되는 것이 아니라, DBN과 비슷해진다는 사실을 발견하게 된다. 이 발견은 DBN학습을 단순히 여러 개의 RBM 학습으로 환원시킴으로써(앞서서 얘기했듯이 RBM은 학습이 어렵지 않다) 어려운 DBN 학습의 복잡도를 층의 개수에 비례하는 복잡도로 낮추었다. 이는 선행학습(Pre-training)과 미세조정(fine-tuning)의 새로운 학습 패러다임으로 발전하게 된다.

RBM 훈련 과정에서의 가중치 갱신은, 다음의 식을 기반으로 경사 하강법을 통해 이루어진다.

$$\Delta w_{ij}(t+1) = w_{ij}(t) + \eta \frac{\partial \log(p(v))}{\partial w_{ij}} \quad \text{여기서 } p(v) \text{는 다음의 식으로 주어지는 가시 벡터의 확률을 나타내고,}$$

$$p(v) = \frac{1}{Z} \sum_h e^{-E(v,h)} \quad Z \text{는 정규화(normalization)를 위해 사용되는 분배함수(partition function)이며, } E(v,h) \text{는 신경망의 상태}$$

(state)에 부여되는 에너지 함수이다. 에너지가 낮을수록 해당 신경망이 더 적합한 상태임을 의미한다. 기울기 $\frac{\partial \log(p(v))}{\partial w_{ij}}$ 는 좀

더 간단하게 $\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}}$ 로 표현이 되는데 이 때 $\langle \dots \rangle_p$ 는 분포 p 에 관한 평균을 의미한다. $\langle v_i h_j \rangle_{\text{model}}$ 을 샘플링 하

려면 alternating Gibbs sampling을 매우 많이 반복해야 하기 때문에 CD에서는 alternating Gibbs sampling을 n 회만 반복한다.(실험을 통해 $n = 1$ 일 경우에도 충분한 성능이 나온다고 알려져 있다). n 회 반복 후 샘플링 된 데이터가 $\langle v_i h_j \rangle_{\text{model}}$ 를 대신한다. CD 수행 과정은 다음과 같다:

1. 가시 유닛들을 훈련 벡터로 초기화 한다.
2. 다음의 가시 유닛들이 주어지면 은닉 유닛들을 모두 동시에 갱신한다: $p(h_j = 1 | \mathbf{V}) = \sigma(b_j + \sum_i v_i w_{ij})$. σ 는 시그모이드 함수(sigmoid function)를 의미하며, b_j 는 h_j 의 편향이다.
3. 다음의 은닉 유닛들이 주어지면 가시 유닛들을 모두 동시에 갱신한다: $p(v_i = 1 | \mathbf{H}) = \sigma(a_i + \sum_j h_j w_{ij})$. a_i 는 v_i 의 편향이다. 이 부분을 "복원(reconstruction)"이라고 부른다.
4. 복원된 가시 유닛들이 주어지면 2에서 쓰인 식을 이용하여 은닉 유닛들을 모두 동시에 다시 갱신한다.
5. 다음과 같이 가중치를 갱신한다: $\Delta w_{ij} \propto \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{reconstruction}}$.

심층 신뢰 신경망 (Deep Belief Network, DBN)

심층 신뢰 신경망(Deep Belief Network, DBN)이란 기계학습에서 사용되는 그래프 생성 모형(generative graphical model)으로, 딥 러닝에서는 잠재변수(latent variable)의 다중계층으로 이루어진 심층 신경망을 의미한다. 계층 간에는 연결이 있지만 계층 내의 유닛 간에는 연결이 없다는 특징이 있다.

DBN은 생성 모형이라는 특성상 선행학습에 사용될 수 있고, 선행학습을 통해 초기 가중치를 학습한 후 역전파 혹은 다른 판별 알고리즘을 통해 가중치의 미조정을 할 수 있다. 이러한 특성은 훈련용 데이터가 적을 때 굉장히 유용한데, 이는 훈련용 데이터가 적을수록 가중치의 초기값이 결과적인 모델에 끼치는 영향이 세지기 때문이다. 선행학습된 가중치 초기값은 임의로 설정된 가중치 초기값에 비해 최적의 가중치에 가깝게 되고 이는 미조정 단계의 성능과 속도향상을 가능케 한다.

DBN은 비지도 방식으로 계층마다 학습을 진행하는데 이때 각각의 계층은 보통 RBM의 형태를 띄고 있다. RBM들을 쌓아 올리면서 DBN을 훈련시키는 방법에 대한 설명은 아래에 제공 되어 있다. RBM은 에너지 기반의 생성 모형으로 가시 유닛과 은닉 유닛으로 이루어진 무방향 이분 그래프 형태이다. 가시 유닛들과 은닉 유닛들 사이에만 연결이 존재한다.

RBM이 훈련되고 나면 다른 RBM이 그 위에 쌓아 올려짐으로써 다중 계층 모형을 형성한다. RBM이 쌓아 올려질 때마다, 이미 훈련된 RBM의 최상위 계층이 새로 쌓이는 RBM의 입력으로 쓰인다. 이 입력을 이용하여 새 RBM이 훈련되고, 원하는 만큼의 계층이 쌓일 때까지 해당 과정이 반복된다.

실험결과에 따르면, CD의 최대가능도 근사가 굉장히 투박함에도 불구하고, 심층 신경망 구조를 학습하기에는 충분히 효과적인 방식이라고 한다.

심층 Q-네트워크(Deep Q-Networks)

2015년 2월, 강화 학습을 위한 가장 최신 딥 러닝 모델이 네이처에 소개되었다.^[27]

응용

자동 음성 인식

아래 표는 TIMIT 데이터에 대한 자동 음성 인식 결과를 보여준다. 이 데이터셋은 딥 러닝의 초창기 평가를 위한 일반적인 데이터로서, 미국의 8가지 방언을 사용하는 총 630명의 사람이 읽은 10가지 문장으로 이루어져 있다.^[28] 데이터의 크기가 작기 때문에 다양한 설정을 효과적으로 적용할 수 있다. 더 중요한 점은 TIMIT에서 음소 순서 인식(phone-sequence recognition)을 고려한다는 점이다. 따라서, 단어 순서 인식(word-sequence recognition)과는 달리 아주 약한 "언어모델"을 허용하고 음성 인식에서의 음향 모델 측면을 더 쉽게 분석할 수 있다. 2009 ~ 2010년 무렵, 크고 작은 범위의 음성인식에 대한 딥 러닝 기술 활용을 위해서 많은 투자가 있었는데, Li Deng과 그의 동료들은 TIMIT에서의 GMM 과 DNN 모델을 비교하는 실험을 수행하였다.^{[29][30]} 결국 그들은 음성인식에서의 딥 러닝 활용에 있어서 가장 앞서나가게 되었다. 이 분석은 먼저 식별적 DNN과 발생적 모델 사이의 성능 비교(1.5% 이하의 오차율)로 수행되었다. 아래 표에 나타난 오차율은 앞에서 말한 초창기 실험을 포함하여 과거 20 여 년 간 수행된 실험들의 음소 오차율(Phone error rate)을 요약한 것이다.

방식	PER (%)
Randomly Initialized RNN	26.1
Bayesian Triphone GMM-HMM	25.6
Hidden Trajectory (Generative) Model	24.8
Monophone Randomly Initialized DNN	23.4
Monophone DBN-DNN	22.4
Triphone GMM-HMM with BMMI Training	21.7
Monophone DBN-DNN on fbank	20.7
Convolutional DNN ^[31]	20.0
Convolutional DNN w. Heterogeneous Pooling	18.7
Ensemble DNN/CNN/RNN ^[32]	18.2
Bidirectional LSTM	17.9

TIMIT로부터 대량 어휘 음성인식(large vocabulary speech recognition)으로의 딥 러닝의 확장은 2010년 산업계 연구자들에 의해 성공적으로 수행되었다.^{[33][34]} 자동 음성 인식 분야의 2014년 10월까지의 최신 동향은 마이크로소프트 리서치의 책 (<http://link.springer.com/book/10.1007%2F978-1-4471-5779-3>) 에 잘 정리되어있다.^[35] 또한 자동 음성인식과 관련된 배경 지식과 다양한 기계 학습 패러다임의 영향을 잘 정리한 글을 참고할 수 있다.^[36] 대용량 자동 음성인식은 최근 딥 러닝의 역사에서 산업계와 학계를 모두 아우르는 처음이자 가장 성공적인 케이스라고 할 수 있다. 2010년부터 2014년까지, 신호처리와 음성인식에 대한 주요 학술 회의인 IEEE-ICASSP 와 Interspeech는 음성인식을 위한 딥 러닝 분야의 합격 논문 개수에 있어서 거의 기하급수적인 성장을 보여 주었다. 더 중요한 것은, 현재 모든 주요 상업 음성인식 시스템(MS 코타나, 스카이프 번역기, 구글 나우, 애플 시리 등등)이 딥 러닝 기법에 기반하고 있다는 점이다.^{[37][38][39]} 뉘앙스 커뮤니케이션즈의 CTO의 최근 인터뷰도 참고할만하다.^[40]

영상 인식

일반적으로 이미지 분류를 위한 평가 데이터로서 MNIST 데이터베이스 데이터가 이용된다. MNIST는 손으로 쓴 숫자들로 구성되어 있으며, 60000개의 학습 예제들과 10000개의 테스트 예제들을 포함한다. TIMIT와 유사하게, 적은 용량의 MNIST 데이터는 복수의 테스트 환경설정이 가능하게 해준다. MNIST 데이터에 대한 종합적인 결과들을 [123]에서 확인할 수 있다. 현재까지 MNIST 데이터에 대한 가장 우수한 결과는 Ciresan 등이 작성한^[41]에서 달성되었으며, 오차율 0.23%를 기록했다.

제프리 힌튼과 그의 제자들은 2012년 가을에 대규모 ImageNet 대회에서 당시 최신 기계 학습 방법들의 성능을 훌쩍 뛰어넘는 결과를 보여주며 우승하였다. 이로 인해 컴퓨터 비전의 주요 분야인 영상 인식 및 사물 인식 분야에서의 딥 러닝의 중요성이 대두되었다. 그 당시, 대규모 음성인식에 딥 러닝이 상당히 잘 작동한다는 것을 알고 있었던 그들은, 20년 전에 고안된 심층 합성곱 신경망 구조를 대규모 작업에 맞도록 대규모로 사용하였다. 2013년부터 2014년에 이르기까지, 딥 러닝을 이용한 ImageNet 과제 결과의 오차율은 대규모 음성인식 분야와 추세를 같이하며 빠르게 줄어났다.

자동 음성인식 분야의 자동 음성 번역 및 이해 분야로의 확장과 마찬가지로, 이미지 분류 분야는 자동 영상 캡션링(captioning)이라는 더욱 도전적인 분야로 확장되었다. 자동 영상 캡셔닝은 딥 러닝을 핵심 기반 기술로 사용하는 분야이다.^{[42] [43] [44] [45]}

적용 사례로는 360°카메라 화면을 이해할 수 있도록 딥 러닝을 통해 학습된 자동차 탐재용 컴퓨터 등이 있다.^[46]

자연어 처리

2000년대 초부터 인공신경망은 언어 모형을 구현하기 위해 사용되어 왔다. 이 분야에서의 핵심 기법은 negative sampling과 단어 표현(word embedding)이다. word2vec과 같은 단어 표현은 데이터집합으로 주어진 단어들 사이의 관계를 학습하는 인공신경망을 이용하여 단어를 벡터 공간 상에 나타내는 것이라고 할 수 있다. 단어표현을 재귀 신경망(recursive neural network)의 입력 계층으로 이용하면 해당 신경망이 *compositional vector grammar*를 통해 문장과 구(phrase)를 분석하도록 학습시킬 수 있다. 이 *compositional vector grammar*는 재귀 신경망으로 구현된 probabilistic context free grammar (PCFG) 라고 할 수 있다. 단어표현을 기반으로 구성된 Recursive autoencoder는 문장 간의 유사도 판단과 의역 탐지를 하도록 훈련이 가능하다. 이러한 심층 인공신경망 구조들은 자동 번역(machine translation), 감정 분석(sentiment analysis), 정보 검색(information retrieval)을 비롯한 다양한 자연어처리 관련 연구에서 최첨단 기술로서 쓰이고 있다.

약물 발견과 독성학

제약 산업에서 많은 수의 약들은 시장에 출시되지 못한다. 이렇게 실패하는 이유중 하나는 생각한 만큼의 효험을 보이지 못하거나, 예상치 못한 다른 작용을 일으키기 때문이다. 이를 해결하기 위해 2012년 George Dahl의 팀은 다중 DNN을 약의 효험을 예측하는 데에 활용하여 "Merck Molecular Activity Challenge"에 우승하였고^{[47][48]}, 2014년 Sepp Hochreiter의 팀은 예상치 못한 다른

작용을 사전에 탐지하기 위해 딥 러닝을 활용하여 "Tox21 Data Challenge"에서 우승하였다.^{[49][50]} 이러한 성공은 딥 러닝이 약학에 있어 가상 실험 방법(Virtual Screening Method)에 적합함을 보여준다.^{[51][52]} 현재 많은 수의 연구자들이 여러 가지 데이터를 조합하여 약물을 발견하는 데에 딥 러닝을 활용하고 있다.^[53]

고객 관계 관리

최근 직접 마케팅 기획, 고객 관계 관리 자동화를 위한 수단 적합성 산출 등에 직접적으로 심층 강화 학습을 활용하여 성공하는 사례들이 알려지고 있다. 인공 신경망은 RFM으로 정의된 고객들에 대한 활용 가능한 마케팅 활동의 값을 예측할 때 사용되었다. 예측된 함수는 고객 생명 주기 값(customer lifetime value, CLV)으로 변환되어 사용되었다.^[54]

딥 러닝과 인간 두뇌

딥 러닝은 인지신경과학자(Cognitive neuroscientist)들이 1990년대 초에 제안한 뇌 발달(Brain development)과 밀접한 관련이 있다.^{[55][56][57]} 이러한 발달 관점의 이론들이 도입되면서, 순수한 전산 기반의 딥 러닝 모델들을 위한 기술적인 기반도 마련되었다. 발달 이론에서는 신경발달요인(Nerve growth factor)의 물결과 같은 뇌에서의 다양한 학습 역학이 결국은 서로 연관된 신경망들의 자기조직화를 도와준다는 특징이 있다. 이는 나중에 전산 기반의 딥 러닝 모델에서도 활용되어서, 인공신경망의 계층적인 필터 구조(각 동작 환경에서 필요한 정보만 걸러내는 다중 계층 구조)가 실제 뇌의 피질과 유사해보이게 되었다. 이러한 과정을 통해 자기조직적인 변환기의 계층구조가 만들어지고 각 환경에 맞도록 조율된다. 1995년에 뉴욕 타임스에서는 다음과 같이 말하였다. "...유아의 두뇌는 영양적인 요인의 영향으로 스스로 조직화되는 것 같다... 뇌에서 한 층의 조직이 먼저 성숙되고 다른 부분과 순차적으로 연결되는 방식으로 전체 뇌가 성숙될 때까지 반복된다."^[58]

인간의 인식 발달 및 진화와 관련하여 딥 러닝의 중요성은 많은 과학자들의 관심을 끌고 있다. 가까운 영장류 동물들과 인간이 차별화되는 부분 중 하나는 바로 발달 시기이다.^[59] 다른 영장류 동물들의 뇌가 출산 전에 거의 완성되는 반면에, 인간의 뇌는 비교적 출산 후에도 계속 발달하는 편이다. 그래서 인간의 경우 뇌가 발달되는 중요한 시기 동안 세상 밖의 훨씬 더 복잡한 경험에 노출될 수 있다. 이러한 현상은 인간이 다른 동물들에 비하여 빠르게 변화하는 환경에 더 잘 적응할 수 있도록 만든다. 이러한 변화의 정도는 대뇌 피질 발달에 반영되기도 하고, 또한 두뇌의 자기조직화 시기에 자극적인 환경으로부터의 정보 추출에 변화를 준다. 물론, 이러한 유연성은 인간이 다른 동물들에 비해 긴 미성숙기(보호자에게 도움을 받고 훈련을 받아야 하는 의존적인 시기)를 가지게 된 원인이기도 하다. 딥 러닝의 이런 이론들은 결국 인간 진화의 기본적인 조건으로서 문화와 인식의 공진화를 보여준다.

영리 활동

딥 러닝은 종종 강인공지능의 현실화를 향한 발걸음으로 묘사되기도 하기 때문에 많은 기관에서 딥 러닝의 활용에 관심을 보이고 있다. 2013년 3월, 제프리 힌튼과 두 명의 제자(Alex Krizhevsky and Ilya Sutskever)가 구글에 채용되었다. 이들은 구글의 기존 기계 학습 제품들을 향상시키고 구글이 가지고있는 데이터를 더 잘 활용할 수 있도록 돕는 일을 할 것이다. 구글은 또한 힌튼의 회사 DNNresearch를 인수하였다. 2013년 12월에는 페이스북이 캘리포니아, 런던, 뉴욕에서 운영될 인공지능 연구실의 수장으로 얀 르쿤을 채용했다고 발표하였다. 이 인공지능 연구실에서는 딥 러닝 기법들을 연구 하여 페이스북의 사진에 자동으로 사람의 이름을 연결시키는 등의 기능을 구현할 예정이다.^[60] 2014년, 구글은 영국의 스타트업기업인 딥마인드 테크놀로지스를 인수하였다. 딥마인드에서는 화면의 정보만을 가지고 비디오 게임을 어떻게 플레이하는지를 스스로 학습하는 시스템을 개발하였다. 바이두는 앤드류 응을 그들의 새로운 실리콘밸리 딥 러닝 연구실의 수장으로 채용하였다.

부정적 시각과 코멘트

딥 러닝에 대한 부정적 시각중 하나로 딥 러닝에 사용되는 방법들의 이론적인 뒷받침이 빈약하다는 것이 있다. 대부분의 딥 러닝 알고리즘은 경사 하강법에 기초를 두고 있다. 그런데 경사 하강법 자체는 이론적으로 이해가 잘 되었지만, 이와 함께 사용하는 다른 알고리즘은 이론적인 검증이 빈약하다. 그러한 알고리즘 중 하나인 contrastive divergence는 이 알고리즘이 실제로 정말 수렴하는지, 수렴한다면 얼마나 빠르게 수렴하는지에 대한 부분이 현재 분명하게 증명되지 않았다. 이와 같이, 딥 러닝에 사용되는 방법들은 이론적이기보다는 경험적으로 검증된 방법들을 사용하기 때문에 종종 블랙박스로 이해되기도 한다.

또다른 사람들은 보통 사람들이 딥 러닝을 모든 것을 해결해주는 솔루션으로 바라보는 것에 대해 딥 러닝은 유일한 솔루션이 아니고, 강한 인공 지능을 실현하는 데에 한 걸음 더 다가갈 수 있게 해주는 것으로 봐야한다고 주장한다. 딥 러닝을 모든 것을 해결해주는 솔루션으로 보기에는 기능적으로 여전히 부족한 점이 많기 때문이다. 심리학자인 게리 마커스는 다음과 같이 말했다.

"현실적으로, 딥 러닝은 지능형 기계를 구축하는 것의 일부분에 불과합니다. 딥 러닝에 쓰이는 테크닉들은 인과 관계를 나타내는 능력이 부족하고 논리적 추론을 하는 것에 확실한 방법을 가지고있지 않습니다. 그리고 이 방법들은 어떤 물체가 무엇인지, 용도가 무엇인지 등과 같은 추상적인 정보를 인식하는 데에 오랜 시간이 걸립니다. 왓슨과 같이 현재 가장 강력한 인공지능 시스템은 딥 러닝을 복잡한 여러 개의 테크닉 중 단지 한 부분을 위해 사용하고 있습니다."^[61]

딥 러닝 소프트웨어 라이브러리

- Torch (<http://torch.ch/>)

- Theano (<http://www.deeplearning.net/software/theano/>) Archived (<https://web.archive.org/web/20160531112558/http://www.deeplearning.net/software/theano/>) 2016년 5월 31일 - 웨이백 머신
- Deeplearning4j (<http://deeplearning4j.org/>), JVM에서의 분산 딥 러닝 라이브러리.
- ND4J (<http://nd4j.org/>) Archived (<https://web.archive.org/web/20200601013308/http://nd4j.org/>) 2020년 6월 1일 - 웨이백 머신
- NVIDIA cuDNN (<https://developer.nvidia.com/cuDNN>) DNN의 프리미티브를 제공해주는 라이브러리.
- DeepLearnToolbox (<https://github.com/rasmusbergpalm/DeepLearnToolbox/graphs/contributors>), MATLAB/Octave 딥 러닝을 위한 툴박스.
- convnetjs (<https://github.com/karpathy/convnetjs>), 자바스크립트로 구현된 딥 러닝 라이브러리.
- Gensim (<https://radimrehurek.com/gensim/>) 자연언어처리를 위한 킷.
- Caffe (<http://caffe.berkeleyvision.org/>) Berkeley Vision and Learning Center (BVLC)에서 만든 딥러닝 프레임워크
- Tensorflow (<http://www.tensorflow.org/>) 구글에서 발표한 기계학습 오픈소스 library
- CNTK Microsoft에서 발표한 오픈소스 library

각주

1. Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," IEEE Trans. PAMI, special issue Learning Deep Architectures, 2013
2. How Many Computers to Identify a Cat? 16,000 http://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html?_r=2&_amp&
3. <http://cs.stanford.edu/people/ang/?portfolio=deep-learning-and-unsupervised-feature-learning>
4. <http://techcrunch.com/2013/12/09/facebook-artificial-intelligence-lab-lecun/>
5. MIT, 올해의 10대 혁신기술 선정 <http://news.donga.com/3/all/20130426/54713529/1>
6. 가트너, 2014 세계 IT 시장 10대 주요 예측 발표 <http://www.acrofan.com/ko-kr/commerce/news/20131013/00000015>
7. Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position <http://www.cs.princeton.edu/courses/archive/spr08/cos598B/Readings/Fukushima1980.pdf>
8. P. Werbos., "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," PhD thesis, Harvard University, 1974.
9. LeCun et al., "Backpropagation Applied to Handwritten Zip Code Recognition," Neural Computation, 1, pp. 541–551, 1989. <https://www.ics.uci.edu/~welling/teaching/273ASpring09/lecun-89e.pdf> Archived (<https://web.archive.org/web/20150529000503/https://www.ics.uci.edu/~welling/teaching/273ASpring09/lecun-89e.pdf>) 2015년 5월 29일 - 웨이백 머신
10. S. Hochreiter., "Untersuchungen zu dynamischen neuronalen Netzen," Diploma thesis. Institut f. Informatik, Technische Univ. Munich. Advisor: J. Schmidhuber, 1991.
11. S. Hochreiter et al., "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," In S. C. Kremer and J. F. Kolen, editors, A Field Guide to Dynamical Recurrent Neural Networks. IEEE Press, 2001
12. http://www.wired.com/2013/03/google_hinton/
13. G. E. Hinton., "Learning multiple layers of representation," Trends in Cognitive Sciences, 11, pp. 428–434, 2007.
14. G. Dahl et al., "Improving DNNs for LVCSR using rectified linear units and dropout," ICASSP, 2013.
15. Y. Bengio, A. Courville, and P. Vincent., "Representation Learning: A Review and New Perspectives," IEEE Trans. PAMI, special issue Learning Deep Architectures, 2013
16. J. Schmidhuber, "Deep Learning in Neural Networks: An Overview" <http://arxiv.org/abs/1404.7828>, 2014
17. Szegedy, Christian, Alexander Toshev, and Dumitru Erhan. "Deep neural networks for object detection." Advances in Neural Information Processing Systems. 2013.
18. T. Mikolov et al., "Recurrent neural network based language model," Interspeech, 2010.
19. LeCun, Y.; 외. "Gradient-based learning applied to document recognition". 《Proceedings of the IEEE》 **86** (11): 2278–2324. doi:10.1109/5.726791 (<https://dx.doi.org/10.1109%2F5.726791>).
20. T. Sainath et al., "Convolutional neural networks for LVCSR," ICASSP, 2013.
21. G. E. Hinton et al., "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The shared views of four research groups," IEEE Signal Processing Magazine, pp. 82–97, November 2012.
22. Y. Bengio et al., "Advances in optimizing recurrent networks," ICASSP, 2013.
23. G. Dahl et al., "Improving DNNs for LVCSR using rectified linear units and dropout," ICASSP, 2013.
24. G. E. Hinton., "A Practical Guide to Training Restricted Boltzmann Machines," Tech. Rep. UTML TR 2010-003, Dept. CS., Univ. of Toronto, 2010.
25. <http://www.cs.toronto.edu/~kriz/conv-cifar10-aug2010.pdf>

26. A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, J. Schmidhuber. A Novel Connectionist System for Improved Unconstrained Handwriting Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, 2009.
27. Mnih, Volodymyr et al. (2015). "Human-level control through deep reinforcement learning" (<http://www.nature.com/nature/journal/v518/n7540/pdf/nature14236.pdf>) (PDF) 518. pp. 529–533.
28. TIMIT Acoustic-Phonetic Continuous Speech Corpus Linguistic Data Consortium, Philadelphia.
29. Deng, L.; Hinton, G.; Kingsbury, B. (2013). "New types of deep neural network learning for speech recognition and related applications: An overview (ICASSP)".
30. NIPS Workshop: Deep Learning for Speech Recognition and Related Applications, Whistler, BC, Canada, Dec. 2009 (Organizers: Li Deng, Geoff Hinton, D. Yu).
31. Abdel-Hamid, O. et al. (2014). "Convolutional Neural Networks for Speech Recognition". *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22 (10): 1533–1545. doi:10.1109/taslp.2014.2339736 (<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6857341>)
32. Deng, L.; Platt, J. (2014). "Ensemble Deep Learning for Speech Recognition". *Proc. Interspeech*.
33. Yu, D.; Deng, L. (2010). "Roles of Pre-Training and Fine-Tuning in Context-Dependent DBN-HMMs for Real-World Speech Recognition". *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*.
34. Deng L., Li, J., Huang, J., Yao, K., Yu, D., Seide, F. et al. Recent Advances in Deep Learning for Speech Research at Microsoft. ICASSP, 2013.
35. Yu, D.; Deng, L. (2014). "Automatic Speech Recognition: A Deep Learning Approach (Publisher: Springer)".
36. Deng, L.; Li, Xiao (2013). "Machine Learning Paradigms for Speech Recognition: An Overview". *IEEE Transactions on Audio, Speech, and Language Processing*.
37. L. Deng and D. Yu (2014) "Deep Learning: Methods and Applications" <http://research.microsoft.com/pubs/209355/DeepLearning-NowPublishing-Vol7-SIG-039.pdf>
38. McMillan, R. "How Skype Used AI to Build Its Amazing New Language Translator", *Wire*, Dec. 2014.
39. Hannun et al. (2014) "Deep Speech: Scaling up end-to-end speech recognition", arXiv:1412.5567.
40. Ron Schneiderman (2015) "Accuracy, Apps Advance Speech Recognition --- Interview with Vlad Sejnoha and Li Deng", *IEEE Signal Processing Magazine*, Jan, 2015.
41. D. Ciresan, U. Meier, J. Schmidhuber., "Multi-column Deep Neural Networks for Image Classification," *Technical Report No. IDSIA-04-12*, 2012.
42. Vinyals et al. (2014). "Show and Tell: A Neural Image Caption Generator," arXiv:1411.4555.
43. Fang et al. (2014). "From Captions to Visual Concepts and Back," arXiv:1411.4952.
44. Kiros et al. (2014). "Unifying Visual-Semantic Embeddings with Multimodal Neural Language Models," arXiv:1411.2539.
45. Zhong, S.; Liu, Y.; Liu, Y. "Bilinear Deep Learning for Image Classification". *《Proceedings of the 19th ACM International Conference on Multimedia》* 11: 343–352.
46. Nvidia Demos a Car Computer Trained with "Deep Learning" (<http://www.technologyreview.com/news/533936/nvidia-demos-a-car-computer-trained-with-deep-learning/>) (2015-01-06), David Talbot, *MIT Technology Review*
47. "Announcement of the winners of the Merck Molecular Activity Challenge" <https://www.kaggle.com/c/MerckActivity/details/winners>.
48. Dahl, G. E.; Jaitly, N.; & Salakhutdinov, R. (2014) "Multi-task Neural Networks for QSAR Predictions," ArXiv, 2014.
49. "Toxicology in the 21st century Data Challenge" <https://tripod.nih.gov/tox21/challenge/leaderboard.jsp>
50. "NCATS Announces Tox21 Data Challenge Winners" <http://www.ncats.nih.gov/news-and-events/features/tox21-challenge-winners.html> Archived (<https://web.archive.org/web/20150228225709/http://www.ncats.nih.gov/news-and-events/features/tox21-challenge-winners.html>) 2015년 2월 28일 - 웨이백 머신
51. Unterthiner, T.; Mayr, A.; Klambauer, G.; Steijaert, M.; Ceulemans, H.; Wegner, J. K.; & Hochreiter, S. (2014) "Deep Learning as an Opportunity in Virtual Screening" (<http://www.bioinf.jku.at/publications/2014/NIPS2014a.pdf>). Workshop on Deep Learning and Representation Learning (NIPS2014).
52. Unterthiner, T.; Mayr, A.; Klambauer, G.; & Hochreiter, S. (2015) "Toxicity Prediction using Deep Learning" (<http://arxiv.org/pdf/1503.01445v1>). ArXiv, 2015.
53. Ramsundar, B.; Kearnes, S.; Riley, P.; Webster, D.; Konerding, D.; & Pande, V. (2015) "Massively Multitask Networks for Drug Discovery". ArXiv, 2015.
54. Tkachenko, Yegor. Autonomous CRM Control via CLV Approximation with Deep Reinforcement Learning in Discrete and Continuous Action Space. (April 8, 2015). arXiv.org: <http://arxiv.org/abs/1504.01840>
55. J. Elman, et al., "Rethinking Innateness," 1996.
56. J. Shrager, MH Johnson., "Dynamic plasticity influences the emergence of function in a simple cortical array," *Neural Networks*, 9 (7), pp. 1119–1129, 1996
57. SR Quartz and TJ Sejnowski., "The neural basis of cognitive development: A constructivist manifesto," *Behavioral and Brain Sciences*, 20 (4), pp. 537–556, 1997.
58. S. Blakeslee., "In brain's early growth, timetable may be critical," *The New York Times*, Science Section, pp. B5–B6, 1995.

59. {BUFILL} E. Bufill, J. Agusti, R. Blesa., "Human neoteny revisited: The case of synaptic plasticity," American Journal of Human Biology, 23 (6), pp. 729–739, 2011.
60. C. Metz., "Facebook's 'Deep Learning' Guru Reveals the Future of AI," <http://www.wired.com/wiredenterprise/2013/12/facebook-yann-lecun-qa/> Wired, 12 December 2013.
61. G. Marcus., "Is 'Deep Learning' a Revolution in Artificial Intelligence?" *The New Yorker*, 25 November 2012.

외부 링크

- java deep neural networks with gnu acceleration (<https://github.com/ivan-vasilev/neuralnetworks>): 딥러닝 구현
 - <http://neuralnetworksanddeeplearning.com/>
 - <http://www.deeplearning.net/tutorial/> Archived (<https://web.archive.org/web/20140704202655/http://www.deeplearning.net/tutorial/>) 2014년 7월 4일 - 웨이백 머신
 - <http://theanalyticsstore.com/deep-learning/>
 - <http://devview.kr/2013/detail.nhn?topicSeq=39>
 - <http://www.toptal.com/machine-learning/an-introduction-to-deep-learning-from-perceptrons-to-deep-networks>
 - 인공지능과 딥러닝 빅데이터 안고 부활하다 (<http://www.mt.co.kr/view/mtview.php?type=1&no=2015022514525399240&outlink=1>)
 - 상상 속 기계가 스스로 학습 기계로 (<http://www.mt.co.kr/view/mtview.php?type=1&no=2015022514490867206&outlink=1>)
 - 폭넓은 관심과 남다른 통찰로 딥러닝 시대 열어 (<http://www.mt.co.kr/view/mtview.php?type=1&no=2015022514293292353&outlink=1>)
-

원본 주소 "https://ko.wikipedia.org/w/index.php?title=딥_러닝&oldid=33505639"