텐서플로를 이용한 OCR 시스템 개발

Development of OCR System Using Tensorflow

저자 (Authors)

윤병일, 최재성, 김병만, 이해연

Byung-Il Yoon, Jae-Sung Choi, Byung-Man Kim, Hae-Yeoun Lee

출처 (Source)

Proceedings of KIIT Summer Conference , 2017.6, 334-336 (3 pages)

발행처 (Publisher)

한국정보기술학회

Korean Institute of Information Technology

URL

http://www.dbpia.co.kr/Article/NODE07182748

APA Style

윤병일, 최재성, 김병만, 이해연 (2017). 텐서플로를 이용한 OCR 시스템 개발. Proceedings of KIIT Summer Conference, 334-336.

텐서플로를 이용한 OCR 시스템 개발

윤병일*, 최재성*, 김병만*, 이해연*

Development of OCR System Using Tensorflow

Byung-Il Yoon*, Jae-Sung Choi*, Byung-Man Kim*, and Hae-Yeoun Lee*

요 약

텐서플로 기반의 한글 글자 영상을 인식하는 기계 학습 모델을 개발하였다. 사용자가 입력한 영상은 모델 내부 레이어를 통과하며 특징 값을 추출하여 글자로 인식된다. 최초 모델 구성 이후에도 사용자의 오류 제시를 통한 추가 학습 영상 입력이 가능하다. 따라서 별도 개발 없이 모델의 학습만으로 정확도를 개선할 수 있으며, 학습 데이터에 따라 한글 외 다른 언어의 OCR도 가능하다.

Abstract

Using Tensorflow, a machine learning model is developed to recognize korean character. Input images pass through the internal layer which extract features and are recognized as characters. After setting up the initial model, it is additionally trained using the notification of error by users. Therefore, the accuracy can be improved without additional developing and applied to other languages by changing training data.

Key words
optical character reader, tensorflow, Korean character

I. 서 론

한글에 대한 광학 문자 인식률은 타 언어에 비해 상대적으로 미흡한 편이다. 영어권에서는 광학 문자 인식을 다용도로 사용하고 있으나, 한국어는 비교적 미흡한 편이라 그 응용에 한계가 있다. 또한 표음 문자 중에서도 자음과 모음을 조합해서 글자를 만드는 특성이 OCR에 불리하게 작용한다. 따라서 텐서플로를 기반으로 한 머신러닝을 통해 인식의 정확도를 개선하고, OCR이 필요한 사용자들에게 유용하게 쓰일 수 있는 한글 OCR 시스템을 제안한다.

II. 제안하는 시스템

2.1 시스템 구조도

그림 1과 같이 서버는 사용자에게 입력받은 한글 영상의 정규화 과정을 거친 이후 해당 문장을 한 글자 단위로 잘라낸 후 각 글자를 인식한다. 인식이 잘못된 경우, 사용자는 시스템을 통해 데이터베이스에 오류 내용을 등록할 수 있다.

* 금오공과대학교 컴퓨터소프트웨어공학과

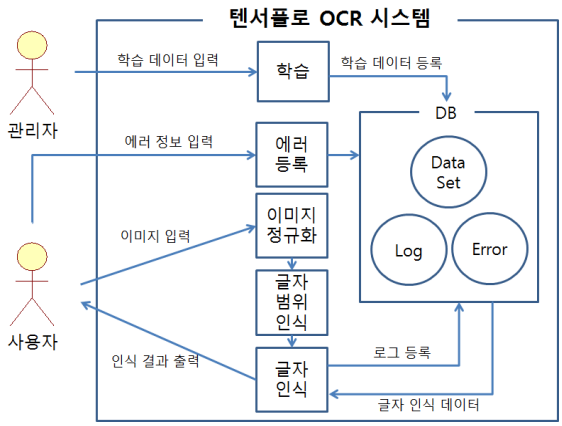

그림 1. 시스템 구조도

2.2 Adaptive Threshold

미리 정해둔 임계값에 대하여 전체 영상을 이진화하는 것이 아닌, 픽셀 주변 영역의 밝기 평균에 일정한 상수를 빼서 결정된 값을 임계값으로 사용하는 방법이다.

2.3 컨볼루션 신경망

전체 영상은 바로 글자로 인식할 수 없기 때문에 영상 특징 추출, Max Pooling이 필요하다. 영상 특징 추출은 영상에 Filter를 적용하여 Feature를 형성하는 과정이다. Max Pooling은 Feature에 혼재한 Noise의 영향력을 줄이기 위해 Feature의 최댓값을 추출하여 보다 작은 규모의 Feature로 축소하는 과정이다[1]. Filter의 임계값은 Convolution 네트워크 학습에서 자동으로 최적화된다.

2.4 학습 모델 구성

학습 모델은 총 10개의 층으로 구성한다. 시스템에 입력한 영상은 정규화 과정을 거친 후 모델로 입력한다. 모델 입력 영상은 그림 2에 해당하는 층을 거쳐 출력 값 중 하나를 선정하게 된다.

각 층은 ReLU 활성화 함수를 사용하고 출력 값을 선택하는 최종 Desne 레이어만 Softmax 활성화 함수를 사용한다. 출력 값은 KS1001에서 사용하는 한글 2350자와 숫자 10개를 합친 2360자 중 1글자로 한다. 모델 구성은 KERAS을 사용하였다[2].

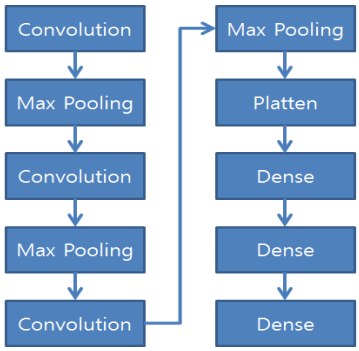

그림 2. 신경망 구성도

III. 개발 결과 및 결론

사용자는 웹페이지를 통해 영상을 서버에 업로드하고, 서버는 입력받은 영상을 이진화하고 잉여공백을 제거한 후, 각각 한 글자 단위로 나누고 글자 인식을 시작한다. 인식이 완료되면 웹페이지를 통해 사용자에게 출력해준다. 출력 글자에 오류가 존재할 시에는 사용자가 직접 수정하여 서버의 데이터베이스에 등록한다. 그림 3에 인식 예를 도시하였다.

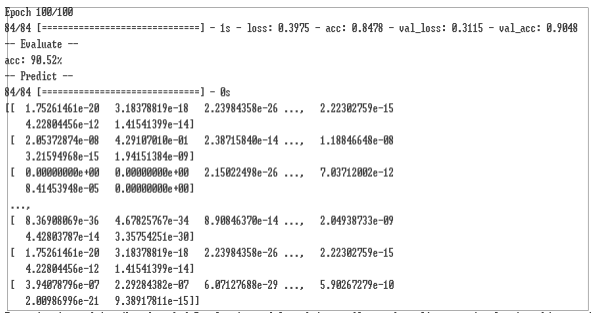

그림 3. 시스템 결과 창

최초 학습 후 검증 데이터의 인식률이 90.52%였다. 그러나 인식률 상승이 학습 과정 전반이 아니라 후반에 집중되는 모습과 별도 데이터 검증 결과 30% 가량의 인식률을 보여서 과적합 발생을 확인하였다. 그 후 Convolution 레이어의 필터 수를 추가하고, DropOut 레이어를 추가한 결과 검증 데이터에 대해서 인식률이 85% 가량으로 감소하였으나 별도 데이터에 대해서도 70% 가량의 인식률을 보여주어 과적합의 감소를 확인할 수 있었다. 이후 연구로는 양질의 학습 데이터 이용 및 모델의 정교화를 통한 인식률 향상을 계획하고 있다.

참 고 문 헌

[1] 잔카를로 자코네, 텐서플로 입문, 에이콘출판, 2016.10

[2] Keras Documentaton, https://keras.io/