

일반 필기데이터와 CNN 을 이용한 온라인 서명인식

박민주*, 윤희용**

*성균관대학교 소프트웨어대학 컴퓨터공학과

**성균관대학교 정보통신대학 전자전기컴퓨터공학과

wyuinche.mido@gmail.com, youn7147@skku.edu

Online Signature Verification using General Handwriting Data and CNN

MINJU PARK*, HEE YONG YOUN**

*Dept. of Computer Science and Engineering, Sungkyunkwan University

**Dept. of Electrical and Computer Engineering, Sungkyunkwan University

요 약

본 논문에서는 대표적인 이미지 분류 모델인 CNN(Convolutional Neural Network)과 시간에 따른 이미지의 변화를 학습할 수 있는 LSTM(Long Short-Term Memory) 기반의 온라인 서명인식 모델을 제안한다. 실제로는 위조서명을 미리 구하기 어렵다는 사실을 고려해 서명검증 대상자가 아닌 타인의 진서명과 대상자의 일반 필기 데이터를 음의 데이터로서 학습에 사용하였다. 실험 결과, 전체 이미지 중 서명 부분의 비율에 따라 좋은 성능을 보이는 검증 모델이 다르며 Accuracy 성능지표를 통해 이 비율이 높거나 낮을 경우 CNN-LSTM 이, 중간일 경우 CNN 이 적합하다는 것을 확인하였다.

1. 서론

최근 온라인 상의 금융거래, 핀테크가 활성화되고 있다.[1] 그에 따라 생체인식 등을 통한 본인인증에 대한 수요 또한 높아지고 있다. 서명인식의 경우 단순히 정적 데이터만을 사용하는 오프라인 방식은 위조가 쉽기 때문에 최근에는 개인의 동적 데이터를 포함한 온라인 방식에 대한 연구가 활발히 이루어지고 있다. 특히 서명검증에는 딥러닝과 SVM 을 이용한 사례가 많아지고 있는데 그 중 CNN 에 시간에 따른 이미지의 변화를 학습할 수 있는 LSTM 모델을 도입한 연구도 증가하고 있다. [2][3][4][5][6]

서명에 포함된 동적 데이터로는 특정 순간의 서명 완성도, 펜의 위치나 압력, 작성 속도 등이 있다. 서명을 그대로 따라 그리는 방법으로 원본에 가깝게 위조할 수는 있으나 실제 사용자의 서명 속도 등 개인적인 특징까지 따라하기는 어렵다. 또한 실제 서명을 작성하는 과정을 그대로 지켜보지 않는 한 서명의 동적 정보를 얻기는 힘들기 때문에 온라인 방식이 오프라인 방식에 비해 보안상 더 안전하다고 할 수 있다.

본문에서는 서명검증에 많이 쓰이는 CNN 모델과 그에 동적 데이터를 학습할 수 있는 LSTM 을 결합한 새로운 CNN-LSTM 모델을 함께 제안하여 서명 분류 성능을 관찰하고 각 모델의 성능을 비교한다. CNN 의 경우 하나의 이미지에 동적 데이터를 포함할 수 있도

록 최승호, 정성훈의 논문에서 사용한 방법을 참고하였다.[2][4] 서명 습득 과정과 가공 방법은 본문에 자세히 서술한다.

서명은 스타일러스 펜을 사용하여 필기압력을 적용한 펜으로 작성한 서명과 그렇지 않은 서명을 구분해 얻었으며 이를 CNN, CNN-LSTM 에 기반한 모델에 학습한 결과를 비교하였다. CNN 모델은 중간 정도의 밀도를 가진 서명에 강한 검증 성능을 보였으며 CNN-LSTM 의 경우 과소적합을 일으키기 쉬우나 Dropout Layer 를 더 추가한 모델의 경우 밀도가 낮거나 높은 서명 검증에서 좋은 결과를 보일 것이라고 기대되었다. 여기서 밀도는 전체 서명 이미지에서 서명 부분이 차지하는 비율을 의미한다.

2. 관련 연구

서명 검증에는 오프라인 방식과 온라인 방식이 있다.[7] 오프라인은 서명이 완전히 완료된 시점의 이미지를 이용하는 방법이며 온라인은 시간에 따른 펜의 압력, 획의 수 등 동적 데이터를 함께 이용하는 방법이다. 특히 복잡한 이미지의 분류를 다루기 위해 많은 연구에서 Krizhevsky 가 제안한 AlexNet 이 적용되었다. 처음 제안된 AlexNet 에는 두개의 LRN(Local Response Normalization) Layer 가 포함되어 있으나 이후 몇몇 연구에서는 Batch Normalization 이 대신 사용되

나 아예 삭제되기도 하였다.[8][9] AlexNet 에 Dropout Layer 을 더 추가한 필기 인식 성능 향상 모델을 제안한 R.Almodfer 의 연구에서는 ReLU 와 tanH 활성화 함수를 사용한 두 경우 모두 92% 이상의 분류 정확도를 달성하여 적절히 변형된 AlexNet 도 필기인식에서 좋은 성과를 낼 수 있다는 것을 보여주었다.[10]

최승호, 정성훈에 의해 제안된 연구에서는 CNN 모델에 시간에 따른 서명 이미지의 변화를 학습하기 위해 서명을 동영상의 형태로 얻어 등간격의 10 프레임으로 나눈 뒤 이를 하나로 합성하여 사용하였다.[2][4] 그러나 이 방식은 합성된 이미지가 지나치게 커지거나 특정 프레임의 영향이 작아지는 단점이 있다. 또한 해당 연구에서는 모델 학습 시 진서명과 위조서명의 수를 같은 비율로 설정하였으나 실제로는 위조서명을 사전에 충분히 얻기 힘들다는 문제가 있다.

부족한 위조서명의 수를 보충하기 위해 대상 서명 외의 필기데이터를 이용한 연구들이 있다. CY. Park 외 2 인이 연구한 서명검증 모델에서는 일반적인 위조서명과 함께 서명검증 대상자 본인이 스스로 위조한 듯 작성한 서명을 필기데이터로서 이용하였다.[9] SVM 기반의 온라인 서명검증 기법을 연구한 최훈, 허경용의 논문에서는 타인의 진서명을 음의 데이터로 사용한 실험에서 더 높은 분류 성능을 보였다.[5]

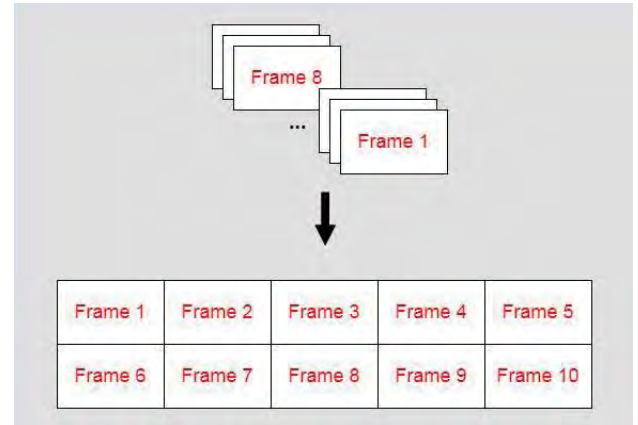
3. 서명 데이터

<표 1> 서명집합의 구성

	Original	Forgery	Others	Written Data
Normal	50 EA	30 EA	4 EA	1 EA
Pen	50 EA	30 EA	4 EA	1 EA

서명집합은 최승호, 정성훈의 연구와 동일한 방식으로 수집하였다.[2][4] 스타일러스 펜을 사용할 수 있는 안드로이드 스마트폰으로 서명 영상을 얻어 서명의 시작과 끝을 기준으로 10 프레임으로 나눈 것을 하나의 서명집합으로 상정하였다. 표 1 은 사용자 1 명에 대한 서명집합의 구성을 표현한 것이다. 표 1 의 서명들을 행을 기준으로 하나의 서명 집합으로 묶어 모델 학습 및 판별 테스트를 진행한다. 위조서명(Forgery)은 특정 5 명 중 서명 판별의 대상자를 제외한 4 명이 대상자의 진서명을 약 20 회 정도 연습한 후 각각 7~10 개씩 작성한 것을 사용하였으며 해당 4 명의 진서명(Others)은 학습 시 음의 데이터로 사용된다. 또한 검증 대상자가 자신의 서명을 위조한 듯이 작성한 필기 데이터(Written Data)도 음의 데이터로 분류된다. 최종적으로는 같은 사용자에 대해 필기압력이 적용된 경우의 서명과 그렇지 않은 경우 두 가지의 서명집합

(Normal, Pen)을 각각 수집하였다. 연구는 총 2 명의 사용자, CASE1, CASE2 에 대해 각 2 개씩 총 4 개의 서명집합을 가지고 진행하였다. 이 4 개의 서명집합을 CASE1-normal, CASE1-pen, CASE2-normal, CASE2-pen 으로 표기하겠다.



(그림 1) 서명 이미지 합성 방법

그림 1 은 CNN 모델에 사용하는 서명 이미지의 합성 방법을 나타낸 것이다. CNN-LSTM 에 기반한 모델은 서명 프레임을 합성하지 않고 각 프레임을 230X160 픽셀로 조정한 것을 사용하였으며 CNN 에 기반한 모델에는 한 서명의 10 프레임을 시간 순서에 따라 왼쪽 위에서 오른쪽 아래 방향으로 합성 후 1225X340 픽셀로 조정한 것을 사용하였다.

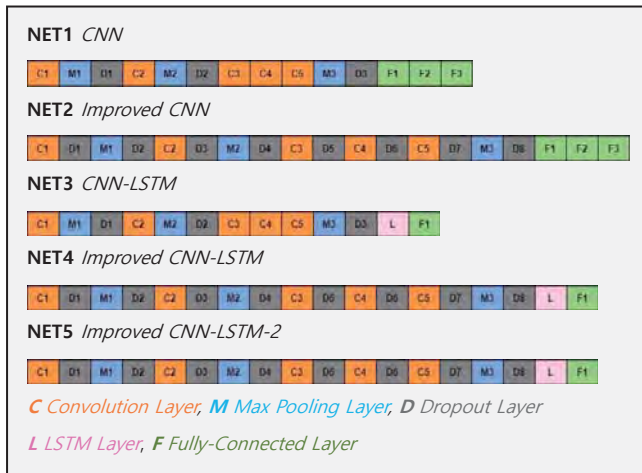
각 서명 이미지는 서명의 펜 선보다 배경의 면적이 훨씬 넓기 때문에 데이터를 배열의 형태로 변환한 경우 최대 휘도인 255(흰색)의 원소가 많아진다. 그러나 배경부분의 값은 실제 서명 학습에 필요한 정보가 아니므로 각 원소의 휘도값 I 에서 255 를 뺀 뒤 결정되는 회소행렬을 사용하는 것이 더 효율적이다. 그 다음 각각을 0~1. 사이의 값으로 표준화하기 위해 -255 로 나누었는데 이를 식으로 나타내면 (1)과 같다.

$$I'(x,y) = (I(x,y) - 255) \times (-\frac{1}{255}) \quad (1)$$

서명 이미지는 배경 부분을 투명하게 만든 뒤 사용하였으며 모델 학습에서는 RGBA 형식으로 이미지를 처리하였기 때문에 실제로 배경원소가 255 의 값을 가지는 것을 확인할 수 있었다.

4. 제안 방법

본 논문에서는 온라인 서명 검증을 위해 5 가지 DL(Deep Learning) 모델을 제안한다. 각 모델은 이미지 분류의 대표적인 모델 중 하나인 AlexNet 을 기반으로 하였으나 Krizhevsky 가 제안한 모델에서 LRN Layer 를 삭제한 것을 사용하였다.[8]



(그림 2) 제안모델의 구조

그림 2 는 제안모델의 구조를 간략히 나타낸 것이다. 모든 모델에서 활성화함수는 ReLU 함수가 쓰였으며 Max Pooling 의 Size 는 (3, 3), Stride 는 (2, 2)로 고정되었다. Convolution Layer 의 경우 첫번째 층은 Stride 가 (4, 4), 그 외에는 전부 (1, 1)로 설정하였다. 마지막 Fully-Connected Layer 의 활성화 함수로는 Softmax 를 사용하였다. NET1-NET2 의 Fully-Connected Layer 는 각각 1024, 1024, 2 개의 뉴런을 가지고 있으며 NET3-NET5 의 경우 128 개의 뉴런을 가지는 LSTM Layer 와 2 개의 뉴런을 가진 Fully-Connected Layer 로 구성되어 있다.

NET1, NET3 와 달리 NET2, NET4, NET5 에서는 과대적합을 줄이기 위한 Dropout Layer 가 각 Convolution Layer 뒤에 하나씩 더 추가가 됐다는 점에서 차이가 있다. Dropout Layer 마다 설정된 Probability 값은 표 2 에 나타냈다.

<표 2> Dropout Layer 의 Probability

	D1	D2	D3	D4	D5	D6	D7	D8
NET1	0.1	0.3	0.5	-	-	-	-	-
NET2	0.0	0.1	0.1	0.3	0.2	0.3	0.4	0.5
NET3	0.1	0.3	0.5	-	-	-	-	-
NET4	0.0	0.1	0.1	0.3	0.2	0.3	0.4	0.5
NET5	0.0	0.1	0.0	0.2	0.1	0.2	0.2	0.3

모델 학습은 epoch 100, batch size 10 로 진행하였으며 진서명 30 개와 타인의 진서명 4 개, 위조서명 10 개, 검증 대상자의 일반 필기데이터 1 개를 사용한다. 실제로는 위조서명을 얻기 힘들다는 점을 고려해 위조서명은 진서명 30 개에 비해 적은 10 개가 사용되었는데 부족한 위조서명 데이터를 보충하기 위해 타인의 진서명과 대상자의 일반 필기데이터를 사용하였다. 진서명 30 개만을 양의 데이터인 1 로 설정하였으며 나머지는 모두 음의 데이터인 0 으로 취급하였다. 위

조서명은 NET1-NET5 에서 모두 고정된 객체를 사용하였으나 진서명은 학습마다 랜덤으로 선택하였다. 이는 위조서명은 습득하기 어려운 반면 진서명은 일반적으로 학습에 사용가능한 데이터가 충분한 상황을 반영한 것이다.

5. 실험 결과

표 3 은 4 개의 서명집합에 대한 NET1-NET5 의 테스트 결과를 나타낸 것이다. 테스트 시에는 서명집합의 진서명 50 개와 위조서명 30 개에 대해 검증하였으며 성능평가는 ACC(Accuracy)를 사용하였다.

<표 3> 제안모델(NET1-NET5)의 테스트 결과

ACC Value (Range: 0~1)					
CASE1	NET1	NET2	NET3	NET4	NET5
Normal	0.3750	0.3750	0.6250	0.6250*	0.3750
Pen	0.6250	0.6250	0.6250	0.6250	0.6250
CASE2	NET1	NET2	NET3	NET4	NET5
Normal	0.9625	0.3750	0.6250	0.6750	0.6250
Pen	0.3750	0.3750	0.6250	0.5500	0.5625

대부분의 모델이 0.3750 혹은 0.6250 의 ACC 값을 가지고 있다. 각 서명에 대한 예측 결과를 분석하니 0.6250 은 모든 서명을 진서명으로, 0.3750 은 반대로 모두 위조서명으로 판단한 경우로 확인되었다. 즉, ACC 가 0.6250 인 것은 과소적합으로, 0.3750 인 것은 과대적합으로 학습된 상황을 의미한다. 예외적으로 CASE1-Normal 로 학습한 NET4 의 경우 서명의 예측 결과가 일관적인 것은 아니나 단지 우연히 제대로 검증된 서명의 수가 전체 80 개 중 50 개인 것으로 확인되었다. 표 3 의 결과만 보았을 때에는 CASE2-Normal 서명집합의 NET1 을 제외하고는 전부 제대로 검증을 수행하지 못한 것으로 보인다.

이 결과를 더 면밀히 분석하기 위해 각 서명집합의 서명 밀도를 측정하였다. 이때 밀도는 50 개의 진서명을 grayscale 의 230X160 크기로 추출해 배경을 제외한 서명 부분의 픽셀 수를 모두 더한 뒤 전체 픽셀수인 230X160X50 으로 나눈 결과를 말한다. CASE1-normal 이 0.09 로 밀도가 가장 낮았으며 CASE1-pen 과 CASE2-normal 은 각각 0.44 와 0.47 로 중간 정도, CASE2-pen 은 0.69 로 밀도가 가장 높았다.

서명 밀도가 낮은 집합인 CASE1-normal 의 경우 CNN 과 CNN-LSTM 모델 종류에 상관없이 전반적으로 과대적합 되는 경향이 높은 것으로 나타났다. 이는 전체 서명 이미지 중 서명 획이 차지하는 비율이 작아 학습에 필요한 정보의 양이 부족하기 때문인 것으로 보인다. 하지만 NET4 에서는 ACC 값이 아주 높지는 않으나 과소적합이나 과대적합이 발생하는 경향

은 상대적으로 작은 것을 알 수 있다. 이 경우 본 논문에서 제안한 100 보다 높은 epoch 를 설정하여 학습하면 정확도가 향상될 것으로 보인다.

중 수준 밀도의 서명인 CASE1-pen 과 CASE2-normal 에서는 전체 모델 중 ACC 값이 0.6250 인 과소적합이 발생할 확률이 0.8 로 매우 높았다. 다만 CNN 을 사용한 NET2 에서는 반대로 과대적합이 발생하였으며 CASE2-normal 의 NET1 에서는 ACC 가 0.9625 로 거의 대부분의 서명을 정확하게 검증하였다. 이는 CASE2-normal 과 비슷한 밀도를 가진 서명의 경우 NET3-NET5, 즉 CNN-LSTM 을 이용한 모델에서는 과소적합이 발생할 경향이 크며 오히려 일반적인 CNN 모델에서 좋은 검증 성능을 보인다는 것을 의미한다.

마지막으로 CASE2-pen 에 대한 모델 검증 결과에 미루어 볼 때 서명 밀도가 높은 경우 CNN 을 이용한 모델(NET1-NET2)에서는 과대적합이 발생할 확률이 높으며 3 개의 Dropout Layer 만을 포함한 일반적인 수준의 CNN-LSTM(NET3)에서는 과소적합이 발생할 수 있다는 사실을 알려준다. 반면 6-8 개의 Dropout Layer 가 포함된 CNN-LSTM(NET4-NET5)에서는 CASE1-normal 을 사용한 NET4 의 경우와 마찬가지로 과대/과소적합이 발생하는 경향이 낮은 것을 보아 epoch 를 높인 학습에서 좋은 분류 성능이 기대된다.

6. 결론

밀도가 다른 각 서명 집합을 일반적인 CNN, CNN-LSTM 모델과 Dropout Layer 를 더 추가한 모델에 학습, 검증한 결과 서명의 밀도에 따라 검증에 적합한 모델이 달라지는 것을 알 수 있었다. 서명의 밀도가 낮거나 반대로 높은 경우 6 개 이상의 Dropout Layer 를 포함한 CNN-LSTM 모델에 학습할 때 과대/과소적합이 발생하는 경향이 낮았으며 이때 정확도가 낮은 것은 충분히 높은 epoch 를 설정하면 해결할 수 있을 것으로 기대된다. 서명의 밀도가 중간정도 수준일 때는 반대로 적은 수의 Dropout Layer 를 포함한 CNN 모델에서 거의 정확한 수준의 검증 성능을 보였다. 다만 실생활에서 서명 학습을 위한 데이터를 얻기가 매우 제한적이며 같은 사람이라도 서명을 할 때마다 결과가 조금씩 달라진다는 사실을 고려해볼 때 적은 수의 서명으로도 정확도를 충분히 높일 수 있는 서명 검증 연구가 필요할 것이다.

참고문헌

- [1] Kim, Hyun-Woo and Seung-In Kim. "A study on User experience of Fintech Application Service-Focused on Toss and Kakaobank." *Journal of Digital Convergence* 18.1: 287-293. (2020)
- [2] Choi, Seoung-Ho and Sung Hoon Jung. "Fake Discrimination using Time Information in CNN-based Signature Recognition." *Proceedings of the Korean Society of Computer Information Conference*. Korean Society of Computer Information, 2017. p. 293-294.
- [3] Zeng, Zihan and Jing Tian. "Deep Learning Methods for Signature Verification." *arXiv preprint arXiv:1912.05435*. (2019)
- [4] Choi, Seoung-Ho and Sung Hoon Jung. "Performance improvement of fake discrimination using time information in CNN-based signature recognition." *Journal of Digital Contents Society* 19.1: 205-212. (2018)
- [5] Choi, Hun and Gyeongyoung Heo. "Online Signature Verification Method using General Handwriting Data and 1-class SVM." *Journal of the Korea Institute of Information and Communication Engineering* 22.11: 1435-1441. (2018)
- [6] Agarap, Abien Fred. "An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification." *arXiv preprint arXiv:1712.03541*. (2017)
- [7] Hafemann, Luiz G., Robert Sabourin, and Luiz S. Oliveira. "Learning features for offline handwritten signature verification using deep convolutional neural networks." *Pattern Recognition*, 70: 163-176. (2017)
- [8] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*, 2012. p. 1097-1105.
- [9] Park, Chan-Yong, Han-Gyu Kim, and Ho-Jin Choi. "Robust Online Signature Verification Using Long-term Recurrent Convolutional Network." 2019 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2019. p. 1-6.
- [10] Almodfer, Rolla, et al. "Enhancing AlexNet for Arabic Handwritten words Recognition Using Incremental Dropout." 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, 2017. p. 663-669.