

MCDNN을 이용한 한글 필기체 인식

Handwritten Hangeul Recognition using Multi-column Deep Neural Networks

박경화¹ · 김병희² · 한동식² · 손성호² · 강우영² · 장병탁^{1,2}

Kyung-Wha Park · Byoung-Hee Kim · Dong-Sig Han · Seongho Son · Woo-Yung Kang · Byoung-Tak Zhang

¹서울대학교 뇌과학협동과정

Interdisciplinary Program in Neuroscience, Seoul National University

²서울대학교 컴퓨터공학부

Department of Computer Science and Engineering, Seoul National University

Abstract

한글 필기체 인식은 영문, 숫자 필기체 인식에 비해 매우 어려운 문제로 알려져 있다. 최근에는 기본적인 컨벌루션 인공신경망 형태의 딥러닝을 적용하여 성능이 향상된 연구 결과가 보고되었으나, 한글의 조합 가능한 가짓수에 비해 데이터는 부족한 문제는 여전히 남아 있다. 본 논문에서는, 왜곡을 이용하여 학습 과정에서 데이터를 자체적으로 증식시키는 다중 컬럼 딥 인공신경망(MCDNN)을 이용하여 한글 필기체 인식 결과를 보인다. MCDNN의 한글 필기체 인식 성능은 데이터 증식 과정이 없는 컨벌루션 인공신경망(CNN)에 비해 인식 성능이 뛰어난 것을 확인하였다.

Keywords : 인공지능, 기계학습, 딥러닝, 컨벌루션 인공신경망, 필기체 인식, 한글 필기체 인식

1. 서론

한글 필기체 인식(이하 HHR)은 유사하게 생긴 글자가 많고, 이어서 쓰는 필기체 형태가 다양하며, 필기체의 형태가 복잡할 뿐만 아니라, 같은 자음을 초성과 중성에 반복하여 사용할 수 있기에 매우 어려운 문제로 알려져 있다. 마찬가지로 분류해야 될 클래스 수가 많아 어려운 데이터인 중국어 필기체에서 최고의 성능을 보인 모델로도 한글 필기체에 대해서는 만족스러운 성능을 내지 못했다. 최근 딥러닝 알고리즘이 적용되어 필기체 이미지의 픽셀 정보 입력만으로도 매우 큰 성능 향상을 이뤘다[1]. 그러나, HHR 성능 향상을 위해서는 여전히 다음의 두 가지 근본적인 문제점에 대한 해결이 필요하다. 첫째, 다른 언어에 비하여 상대적으로 부족한 데이터셋의 보충이 필요하고 둘째, 앞서 언급한 한글 고유의 특성을 잘 반영하는 인식 모델이 개발되어야 한다.

한글 필기체 인식을 위해 기존에 구축되어 있는 데이터셋의 각 구성은 [표 1]과 같이 일관되어 있지 않다[2]. PE92는 표준 완성형 글자 2350자 전체를 클래스로 구성한 반면, 다른 데이터셋은 일부 글자만을 모았다. 각 클래스 별 데이터의 수가 최대 1000개인 점은 대표적 필기체 데이터인 MNIST 사례(클래스 당 6000개)와 비교하여 부족하며, 중국어에 비교하면 최근까지 새로운 데이터셋이 만들어지고 과거 데이터셋도 새롭게 합친 혼합형 데이터셋이 지속적으로 나오고 있는데[3][4] 비하면 상대적으로 부족하기 때문에, 이를 보충(augmentation)할 수단이 필요하다. 데이터를 추가적

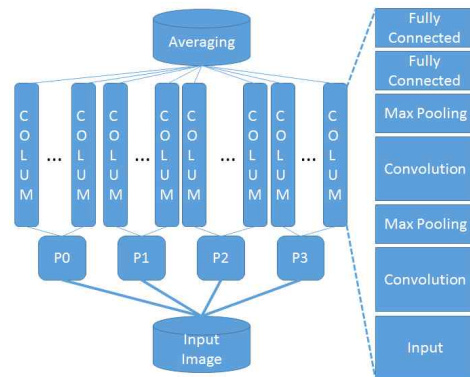


그림 1 MCDNN의 구조.

Fig 1 Structure of MCDNN

으로 수집하여 기존 데이터와 합치거나, 생성형 모델(generative model)을 이용한 샘플 데이터 추가가 일반적으로 고려하는 방법이다.

본 논문에서는, HHR 문제를 해결하기 위해 데이터의 왜곡을 통해 증식하여 필기체 인식에서 뛰어난 성능을 보인 다중 컬럼 딥 인공신경망(이하 MCDNN)을 이용하여 한글 필기체 인식을 시도한다. 그 결과로 데이터를 증식하지 않는 일반 컨벌루션 인공신경망(이하 CNN)보다 더 뛰어난 성능을 보인다.

2. MCDNN

MCDNN은 [그림 1]과 같은 형태를 가지는 딥 인공신경망으로, 2013년 중국어를 비롯한 필기체 인식 대회에서 뛰어난 성능을 보인 모델이다[5]. 그 구조는 이미지가 입력으로 주어지면 전처리 층인 Pn 층에서 입력 이미지를 서로 다르게 왜곡 시킨다. 그리고 전처리 층마다 복수의 컨벌루션 인공신경망(이하 CNN)을 학습시키며, 그 최종 결과의 평균으로 입력의 클래스를 결정하는 형태이다. 컬럼 하나를 이루는 CNN은 입력이 들어오면 이를 컨벌루션하는 과정을 거

표 1. 한글 필기체 인식 데이터셋 종류와 그 형태
Table 1. Hangeul handwritten datasets and its formation

Name	number of classes	instances per class
KU-1	1500	1000
SERI95	520	1000
PE92	2350	100

처 그 중 최댓값만을 취하는 max pooling 층을 거치며 이 두 과정을 한 번 더 반복하여 추상화된 데이터를 다층 퍼셉트론과 비슷한 Fully connected 층 2개로 출력을 내게 된다.

MCDNN은 실험에 사용하는 데이터에 따라 구조를 다양하게 바꿀 수 있어 가변성이 뛰어나며, 하나의 데이터를 서로 다른 측면에서 학습하는 효과가 있어 그 성능이 뛰어나다[3].

3. 실험

3.1 데이터

실험에 사용한 데이터는, 앞서 [표 1]에서의 SERI95 데이터를 사용했다. 클래스 수는 상대적으로 적지만, 전체 데이터 수가 52만개로 많고, 클래스 당 데이터 수도 많은 편이기에 MCDNN 학습에 좀 더 유효하기에 선택하였다. 전체 52만개 중에 불명확하거나 유실된 데이터를 제외하고, 클래스 당 데이터 개수를 최대한 균등하게 맞췄으며, 빠른 실험을 위해 데이터를 축소하여 21만여 개만을 사용하였다.

3.2 실험 MCDNN 구조

본 논문에서 구성한 MCDNN의 구조는 [그림 1]과 같이 4개의 전처리 노드(P0, P1, P2, P3)와 각각의 전처리 노드에 5개의 CNN 컬럼으로 구성된 형태이며, 각 전처리 노드에서는 입력 데이터의 가로축인 폭을 100%, 30%, 50%, 70%로 가변하여 각 CNN 컬럼에 입력으로 전달한다. 각 CNN 컬럼의 구조는 1x60x60-25C4-MP3-50C5-MP3-256N-520N이다. 1x60x60은 가로세로가 60*60인 입력 데이터 층을 의미하고, 25C4는 25개의 필터를 가지는 4*4의 컨벌루션 윈도우의 형태를 의미하며, MP3는 3*3 윈도우 안에서 최댓값을 선택하는 max pooling 층을 의미한다. 256N은 직전의 max pooling 층의 결과를 다중퍼셉트론과 유사하게 256개의 노드로 이뤄진 은닉층을 의미하고, 520N은 마지막 520개의 출력 노드를 의미한다. 520은 데이터셋의 클래스 수와 같으며, 여기서 가장 확률이 높은 노드의 결과를 분류 결과로 결정한다.

3.3 실험 환경

Ubuntu 12.04, AMD Opteron(TM) processor 6212*16 cores, 256 GiB RAM 환경에서 실험했다. GPU 병렬처리는 하지 않았다.

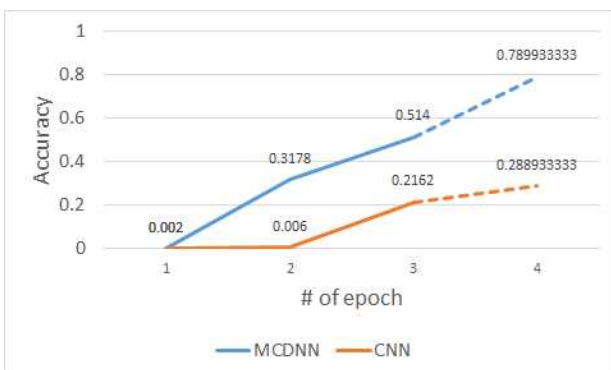


그림 2 MCDNN과 CNN의 성능 비교, 점선으로 표시된 부분은 추정치임.

Fig 2 Performance comparison between MCDNN and CNN, dashed line is projection.

4. 실험 결과

SERI95 데이터셋을 MCDNN과 같은 CNN 컬럼 구조를 1개만 가지는 일반 CNN에 학습시켜, 학습에 사용하지 않은 검증 데이터로 성능을 검증했다. 그 결과를 [그림 2]에서 나타내고 있다. 가로축은 전체 학습 데이터를 학습시킨 횟수를 나타내며, 세로축은 전체 클래스의 분류 정확도를 나타낸다. 1회 학습 당 약 20시간이 소요되며, 3회까지 학습시키는 데 약 3일이 소요됐다. MCDNN이 일반 CNN에 비해 같은 학습 횟수이면 약 30%의 더 좋은 성능을 나타내는 것을 볼 수 있다.

5. 결론 및 논의

본 논문에서는, 다른 언어의 필기체 인식에 뛰어난 성능을 보인 MCDNN을 이용하여, 데이터 왜곡을 통한 데이터 증식을 통해 다양한 방향에서의 데이터를 학습시키는 것이 그렇지 않은 것 보다 더 좋은 성능을 낸다는 것을 보였다. 이는 한글이 가지는 구조적 특징을 잘 대변할 수 있는 구조이기에 더욱 유효한 기계학습 모델 선택이다.

그러나, 일반 딥러닝 알고리즘과 마찬가지로 학습 시간이 매우 오래 걸리며, GPU 병렬처리를 활용한다고 해도 복수의 CNN을 학습시키기 때문에 일반 CNN에 비하여 학습 시간 면에서는 손해를 본다. 데이터를 왜곡시켜 증식시키는 방식을 선택하기 이전에 한글 필기체 데이터셋의 추가적인 수집이 필요하다. 그럴 경우 발생하는 서로 다른 데이터셋 간의 데이터 혼합(augmentation)의 문제는 앞으로 추가적인 연구가 필요하며, 이것이 곤란할 경우 기계학습을 통한 데이터를 증식(generation)시켜 학습하는 전체 프레임워크의 연구가 필요하다.

참 고 문 헌

- [1] I. J. Kim and X. H. Xie, "Handwritten Hangul recognition using deep convolutional neural networks," *Int. J. Doc. Anal. Recognit.*, vol. 18, no. 1, pp. 1 - 13, 2015.
- [2] D.-S. Ham, D.-R. Lee, I. Jung, and I.-S. Oh, "Construction of Printed Hangul Character Database PHD08," *J. Korea Contents Assoc.*, vol. 8, no. 11, pp. 33 - 40, 2008.
- [3] L. Jin, Y. Gao, G. Liu, Y. Li, and K. Ding, "SCUT-COUCH2009-a comprehensive online unconstrained Chinese handwriting database and benchmark evaluation," *Int. J. Doc. Anal. Recognit.*, vol. 14, no. 1, pp. 53 - 64, 2011.
- [4] C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang, "Online and offline handwritten Chinese character recognition: Benchmarking on new databases," *Pattern Recognit.*, vol. 46, no. 1, pp. 155 - 162, 2013.
- [5] D. Ciresan, U. Meier, and J. Schmidhuber, Multi-column Deep Neural Networks for Image Classification, *Computer Vision and Pattern Recognition (CVPR)*, pp. 3642-3649, 2012.