CNN을 이용한 문자열 CAPTCHA 공격

String CAPTCHA attack using CNN

이상헌[1], 우상명[2], 이지형†
SangHeon. Lee[1], SangMyeong. Woh[2], Jee-Hyong. Lee†

[1,2,†] 성균관대학교 전자전기컴퓨터공학과
Department of Electrical and Computer Engineering, Sungkyunkwan University

요 약

CAPTCHA는 인터넷상에서 서비스 대상자의 사람 여부를 판단하는 시스템으로, 문자열 CAPTCHA 혹은 이미지 CAPTCHA 등 다양한 형태의 CAPTCHA가 사용되고 있다. 널리 쓰이는 문자열 CAPTCHA는 영문 혹은 숫자로 이루어진 문자열 이미지에 잡음을 섞음으로써 사람이 아닌 대상이 문자열의 의미를 판독하지 못하도록 한다. 본 논문은 현재 사용되고 있는 문자열 CAPTCHA 내 잡음을 image processing을 통해 제거하고, 단일 문자 이미지 데이터를 이용하여 CNN을 학습시킴으로써 문자열을 판독할 수 있도록 한다.

1. 서 론

CAPTCHA는 인터넷상에서 서비스 대상자의 사람 여부를 판단하는 시스템으로, 계속되는 로그인 시도 및 우회 경로를 통한 회원가입을 방지하는 데 사용되고 있다[1].

CAPTCHA는 그 형태에 따라 많은 종류가 있고, 문자열 CAPTCHA 또는 이미지 CAPTCHA가 널리 쓰이고 있다. 문자열 CAPTCHA는 영문 혹은 숫자로 이루어진 문자열이 담긴 이미지에 잡음을 섞음으로써 사람이 아닌 대상이 문자열의 의미를 판독하지 못하도록 한다.

본 논문에서는 국내 티켓 예매 사이트인 “인터파크 티켓”에서 사용되는 문자열 기반 CAPTCHA를 인식하는 모델을 제안한다. Image processing을 통해 잡음 제거 및 단일 문자로 분리하고, 분리한 단일 문자 데이터를 이용하여 CNN(Convolutional Neural Network)을 학습함으로써 CAPTCHA 문자열의 의미를 판독할 수 있도록 한다.

2장에서는 기계학습 모델인 CNN을 설명한다. 3장에서는 인터파크 티켓에서 사용하는 문자열 CAPTCHA의 특징을 소개하며, 4장에서는 제안 기법에 대해 서술한다. 5장에서는 결과를 서술 및 분석하고, 6장의 결론 및 소감으로 본 논문을 마친다.

† 교신저자

감사의 글 : 이 논문은 2014년도 과학기술정보통신부의 재원으로 한국연구재단-차세대정보 컴퓨팅기술 개발사업의 지원을 받아 수행된 연구임 (No. NRF-2014M3C4A7030503). 또한, 이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단-차세대정보컴퓨팅기술개발사업의 지원을 받아 수행된 연구임 (No. NRF-2017M3C4A7069440).

2. Convolutional Neural Network

CNN(Convolutional Neural Network)은 패턴 인식 분야에서 좋은 성능을 보이고 있어 이미지, 텍스트 등의 데이터를 분류, 인식하는데 주로 사용되는 다층신경망이다[2]. CNN은 크게 이미지 전처리 단계와 분류 단계로 나누어져 있다. 전처리 단계는 input 데이터의 특징을 추출하는 convolution 단계와 convolution을 통해 추출된 특징에서 가장 핵심적인 부분을 추출하는 pooling 단계로 이루어진다. 분류 단계에서는 전처리한 데이터를 이용하여 fully-connected 다층신경망을 통해 이미지를 분류, 인식한다. 그림 1은 CNN의 일종인 LeNet-5의 구조이다[3].

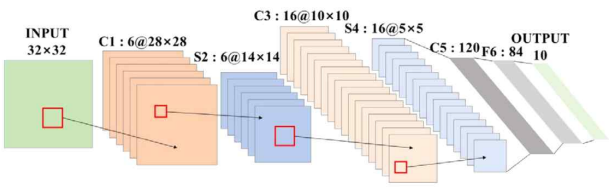

그림 1. CNN 구조

Fig 1. CNN sturcture

3. 인터파크 티켓 CAPTCHA의 특징

국내 티켓 예매 사이트인 “인터파크 티켓”은 서비스 이용 대상자가 사람인지를 판별하기 위해 예매 과정에서 문자열 CAPTCHA를 사용하는 “안심예매” 시스템을 적용하고 있다. 그림 2는 “안심 예매” 시스템에서 사용하는 문자열 CAPTCHA이다.

그림 2. 인터파크 티켓의 문자열 CAPTCHA
Fig 2. String CAPTCHA used in Interpark Ticket

인터파크 티켓에서 사용되는 문자열 CAPTCHA의 특징은 다음과 같다.

- 문자열은 A, B, C, D, E, K, L, M, N, P, Q, R, S, T, U, W, X, Z의 18개 대문자 영문자로 이루어져 있고, 길이는 6이다.
- 한 CAPTCHA 내 문자들의 색은 서로 같으며, 문자열 뒤에 단색의 배경이 있다.
- 문자열 주위로 작은 점들이 분포되어 있고, 문자열 가운데에 직선이 있다.
- CAPTCHA 이미지의 크기는 고정되어 있고, 배경 색과 문자열의 색은 재사용된다.

4. 제안 기법

4.1 Image processing

Image processing은 CAPTCHA 이미지 파일에서 단일 문자 이미지를 추출하는 과정이다. Image processing은 크게 두 단계로 구분되는데, 이는 CAPTCHA 이미지 내의 잡음을 제거하는 단계와 잡음이 제거된 문자열 이미지에서 단일 문자 이미지로 분리 및 저장하는 단계이다. Image processing 과정은 C++ 환경에서 OpenCV(Open Source Computer Vision) 라이브러리를 이용하여 구현하였다.

CAPTCHA 이미지 내의 잡음을 제거하는 단계는 morphology 연산을 통해 구현하였다. Morphology 연산은 영상의 밝은 영역이나 어두운 영역을 축소, 확대함으로써 특정 객체의 형태를 변형시키는 영상 처리 기법이다. 열림 연산은 영역에 나타난 미세한 조각들을 제거하는 역할을, 닫힘 연산은 영역에 생긴 미세한 틈을 메우는 역할을 하기 때문에, 닫힘과 열림 연산을 차례로 수행함으로써 이미지 내의 잡음을 제거할 수 있다. 그림 3은 0(black) 또는 255(white)의 픽셀 값을 갖는 이미지에 이를 적용하여 잡음을 제거한 모습이다.

문자열에서 단일 문자 이미지로 분리 및 저장하는 단계는 영상 내의 특정 객체의 경계선(contour)을 추출하여 배열의 형태로 저장하는 findContours() 함수를 이용하였다. 함수를 통해 각 단어가 포함된 최소 사각형 경계선을 저장하고, 경계선을 따라 새로운 이미지로 잘라내어 저장함으로써 단일 문자 이미지를 얻었다.

Image processing을 통해 얻은 단일 문자 이미지는 32*32 크기의 GrayScale, JPG 형식으로 저장하였다. 분리 및 저장된 단일 문자 이미지는 그림 4와 같다.

그림 3. 닫힘과 열림 연산을 통한 이미지 잡음 제거
Fig 3. Eliminate image noise with open and close operations

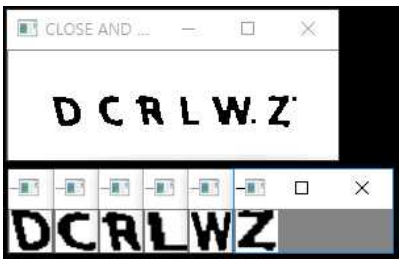

그림 4. 단일 문자 영역 분리
Fig 4. Separate single character

4.2 Image Augmentation

Image augmentation은 이미지 데이터에 특정 확률로 왜곡을 추가함으로써 데이터의 개수를 증폭시키는 작업이다. Image Processing 과정을 통해 단일 문자 이미지 3000개를 저장하였고, image augmentation 과정을 통해 각 영문자 label당 1500개씩 총 27000개의 이미지 데이터를 얻었다[4].

4.3 이미지 학습 및 인식

Image Processing 및 image augmentation 과정을 통해 저장한 27000개의 단일 문자 이미지 데이터는 학습을 위한 dataset으로 분류하였다. 또한 학습 dataset에 포함되지 않는 1800개의 CAPTCHA

이미지에 대해 image processing 과정을 적용하여 얻은 10800개의 단일 문자 이미지 데이터는 학습 모델의 인식 성능을 알아보기 위한 dataset으로 분류하였다. 학습은 CNN 모델 중 Google의 Inception V3 모델을 이용하였다[5].

5. 결과 분석

5.1 Image Processing 결과

CAPTCHA 내 잡음을 제거하고 문자열에서 단일 문자로 분리하는 image processing 과정 결과, 제대로 제거되지 않은 잡음이 이미지에 포함된 것을 확인할 수 있었다. 이는 morphology 연산만으로는 잡음을 확실히 제거하지 못한다는 것을 의미한다. 제거되지 않은 잡음에 대해 픽셀 단위의 연산을 수행하여 해당 픽셀이 잡음인지 아닌지를 판단하고, 잡음에 해당하는 경우 이를 제거하는 새로운 알고리즘을 필요로 한다.

5.2 이미지 학습 및 인식 결과

이미지 인식 실험 결과 총 1800개의 서로 다른 문자열을 포함하는 CAPTCHA 중 1531개의 CAPTCHA를 성공적으로 인식하였다. 또한 단일 문자 이미지 기준으로 10800개의 데이터 중 10498개의 데이터를 성공적으로 인식하였다. CAPTCHA 기준의 인식 성공은 CAPTCHA의 문자열 내 모든 문자를 성공적으로 인식한 경우에만 해당되므로, 한 문자라도 인식하는데 실패하였다면 CAPTCHA 인식에 실패하였다고 간주하였다. CAPTCHA 기준 전체 인식률은 85.06%이며, 단일 문자 기준 인식률은 97.20%이다. 단일 문자 기준 인식률이 매우 높음에 따라 image processing을 통해 분리된 각 문자에 대한 특징 추출이 성공적임을 알 수 있다.

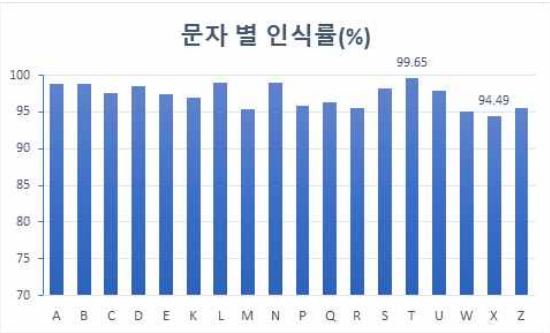

그림 5. 영문자의 종류에 따른 인식률

Fig 5. Recognition rate according to type of alphabet

그림 5는 인터파크 티켓에서 사용되는 CAPTCHA를 이루는 18개의 영문자에 따른 인식률이다. 가장 높은 인식률을 보인 문자는 T와 N(각 99.65%, 99.00%의 인식률)이며, 가장 낮은 인식률을 보인 문자는 X와 W(각 94.49%, 95.02%의 인식률)였다.

6. 결 론

본 논문은 현재 국내 티켓 예매 사이트에서 사용되고 있는 문자열 기반 CAPTCHA에 대해 높은 확률로 이를 인식하는 모델을 구현함으로써 취약점을 입증하였다. CAPTCHA 내 잡음을 제거하고 단일 문자로 분리하는 image processing 과정을 제안하였고, 문자 인식에는 CNN의 Inception V3 모델을 사용하였다. 인식 실험 결과 CAPTCHA 기준 85.06% 및 단일 문자 기준 97.20%의 높은 인식률을 나타내었다. 특히 단일 문자 기준 인식률이 높다는 것은, 단순히 CAPTCHA 내 문자열의 길이를 늘이는 방식으로는 제안된 모델의 CAPTCHA 인식률을 낮추기 힘들다는 것을 의미한다. 따라서 해당 사이트는 문자에 새로운 왜곡을 추가하거나, 문자 사이의 거리를 조절하는 등의 새로운 알고리즘을 통해 보완된 CAPTCHA를 도입해야할 것을 권고한다.

참 고 문 헌

[1] 김재환, 김수아, 김형중, “특징 분리를 통한 자연 배경을 지닌 글자 기반 CAPTCHA 공격,” Journal of The Korea Institute of Information Security & Cryptology, Vol. 25, No. 5, pp. 1011-1019, Oct. 2015.

[2] 이우영, 고광은, 김종우, “HS 알고리즘을 이용한 CNN의 Hyperparameter 결정 기법,” Journal of Korean Institute of Intelligent Systems, Vol. 27, No. 1, pp. 22-28, Feb. 2017

[3] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, Vol. 86, No. 11, pp. 2278-2324, 1998

[4] Image augmentation OpenSource in Github, https://github.com/mdbloice/Augmentor

[5] CNN Inception V3 OpenSource in Github, https://github.com/ArunMichaelDsouza/tensorflow-image-detection