

필기체 인식을 위한 CNN 구현에서 입력단 필터의 최적화

*윤희경, **이순진, ***한종기

세종대학교

*yhg1yhg1@naver.com, **sincere0831@naver.com, ***hjk@sejong.edu

Optimization of fore-end filter for CNN to recognize the handwriting

*Hee-kyeong Yoon, **Soon-Jin Lee, ***Jong-Ki Han

Sejong University

요약

영상 신호에 대해 인공지능적인 프로세스를 수행하는 방법들 중에 우수한 성능을 나타내면서 주목을 끌고 있는 방법으로 Convolution Neural Network(CNN)이 있다. 이를 구성할 때 전반부는 convolution network로 구현되고, 후반부는 Neural Network(NN)로 구현된다. 이때, 전반부에서 convolution 과정을 수행하기 위해 다양한 필터가 사용되는데, 이 필터들의 초기값에 따라 CNN의 성능이 달라지게 된다. 본 논문에서는 CNN의 성능을 향상시키기 위해 convolution network의 초기값을 설정하는 방법에 대해 제안하며, 이를 컴퓨터 실험을 통해 증명하기 위해 필기체 인식이라는 응용 알고리즘을 구현하였다.

1. 서론

얼마 전 세간의 주목을 끌었던 알파고의 등장은 우리 사회에 인공지능 및 Deep Learning에 대한 연구 동력을 제공하고 있다. 이 알파고 시스템의 기본 구조는 CNN 또는 RNN 등으로 구현되어 있다고 알려져 있는데, 이는 인간 두뇌의 신경망을 흉내 내어, 입력 정보들로부터 반복적인 학습을 거쳐서 올바르게 목표값을 인식하게 하는 방법이다. Convolution Neural Network는 영상에 대해 convolution 과정을 수행하여 대푯값을 추출하고, 이를 Neural Network의 입력 파라미터로 사용한다. Neural network는 그것으로 결과를 추출하고, 이에 대해 학습 알고리즘을 반복한다. 이때 CNN의 전반부에서 convolution 수행에 필터가 사용되어지는데, 이 필터의 초기 값에 따라 성능 결과가 다양하기 때문에 이 필터들의 초기 설정이 매우 중요하다.

본 논문의 2장에서는 CNN의 기본 구조에 대해서 설명하고 있으며, CNN에서 사용되는 convolution 필터의 구조에 대해서는 3장에서 설명하고 있다. 4장에서는 컴퓨터 프로그램을 이용한 실험 결과를 제시하고 있으며, 5장에서 결론을 맺는다.

2. Convolution neural network의 구조

그림 1에서 설명하는 CNN의 구조를 설명하면 다음과 같다. CNN 구조는 전반부를 구성하는 convolution network와 후반부를 구성하는

NN로 구성되어 있다.

전반부인 convolution network에서는 입력의 일부 영상을 필터와 convolution 연산을 하여 feature map이 생성된다. 이때 필터들의 초기값이 중요한데, 이를 random으로 정하면 특징이 잘 추출되지 않고 오히려 역효과를 보일 수 있다. 따라서 본 논문에서의 필터의 초기값은 손글씨 추출 과정의 특징을 더 잘 반영할 수 있는 필터를 사용하여 구현하였다. Convolution 과정을 통해 만들어진 feature map에 활성화 함수 Rectified Linear Unit를 사용하여 계산된 값을 적절하게 잘라내기(clipping)을 수행한다. 활성화 함수까지 거친 feature map의 크기를 줄이기 위해 subsampling을 진행한다. 이 과정을 max-pooling이라고 한다. 이때 가장 자국이 큰 곳을 고르는 의미로 2 by 2 max pooling을 사용하여 진행 하였다. 위와 같은 과정이 여러 단을 거쳐서 반복 수행 되면 CNN의 후반부에는 공간해상도가 작아진 특징 정보들이 모이게 된다. 이 값들을 CNN의 후반부인 NN의 입력으로 사용한다.

CNN 구조에서 앞단의 convolution network와 뒷단인 Neural network에는 많은 뉴런 층들과 이를 연결하는 시냅스 층들이 존재한다. 이 시냅스들을 연결하는 연결선에는 각각의 가중치가 존재한다. 이렇게 연결된 시냅스들의 가중치와 연산을 하여 다음 층의 뉴런에게 값을 전달해주는 feed forward가 일어난다. CNN의 앞단인 convolution network에서는 활성화 함수로 Rectified Linear Unit를 사용하는 것과 달리, 후반부의 NN에서는 활성화 함수로 sigmoid 함수를 사용하여 feed forward 연산 값을 조절하고, 이 값들이 다음 층의 뉴런에게 전달되는 것이 반복된다.

위의 과정이 모두 끝나면, Back propagation 학습 알고리즘을 수행하여 학습시킨다. Back propagation 알고리즘은 마지막 뉴런 층의

1) 이 논문은 2015년도 미래창조과학부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2015R1A2A2A01006193).

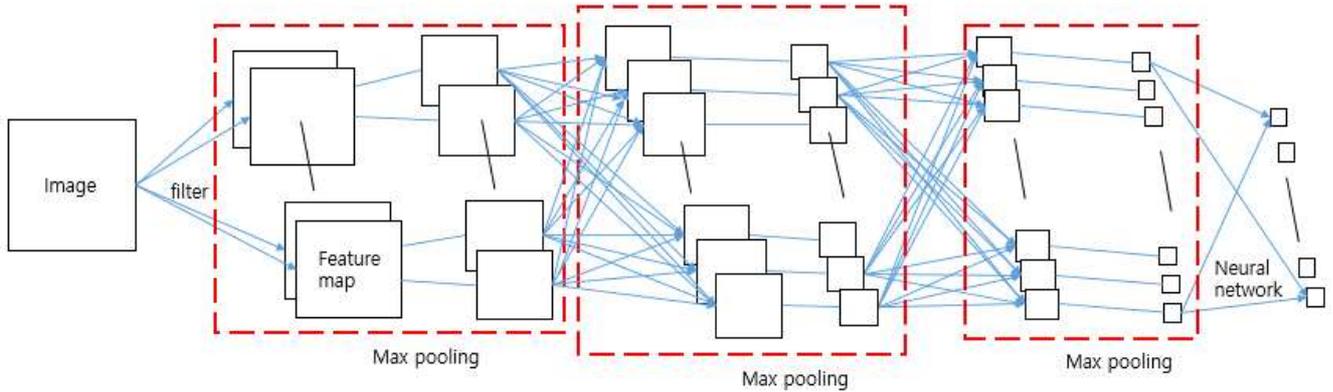


그림 1. 필기체 인식을 위한 CNN 구조

값들과 설정해 놓은 desire 값들과의 차이의 총합, 즉 전체 error가 최소화 되게 시냅스의 가중치를 수정해 가는 것이다. error의 총합을 나타내는 식인 cost function을 미분하여 0이 되는 지점이 최소값을 갖게 되기 때문에, Newton's Method를 사용하여 필터의 계수 및 시냅스를 연결하는 가중치 계수를 update한다. 이러한 반복 과정은 에러율이 정해놓은 만큼 아주 작아졌을 때 중단된다. [1][2]

3. CNN에서 입력단 필터의 초기값

본 논문에서는 CNN에서 입력단에서 사용되는 convolution filter들에 대해 설명하고 실험을 통하여 이 필터들의 초기값의 중요성을 보여주고자 한다. 먼저 Low pass filter는 저주파 차원에서 저주파는 통과시키고 고주파는 제한하는 필터이다. 이때 cut off 주파수를 사용하여 통과시킬 저주파 대역을 정한다. cut off 주파수가 작은 low pass filter 일수록 영상이 더 흐릿하게 된다. 다음으로 High pass filter는 low pass filter와는 반대로 특정 주파수, cut off 주파수 보다 높은 신호만을 통과시키는 필터이다. 이 필터는 윤곽, 경계선 부분을 검출하여 강조하는데 사용된다.

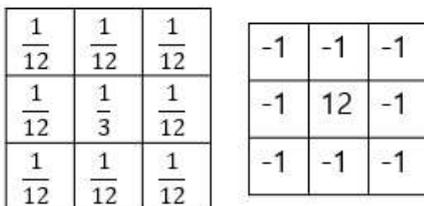


그림 2. Low pass filter(왼쪽)과 High pass filter(오른쪽)

Gaussian filter는 영상의 경계, 윤곽을 흐리게 하는 필터이다. 이를 영상에 적용하면 정규분포, 확률분포에 의해 생성된 잡음을 제거할 수 있다.

4. 실험 결과

본 논문에서는 제안하는 방법을 실험하기 위해 여러 가지 숫자들

을 손글씨로 쓴 후, 이를 훈련 영상으로 만들어 사용하였다. 본 실험에서는 숫자 세 가지를 각각 50장씩 실험 영상으로 사용하였다. 50장은 모두 손글씨로 서로 다른 사람에 의해서 만들어졌다. 이 150장의 영상으로 training을 한 후, training 영상 집합에 포함되지 않는 손글씨 숫자 영상들 30장으로 test를 수행하였다. Convolution neural network의 필터들의 값이 위에 설명한 필터를 포함하여 특정한 초기값일 때와 random한 필터 초기값을 사용하여 영상 인식을 실험하고, 이들 error 값을 비교해 보았다.

Convolution neural network는 C언어를 사용하여 구현하여 실험하였다. Convolution neural network는 convolution 층 4개, 후의 neural net은 3개의 뉴런 층을 사용하였으며 뒷부분의 Neural network는 3개의 뉴런 층을 사용하였고, desire 값은 3개로 설정하였다.

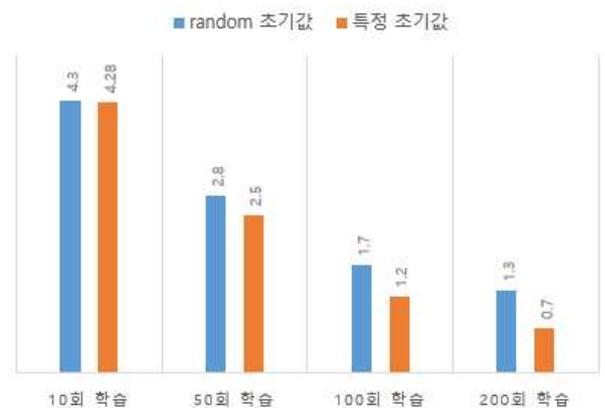


그림 3. CNN의 입력단 필터 최적화에 따른 에러율

실험 결과 Convolution neural network의 필터 초기값을 random으로 하였을 때 보다 특정한 값을 사용하였을 때 반복 횟수에 비해 error 감소율이 훨씬 큰 것을 볼 수 있었다. 비록 Back propagation 학습 알고리즘을 통하여 필터 값이 계속 변화하긴 하지만, 초기 필터 값에 따른 실험 결과를 볼 때 이는 매우 중요한 것임을 알 수 있다.

5. 결론

본 논문에서는 필기체 인식에 대해 최적화 된 Convolution neural network 입력단 필터를 연구하는 실험을 하였다. 이때 Convolution Neural network의 입력단의 초기 필터 값이 random일 때보다 특정한 초기값을 설정하는 것이 반복 횟수에 비해 에러 감소율이 훨씬 큰 것을 알 수 있었다. 위에 언급한 필터 뿐만 아니라, 실험 영상에 따른 최적의 필터를 찾아 Convolution neural network를 구현하는 것이 필요하다고 판단된다.

참 고 문 헌

- [1] Meiyin Wu, Li Chen, “Image Recognition Based on Deep Learning”, Chinese Automation Congress, pp. 542~544, OCTOBER, 2015
- [2] Y. Lecun, L.Bottou, Y.Bengio, P.Haffner, “Gradient-based learning applied to document recognition”, Proceedings of the IEEE, pp. 2294 ~ 2295, NOVEMBER, 1998