이미지 및 텍스트 정보를 이용하여 딥러닝 모델기반 쇼핑몰 상품 카테고리 분류

명정연

고려대학교 컴퓨터정보통신대학원

audwjd@korea.ac.kr

Classify products in shopping malls based on deep learning models using image and text information

Jeongyeon Myeong

Korea University Graduate School of Computer & Information Technology

요 약

본 논문에서는 인터넷 쇼핑몰의 상품의 이미지와 텍스트 정보를 이용해서 상품을 해당 카테고리로 분류하는 방법을 제안한다. 기존에 주로 한 종류의 데이터만 이용하여 카테고리를 분류하는 모델의 유형에서 벗어나 이미지 및 텍스트 정보를 모두 활용하여 카테고리를 분류하는 모델을 구축하고자 한다. 이미지 데이터는 합성곱 신경망(Convolution neural network), 텍스트 데이터는 장단기메모리(Long Short-Term Memory)로 딥러닝 모델을 적용했다. 초기에는 학습 데이터가 부족한 원인으로 모델의 정확도가 낮았으며, 이미지 데이터에 사전에 학습한 값을 초기설정으로 사용하는 전이학습(Transfer learning)을 적용하여 92% 정확도로 이전보다 높은 성능을 갖는 모델을 구성했다. 방대한 상품정보를 짧은 시간에 정확한 분류를 진행할 수 있는 모델로 업무 환경에 긍정적인 효과를 가져다 줄 것으로 예상된다.

1. 서 론

온라인을 활용한 '언택트 쇼핑'이 활성화되면서 인터넷 쇼핑몰에서는 백화점, 아웃렛, 해외직구 상품, 산지 직송 농산물, 축산물 등 다양한 상품을 시간과 장소에 구애받지 않고 편리하게 구매할 수 있다. 매경LUXMEN 제118호(2020년 7월) 온택트 전성시대에 따르면 국내 유명한 포털 사이트의 쇼핑몰인 네이버 쇼핑에서 판매되고 있는 등록 상품 수는 8억 개, 매일 700만 개의 새로운 상품이 올라온다고 한다. 상품의 카테고리 종류는 매우 다양하여 매일 실시간으로 쏟아지는 방대한 상품 정보를 사람들의 손수 작업으로 진행하기에는 상당한 비용 및 시간이 필요하다.

본 논문은 상품의 이미지와 텍스트 정보로 새롭게 등록되는 상품을 해당 카테고리로 분류하는 모델 연구를 진행하고자 한다. 해당 모델이 업무에 적용된다면 상품의 대분류 카테고리로 자동 분류할 것이다. 다른 산업 분야에서도 소요되는 시간과 비용이 절약되는 상품 분류 자동화 시스템으로 확장될 것으로 예상된다.

2. 관련 연구

기존 연구들은 주로 이미지 또는 텍스트 정보만을 이용하여 상품 분류[1];[2]를 진행하였다. 한 종류 데이터로 대상의 특성을 분석하여 분류를 진행하기에는 대상을 인식하는데 어려운 특성들로 높은 성능을 얻는 것에 한계가 있다. 예를 들어 '콜링우드 홈세트 14P'라는 상품명은 사진을 보기 전에 텍스트만으로 식기세트라는 것을 인식하기가 어렵다. 본 논문은 상호보완적인 관계로 이미지와 텍스트 두 가지 데이터의 특성을 모두 이용하여 해당 카테고리로 정확하게 분류할 수 있는 방법을 제안하고자 한다.

3. 제안하는 방법

본 연구에서 데이터의 특성에 따라 적용되는 모델이 나누어진다. 이미지 데이터는 합성곱 신경망(Convolution neural network)과 전이학습(Transfer learning) 기반 Resnet18이 적용되며 텍스트 데이터는 장단기메모리(Long Short-Term Memory)로 딥러닝 모델을 이용했다.

이미지 데이터에서 사용하는 모델인 CNN은 이미지의 특징을 filter를 이용하여 추출하며 여러 Convolution layer을 쌓으면서 class label과 연관된 특징을 검출하여 분류할 수 있는 모델이다. 하지만 입력한 학습 데이터의 양이 충분하지 않아 모델의 정확도가 낮아질 위험이 있다. 사전에 대규모의 데이터 셋에서 합성곱 신경망을 미리 학습한 후, 그 값을 초기설정으로 사용하는 전이학습을 적용한 Resnet18 모델을 사용했다. 텍스트 데이터 분류 시, 기본 순환 신경망(RNN)은 초기의 입력

값이 점차 사라지는 gradient vanishing problem이 예상된다. 이와 같은 문제를 해결하기 위해 time-step 사이에 은닉 상태뿐만 아니라 셀 상태도 함께 전달하는 LSTM[3]을 이용하여 입력 값이 유지되도록 고려했다.

4. 실험

4.1 데이터 셋

2020년 6월 11일, 쇼핑몰 11번가의 베스트 200 메뉴에서 데이터를 추출했다. 상품 종류는 바지, 원피스, 가방, 신발, 티셔츠 5개이며 각 카테고리 별로 500개의 데이터, 총 데이터의 수는 2천 개이다. 각 카테고리 별로 학습데이터는 400개, 테스트 데이터는 100개씩으로 나누었다. 데이터 셋은 상품 이미지와 상품을 설명하는 텍스트, 카테고리 명으로 구성했다.

4.2 실험 모델

1)CNN: 이미지의 특징을 추출하고 학습하는 지도학습 방법이다. 본 연구에서는 입력 이미지에서 3개의 input channel에 6개의 output channel, 5*5 커널의 크기로 구성된 filter연산으로 convolution layer을 만들고 max pooling을 수행한다. 5*5 커널 사이즈와 16개의 output channel을 구성한 filter로 convolution과 max pooling 수행한 후 마지막으로 fully connected layer을 구성하여 5개의 label로 분류 작업을 한다.
2)Resnet18: 깊게 neural network를 쌓으면서 발생하는 error 문제를 해결한 모델이다. output layer에서 다음 layer로 들어갈 때 이전의 input을 더해주는 형태이다.
3)LSTM: RNN의 주요 모델 중 하나로 장기의존성 문제를 해결할 수 있는 딥러닝 모델이며 본 연구에서는 embedding layer의 벡터 크기를 LSTM의 input차원으로 넣어 학습하였다.
4)LDA: 구조화되지 않은 방대한 문서 집합에서 어떤 주제가 존재하는지 단어 수 분포를 분석하여 예측하는 확률적인 토픽모델링의 알고리즘이다.

4.3 실험절차

실험은 정보수집 단계, 전처리 단계, 모델링 단계, 성능검증 단계로 나누어진다. 정보수집 단계는 BeautifulSoup이라는 웹 크롤링하는 python 패키지를 이용하여 웹 페이지의 html tag로 데이터를 추출했다.

전처리 단계에서는 이미지는 정규화를 진행하여 벡터로 구성한다. 텍스트는 특수문자를 제거하고 tokenize 수행 후, 빈도수를 기준으로 단어에 대한 인덱스를 부여 및 word embedding 작업을 진행한다.

모델링 단계는 colab이라는 구글 클라우드의 가상서버를 이용하여 python으로 머신러닝 및 딥러닝 모델을 구성했다. 전처리한 embedding을 딥러닝 모델의 입력 값으로 넣고 모델 구성요소의 parameter를 변경하며 학습을 진행한다. 학습한 모델에 테스트 데이터를 넣어 Accuracy와 Loss를 기반으로 성능이 좋은 모델을 선정한다. 상품의 카테고리 분류 시, 이미지와 텍스트 분류 결과 값이 동일한 상품 대상으로 우선적으로 분류가 진행된다. 분류 결과가 다른 경우, 성능이 높은 모델 기준으로 우선 분류한 후 직접 정보를 확인하여 재 분류를 진행한다.

4.4 실험결과

CNN으로 이미지 데이터에 대해 분류를 예측한 결과, 입력 데이터를 20번 반복 학습하여 loss를 줄이려 했지만 Accuracy가 0.07로 성능이 매우 낮았다. 텍스트 데이터를 LSTM 모델을 이용하여 예측한 결과, Accuracy는 거의 0에 가까운 성능이 안 좋은 모델이었다. 텍스트 데이터 훈련 시 아래 그림 1와 같이 학습 데이터의 loss는 epoch이 증가함에 따라 줄어들었다. 테스트 데이터의 loss는 7을 넘어가는 매우 높은 값으로 수렴되지 않고 위아래로 요동치는 모습으로 나타났다. 그래프로부터 모델이 과적합된 상태라는 것을 알 수 있다.

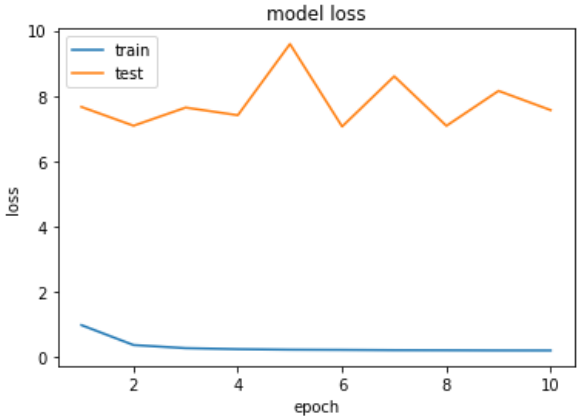

그림 1. 텍스트 데이터의 LSTM 모델 loss 그래프

CNN은 훈련할 많은 데이터를 필요로 하며 적은 데이터의 경우 과적합 문제가 발생될 수 있다[4]. 모델의 정확도가 낮은 원인으로 대상 데이터 셋이 매우 적은 문제로 판단된다. 이미 학습이 어느 정도 진행된 네트워크를 초기값으로 사용하는 전이학습(Transfer learning)은 데이터 셋에서 상당히 표본 수가 적은 경우를 다루기 위한 많은 분야 안에서 인기 있는 방법이다[5]. 이미지 데이터의 경우, 전이학습을 적용한 Resnet18 모델로 훈련한 결과 모델의 Loss: 0.2017, Accuracy: 0.9260로 CNN보다 성능이 향상된 것을 확인할 수 있었다.

표 1. 카테고리 분류 모델 정확도

모델	*Accuracy*
CNN	0.07
Transfer learning	0. 9260
LSTM	0

아래 카테고리가 바지(Pants)인 그림 2의 이미지로 카테고리를 분류할 때, 표 2과 같이 결과가 나타나며 CNN와 Resnet18의 예측 값을 비교할 수 있다.

그림 2. 테스트 데이터의 이미지

표 2. 그림 2에 해당하는 이미지에 대한 각 모델의 예측 결과

모델	*예측 값*			
CNN	Onepiece	Tshirt	Tshirt	Shoes
Resnet18	Pants	Pants	Bags	Pants

4.5 실험 분석 및 보완 방향 제안

본 연구에서는 텍스트 데이터를 학습한 결과, 분류 정확도가 매우 낮았다.

첫 번째 원인으로 텍스트 전처리의 문제로 판단된다. 적용된 모델은 어절단위의 tokenizer을 수행했다. 하나의 예로 상품명 중에 '트로피칼반팔티셔츠' 경우, 띄어쓰기가 잘 안되어 있어 카테고리를 예측하는데 어려움이 있었다. 한국어 형태소 분석기로 tokenizer를 이용하여 ['트로피', '칼', '반', '팔', '티셔츠']로 띄어쓰기가 이루어진다면 '티셔츠'로 카테고리를 쉽게 분류하여 성능이 개선될 수 있다.

두 번째 원인으로는 학습할 때 적은 데이터 셋으로 인한 과적합이 발생된 것으로 판단된다. 이와 같은 경우 한국어 데이터에 대해 사전에 훈련된 모델을 이용하여 분류 모델링을 진행할 필요가 있다.

추후 성능 개선이 되지 않은 경우, 텍스트 분류의 다른 방법으로 토픽모델링을 생각해 본다. 카테고리를 대표하는 핵심 키워드로 카테고리를 분류하는 방법이다. 아래는 표 3은 토픽모델링 방법 중 LDA(Latent Dirichlet Allocation)를 이용하여 텍스트 데이터에서 해당 카테고리의 잠재적인 토픽을 추출하여 토픽 명 상위 5개를 나타내었다. 카테고리 안에서 상품을 대표하는 명칭들이 상위권으로 나타나 해당 카테고리를 잘 나타내는 단어로 보인다. LDA로 카테고리를 대표하는 핵심 키워드를 추출하여 SVM을 이용한 카테고리를 분류하는 모델을 구성하는데 사용될 수 있다.

표 3. LDA로 추출한 카테고리의 토픽 명 TOP 5

	바지	티셔츠	운동화	가방	원피스
1	밴딩	티셔츠	스니커즈	크로스백	원피스
2	여름	반팔	운동화	에코백	갤러리아
3	부	종	여성	숄더백	관
4	슬랙스	남성	남성	토트백	롱
5	팬츠	갤러리아	신발	가방	대구 백화점

5. 결 론

본 논문에서는 한 종류의 데이터로 상품의 카테고리를 분류하는 모델이 업무에서 사용하기에는 한계가 있다고 판단되었다. 데이터 안에서 구분하기 어려운 특성으로 인하여 이미지와 텍스트 데이터 모두 이용한 모델링을 제안했다. 학습을 진행한 결과, 이미지 데이터의 모델은 정확도가 92%가 넘는 높은 성능을 보였지만 텍스트 데이터의 모델은 0%에 가까운 매우 낮은 성능을 보였다. 낮은 성능의 원인으로는 적은 데이터 셋으로 학습되어 과적합이 발생된 것으로 판단되었고 향후 카테고리 분류 모델의 보완이 필요한 상황이다. 우선적으로 데이터를 추가로 확보하여 모델 학습이 이루어져야 하며 앞서 4.5에서 제시한 보완 방향으로 모델 개선이 필요하다.

본 연구 모델은 대규모의 카테고리에 해당하는 상품 분류에 초점을 맞춰 모델링을 진행했다. 향후 세부적인 카테고리 분류 방법을 고안하여 업무에 활용도를 높여야 한다. 그리고 이미지와 텍스트 데이터를 각 모델에서 따로 수행하는 번거로움이 있는데 하나의 통합된 모델로 수행하는 카테고리 자동화 시스템 개발이 필요하다. 아쉬운 부분들이 개선되어 여러 산업 분야에서도 널리 활용되는 긍정적인 효과를 나타내는 모델이 되기를 기대한다.

References

[1] Sung, Jae-Kyung, et al. "Deep learning-based product image classification system and its usability evaluation for the
O2O shopping mall platform." The Journal of The Institute of Internet, Broadcasting and Communication 17.3, 227-234, 2017.

[2] 김진삼. 딥러닝을 이용한 형태소 분석 기반의 상품 카테고리 분류 기법. Diss. 한양대학교, 2017.

[3]S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation Vol.9, No. 8, pp. 1735-1780, 1997.

[4] Lu, Changchong, and Weihai Li. "Ship classification in high-resolution sar images via transfer learning with small training dataset." Sensors 19.1 (2019): 63, 2019

[5] D'souza, Rhett N., Po-Yao Huang, and Fang-Cheng Yeh. "Structural analysis and optimization of convolutional neural networks with a small sample size." Scientific reports 10.1, 1-13, 2020