

딥러닝 기반 나이 예측과 자연어처리를 활용한 의류추천 시스템

민선옥, 최영인, 김동균, 이주현
한양대학교, 한양대학교, 한양대학교, 한양대학교

rosuaol@hanyang.ac.kr, yic0704@hanyang.ac.kr, ehdrbstnl@hanyang.ac.kr,
joohyunlee@hanyang.ac.kr

A Clothing Recommendation System Using Deep Learning-Based Age Estimation and Natural Language Processing

Sunok Min, Youngin Choi, Dongkyun Kim, Joohyun Lee
Hanyang Univ., Hanyang Univ., Hanyang Univ., Hanyang Univ.

요약

본 논문은 인공지능 학습과 자연어 처리, 추천 시스템을 활용하여 사용자의 나이에 따른 의류를 추천하는 방법을 제시한다. CNN 기반의 회귀 모델을 사용하여 사용자의 나이를 예측한 뒤, BERT(Bidirectional Encoder Representations from Transformers)를 기반으로 한 자연어 처리 모델로 Women's Clothing E-Commerce Reviews 데이터의 긍정, 부정 문장을 나눠 의류 상품의 평점을 재설정 하였다. 이를 토대로 사용자의 나이에 맞는 의류를 추천하여 쇼핑의 편의성을 높였다.

I. 서론

연구동기

서비스업에서 사용자가 좋아할 만한 상품을 추천하는 것은 사용자의 편의성과 직결된다. 이러한 추천 알고리즘을 딥러닝 모델과 결합하여 적절한 옷을 선택하는데 많은 어려움을 겪는 사람들을 위한 추천 알고리즘을 만들고자 하였다.

문제설정

본연구의 목적은 사용자의 나이대에 맞는 의류를 추천해주는 것이다. CNN(Convolutional Neural Network)기반의 회귀모델에 사용자의 안면 이미지를 입력하여 사용자의 나이대를 예측한다. 나이대를 예측하는 회귀 모델의 시스템은 그림.1 과 같다.



[그림.1]

$$WeightedAverage(WR) = \left(\frac{v}{v+m} \times R\right) + \left(\frac{m}{v+m} \times C\right) \quad (1)$$

추천시스템은 단순추천시스템(Simple Recommender)인 인기차트, 베스트 댓글, 검색시스템을 이용하여 높은 평점 순으로 의류를 추천한다. 그 후 사용자의 안면 이미지로 나이대가 예측되면, Womens Clothing E-Commerce Reviews 데이터를 BERT 기반 자연어처리모델로 리뷰의 긍정, 부정을 판단한 후 새로운 평점을 설정한다. 이때, (1)와 같은 가중평균식을 이용하여 각 아이템의 평점을 재설정한다.

II. 본론

나이예측

나이 예측에는 CNN 기반의 회귀분석(Regression Analysis) 모델을 사용하였다. 나이를 Class label 로 만들어 예측하는 것이 아니라, 연속된 숫자인 나이를 예측하도록 설계하였다. Loss 함수는 회귀에 적합한 MSE(Mean-Square Error)를 사용하였고, 여러 종류의 CNN 이 예측한 값의 평균을 최종 OUTPUT 으로하는 앙상블 모델을 설계하였다. 이때 CNN Model 로 VGG-19, SE-Resnext50 그리고 Inception 모델을 사용하였다.

$$V(aX_1 + bX_2) = E[(aX_1 + bX_2) - (a\mu_1 + b\mu_2)]^2 \quad (2) \\ = a^2Var(X_1) + b^2Var(X_2) + 2abCov(X_1, X_2)$$

위 식은 선형결합 변수의 분산 식이다. 각 변수(X)를 CNN Model 들의 MSE Loss 라고 하면, 여러 인공지능의 예측 값의 평균은 하나의 인공지능이 예측하는 값보다 분산이 줄어들음을 알 수 있다.

리뷰 데이터 자연어처리

Womens Clothing E-Commerce Reviews 에 각 문장별 긍정, 부정을 표기한뒤, 사전 학습된 Google 의 BERT 모델에 문장의 긍정, 부정을 분류하는 Down stream task 를 수행하도록 학습시켰다.

특수문자 제거와 띄어쓰기 교정 등 전처리과 정을 통해 데이터를 정제하였고, BERT 의 긍정, 부정 문장 판별 결과에 따라 기존에 책정되어 있는 리뷰 점수를 아래와 같은 방식을 이용하여 재설정하였다.

$$\frac{\text{긍정문장 개수} - \text{부정문장 개수}}{\text{전체 문장 개수}} + \text{원래 리뷰 평점} = \text{새로운 평점} \quad (3)$$

추천시스템

사용자와 비슷한 나이대에 구매한 아이템 중 리뷰 평점이 높은 순으로 추천하는 시스템을 구현하고자 하였다. 1 명이 5 점을 준 옷보다 100 명이 4 점을 준 옷이 더 좋을 가능성이 높기 때문에 인기도에 기반하여

추천하는 알고리즘인 단순추천 시스템 (Simple Recommender)에 아래 (4)의 가중평균 식을 적용하여 평점을 보정하였다.

$$WeightedAverage(WR) = (\frac{v}{v+m} \times R) + (\frac{m}{v+m} \times C) \quad (4)$$

위 식에서 v 는 각 아이템 리뷰 수를 의미하고, m 은 추천리스트에 등재되기 위한 최소 리뷰 수, R 은 상품 평점, C 는 전체 평균 리뷰 수를 의미한다.

전체적인 추천시스템 과정은 입력 받은 Face image 나이를 예측하여, 리뷰작성자의 나이가 예측 나이 ± 3 이내인 리뷰데이터만 추출한 뒤 추출한 데이터 내의 각 아이템 평균 평점, 리뷰 개수, 리뷰 작성자의 나이 평균을 계산한 새로운 DataFrame 을 생성하여 그 data 를 가중 평균을 통해 평점을 보정하고 평점 높은 순으로 상위 5 개의 아이템을 추천하는 방식이다.

III. 결론

Face image 의 나이를 분석하는 인공지능의 정확도는 (실제나이-예측한 나이)의 절대값이 5 이하인 경우 정답으로 처리하여 계산하였고, 그 결과 약 79%의 정확도를 얻었다.

NLP 모델의 긍정, 부정 분류 정확도는 약 94%의 결과를 얻을 수 있었고, 학습된 모델로 Women e-commerce review 데이터의 리뷰를 긍정, 부정을 판단하여 평점을 재설정하였다.

최종적으로 Face image 의 나이를 예측하여 비슷한 나이대의 리뷰데이터를 추출한 다음, 가중평균을 이용한 추천시스템을 통해 평점이 가장 높은 5 개의 아이템 리스트를 화면에 출력한다.

리뷰 데이터의 자연어처리 유무에 따라 추천리스트의 순위가 바뀌었음을 알 수 있었다. 기존의 리뷰 데이터에 나이와 각 문장의 긍정, 부정을 추가로 고려하여 개선된 추천 알고리즘을 개발하였다.

	Clothing ID	Department Name	Class Name	Rating	Age
1	1099	Dresses	Dresses	4.87	49.9
2	940	Tops	Sweaters	4.759	49.1
3	984	Jackets	Jackets	4.752	49.7
4	1044	Bottoms	Pants	4.75	49.1
5	1025	Bottoms	Jeans	4.745	49.1

[그림.4] 개선 한 추천시스템 결과

참 고 문 헌

- [1] 박상길, BERT 톨아보기, 2018, <http://docs.likejazz.com/bert/>
- [2] 네이버 영화 리뷰 감정 분석, https://colab.research.google.com/drive/1tlf0Ugdqg4qT7gcxi-a3tL7und64Rv1dP#scrollTo=P58qy4--s5_x
- [3] 니컬 부두마, "딥러닝의 정석 (Fundamental of Deep Learning)", 2018
- [4] Vishnu Subramanian, "Deep Learning with Pytorch", 2018
- [5] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

	Clothing ID	Department Name	Class Name	Rating	Age
1	1099	Dresses	Dresses	4.87	49.9
2	940	Tops	Sweaters	4.759	49.1
3	984	Jackets	Jackets	4.752	49.7
4	1044	Bottoms	Pants	4.75	49.1
5	1025	Bottoms	Jeans	4.745	49.1

[그림.2] 추천시스템 실행 결과

	Clothing ID	Department Name	Class Name	Rating	Age
1	1099	Dresses	Dresses	4.629	49.9
2	1023	Bottoms	Jeans	4.513	49.1
3	1044	Bottoms	Pants	4.514	49.1
4	940	Tops	Sweaters	4.512	49.1
5	984	Jackets	Jackets	4.493	49.7

[그림.3] 기존의 추천시스템 결과