

순환신경망을 이용한 한글 필기체 인식

김병희^o 장병탁
 서울대학교 컴퓨터공학부
 {bhkim, btzhang}@bi.snu.ac.kr

Hangul Handwriting Recognition using Recurrent Neural Networks

Byoung-Hee Kim^o Byoung-Tak Zhang
 School of Computer Science and Engineering, Seoul National University, South Korea

요 약

온라인 방식의 한글 필기체 인식 문제를 분석하고 순환신경망(recurrent neural networks, RNN) 기반의 해법을 모색한다. 한글 낱글자 인식 문제를 순서데이터 레이블링의 관점에서 세 단계로 구분하여 각각에 대한 해법을 살펴보고, 특히 한글의 구성 원리에 따른 해결 방안을 정리한다. 한글 2350글자에 대한 온라인 필기체 데이터에 순환신경망 모델 중 최근에 발표된 GRU를 적용한 결과, 낱글자 인식을 67.3%, 초·중·종성 구성에 따른 6가지 유형 분류 정확도 95.2%를 확인하였다.

1. 서 론

필기체 인식은 컴퓨터가 다양한 형태의 필기 입력을 인식하는 문제로서, 인공지능 분야의 대표적인 응용문제 중 하나로 여전히 활발히 연구되는 주제이다. 이미 쓰인 필기 정보를 입력으로 받는 오프라인 방식과, 펜촉의 움직임 정보(좌표, 필기면 접촉 여부)를 기반으로 하는 온라인 방식으로 크게 구분된다.

필기체 글자 인식의 경우, 기존에는 연구자별로 다양한 특성(feature)을 추출하여 은닉마코프모델(HMM) 기반의 알고리즘으로 인식하는 것이 일반적이었으나, 최근 딥러닝 기법으로 공통의 프레임 하에서 해법을 모색하는 것이 가능해졌다. 컨벌루션 신경망(CNN)과 순환신경망(RNN)이 숫자 인식[1], 영어 인식[2], 중국어 인식[3], 아라비아어 인식[4] 등에서 낱글자와 문장 인식 모두 기존 방법 대비 큰 성능 향상을 보인 바 있으며, 90% 중후반대의 높은 성능을 보이기도 한다. 이러한 동향과는 달리, 한글 필기체 인식 문제의 경우 학계에서의 연구는 다른 언어에 비해 상대적으로 정체되어 있는 것으로 보인다. 은닉마코프모델(HMM) 또는 동적 베이지안망(DBN) 기반의 인식 기법이 주로 사용되는 와중에[5][6] 최근 컨벌루션 신경망을 이용한 오프라인 인식 결과가 발표되고 있다[7][8].

본 논문에서는 한글 필기체 인식 문제에서 온라인 방식의 해법 체계를 정리하고, 순환신경망 기반의 딥러닝 기법의 가능성을 예비 실험을 통해 확인한 결과를 보고한다.

2. 한글필기체 인식 문제

이 절에서는 필기체 인식 문제에 대한 일반적인 분석과 한글 필기체 인식의 특성을 정리한다. 필기체 인식 문제의 핵심 요소는 순서데이터 레이블링(sequence labeling) 문제이며, 다음과 같이 세 종류의 문제로 구분된다[9].

- 서열 분류(sequence classification): 하나의 서열 데이터별로 하나의 레이블이 부여되는 문제. 한 순서데

이터에 대한 레이블의 길이가 1로 제한된다.

- 구간 분류(segment classification): 순서데이터 내에서 구간별로 레이블이 주어짐. 전후 문맥(context)이 중요한 단계이며, 구간과 해당 레이블이 맞춤(align)이 되어 있어야만 하는 종류의 문제이다.
- 시간별 분류(temporal classification): 레이블도 서열 형태로 주어지며, 레이블 서열의 길이가 입력 데이터 서열의 길이 이하라는 조건 외에는 레이블 서열에 대한 조건이 없는 설정이다. 레이블이 부여되는 구간의 경계가 사전에 알려져 있지 않다.

온라인 방식의 한글 낱글자 인식 문제에서, 낱글자 전체 획(strokes) 정보를 입력으로 하고, 글자 또는 글자의 종류 중 하나를 출력으로 하는 경우 첫 번째 설정인 서열 분류 문제에 해당된다. 초성, 중성, 종성별로 각 획이 구분되고 해당 자음 또는 모음이 명시적으로 부여된 데이터가 존재하는 경우 두 번째 설정에 해당하는 구간 분류 문제이다. [6]에서 이와 같은 사례를 다룬 바 있으나, 구간을 명확히 구분하는 작업에 사람이 개입을 해야 하기 때문에 학습 데이터를 구성하고 필기체 인식기를 학습하려면 비용이 많이 든다. 구간 정보가 사전에 알려지지 않더라도 순서대로 초성, 중성, 종성에 해당하는 자음 또는 모음을 레이블로 사용하는 경우 세 번째 시간별 분류 설정의 해법을 적용할 수 있다.

3. 한글필기체 인식 방법

이 절에서는 한글필기체의 온라인 인식에 적용 가능한 알고리즘을 정리한다. 딥러닝이 표상 학습(representation learning)을 가능하게 하고 end-to-end 방식의 해법을 추구하는 점을 고려하여 특성(feature) 추출에 대한 논의는 제외한다.

순서가 있는 데이터에 레이블을 예측하는 문제에는 전통적으로 은닉마코프모델이 널리 사용되었으며, CRF(conditional random field)도 좋은 선택지 중 하나이다. HMM 모델에서는 은닉 상태가 주어진 경우, 시간에 따른 입력값은 무관하다는 가정을 하기 때문에, 서열상의

거리가 먼 입력값이 원활하게 반영되지 않는 한계가 있다. 이에 대한 대안으로 순환신경망 계열의 알고리즘이 연구되었다[10]. 싸이클이 있는 망 구조를 허용하는 고전적인 RNN[11][12]은 안정적인 학습이 어려운 문제가 있어 응용에 제한이 있으며, 이를 해결한 대표적 기법인 LSTM (Long Short-Term Memory)[13]이 최근 딥러닝 모델의 핵심 도구 중 하나로 각광을 받고 있다. 서열의 순방향뿐만 아니라 역순 정보까지 고려하는 양방향(bi-directional) 순환신경망 모델은 음성인식, 영어 필기체 문장 인식 등의 문제에서 공통적으로 적용하는 해법이다. 특히, ‘시간별 분류’ 문제의 경우 CTC (connectionist temporal classification) 출력층의 도입은 필수적인 선택지가 되고 있다[14][15].

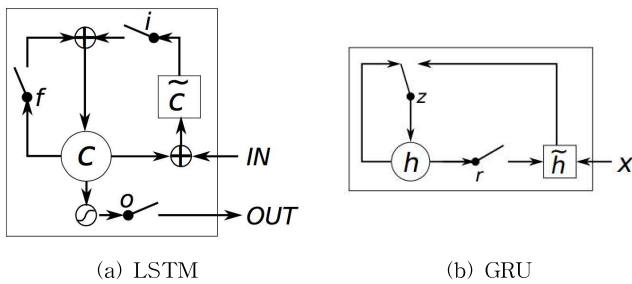


그림 1. 대표적인 두 순환신경망의 hidden unit 구조. LSTM은 입력(i), 망각(f), 출력(o) 세 개의 ‘게이트’를 두고, GRU는 리셋(r), 업데이트(z) 두 개의 게이트로 내부 메모리의 값(C , h)을 조절한다.

한글 낱글자에 대한 온라인 방식 인식에 대한 해법으로, ‘서열 분류’ 단계의 설정에는 LSTM 및 구조를 간소화한 GRU (Gated Recurrent Unit)[16]를 적용할 수 있으며(그림 1), 양방향 구조를 추가하여 성능 향상을 도모할 수 있다. 그러나, 초성·중성·종성의 순서가 있는 한글의 구조적 특성을 반영하여, 자음, 모음, 받침용 자음의 순서를 부분적 레이블로 활용한다면, ‘시간별 분류’ 설정의 해법을 적용할 수 있으며, CTC 기법을 활용할 수 있다. 낱글자의 초·중·종성을 파악하여 레이블로 활용하는 작업은 간단히 자동화할 수 있기 때문에 데이터를 준비하는데 들어가는 비용은 서열 분류 단계와 거의 차이가 나지 않는다.

한편 온라인 필기체 데이터의 좌표 기반 정보를 기초로 글자 데이터를 정규화된 이미지로 변환하고, 오프라

인 방식으로 CNN 기반 인식을 병행하는 것도 가능하다. CNN/RNN 하이브리드 모델은 성능 향상을 기대할 수 있지만, 학습 단계와 테스트/인식 단계 모두 시간적 비용이 커질 우려가 있다.

4. 실험 및 결과

이 절에서는 한글 낱글자 인식 문제를 ‘서열 분류’ 단계에서 단순한 순환신경망 모델로 성능 평가를 수행한 결과를 정리한다.

실험에 사용한 한글 필기체 데이터셋은 KS완성형 2350가지 낱글자 각각에 대해 500개씩의 샘플로 구성되어 있다. 각 샘플은 online 방식의 필기체 정보, 즉 펜의 위치(x, y)와 상태 정보(up/down)를 담고 있다. 전체 데이터를 학습 데이터와 테스트 데이터를 3:1의 비율로 분리하여 순환신경망 기반 인식 성능을 평가한다.

실험은 2350가지 한글 낱글자 인식 문제와, 한글의 초성·중성·종성 구성에 따른 6가지 유형 분류 문제(그림 2)로 구성하였다. 인식 알고리즘으로는 LSTM보다 간결한 GRU를 적용하였으며, 레이블이 많은 낱글자 인식 문제에 보다 복잡한 모델을 적용하였다. 입력 서열 데이터에 대해 ‘Masking’ 레이어를 통과시켜 지정 길이보다 긴 부분은 삭제하고, 지정 길이보다 짧은 데이터의 경우 데이터 이외의 뒷부분은 0으로 채웠다. [표 1]에 두 종류 실험 결과에 대한 종합 정보를 정리하였다. [그림 3]에서 학습 곡선을 확인할 수 있다.

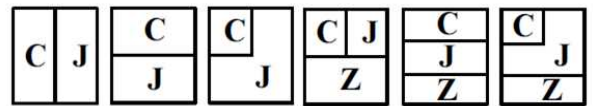


그림 2. 한글의 초성(C),중성(J), 종성(Z) 구성에 따른 여섯 가지 유형 구분(그림 출처:[6])

낱글자 인식의 경우 현재 설정에서는 테스트 정확도가 65 ~ 67%대에 수렴하며, 여섯 가지 유형 인식의 경우 95% 전후로 수렴한다. 수렴 또는 과적합이 발생하는 시점까지 학습에 소요되는 시간은 유형 인식(40 epoch 전후)에 비해 낱글자 인식(220 epoch 전후) 모델이 상대적으로 컸다.

인식 성능을 위해서는 모델의 복잡도를 늘리는 것이 필요해 보인다. GRU의 출력 차원 증가, 완전 연결망(dense) 단계에서 drop-out 옵션 추가, 순환망 또는 완전 연결망의 다층화 등을 고려할 수 있다. 한편, 최근 발표

표 1. 순환신경망을 이용한 한글 필기체 인식 결과(실험 환경: [S/W] Ubuntu 14.04, Keras 1.0.3, backend=Theano 0.8.2, [H/W] Intel i5-6600, Titan X D5 (VRAM 12G) * 1 node, RAM DDR 4 64G).

문제 종류	모델 구성 (# parameters)	학습 설정(공통)	학습 설정	최고 인식 정확도(accuracy)		학습소 요시간 (시간)	테스트 소요 시간
				학습단계	테스트단계		
낱글자 인식	Masking- GRU[24]-Dense (60941)	optimizer=adam obj. function = softmax	240 epoch	67.6%	67.3%	11.5 hr/ 100 ep	32s, 0.115 ms/char
6유형 인식	Masking- GRU[16]-Dense (41029)	cross-entropy, batch size = 1024	100 epoch	95.8%	95.2%	8.56	24s, 0.081 ms/char

된 순환신경망의 batch normalization [17]을 적용하면, CNN의 경우에서와 같이 학습 속도와 성능이 향상될 것으로 기대한다[18].

5. 결론

본 논문에서는 한글 필기체의 온라인 방식 문제에 대한 분석과 순환신경망 기반의 해법을 정리하고, 예비실험을 통해 2350글자에 대한 낱글자 인식률 67%, 6가지 한글 유형 인식률 95%의 성능을 확인하였다.

보다 실용적인 수준의 인식률(낱글자 90% 이상, 유형 99% 이상)을 목표로 순환신경망 모델의 자연스러운 확장인 양방향(bi-directional) 또는 다층(multi-layer) 구조를 적용하여 후속 실험을 수행하고, 한글 필기체 인식을 위한 딥러닝 기반의 보다 정교한 방법론을 개발하고자 한다.

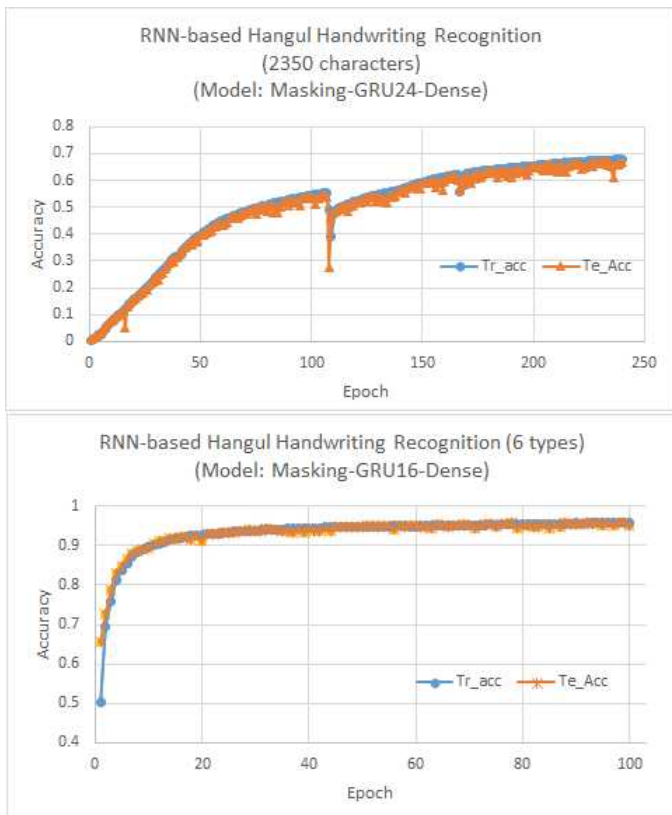


그림 3. 순환신경망을 이용한 한글 낱글자 인식 및 유형 인식 결과 (학습 단계에서는 mini batch 단위로 평가하여 평균을 내기 때문에, test 단계에 비해 성능이 낮게 나오기도 함)

감사의 글

이 논문은 2016년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터(R0126-16-1072-SW스타랩)와 한국산업기술평가관리원(10044009-HRI.MESSI, 10060086-RISF)의 지원을 받아 수행된 연구이며, 삼성전자의 지원을 일부 받았음.

참고문헌

[1] D. Cirestan, U. Meier, J. Schmidhuber, Multi-column deep

neural networks for image classification, In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012)*, pp. 3642-3649, 2012.

[2] M. Liwicki, A. Graves, S. Fernández, H. Bunke, J. Schmidhuber, A Novel Approach to On-Line Handwriting Recognition Based on Bidirectional Long Short-Term Memory Networks, In *Proceedings of the 2007 International Conference on Document Analysis and Recognition (ICDAR 2007)*, pp. 367-371, 2007

[3] F. Yin, Q.-F. Wang, X.-Y. Zhang, and C.-L. Liu, ICDAR 2013 Chinese Handwriting Recognition Competition, In *Proceedings of the 2013 International Conference on Document Analysis and Recognition (ICDAR 2013)*, 1464-1470, 2013.

[4] A. Graves, *Offline Arabic Handwriting Recognition with Multidimensional Neural Networks*. Book Chapter, Guide to OCR for Arabic Scripts, Springer, 2012.

[5] B.-K. Sin and J. Kim, Ligature modeling for online cursive script recognition, *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 19(6):623-633, 1997.

[6] S. Cho and J. Kim, Bayesian Network Modeling of Hangul Characters for On-line Handwriting Recognition, In *Proceedings of the 2003 International Conference on Document Analysis and Recognition (ICDAR 2003)*, pp. 207-211, 2003.

[7] I. J. Kim and X. H. Xie, Handwritten Hangul recognition using deep convolutional neural networks, *International Journal on Document Analysis and Recognition*, 18(1):1-13, 2015.

[8] I. J. Kim, C.B Choi, and S.H. Lee, Improving Discrimination Ability of Convolutional Neural Networks by Hybrid Learning, *International Journal on Document Analysis and Recognition*, online published, 2015.

[9] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*. Textbook, Studies in Computational Intelligence, Springer, 2012.

[10] 윤현구, 정동석, 정교민, 순환 신경망의 구조와 필기체 인식, *정보과학회지* 33(9):42-48, 2015.

[11] J. L. Elman, Finding Structure in Time, *Cognitive Science*, 14:179-211, 1990.

[12] M. I. Jordan, Attractor Dynamics and Parallelism in a Connectionist Sequential Machine, In *Artificial neural networks*, Joachim Diederich (Ed.), IEEE Press, 112-127, 1990.

[13] S. Hochreiter and J. Schmidhuber, Long Short-Term Memory, *Neural Computation*, 9(8):1735-1780, 1997.

[14] A. Graves, S. Fernandez, F. Gomez, and J. Schmidhuber, Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks, In *Proc. Int. Conf. on Machine Learning (ICML 2006)*, pp. 369-376, 2006.

[15] A. Graves, A. Mohamed, G. Hinton, Speech Recognition with Deep Recurrent Neural Networks, *ICASSP 2013*, pp. 6645-6649, 2013.

[16] K. Cho, B. Van Merriënboer, C. Gülçehre, F. Bougares, H. Schwenk, and Y. Bengio, Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation, arXiv preprint arXiv:1406.1078, 2014.

[17] T. Cooijmans, N. Ballas, C. Laurent, Ç. Gülçehre, A. Courville, Recurrent Batch Normalization, arXiv:1603.09025, 2016.

[18] S. Ioffe and C. Szegedy, Batch Normalization, Accelerating Deep Network Training by Reducing Internal Covariate Shift, In *Proc. Int. Conf. on Machine Learning (ICML 2015)*, 37, 2015.