

## 순환신경망을 이용한 한글 필기체 생성

김병희<sup>o</sup> 장병탁  
 서울대학교 컴퓨터공학부  
 {bhkim, btzhang}@bi.snu.ac.kr

## Generation of Hangul Handwritten Characters using Recurrent Neural Networks

Byoung-Hee Kim<sup>o</sup> Byoung-Tak Zhang  
 School of Computer Science and Engineering, Seoul National University, South Korea

## 요 약

본 논문에서는 한글 필기체 생성을 위한 순환신경망(recurrent neural networks, RNN) 기반의 학습 및 생성 기법을 다룬다. 특히, 인공지능 측면에서 사전 지식 없이 데이터만을 이용해 학습한 모델이 한글의 구조와 같은 추상적 개념의 지식을 발견할 수 있는지에 초점을 둔다. 순서 데이터 생성을 위한 기법으로 1) 기존에 알려진 순환신경망과 혼합 확률 밀도망(mixture density network)의 조합과, 2) 최근 발표된 HyperRNN을 적용한다. 제안한 기법을 한글 2350글자에 대한 온라인 필기체 데이터에 적용하여 학습 및 생성한 결과 자음의 구분과 조합을 대략적으로 재현함을 확인하였다.

## 1. 서 론

문서, 음악, 필기체 등과 같이 인간이 창작한 순서 데이터를 학습하여 자동으로 유사한 데이터를 생성하는 문제는 인공지능의 대표적인 연구주제중 하나이다[1][2]. 문서 생성, 음악생성 등에 공통적으로 적용되는 대표적 해법 설정은, 누적 관측값의 패턴에 근거한 다음 관측값 예측 확률 모델, 즉  $Pr(x_{t+1}|x_{1:t})$ 에 대한 모델링이다. 관측값이 이산적인 값을 가지는 경우 통상적으로 언어 모델(language model)로 불리며, [3]이 등장하기 전에는  $n$ -그램 방식의 모델이 주를 이루었고, [3] 이후 순환신경망에 기반한 모델이 최신 연구의 주류를 이루고 있다.

온라인 방식의 필기체 생성 문제의 경우, 2차원 평면 상에서 실수값을 가지는 점의 좌표와 획을 예측해야 하기 때문에, 이산적 값을 다루는 언어 모델을 직접 적용할 수 없다. 영어 필기체의 경우 [4]에서 순환신경망과 혼합 확률밀도망(MDN)을 결합한 생성 기법을 처음 제시하였다. [5]는 이 기법을 정자로 쓴 한자 데이터에 적용한 결과를 보인 바 있다.

인공지능적 측면에서 머신러닝과 딥러닝 모델의 학습을 통해 얻고자 하는 바는 단순한 예측 모델과 데이터 생성 모델뿐만 아니라, 데이터에서 사람의 인지, 언어 수준에서 자연스러운 정도의 추상화된 규칙성과 지식을 발견하는 것이다. 2차원 사진 데이터에 대해서는 딥베이지안망(deep belief net)과 컨벌루션 신경망(CNN)을 이용해 다양한 수준의 추상적 표현(representation) 학습이 됨이 계속 보고되고 있다[6][7]. 그러나 순서 데이터에서의 규칙적 패턴과 직관적 수준의 지식 추출 또는 가시화는 잘 이루어지지 않고 있다. 최근에는 [8]에서 소개된 사례와 같이 순서 데이터에서 ‘개념’을 학습하는 모델에 대한 관심이 커지고 있다. 한글과 한자와 같이 획이 많고 복잡하면서도 제한된 범위의 규칙이 알려진 언어에 대해, 필기체 데이터에서 이러한 규칙을 학습을 통해 발견하는 작업은 이러한 현재 연구의 한계를 확장할 수 있다는 점에서 의미가 있다.

본 논문에서 정리한 연구는 다음과 같은 점에서 기여를 하고자 한다: 1) 학술적으로 대규모 온라인 한글 필기체 생성 사례를 최초로 보고, 2) 기법 측면에서 HyperLSTM +MDN+차등우도 조합을 최초로 시도, 3) 인공지능 연구 측면에서 순환신경망 기반 딥러닝 모델의 학습을 통한 구조적 패턴 인식 가능성 보고를 한다.

## 2. 한글필기체 생성 문제

[4]에서는 RNN 기반의 순차적 예측망으로 실수값을 가지는 영어 필기체 문장을 생성하기 위해, 혼합 확률밀도망(MDN, mixture density network, [9])을 결합하는 아이디어를 제안하고, IAM 온라인 영문 DB로 학습한 망이 다양한 필체를 모사함을 보인 바 있다.

한글 필기체의 경우 영문 필기체와 비교해볼 때 획과 획 사이의 방향과 거리의 변화에 대한 불확실성이 매우 큰 특성이 있다. 사전 지식을 반영하지 않고 데이터만으로 한글의 구조적 특성을 파악하고 생성할 수 있는지 여부는, 머신러닝을 통해 지식의 발견을 추구하는 이론적 측면과 실용적 측면에서 모두 관심을 두게 되는 문제이다. 머신러닝 기반의 한글 필기체 생성 관련 연구 사례는 많지 않으며, 동적 베이지안망으로 초성, 중성, 종성의 구조를 명시적으로 모델링한 연구[10] 이후의 사례는 찾아볼 수 없었다.

[5]에서는 한자를 정자로 쓴 필기체 데이터를 이용하여 유사 한자를 생성하는 사례를 보인 바 있다. 한자 역시 영어에 비해 획의 상하, 중횡의 변화가 큰 특성이 있으며 글자를 구성하는 요소의 구조가 어느 정도 정해져 있는 특징이 있다. 정자로 획의 순서 규칙을 준수하고 획간의 구분을 명확히 하여 ‘또박또박’ 쓴 데이터를 이용하면 한자의 이러한 구조적 특징을 학습할 수 있음을 보였다고 평할 수 있다.

본 논문에서는 [10]과 [5]의 사례보다 구불구불한(cursive) 형태의 한글 필기체 데이터를 이용한 한글 낱글자 필기체 생성을 시도한다. 초/중/종성을 구성하는 자음과 모음의 형태가 매우 가변적이고 경우에 따라 여러 자

음과 모음을 하나의 획으로 이어쓰는 경우가 포함되어 매우 노이즈가 큰 데이터에서도 한글의 구조적 특성을 파악해내고 생성할 수 있는지 여부를 확인하고자 한다.

### 3. 한글필기체 생성 방법

#### 1) RNN + MDN + 차등 우도

혼합 확률밀도망(MDN)은 혼합 확률분포의 파라미터를 신경망의 출력을 이용하여 정하는 기법이다. 혼합 가중치 출력은 소프트맥스 함수로 정규화하고, 다른 출력은 적합한 함수를 추가로 적용하여 값의 범위가 의미에 맞도록 조정한다(예: exponential 함수를 적용하여 양수 범위의 값으로 조정). 순환신경망의 출력값을 혼합 확률 파라미터로 두는 경우, 입력값의 순서를 고려한 출력이 가능하다.

온라인 필기체 데이터가 점의 좌표  $(x_1, x_2)$ 와 획의 종료 여부(end-of-stroke, eos)를 표현하는 이진 변수  $x_3$ 의 순서 정보로 주어질 때, 시간  $t$  시점까지의 입력을 바탕으로  $t+1$  시점에 이어질 점의 위치와 획 종료 여부는 각각 가우시안 혼합 분포(Gaussian mixture, 이하 GM)와 베르누이 분포로 모델링할 수 있다. 이러한 설정에서 RNN의 입력 벡터  $x_t$ 와 출력 벡터  $y_t$ 는 다음과 같이 표현된다.

$$x_t \in \mathbb{R} \times \mathbb{R} \times \{0, 1\}$$

$$y_t = \left( \left\{ \pi_t^j, \mu_t^j, \sigma_t^j, \rho_t^j \right\}_{j=1}^M, e_t \right)$$

$y_t$ 는  $M$ 개 2차원 가우시안 분포의 혼합 비율, 평균, 표준편차(분산의 두 대각선 원소값), 상관계수(분산의 비대각선 원소값에서 도출)와 획이 끝날 확률( $\Pr(z) = \Pr(x_3 = 1)$ )을 의미하는 노드로 구성된다.

필기체 생성을 펜의 다음 시점 위치를 예측하는 문제, 즉  $\Pr(x_{t+1}|x_t)$ 를 구하는 문제로 설정하는 경우, RNN+MDN은 다음과 같이 문제를 설정한다:

$$\Pr(x_{t+1}|y_t) = \sum_{j=1}^M \pi_t^j N(x_{t+1} | \mu_t^j, \sigma_t^j, \rho_t^j) \begin{cases} e_t & \text{if } (x_{t+1})_3 = 1 \\ 1 - e_t & \text{otherwise} \end{cases}$$

$$\Pr(\mathbf{x}) = \prod_{t=1}^T \Pr(x_{t+1}|y_t),$$

망 학습을 가이드하는 목적함수(sequence loss)는  $\Pr(\mathbf{x})$ 의 음의 로그우도값으로 설정한다.

$$L(\mathbf{x}) = - \sum_{t=1}^T \log \Pr(x_{t+1}|y_t) = L(GM) + L(z).$$

Ha는 [5]에서 한자 생성 문제의 경우 획의 종료 시점과 글자의 종료 시점(end-of-character, eoc)을 구분해야만 낱글자 단위의 생성이 가능함을 지적하며, eoc의 빈도가 드문 점을 보완하는 방법으로서, Firth 기법[11]과 유사한 차등 우도를 설정할 것을 제안한다. 이에 따라  $x_3$ 항을 eos, eoc, stay의 세 상태에 대한 one-hot 인코딩 벡터로 조정하고, 다음과 같이 목적함수를 설정하여 빈도가 적은 eos, eoc의 기울기를 크게 설정한다:

$$L(\mathbf{x}) = L(GM) + w(z)L(z),$$

$$w(z) = \begin{cases} w_{stay} = 1 \\ w_{eos} = \sqrt{w_{eoc}} (\gg w_{stay}). \\ w_{eoc} \end{cases}$$

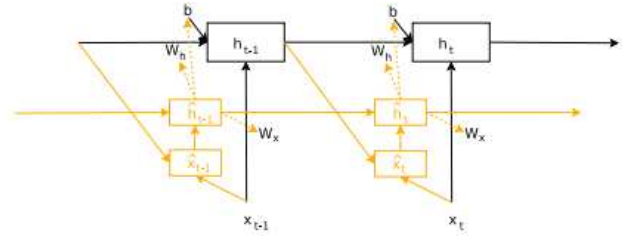


그림 1. HyperRNN의 개념도[12]. RNN의 가중치 행렬을 동적으로 생성하는 미니 RNN을 둔다.

#### 2) HyperLSTM

Ha 등은 순서 데이터 생성 모델을 생성하는 신경망 모델로서, RNN의 가중치 행렬이 매 시간마다 동일한 값으로 공유되는 제한 조건을 완화하는 동적 하이퍼네트워크를 제안하고(그림 2), LSTM에 이 개념을 적용한 HyperLSTM을 소개하였다[12]. HyperLSTM 셀이 LSTM의 세 개의 게이트(입력, 망각, 출력)와 상태값의 후보( $\tilde{C}$ )에 연계된 가중치 행렬과 바이어스를 임베딩하는 방식으로 컨트롤하는 ‘생성 모델을 생성하는 모델’이며, [4]와 비교하여 보다 좋은 품질의 영어 필기체를 생성한 결과가 보고되었다.

### 4. 실험 및 결과

#### 1) 데이터 및 모델 설정

데이터는 완성형 한글 2350글자에 대한 미공개 온라인 필기체 데이터를 사용하였다. 글자별로 다수의 사용자가 쓴 100개의 샘플로 구성되어 있다. 이 데이터에 대해 글자의 크기는 가로축과 세로축 중 긴 축의 길이를 1.0으로 두고 비율은 유지하는 정규화를 적용하고, 좌표값은 연속점 간의 변화량을 계산하여 적용하였다. 펜의 상태를 eos, eoc, stay의 세 가지 상태로 구분하여 3차원 one-hot 인코딩을 하였으며, 한 글자의 정보는  $(\Delta x, \Delta y, eos, eoc, stay)$ 의 5차원 벡터의 순서로 표현되며, 글자 샘플마다 길이와 획의 수는 다르다.

필기체 생성 학습을 위한 모델은 LSTM+MDN+차등우도 적용 모델, HyperLSTM+MDN+차등우도 적용 모델의 두 조합을 적용한다. 각 모델별 설정은 [표 1]에 정리하였으며, 공통적으로 batch\_size는 64, 한 글자의 최대 길이 600, 미분값 상한은 전역적 최대값 10.0을 적용한다.

[5]의 제안을 참조하여 전체 데이터는 랜덤하게 순서를 조정하고, 각 미니배치는 항상 특정 글자의 첫 점에서 시작하도록 하였으며, 미니배치 단위의 학습 때마다 글자 크기 척도를 30% 범위 내에서 랜덤하게 증감시키는 방식의 동적 데이터 보충(augmentation)을 하였다.

표 1. 한글 필기체 생성에 적용한 순환신경망 기반 모델의 설정

모델	구성
LSTM+MDN+차등우도	LSTM: 256 unit * 2 layer MDN: 24 mixture 차등우도: $w_{eoc} = 200$
HyperLSTM+MDN+차등우도	상기 구성에 더해 HyperLSTM: 128 unit, embedding size=4, LSTM 셀에만 출력 및 recurrent dropout 적용(유지확률 0.9)

2) 실험 결과 및 분석

[표 2]와 [그림 2]에 각각, 모델 학습 결과와 학습한 모델에서 연속적 샘플링을 통해 생성한 한글 낱글자 사례를 정리하였다. 두 모델 모두 정형화된 글자까지의 학습은 진행되지 않았으며, 존재하지 않는 글자나 자모음 조합이 나타나기도 한다. 그러나 자음과 모음 단위로는 정확한 구성을 보이는 사례가 많으며, 특히 한글 글자 구성의 여섯 가지 사례를 빈도에 대략 비례하여 생성하는 것을 확인하였다(그림 3). 완성형 한글의 분포(G)와 각 학습 결과 모델별로 1만 건 생성글자군과의 빈도( $M_1$ ,  $M_2$ )의 유사도를 Hellinger 거리[14]로 측정한 결과 각각  $H(G, M_1)=0.0354$ ,  $H(G, M_2)=0.0531$ 로 측정되었다.

표 2. 생성모델 학습 결과. 전체 loss  $L(\mathbf{x})$ 는 글자의 구성과 펜의 상태에 대한 두 loss로 구성된다.

모델		Log Loss		
		$L(\mathbf{x})$	$L(GM)$	$L(z)$
1	LSTM+MDN+차등우도	-5.10	-6.29	1.186
2	HyperLSTM+MDN+차등우도	-5.27	-6.33	1.058

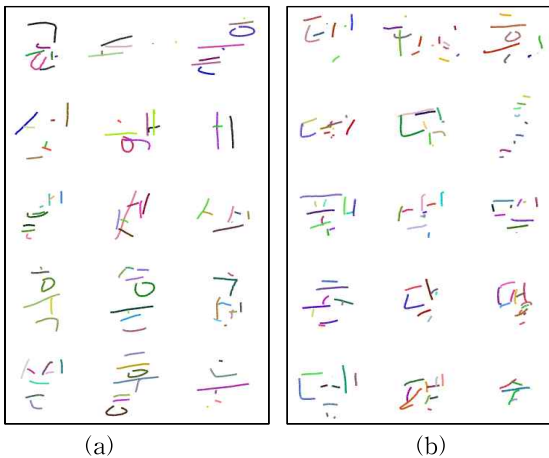


그림 3. 한글 낱글자 생성 모델에서 샘플링한 사례. (a) LSTM+MDN+차등우도 모델(log loss=-5.10) (b) HyperLSTM+MDN+차등우도 모델(log loss=-5.27)을 각각 적용.

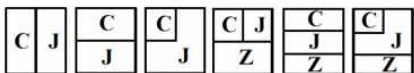
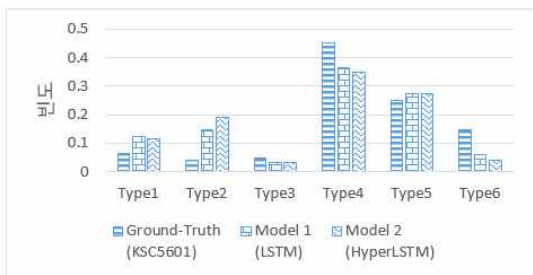


그림 2. [표 2]의 두 모델로 생성한 글자의 정량적 평가. 완성형 한글 2350글자에서 초성(C), 중성(J), 종성(Z) 구성에 따른 여섯 가지 유형[10]의 빈도를 비교한 결과, Type2와 Type6의 빈도가 바뀐 것을 제외하면 유사한 분포를 보임. 글자의 유형 예측은 [13]의 모델을 적용함.

5. 결론

본 논문에서는 순환신경망 기반의 최신 기법의 새로운 조합으로 한글 낱글자 생성하는 기법과 결과를 정리하였다. 알고리즘이 자모음 및 초/중/종성의 정형화된 한글의 구조를 개략적으로 파악하는 것으로 분석하였다. 모델의 학습이 충분히 이루어지고, 최적의 파라미터 조합을 탐색하여 실제 필기체에 보다 가까운 결과를 얻을 수 있을 것으로 기대한다. [4]에서 발표한 ‘지정한 글자’를 필기체로 생성하는 기법을 한글 필기체에 적용 가능하도록 확장하는 연구가 이론적, 실용적으로 모두 의미 있는 후속 연구가 될 것이다.

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] C. Sauper and R. Barzilay, Automatically generating Wikipedia articles: A structure-aware approach, In *Proc. Jt. Conf. 47th Annu. Meet. ACL & 4th Int. Jt. Conf. Nat. Lang. Process. AFNLP*, pp. 208-216, 2009.
- [2] J. D. Fernández and F. Vico, “AI methods in algorithmic composition: A comprehensive survey,” *J. Artif. Intell. Res.*, vol. 48, pp. 513-582, 2013.
- [3] T. Mikolov, M. Karafiat, L. Burget, J. Cernocky, and S. Khudanpur, Recurrent neural network based language model, *INTERSPEECH 2010*, pp. 1045-1048, 2010.
- [4] A. Graves, Generating sequences with recurrent neural networks, *arXiv:1308.0850*, 2013.
- [5] D. Ha, Recurrent net dreams up fake Chinese characters in vector format with TensorFlow, Blog Post, 2015. <http://blog.otoro.net/2015/12/28/recurrent-net-dreams-up-fake-chinese-characters-in-vector-format-with-tensorflow/>
- [6] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, *ICML 2009*, pp. 1-8, 2009.
- [7] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, DeepFace: Closing the gap to human-level performance in face verification, *CVPR 2014*, pp. 1701-1708, 2014.
- [8] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, “Human-level concept learning through probabilistic program induction,” *Science.*, vol. 350, no. 6266, pp. 1332-1338, 2015.
- [9] C. Bishop. Mixture density networks. Technical report, Oxford University, 1994.
- [10] H. Choi, S.-J. Cho, and J. H. Kim, Generation of handwritten characters with Bayesian network based on-line handwriting recognizers, *ICDAR 2003*, pp. 995-999, 2003.
- [11] D. Firth, “Bias reduction of maximum likelihood estimates,” *Biometrika*, 80(1):27-38, 1993.
- [12] D. Ha, A. Dai, and Q.V. Le, HyperNetworks, *arXiv:1609.09106*, 2016.
- [13] 김병희, 장병탁, “순환신경망을 이용한 한글 필기체 인식,” *정보과학회 컴퓨터의 실제 논문지*, (계재 예정).
- [14] D. Blei and J. Lafferty, “A correlated topic model of Science”, *Ann. Appl. Stat.*, 1(1):17-35, 2007.