의류 수요 정보 예측을 위한 멀티모달 기반 딥 뉴럴 네트워크

*김동주[1], 이민식[1,2]
[1]한양대학교 인공지능융합학과, [2]한양대학교 전자공학과
e-mail: willkim4927@hanyang.*ac.kr*, mleepaper@hanyang.*ac.kr*

A multi-modal deep neural network for predicting clothing demand

*Dongjoo Kim[1], Minsik Lee[1,2]
[1]Department of Applied Artificial Intelligence,
[2]Department of Electric and Electronic Engineering
Hanyang University

Abstract

This paper proposes a deep neural network to estimate the quantitative demand information of clothes displayed in an online shopping mall, using images and other miscellaneous information.

In this paper, we construct a clothing demand evaluation database with clothing images collected from an online shopping mall, along with product names, prices, genders, preferred genders, preferred age groups, product numbers, cumulative sales, and product categories. Various modalities are combined in the proposed deep neural network to predict the demand information. To process the images and the text information (product names), we employ ResNet18 and the Multilingual BERT, respectively, in the proposed network.

The experimental results show that the proposed multi-modal network shows reliable performance in predicting the demand of clothing.

I. 서론

온라인 쇼핑 시장은 인터넷 쇼핑이 생긴 이래로 계속 성장해 왔으며, 특히 코로나 19로 인해 사회 전반에 걸쳐 비대면 문화가 확산되며 온라인 시장 규모는 점점 증가하고 있다. 온라인 패션 시장도 예외는 아니다. 이로 인해 많은 패션 브랜드가 시장을 점유하려 했고, 온라인 패션 시장에 자리를 잡았다.

우리는 무신사[1]와 하프클럽, 이랜드 등의 패션 온라인 쇼핑몰을 통해 다양한 의류를 접할 수 있다. 그 중 무신사는 현재 많은 사람들이 대중적으로 이용하는 패션 온라인 쇼핑몰이라 할 수 있다.

소비자들은 무신사에서 매일매일 새로운 의류를 접할 수 있으며, 같은 카테고리 내에서도 다양한 종류의 제품들을 접할 수 있다. 소비자들은 제품의 디자인, 상품명, 남성/여성용, 가격과 같은 특징들에 따라 의류를 구매한다. 따라서 의류를 제작 및 출고하는 브랜드 회사는 효율적인 제품 판매를 위해 제품의 특징에 따른 수요 정보를 정확히 예측하여 위와 같은 특징들을 적절히 관리 및 설정할 필요가 있다.

의류 이미지는 제품의 디자인 정보를 가지고 있고, 디자인이 가지고 있는 시각적인 특징들은 제품에 대한

구매 의사를 결정하는 중요한 요소이다. 상품명 항목은 브랜드, 제품의 주요 정보, 적용되어 있는 이벤트와 광고모델 착용여부 등의 특징을 가지고 있다. 이 특징들 또한 제품에 대한 구매 의사를 결정하는 중요한 요소이다. 따라서 소비자들의 제품 구매 결정에 의류 이미지와 상품명 항목 모두 중요하다고 판단하여 해당 항목을 학습에 사용하기 위해 멀티모달 기법을 도입한다.

본 논문은 임의의 의류 이미지와 여러 정보(상품명, 가격, 남성/여성용, 카테고리)를 바탕으로 제품의 판매량, 조회수, 선호 연령대, 선호 성별을 예측할 수 있는 멀티모달 기반의 딥 러닝 기술을 개발하여 이 문제를 해결한다.

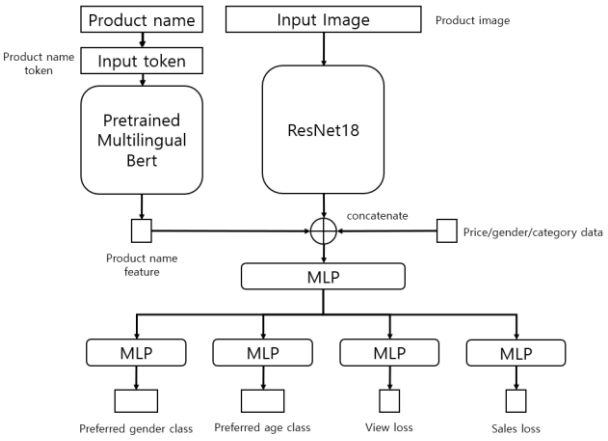

그림 1. 의류 수요 정보 예측 네트워크

II. 본론

2.1 의류 수요 정보 예측 데이터베이스

본 논문은 네트워크 구현을 위해 입력 값과 라벨로 구성된 의류 수요 정보 예측 데이터베이스를 구축한다. 해당 데이터베이스는 무신사의 상품 랭킹 페이지에서 웹 스크래핑을 통해 수집한 이미지와, 상품명, 가격, 남성/여성용, 선호 성별, 선호 연령대, 상품번호, 누적판매량과 16종의 상품 카테고리로 이루어져 있다. 이 중 선호 성별, 선호 연령대, 조회수, 누적판매량을 라벨로 취한다. 데이터베이스는 총 59,998장의 데이터로 구성되어 있으며, 이 중 44,998장은 학습 데이터, 15,000장은 테스트 데이터로 지정한다. 원활한 학습 및 테스트를 위해, 학습 데이터와 테스트 데이터의 클래스 라벨 비율을 비슷하게 설정한다. 데이터베이스로 사용된 무신사 의류 이미지들은 저작권의 문제로 본 논문에 첨부하지 못하였다.

2.2 의류 수요 정보 예측 네트워크

그림 1에서는 제안하는 의류 수요 정보 예측 네트워크를 보여준다. 네트워크는 125 x 125 크기의 RGB 영상을 입력으로 받아, ResNet18[2]과 1개의 완전 연결 계층을 거친 다음, 총 4개의 완전 연결 계층으로 분기하여 의류의 수요 정보를 예측한다. 상품명 데이터는 pretrained Multilingual Bert[3]를 이용하여 feature extraction을 한 뒤, 네트워크의 완전 연결 계층 직전에 상품명 feature, 가격, 성별, 카테고리를 병합한다. ResNet은 네트워크 학습 종료시까지 계속 학습하고, Bert는 feature extraction을 수행한 뒤 추가로 학습하지 않는다. 남성/여성용에 대한 성별 정보를 입력으로 받고 선호 성별을 다시 예측하는 이유는 표 1과 같이 특정 성별을 위해 제품을 생산하였지만 선물용 혹은 제품의 편리성과 같은 여러 가지 이유로 인해 선호 성별이 기존에 설정된 성별과 다른 제품들이 있어 제품의 수요 정보에 대한 정확한 예측을 위해 위와 같이 설정하였다. 네트워크의 과적합 현상을 피하기 위해 완전 연결 계층의 각 단에 dropout을 0.5의 확률로 적용하였고, 성능을 높이기 위해 data augmentation[4] 기법을 사용하였다. 영상의 90도와 180도 회전, 밝기, 대비, 채도 변경, 좌우 반전과 상하 반전 중 랜덤으로 data augmentation을 수행하도록 하였다. 네트워크 학습에 사용한 옵티마이저는 Adam[5] 기법을 사용하였다. 손실 함수는 선호 성별에 대해서는 Cross Entropy loss를 활용하였고, 선호 연령대는 클래스 불균형이 있어 이를 보정하기 위해 Focal loss[6]를 활용하였다. 조회수와 판매량은 아웃라이어와 스케일에 영향을 덜 받게 하기 위해 각각 log를 취해 스케일링 하여 학습하며, MSE loss를 활용하였다. 총 손실 함수는 아래와 같다.

$$L_{total} = CE_{gender} + FL_{age} + \alpha MSE_{view} + \beta MSE_{sales} \quad (1)$$

식 (1)에서 L_{total}은 총 손실 함수이며 CE_{gender}는 선호 성별에 대한 Cross Entropy Loss, FL_{age}는 선호 연령대에 대한 Focal loss이다. 또한 MSE_{view}와 MSE_{sales}는 각각 조회수와 판매량에 대한 MSE loss이다. α와 β는 식 (1)을 효과적으로 학습하기 위해 total loss에 대한 각 항의 효과를 결정하는 하이퍼파라미터이다. 본 논문에서는 각각 0.01로 설정하였다.

learning rate는 0.0001로 설정하였고 weight decay는 0.00001로 설정하였다. 학습과정은 64개의 batch를 가지고 3,000에폭 동안 수행하였다. 본 논문에서 제안하는 방법은 pytorch 라이브러리[7]을

이용하며 python 언어로 작성하였다.

표 1. 남성/여성용과 선호 성별의 데이터 불일치 정도

	일치	불일치
데이터 개수	25,834	34,164

표 2. 의류 수요 정보 예측 네트워크 성능

네트워크	선호성별	선호연령대	조회수	누적판매량
Proposed	84.5%	73.7%	0.092	0.054
No word	83.9%	73.5%	0.09	0.058

Ⅲ. 실험

데이터베이스의 테스트 데이터를 이용하여 학습한 네트워크의 성능을 평가한다. 또한 제안하는 멀티모달 기법이 네트워크에 어떠한 영향을 끼치는지 확인하기 위해 ablation study를 수행한다.

3.1 성능 평가

표 2의 Proposed는 본 논문에서 제안한 멀티모달 기법으로, 테스트 데이터에 대한 실험 결과를 보여준다. 본 논문에서 제안하는 기법을 사용하여 44,998장의 의류 이미지와 상품명, 가격, 성별 등의 추가 정보를 입력으로 받는 네트워크 Proposed는 각각 선호 성별에서 84.5%, 선호 연령대에서 73.7%의 정확도를 보이며, 조회수에서 0.092, 누적판매량에서 0.054의 MSE 오차를 보인다. MSE 오차를 실제 비율로 변환하면 약 1.36, 1.26으로, 의류 이미지와 추가 정보에 대한 특성을 충분히 학습한 것으로 판단할 수 있다.

3.2 Ablation study

표 2는 ablation study의 실험 결과를 보여준다. No word는 word feature를 사용하지 않은 기법이다. Bert에서 추출된 word feature를 사용했을 때 선호 성별에 대한 성능은 약 0.6%, 선호 연령대에 대한 성능이 약 0.2% 향상된 것을 확인할 수 있다. MSE 오차에 해당하는 성능은 조회수에서 약 0.002만큼 하락하고, 누적판매량에서 약 0.004만큼 향상된 것을 확인할 수 있다. 이것을 실제 비율로 변환하면 각각 약 0.005, 0.011이다.

실험 결과는 word feature를 사용하는 제안하는 기법이 조회수를 제외한 나머지 항목에서 더 좋은 성능을 보임을 알 수 있다. 이것은 상품명을 word feature로써 사용하여 학습했을 때 이미지와 같은 하나의 데이터 형태를 학습했을 때보다 더 정확한 네트워크가 되었다고 판단할 수 있다. 조회수 항목에서 제안하는 기법이 낮은 성능을 달성했으나, 해당 지표의 성능 하락에 비해 다른 지표(선호 성별, 선호 연령대, 누적판매량)에서 비교적 더 큰 성능 향상을 이루었기 때문에 제안하는 기법이 더 효과적이라 할 수 있다.

Ⅳ. 결론 및 향후 연구 방향

본 논문에서는 무신사 웹 사이트에서 의류의 이미지와 상품명, 가격, 성별, 선호 성별, 선호 연령대, 상품번호, 누적판매량과 상품 카테고리를 활용하여 의류의 수요 정보를 예측할 수 있는 멀티모달 기반의 네트워크를 구현 및 학습하였다. 실험 결과는 제안하는 네트워크가 의류 이미지와 추가 정보를 바탕으로 무신사 상품에 대한 수요 정보를 적절히 예측했다고 볼 수 있다. 또한, 제안한 멀티모달 기법이 그것을 사용하지 않았을 때보다 성능이 대체로 높아 해당 기법은 효과적이라 할 수 있다. 본 논문보다 더 뛰어난 양질의 데이터베이스를 구축할 수 있다면, 해당 문제에 대한 성능 향상에 도움이 될 것으로 기대한다.

사 사

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임. (No.RS-2022-00155885, 인공지능융합혁신인재양성(한양대학교 ERICA))

참고문헌

[1] https://www.musinsa.com/app/

[2] Kaiming He, XiangYu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition", CVPR, 2016

[3] Jacob Delvin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding", NAACL, 2019

[4] SC. Wong, A. Gatt, V. Stamatescu, and MD. McDonnell, "Understanding data augmentation for classification: when to warp?", DICTA, 2016

[5] diederik P. Kingma, Jimma Ba, "Adam: A Method for Stochastic Optimization", ICLR, 2015

[6] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár, Focal Loss for Dense Object Detection, ICCV, 2017

[7] A. Paszke, and *et. Al.*, "Automatic differentiation in Pytorch", NIPS, 2017