

유전 알고리즘을 이용한 신경망 구조 최적화

손병욱, 최지원, 이화경, 조한희, 박희민*

상명대학교 소프트웨어학과

상명대학교 소프트웨어학과 교수*

e-mail: heemin@smu.ac.kr

Neural Network Architecture Optimization using Genetic Algorithm

Byung-Wook Son, Ji-won Choi, Hwa-Kyung Lee, Han-Hee Cho, Hee-min Park*

Dept of Software, Sang-Myung University

Professor, Dept of Software, Sang-Myung University*

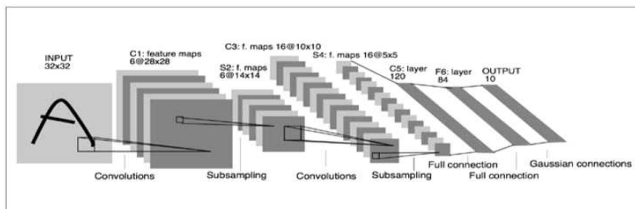
요 약

신경망 학습에 있어서 하이퍼파라미터를 설정하는 것과 신경망의 구조를 설계하는 것은 신경망의 성능에 있어서 매우 핵심적인 부분이다. 하지만 사람이 하이퍼파라미터와 신경망의 구조를 어떻게 정해주어야 신경망의 성능이 좋을지 알기도 어려울뿐더러 많은 시간이 요구된다. 현재 대부분의 신경망 구조는 전문가의 경험적 지식에 의존하여 직관적인 판단에 의해 설계된다. 이에 본 논문에서는 유전 알고리즘을 이용하여 하이퍼파라미터와 신경망 구조를 최적화하는 방법에 대해 소개한다.

1. 서 론

현재 신경망(Neural Network) 학습에 있어서 하이퍼파라미터를 설정하는 것과 신경망의 구조를 설계하는 것은 신경망의 성능에 있어서 매우 핵심적인 부분을 차지한다. 여기서 하이퍼파라미터란 훈련 데이터와 학습 알고리즘에 의해 자동 획득되는 매개변수와 달리, 사람이 직접 설정해야 하는 매개변수이다[1]. 대부분의 신경망을 학습시킬 때 많은 하이퍼파라미터를 설정은 사람이 직접 설정하는 경우가 많다. 하지만 사람이 하이퍼파라미터와 신경망 구조를 어떻게 정해야 신경망의 성능이 좋을지 알기 어렵고 많은 시간을 필요로 한다. 본 연구는 유전 알고리즘을 이용해 신경망 구조 최적화가 사람에게 비해 적은 시간을 소비하면서도, 유사하거나 더 좋은 결과를 도출할 수 있음을 보이는 것을 목표로 한다.

2. CNN (Convolution Neural Network)

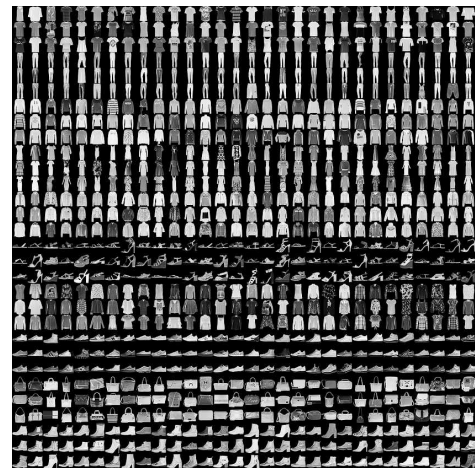


<그림 1. CNN구조>

합성곱 신경망(CNN, Convolution Neural Network)은 시각적인 이미지를 분석하는데에 가장 흔하게 쓰이는 신경망이다. 합성곱층(convolution layer)과 풀링층(pooling layer)으로 구성되어 있으며 마지막 출력층과 가까운 층은 완전연결층(fully-connected)으로 구현하는 것이 일반적인 CNN 구조이다. 이미지를

분석할 때 이전의 완전연결 계층으로 구성된 신경망으로 모델을 만들면 3차원 데이터인 이미지를 1차원으로 펼쳐주어야 하는데 이 과정에서 3차원 데이터의 형상이 무시되면서 공간적 특징을 잃어버리게 된다. 하지만 합성곱 신경망은 입력 데이터의 형상을 그대로 유지하여 이미지처럼 형상을 갖는 데이터를 제대로 학습할 가능성이 높다. 합성곱 신경망의 층 깊이가 깊어지면 역전파를 하는 과정에서 가중치 기울기가 사라지는 경사 사라짐(Vanishing Gradient)문제가 발생할 수 있다.

2-1. FASHION MNIST

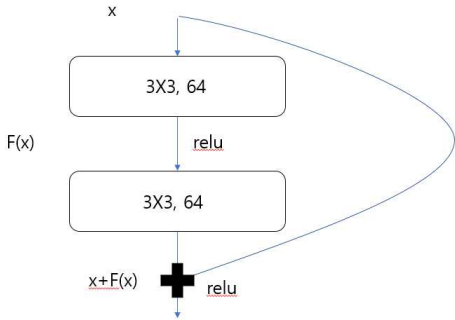


<그림 2. FASHION MNIST 데이터>

본 논문은 Fashion MNIST 데이터를 사용하는 CNN 구조를 최적화한다. Fashion MNIST는 10개의 카테고리과 7만 장의 흑백 이미지로 구성된 데이터셋이다. 학습 이미지가 6만 장, 테스트 이미지가 1만 장으로 구성되어 있고, MNIST와 동일하

게 이미지 데이터의 해상도는 28x28 픽셀이다. 카테고리 레이블은 0에서 9까지의 정수 배열이고, 각각 T-shirt, Dress, Coat, Sandal, Bag 등의 10가지 클래스로 구성되어 있다.

3. ResNet



<그림3. short cut connection>

ResNet은 마이크로소프트 팀이 개발한 네트워크로, 층이 깊어질수록 발생하는 학습오류를 줄이기 위해 3x3의 합성곱 레이어 두 개를 거친 결과에 거치기 전 결과를 더하는 skip connection을 도입한 네트워크 모델이다[1]. 2015년 이미지넷 이미지 인식 대회인 ILSVRC에서 우승한 우수한 네트워크 모델이므로 본 논문 시스템에서 skip connection을 고려한 모델을 최적화한다.

4. 유전 알고리즘

유전 알고리즘이란 생물의 진화 과정을 모방한 것으로, 선택, 교차, 변이 등의 연산을 통해 최적에 가까운 해를 찾아주는 방법론이다[5]. 유전 알고리즘의 구조는 초기 세대의 염색체 구성, 초기 세대의 유전염색체들에 대한 적합도 계산, 적합도가 높은 염색체 선택, 자손 생성의 반복으로 이루어진다. 이 과정에서 교차, 돌연변이 연산을 통해 유전자풀의 다양성을 확보할 수 있다. 유전 알고리즘은 해를 정형화된 유전표현으로 나타내어야 하고, 각각의 개체를 평가하기 위한 적합도 함수를 정의해주어야 한다[6].

유전 알고리즘으로 유전표현들을 최적화시키기 전에 하이퍼파라미터가 CNN의 성능에 직접적으로 영향을 주는지 확인하기 위해 사전 실험을 진행했다. 사전 실험에 사용된 CNN 구조는 skip connection이 없는 일반적인 CNN이다.

<표 1. 각 요소 값 변화에 대한 정확도 변화>

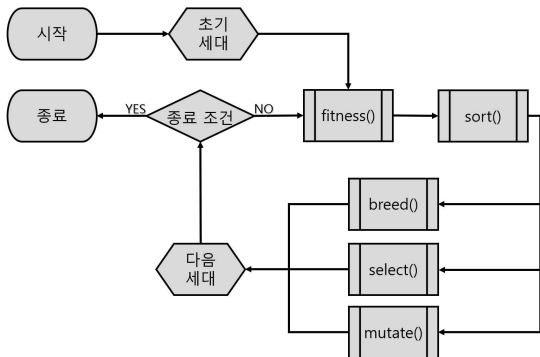
	Activation Function	Optimizer	Weight Init	Conv Layer Dept	Kernel Size	Learning Rate	FC Layer Dept	DropOut Rate	Epochs	Accuracy
Original	relu	adam	Xavier	3	3	0.001	2	0.3	20	0.8819
A	relu	adam	Xavier	3	3	0.001	2	0.3	5	0.8417
B	relu	adam	Xavier	3	3	0.001	2	0.5	20	0.8468
C	relu	adam	Xavier	3	3	0.001	5	0.3	20	0.8666
D	relu	adam	Xavier	3	3	0.01	2	0.3	20	0.7621
E	relu	adam	Xavier	3	7	0.001	2	0.3	20	0.9168
F	relu	adam	Xavier	8	3	0.001	2	0.3	20	0.9199
G	relu	adam	Zero	3	3	0.001	2	0.3	20	0.1000
H	relu	SGD	Xavier	3	3	0.001	2	0.3	20	0.6965
I	sigmoid	adam	Xavier	3	3	0.001	2	0.3	20	0.8321

위 실험결과를 통해 하이퍼파라미터 값과 신경망의 구조가

신경망의 성능에 영향을 미친다는 사실을 알 수 있다.

5. 시스템 설계

본 논문은 유전 알고리즘을 사용해 신경망의 구조 및 파라미터를 최적화하는 것을 목표로 한다. 따라서 본 논문의 유전 알고리즘의 염색체 유전정보에는 learning rate, 가중치 초기값의 종류, optimizer의 종류, 활성화 함수의 종류, 커널 크기, 합성곱 블록의 개수, 한 블록에 들어가는 합성곱 층의 개수, 완전 연결 층의 개수, drop out 비율, epoch 수를 넣어준다. 각 염색체를 평가하기 위한 적합도 함수는 유전정보에 따라 설계된 CNN의 정확도로 설정한다. 위의 유전정보의 형태에 맞게 초기 세대를 정의해주어야 하는데 이때 일정 범위 내의 무작위 값들을 넣어서 정의해준다. 본 논문의 시스템에서는 100개의 개체를 유전 알고리즘의 입력층으로 넣어준다. 임의로 생성된 초기 세대에서 적합도를 구하고 선택, 교배, 변이 연산을 통해 다음 세대로 넘어갈 개체들을 정해준다. 이 과정을 반복하여 최적해에 가까운 적절한 하이퍼파라미터 값과 신경망의 구조를 찾을 수 있다.



<그림 2. 시스템 구조도>

5-1. 적합도 함수

본 유전 알고리즘의 적합도는 유전자에 의해 만들어진 CNN의 성능이다. 따라서 유전 알고리즘을 통해 생성된 하이퍼파라미터와 신경망 구조의 정보를 담고 있는 개체가 입력으로 들어가게 되고, 그 유전정보를 통해 만들어진 CNN에 FASHION M NIST 데이터를 학습시켜 얼마나 좋은 성능을 내는지 정확도를 출력으로 설정하였다. 이렇게 구한 적합도를 sort() 함수를 통해 적합도가 높은 개체부터 내림차순으로 재배열을 해준다.

5-2. 선택

선택 연산에서는 적합도 함수를 거쳐 내림차순으로 정렬된 적합도 값들을 기반으로 다음 세대로 보낼 개체들을 우선적으로 선택한다. 본 시스템에서는 다음 세대로 보낼 개체들을 선택하는 방법으로 상위 30개의 개체를 우선 선택한 후 남은 70 자리를 교배와 변이 연산을 통해 채운다.

5-3. 변이

변이 연산에서는 유전형의 값에 변화를 주어 더 다양한 유전 자풀을 갖게 해주는 연산을 수행한다. 본 시스템의 변이 연산에서는 선택 연산에서 선택된 30개의 개체를 받아와 각 개체에 대해 최소 1개에서 3개의 유전형의 값에 대해 변화를 준다. 이

때, 사용자가 각 요소 값들에 대한 범위를 설정하여 각 요소 값들이 비이상적인 값으로 변화하지 않도록 한다. 이렇게 설정한 범위 안에서 무작위로 값을 받아와 더 다양한 유전자풀을 갖게 해준다. 이후 남은 40자리는 교배연산을 통해 채워진다.

5-4. 교배

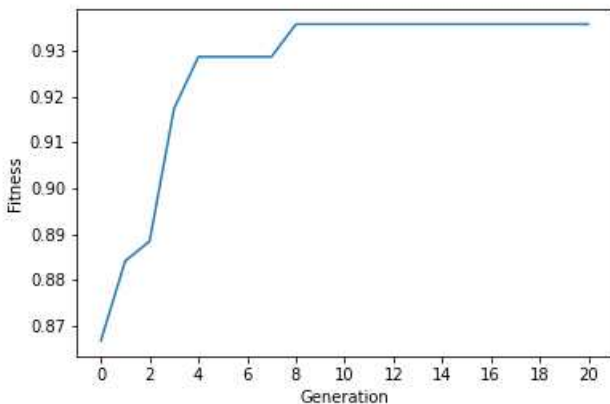
교배연산에서는 두 개체의 염색체를 교차시켜 새로운 개체를 만들어 더 다양한 유전자풀을 갖게 해주는 연산을 수행한다. 본 시스템의 교배연산에서는 선택 연산과 변이 연산을 통해 생성된 개체들 중 두 개의 개체를 뽑아 무작위로 분리점을 정해 염색체를 분리시킨 후 분리된 염색체를 다른 염색체의 것과 서로 바꾸어 새로운 개체를 만든다.

이렇게 선택, 변이, 교배연산을 통해 생성된 100개의 개체가 다음세대로 보내진다.

6. 실험 및 결과

본 논문의 실험에서는 한 세대에 20개의 개체를 20세대 동안 유전 알고리즘을 적용시켜 최적화를 해보았다. 또한, 빠른 최적화를 위해 이전 세대 모델의 가중치를 저장하여 변이 및 교배된 모델에 적용하여 학습시켰다. 실험결과는 그림 3과 같다.

<그림3. 실험결과>



본 실험에서 생성된 첫 세대중 가장 성능이 좋았던 개체의 정확도는 0.8669이며, [learning rate: 0.7747, 가중치 초기값: 'he_uniform', optimizer: 'Adagrad', 활성화 함수: 'relu', 커널 크기: 5, 합성곱 블록의 개수: 1, 한 블록에 들어가는 합성곱 층의 개수: 3, 완전연결 층의 개수: 1, drop out 비율: 0.4033619891363358, epochs: 20]와 같은 유전정보를 가졌다. 이 개체가 20번의 유전 알고리즘을 거쳐 0.9357의 정확도를 기록하였다. 최종적으로 선택된 가장 좋은 개체는 [learning rate: 1.1321, 가중치 초기값: 'he_uniform', optimizer: 'Adadelta', 활성화 함수: 'relu', 커널 크기: 3, 합성곱층의 개수: 4, 한 층에 들어가는 합성곱층의 개수: 1, 완전연결층의 개수: 2, drop out 비율: 0.29, epochs: 15]의 유전정보를 가졌다.

사전 실험에서 단순 하이퍼파라미터 변경 등으로 얻을 수 있는 일반적인 CNN 구조의 가장 좋은 정확도는 91%였으나, 유전 알고리즘으로 최적화하여 93%까지 정확도를 향상시킨 CNN 모델을 얻을 수 있었다. 따라서 유전 알고리즘을 통해 하이퍼

파라미터와 CNN 구조를 최적화하여 좋은 성능의 모델을 구할 수 있다는 결론을 내릴 수 있다.

7. 결론

본 논문의 시스템은 신경망의 구조 및 하이퍼파라미터를 최적화한다. 시스템이 잘 작동한다면 신경망의 구조와 하이퍼 파라미터를 사람이 직접 정해주지 않아도, 기존에 최적이라고 생각되었던 하이퍼파라미터 세팅에 대한 정확도 값에 필적하는 결과가 나올 것이다.

본 논문의 실험은 ResNet을 고려하여 모든 모델에 skip connection이 적용되도록 설계하였으며 비교적 간단한 CNN 구조를 최적화해보았다. 향후 연구로 shortcut 유무와 채널수 등을 돌연변이 사항으로 넣어 더 다양한 신경망 구조를 구현할 수 있는 실험을 진행해볼 것이다.

8. 참고문헌

- [1] 사이토 고키, Deep Learning from Scratch, 한빛미디어, 2017.
- [2] 김인중, Deep Learning: 기계학습의 새로운 트렌드, 한국통신학회지(정보와통신) 31(11), 52-57(6page), 2014.10.
- [3] Hochreiter S., Bengio Y., Frasconi P., and Schmidhuber J. "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," A Field Guide to Dynamical Recurrent Neural Networks. IEEE Press, 2001.
- [4] Y. LeCun, and C. Cortes., MNIST handwritten digit database, <http://yann.lecun.com/exdb/mnist/>, 2010
- [5] M. Srinivas, Lalit M. Patnaik, Genetic algorithms, A Survey. Computer, Vol. 27, No. 6 (June 1994), 17-26, 1994.
- [6] 주형주, 박정은, 전지윤, 김철연, 유전 알고리즘을 이용한 CNN 내 가중치의 최적화, 한국정보과학회 학술발표논문집, 1944-1946, 2019.
- [7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385, 2015. 5