이미지 분류 정확도 향상을 위한 텍스트 활용 이미지 분류 모델

이주혁[1], 김미희[1]★

[1]한경국립대학교 컴퓨터응용수학부, 컴퓨터시스템연구소
e-mail:{xpdlwm99, mhkim}@hknu.ac.kr
★교신저자

Image classification model utilizing text to improve image classification accuracy

Ju-Hyeok Lee[1], Mi-Hui Kim[1]★

[1]School of Computer Engineering & Applied Mathematics, Computer System Institute, Hankyong National University

요 약

컴퓨터 비전 문제 중 이미지 분류는 핵심적인 주제 중 하나이다. 딥러닝의 발전으로 이미지 분류 문제에서 높은 정확도와 성능을 보여준다. 하지만 대부분 이미지 분류 연구에서 시각정보인 이미지 내의 특징에만 의존하고 있다. 그렇기에 이미지의 본질적인 맥락과 함께 있는 텍스트 정보를 활용하지 못하는 경우도 있다. 이에 본 논문은 텍스트 정보를 활용하여 이미지 분류 성능을 개선하는 방식을 제안한다.

1. 서론

이미지 분류란, 컴퓨터가 사람처럼 이미지가 주어졌을 때 그 안에 어떤 객체가 있는지 판단하는 기술을 말한다. 컴퓨터 비전 분야에서 다루는 문제들 중에 이미지 분류는 핵심적인 주제 중 하나이다. 딥러닝의 발전으로 인해 이런 이미지 분류 문제에서 높은 정확도와 성능을 보여주고 있다. 예를 들어, 합성곱 신경망(이하 CNN)을 사용하여 이미지 데이터셋의 특징을 추출한다. 추출한 특징과 유사한 특징을 가진 이미지들이 분류에 활용이 된다. 하지만 대부분 연구들이 시각정보에만 의존하고 있다.[1] 그렇기에 이미지의 본질적인 맥락과 함께 있는 텍스트 정보를 활용하지 못하는 경우도 있다.[2]

본 논문에서는 텍스트 정보를 활용하여 이미지 분류 성능을 개선하는 방식을 제안한다. 이미지와 텍스트의 상관관계를 분석하는 딥러닝 모델을 통해 이미지 분류 정확도를 향상시킨다.

이미지와 관련된 설명, 주석등의 텍스트 데이터가 이미지 자체에서는 얻기 힘든 추가적인 맥락과 시맨틱 정보를 제공할 수 있기 때문이다. 제안 방식을 검증 하기 위해 이미지와 텍스트를 함께 훈련하는 분류 모델을 통해 유효성을 판단한다.

2. 배경 지식

2.1 CNN(Convolutional Neural Network)

CNN은 이미지 인식과 처리에 강점을 보이는 신경망 구조이다.[3] CNN은 합성곱 신경망을 통해서 입력된 이미지에서 패턴을 찾아내어 이미지 인식과 처리한다. 패턴을 찾아내기 위해서 인공 신경망 구조에서 합성곱 계층과 플링 계층 등 여러 계층을 포함한다.

본 논문에서는 이미지 분류 모델의 큰 축이 되어 CNN 구조를 활용한다. CNN 구조를 이용하여 이미지 패턴을 분석하고 LSTM을 통해 얻은 정보와 신경망 구조에서 결합을 통해 최종 이미지 분류에 활용한다.

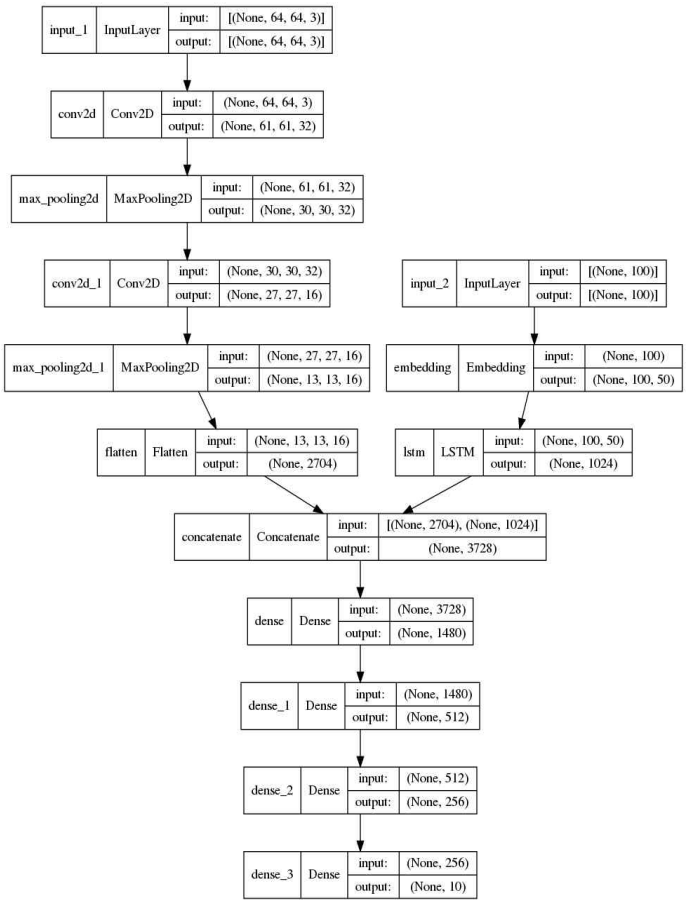
2.2 LSTM(Long Short-Term Memory)

LSTM은 순환 신경망의 한 종류로 RNN의 장기 의존성 문제를 해결하기 위해 설계되었다. 이를 통해서 시퀀스 데이터의 긴 범위 종속성을 모델링하는데 탁월한 성능을 보인다. 긴 범위 종속성을 통해 자연어 처리와 음성 인식 등 다양한 분야에서 활용

한다.[4]

본 논문에서는 텍스트 정보를 활용하기 위해, 이미지와 관련된 설명, 주석 등의 텍스트 데이터를 어떠 정보를 유지하고 버릴지 학습한다. 학습된 정보들은 CNN과 함께 결합하여 최종적으로 이미지 부류에 활용한다.

3. 제안 방식

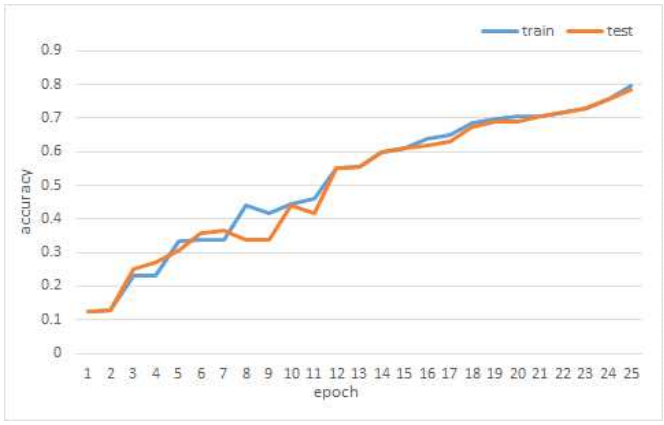


(그림 1) 이미지 분류 모델.

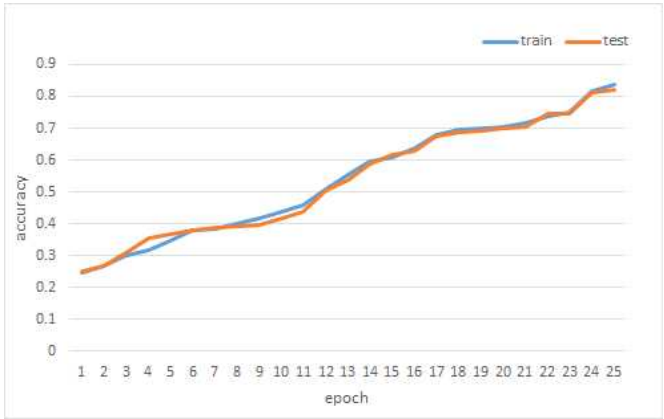
(그림 1)은 본 논문에서 제안 방식의 이미지 분류 모델이다. 입력값으로는 이미지와 텍스트가 동시입력된다. 이미지는 (그림 1)의 input_1로 입력되어 CNN의 구조를 거쳐 이미지의 패턴을 파악한다. 텍스트는 (그림 1)의 input_2로 입력되어 이미지에 대하 설명 중 어떤 데이터를 유지하고 버릴지 학습한다.

(그림 1)의 concatenate층에서 결합을 통해 이미지 데이터와 텍스트 데이터를 결합하여 은닉층의 입력값으로 사용한다. 마지막으로 인공 신경망 구조를 통해 이미지 데이터와 텍스트 데이터를 함께 결합하여 이미지 내의 특징을 학습한다.

4. 실험 결과



(그림 2) 이미지만을 사용한 이미지 분류 모델 결과.



(그림 3) 제안 방식 이미지 분류 모델 결과.

실험에서 사용한 데이터는 COYO dataset[5] 중 랜덤으로 선택한 이미지와 텍스트 데이터 쌍 2,000개를 입력값으로 사용했다. train과 test 데이터는 8대2 비율로 나누어 사용했다.

(그림 2)는 입력값으로 이미지 데이터만을 사용한 이미지 분류 모델 결과가 그래프이다. 정확도는 약 0.7954이다.

(그림 3)는 제안 방식 이미지 분류 모델 결과를 보여주는 그래프다. 모델 정확도 결과 약 0.8356의 값을 얻어 약 5% 상승한 것을 확인할 수 있다.

5. 결론

본 논문에서는 이미지 분류 정확도 향상을 위한 텍스트 활용 이미지 분류 모델을 제안했다. 제안 모델을 통해 이미지 내의 특징 혹은 패턴에 대한 정보를 텍스트를 통해서 추가 학습을 할 수 있게 되었다. 이를 통해 이미지 분류 모델의 정확도 향상을 위해 텍스트 데이터가 효과적임을 확인했다. 향후 연구에서는 이미지 분류 모델의 구조를 강건화를 통해 정확도를 높이고 다양한 기법들과의 성능 비교

평가를 수행하고자 한다.

이미지와 텍스트 간의 상관관계를 분석하여 이미지 설명 생성 기능과 주어진 텍스트 설명에 가장 잘 맞는 이미지를 찾을 수 있는 이미지 검색에 대한 실험도 수행하고자 한다.

참고문헌

[1] Gang Z ,Tao Lei ,Yi Cui, Ping Jiang, "A dual-path and lightweight convolutional neural network for high-resolution aerial image segmentation." ISPRS International Journal of Geo-Information vol . 8.no.12 pp.582, 2019.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding",Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp.4171-4186, 2019.

[3] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "mageNet Classification with Deep Convolutional Neural Networks", Advances in Neural Information Processing Systems 25 (NIPS 2012), 2012.

[4] S. Hochreiter and J. Schmidhuber, "LONG SHORT-TERM MEMORY", Neural Computation, vol. 9, no.8, pp. 1735-1780, 1997.

[5] COYO-700M: Image-Text Pair Dataset "COYO dataset", [Internet], https://github.com/kakaobrain/coyo-dataset