

EfficientNet 모델과 전이학습을 이용한 상품 이미지와 텍스트 데이터의 결합

임수빈¹, 김범윤¹, 김선재¹, 한정우¹, 유동영²

¹홍익대학교 컴퓨터정보통신공학과 학부생

²홍익대학교 소프트웨어융합학과 교수

robin980108@naver.com, kimbumyun98@naver.com, b884013@g.hongik.ac.kr, hjw6191.wh@gmail.com, ydy@hongik.ac.kr

Combination of product image and text data" using EfficientNet model and transfer learning

Soo-Bin IM¹, Bum-Yun Kim¹, Sun Jae KIM¹, Jeong-Woo HAN¹, Dong-Young Yoo¹

¹Department of Software and Communications Engineering, Hongik University

요 약

본 논문에서는 이미지 데이터와 각종 텍스트 기반의 데이터를 적절히 결합하여 유용한 데이터를 만들어 내는 방법을 제안한다. 그 사례로 편의점 상품 이미지와 편의점 프로모션 데이터, 사용자 위치정보 데이터를 적절히 결합하여 사용자가 편의점 상품 전면 이미지를 제공했을 때, 해당 상품이 어떤 편의점 브랜드에서 어떤 프로모션을 진행하고 있는지, 그리고 현재 위치에서 가까운 점포가 어디인지를 사용자에게 제공하는 시스템을 구현한다. 이미지를 어떤 데이터와 결합하는지에 따라 다양한 요구사항에 대응할 수 있다.

1. 서론

이미지 데이터와 텍스트 데이터는 결합 방법에 따라 더욱 유용한 데이터가 되기도, 무용한 데이터가 되기도 한다. 그러나 대부분의 시스템에서 이미지는 그저 텍스트 데이터의 부가적인 추가 데이터의 역할만 할 뿐, 이미지 데이터와 텍스트 데이터를 결합했다고 보기 어렵다. 따라서 본 논문에서는 이미지 데이터와 텍스트 기반의 데이터를 결합하여 유용한 데이터를 생성하는 방법을 제안한다. 또한, 해당 방법을 실험하기 위해 편의점 상품 이미지와 편의점 프로모션 데이터를 결합하여 유용한 데이터를 생성하는 시스템을 제안한다.

2. 관련 연구

전이학습은 대량의 데이터셋으로 사전훈련된 모델(pre-trained model)을 사용하여 적은 양의 데이터셋으로도 작업 성능을 향상시키는 머신러닝 기술로 동일한 조건에서 바닥에서 훈련시키는(train from the scratch) 모델보다 2~3 배의 빠른 속도로 학습이 가능하다.[2]

VGG16, Xception, ResNet50, MobileNet 은 정확도가 높다고 알려진 전이학습 모델 4 종이다[3]. 그러나 높은 정확도를 얻기 위해서 VGG16 모델은 클래스당 수천에서 수만개의 이미지가 필요하고 Xception, ResNet50 모델은 수백~수천, MobileNet 모델은 수십~수백개의 이미지가 필요하다[3]. 위와 같은 이유로 훈련시킬 이미지 데이터의 양이 한계가 있는 경우, 별도의 모델선정이 필요하다.

ImageNet 데이터를 사전 학습한 EfficientNet 은 기존의 SOTA 모델에 비해 훨씬 적은 파라미터 수를 가지고 더욱 높은 정확도를 보여주는 모델이다.

기존 모델들은 성능을 높이기 위해 width, depth, resolution 중 하나에 집중하는 스케일링 방법을 사용했으나[1] EfficientNet 은 세 가지를 복합적으로 고려하며 기존 모델의 한계점을 뛰어넘었다.

3. 제안하는 방법

EfficientNet 모델과 전이학습을 이용해 이미지 데이터와 텍스트 데이터를 결합하는 방법을 제안한다.

전이학습 시 사전훈련(pre-trained)된 모델의 가중치부터 학습을 시작하여 파라미터를 미세 조정하는 방법을 적용한다. 기존의 ImageNet 이미지 분류 모

델로 학습된 EfficientNet 모델을 상품 이미지 분류 문제에 적합하도록 Drop Out Layer와 softmax 활성화 함수를 사용하며 L2 정규화를 진행하는 Output Layer를 추가한다. 이를 통해 적은 양의 훈련데이터, 검증데이터로 학습할 시 발생할 수 있는 과적합 문제를 해결한다.

4. 사례 시스템 설계

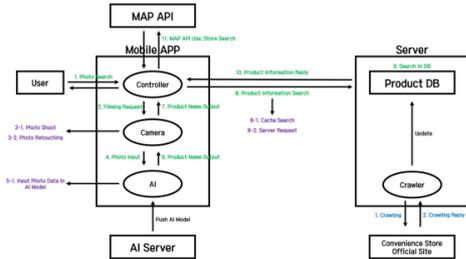


그림 1. 시스템 구성도

앞서 제안한 방법을 적용한 사례로 상품 정면 이미지를 스캔하여 상품을 식별하고, 해당 상품의 정보를 제공하는 시스템의 설계를 제안한다. 그림 1은 제안하는 시스템의 설계이다. 본 시스템은 서버와 클라이언트 어플리케이션으로 나뉜다. 서버에는 상품의 프로모션 정보를 수집하는 크롤러와 해당 정보를 저장하는 데이터베이스(DB)가 포함된다. 클라이언트 어플리케이션은 상품을 스캔하는 카메라, 스캔된 상품을 식별하는 인공지능 모델, 상품 데이터 Cache DB, 그리고 서버 및 지도 API와의 통신과 유저와 상호작용하는 컨트롤러가 포함된다.

크롤러는 각 편의점 공식 홈페이지를 selenium과 chromedriver를 이용해 크롤링한다. 이를 통해 상품의 실시간 프로모션 정보를 수집한다. 수집된 상품 프로모션 정보는 DB에 저장된다. 저장된 상품 정보를 안드로이드 어플리케이션에 전달하기 위해 REST API를 구축한다.

어플리케이션을 통한 상품 인식 과정은 다음과 같다. 우선 사용자가 카메라를 통해 상품을 스캔한다. 카메라 모듈에서는 스캔된 상품을 캡처하여 인공지능 모듈에 전달한다. 인공지능 모듈은 상품을 식별하여 상품명을 컨트롤러에 전달한다. 컨트롤러는 상품명을 이용해 어플리케이션 내부의 Cache DB에서 상품 정보를 검색한다. Cache miss시 상품명은 서버에 전송한다. 서버는 상품명을 통해 DB에서 상품 정보를 검색하고 그 결과를 어플리케이션에 상품 정보를 전송한다. 컨트롤러는 상품 정보를 사용자에게 제공한다. 추가적으로 컨트롤러에서는 사용자에게 오픈 지도 API를 통해 사용자 위치 기준 가까운 편의점 점포를 검색하여 그 결과를 제공한다.

카메라 모듈을 통해 스캔된 상품을 인식하기 위한 모델은 별도의 고성능 컴퓨팅 환경에서 EfficientNet 모델에 대해 상품의 정면 이미지를 전이 학습한 결과를 어플리케이션에 이식한다.

5. 사례 시스템과 기존 시스템의 비교분석

텍스트가 아닌 데이터와 텍스트를 결합해 유용한 데이터를 제공하는 기존 사례로, 식품의약품안전처에서 제공하는 ‘내손안 식품안전정보’ 어플리케이션이 있다. 이 시스템은 식품 포장지에 있는 바코드 데이터를 결합하여 유용한 데이터를 제공한다.

표 1. 내손안 식품안전정보와 비교

	제안하는 시스템	내손안 식품안전정보
결합 대상	상품 전면 이미지 데이터와 프로모션 데이터	식품 바코드 데이터와 식품 영양정보

텍스트 데이터에 추가적인 추가 정보로서 이미지 데이터를 사용하는 기존 시스템으로는 ‘난네판’이 있다. 상품명 검색의 결과로서 프로모션 데이터를 제공하고 추가적으로 상품 이미지를 제공한다.

표 2. 난네판과 비교

	제안하는 시스템	난네판
이미지의 역할	프로모션 데이터와 결합하여 유용한 데이터를 생성함	프로모션 데이터를 부가적으로 설명하는 데이터로 활용함

6. 결론

본 논문에선, 이미지 데이터와 텍스트 데이터를 적절히 결합하여 유용한 데이터를 생성하는 방법을 제안했다. 이 방법을 적용한 사례로 편의점 상품 전면 이미지와 프로모션 데이터를 결합해 유용한 데이터를 제공하는 시스템을 설계했다. 이 외에도 어떤 데이터를 결합하는지에 따라 다양한 시스템을 설계할 수 있다. 가령, 의류 이미지 데이터와 의류의 가격과 재질 등의 데이터를 결합하여 비슷한 분위기를 연출할 수 있는 다른 의류를 추천하는 시스템을 설계하거나, 개인간 중고거래 중개 플랫폼에서 물품 데이터와 물품의 이미지 데이터를 결합하여 물품 이미지로 거래대상을 추천하는 시스템을 설계할 수 있다. 이 외에도 여러 시스템을 설계할 때 본 논문에서 제안하는 방법을 활용할 수 있을 것으로 기대된다.

참고문헌

[1] Tan, Mingxing and Le, Quoc V "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks" International Conference on Machine Learning, 2019, 6105-6114
 [2] Kaiming He Ross Girshick Piotr Dollar "Rethinking ImageNet Pre-training", 2018
 [3] Kornblith Simon and Shlens Jonathon and Le, Quoc V "Do Better ImageNet Models Transfer Better?" International Conference on Computer Vision, 2019