오토인코더 및 컨볼루션 네트워크를 활용한 사용자 인증에서의 개인정보 보호

김현지*, 임세진**, 양유진**, 서화정***†

*한성대학교 IT 융합공학부 (대학원생)

**한성대학교 IT 융합공학부 (대학생)

***† 한성대학교 IT 융합공학부 (교수)

Privacy protection in user authentication using autoencoder and convolutional neural network

Hyun-Ji Kim*, Se-Jin Lim**, Yu-Jin Yang**, Hwa-Jeong Seo***†

*Hansung University, Department of IT Convergence Engineering. (Graduate student)

**Hansung University, Department of IT Convergence Engineering. (Student)

***† Hansung University, Department of IT Convergence Engineering. (Professor)

요 약

최근, COVID-19로 인해 특정 장소 출입 시 QR code 인증이 필수적으로 수행되고 있다. 그러나 QR code 기반의 인증 방식은 디코딩 과정을 필요로 하며, 저장 및 인증 과정에서 사용자의 개인정보가 모두 노출된다. 이러한 보안 취약점을 극복하기 위해 본 논문에서는 디코딩 과정이 필요하지 않은 생체인증 방안을 제안한다. 사용자의 지문 정보가 오토인코더에 입력되면, 지문의 특징 정보만이 추출되며 차원변경을 통해 인증에 사용될 이미지가 생성된다. 검증자는 해당 정보를 인식하여 컨볼루션 네트워크를 통해 사용자 인증을 수행한다. 즉, 검증자는 사용자의 실제 지문 정보를 학습 및 저장할 필요가 없으므로, 기존 방법의 보안 취약점을 극복할 수 있다. 또한, 비콘 등의 BLE 단말과 함께 사용할 경우 다른 장소에서 인증을 수행하는 경우를 방지할 수 있다.

I. 서론

COVID-19가 전 세계적으로 유행함에 따라, 특정 시설 출입 시 방문 기록을 남기기 위해 QR-code 인증 방안이 시행되고 있다. 그러나 QR-code 사용으로 인한 개인 정보 노출, 악성 코드 삽입 등의 보안 취약점이 존재한다.

II. 관련 연구

2.1 QR code 인증의 보안 취약점

QR코드로 인증을 할 때 암호화된 정보를 해독해주는 디코딩(decoding) 과정이 필수적이다. 인증 절차를 수행하기 위하여 디코딩 과정을 거쳐 나온 개인 정보를 인증 서버로 전송하여야 하는데, 이 과정에서 사용자의 정보가 모두 노출된다는 문제점이 있다.

2.2. 오토인코더 (Auto Encoder)

오토인코더는 데이터 레이블 없이 학습시키는 비지도 학습 방법이다[1]. 인코더는 입력된 데이터의 핵심 특징만을 추출하고, 은닉층에서 이 추출한 특징을 학습시키면 디코더에서 추출된 특징 값을 바탕으로 원본 데이터와 근사한 값이 나오도록 재구성해준다. 이때, 인코더와 디코더에 들어가는 노드수보다 은닉층에 들어가는 노드수가 더 적은 손실 압축 방법을 사용한다.

III. 시스템 제안

본 논문에서는 QR code 기반 인증 방법의 보안 취약점을 극복하기 위해 오토인코더 및 컨볼루션 네트워크 (Convolutional Neural Network, CNN) 기반의 본인 인증 방안을 제안한다. 제안 기법을 통해 인증 정보에 대한 디코딩 과정 없이 본인 인증을 수행할 수 있다. 시스템 구성은 그림 1과 같다. 제안 시스템에는 인증 대상자인 사용자와 사용자에 대한 인증을 수행하는 기관인 검증자가 참여한다. 해당 시스템은 tensorflow lite 모델로 변환되어 디바이스 상에서 수행되며, 디바이스에 저장된 지문 정보를 활용하여 인증에 사용될 이미지를 생성한다. 검증자는 생성된 해당 이미지를 통해 사용자 인증을 수행한다.

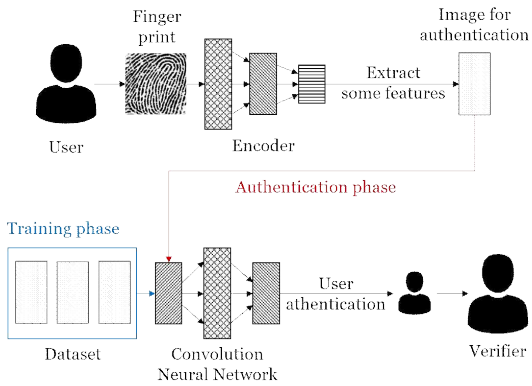

그림 1. 시스템 구성도

3.1 사용자

사용자는 본인 인증에 사용할 지문 정보에 대한 이미지를 생성해야한다. 이를 위해 오토인코더를 사용하여 자신의 지문 정보로부터 특징점을 추출해낸다. 오토인코더는 인코더를 활용하여 주요 특징 정보들을 추출한 후, 디코더를 통해 해당 정보를 기반으로 다시 원본 데이터로 복구한다. 이 과정에서 노이즈가 제거된다. 그러나 지문 정보를 이미지와 같은 원본 데이터의 형태로 재생산하는 것이 아니라, 해당 지문이 갖는 주요 특징만을 가지고 인증 정보를 생성하는 것이므로 제안하는 시스템에서는 디코더 모델이 필요하지 않다. 따라서 학습 시에는 각각 설계된 인코더와 디코더를 결합한 모델을 사용하고, 이후 인증 정보 생성 시에는 학습된 인코더만을 사용하여 추론한다. 이러한 구조를 통해 지문 전체에 대한 정보 노출 없이도 본인인증이 가능하며, 모바일 장치 상에서의 추론 시 배포되는 모델의 용량을 줄일 수 있다.

훈련에 사용되는 오토인코더의 구조 및 하이퍼파라미터는 다음과 같다. conv2D + Maxpooling2D 구조를 두 번 반복한 후 출력층으로 conv2D를 사용한다. max pooling에서 pool size를 2로 설정하였으므로 입력 이미지의 가로, 세로 길이가 총 4배 줄어든다. 또한, 특징 벡터의 차원을 나타내는 latent vector값을 128로 설정하여 인코더의 출력은 (56,56,128)의 모양을 갖게 된다. 인코더의 출력은 디코더의 입력으로 사용된다. 데이터의 크기를 늘려주는 upsampling 과정을 통해 가로, 세로의 크기를 인코더의 입력 데이터의 형태로 복원한다. 학습을 위해 Mean Square Error 손실 함수를 사용하였고, 최적화 함수는 RMSprop, 활성화 함수는 은닉층에 ReLu를, 출력층에 Sigmoid를 사용하였다.

3.2 검증자

검증자는 CNN을 기반으로 하여 인증을 원하는 사용자로부터 인식한 정보를 통해 본인 인증을 수행한다. 학습에 사용되는 데이터는 인코더를 통해 생성되는, 인증을 위한 이미지이다. 즉, 사용자들의 실제 지문 정보를 저장하지 않고, 특징점 추출 후 변형된 상태의 이미지만을 저장하여 학습에 사용한다. 따라서 기존의 QR code 인증과 달리 디코딩 과정에서의 정보 노출 없이 인코딩된 정보 그대로 인증이 가능하다. 또한, 비콘과 같은 BLE 단말을 사용하여 사용자가 해당 위치에 존재함을 증명하고 그 이후에 본인 인증 단계를 수행할 수 있게 할 경우, 사용자가 현재 위치를 속이고 다른 장소에서 인증하는 상황을 방지할 수 있다.

사용자의 본인 인증을 위해 훈련된 모델을 통한 추론을 진행한다. 먼저, 사용자가 생성한 인증 데이터를 훈련된 모델에 입력한다. 다중 클래스 분류 문제이기 때문에, 입력 데이터는 각각의 클래스에 속할 확률 값을 가진다. 해당 값들 중 가장 높은 값이 사전에 설정한 인증 임계값을 넘을 경우 본인으로 인증한다. 만약 A 사용자의 인증 정보가 A 사용자로 분류되었다고 해도 임계값을 넘지 못 할 경우 인증되지 않는다. 또한, B 사용자의 인증 정보를 A 사용자로 분류할 경우, 임계값 초과 여부와 상관없이 인증에 실패하게 된다. 즉, 인증이 가능한 경우는 본인이 본인으로 분류되며, 해당 클래스로 분류될 확률이 임계값을 넘는 경우뿐이다. 사용자 인증을 위한 CNN 모델로는 사전 학습된 모델인 Inception v3 model를 사용하였다.

IV. 성능 평가

본 실험을 위해 클라우드 기반 서비스인 Google Colaboratory를 활용한다. Ubuntu 18.04.3 LTS에서 실행되며, 12GB RAM의 Nvidia GPU로 구성되고 Python 3.6.9, tensorflow 2.2.0-rc, Keras 2.3.1이 사용된다. 실험 결과는 딥러닝 평가 기준 중 하나인 재현율 (recall)과 정밀도 (precision)의 조화평균인 F-measure로 평가한다. 또한, 동일 오류율 (Equal Error Rate, EER)을 통해 본인인증에 대한 성능을 확인한다.

인증 정보를 생성하는 오토인코더의 학습 결과는 그림 2와 같다. 훈련 및 검증 손실이 비슷한 수준으로 감소한 것을 볼 수 있다.

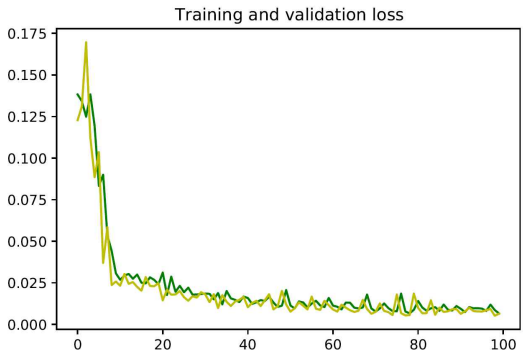

그림 2. Loss of AutoEncoder

그림 3은 사용자 인증을 위한 CNN의 학습 결과를 나타낸다. 또한, 그림 4와 같이 동일 오류율을 계산하고, 결과 값을 정상 서명과 위조 서명을 구분할 임계값으로 설정한다. EER은 타인 수락률 (FAR)과 본인 거부율 (FRR)이 같아지는 지점이며, 타인 수락률과 본인 거부율은 식 1과 같은 과정을 통해 계산한다. 표 1은 제안 시스템의 본인 인증에 대한 성능, EER 및 임계값을 정리한 표이다.

$$FAR = \frac{FP}{FP + TP}, \quad FRR = \frac{FN}{FN + TN} \qquad (1)$$

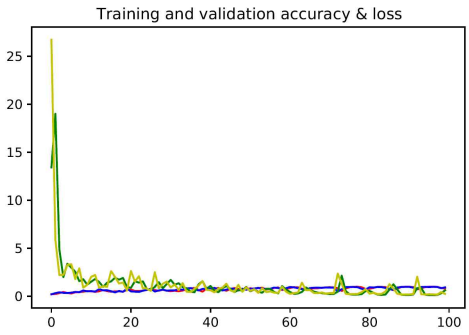

그림 3. Loss of CNN

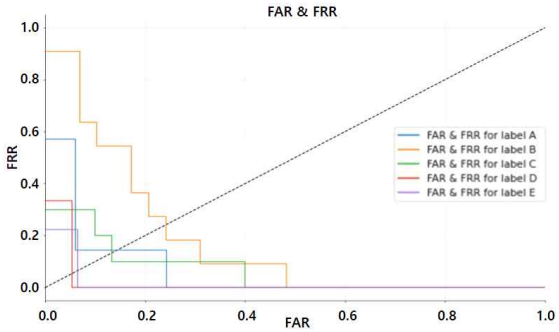

그림 4. 동일 오류율

표 1. F-measure, EER and threshold

F-measure	0.68
EER(average, max)	0.1, 0.225
threshold	0.225

V. 결론

본 논문에서는 오토인코더 및 컨볼루션 네트워크를 기반으로 디코딩 과정이 필요하지 않은 생체인증 방안을 제안하였다. 본 시스템이 위치 인증과 함께 사용될 경우 보안성을 더욱 향상시킬 것으로 기대된다.

VI. Acknowledgment

이 성과는 부분적으로 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2020R1F1A1048478) 그리고 이 성과는 부분적으로 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2018-0-00264, IoT 융합형 블록체인 플랫폼 보안 원천 기술 연구).

[참고문헌]

[1] SBaldi, Pierre. "Autoencoders, unsupervised learning, and deep architectures." Proceedings of ICML workshop on unsupervised and transfer learning. JMLR Workshop and Conference Proceedings, 2012.