

이미지 비유사도 기반의 개인화된 하이브리드 의류 추천 모델

양정원, 백지혜, 김현희
동덕여자대학교 정보통계학과

anna01216@naver.com, jihye9949@naver.com, heekim@dongduk.ac.kr

Personalized Hybrid Outfit Recommendation Based on Image Dissimilarity

Jeong-Won Yang, Ji-Hye Baek, Hyon-Hee Kim
Dept. of Statistics and Information Science, Dongduk Women's University

요 약

기존의 추천시스템은 상품간 혹은 사용자 간의 유사도를 기반으로 작동한다. 하지만 이는 사용자가 유사한 상품 추천 속에 갇히게 되는 필터 버블의 문제와 추천시스템의 고질적인 문제인 데이터 희소성 문제를 피할 수 없게 된다. 따라서 본 연구에서는 사용자의 취향과 체형 정보를 반영하여 사용자의 평점을 예측하는 협업 필터링 기반 딥러닝 추천과 상품간 비유사성을 고려하여 사용자의 평점을 예측하는 내용 기반 추천을 혼합한 하이브리드 추천 모델을 구축하여 기존 추천시스템의 문제점을 해결하였다. 모델의 성능평가를 위해 인터넷 의류 쇼핑몰을 대상으로 유사한 이미지를 활용한 하이브리드 추천 모델과 NDCG 값을 비교하였고 유사도가 낮은 이미지를 활용한 모델이 더 우수한 성능을 보였다. 이는 다른 제품과는 달리 소비자가 의류를 구매할 경우 이미 구매한 상품과 유사한 상품보다는 유사하지 않은 상품을 구매할 가능성이 크다는 것을 보여준다.

1. 서론

추천 시스템은 기본적으로 취향이 유사한 사용자들이 구매한 상품을 추천하거나, 유사한 상품들을 추천하는 등 유사도가 높은 사용자 및 아이템을 활용한다 [1]. 하지만 의류의 경우 이전에 구매한 의류와 매우 비슷한 상품이라면 사용자들이 구매하지 않을 가능성이 더 크다고 판단되므로, 비유사성을 기반으로 한 추천이 필요할 것으로 보인다[2]. 이미지의 비유사도를 이용한 추천 모델은 아직까지 많이 이루어지고 있지 않지만, [3]의 연구에서 추천 시스템에 있어 비유사도를 다루는 것이 중요함을 강조하였다. 비유사도를 고려하게 되면 유사한 제품만을 추천받는 경우 나타나는 필터 버블 문제를 해결할 수 있다.

본 연구에서는 사용자의 취향과 체형정보가 고려된 딥러닝 추천 모델과 상품간 비유사도를 적용한 이미지 추천 모델을 혼합한 하이브리드 추천 모델을 구축하였다. 제안하는 추천 모델은 사용자의 평점 예측을 통해 데이터 희소성(Data Sparsity) 문제를 해결하고 사용자의 구매상품과 유사한 상품이 아닌 참신한 상품 추천으로 필터 버블 문제를 해결하였다. 제안하는 추천 모델의 성능 평가를 위해 딥러닝 추천, 비유사도 기반 이미지 추천, 그리고 하이브리드 추천으로 나누어 NDCG 성능 평가를 실시하였으며, 실험 결과

이미지 비유사도를 기반으로 상품을 추천해주었을 경우 유사도를 기반으로 추천해주었을 때보다 NDCG의 성능이 더 좋은 것을 확인하였다.

본 논문은 다음과 같이 구성된다. 제 2 장에서는 데이터 수집 및 전처리, 제 3 장에서는 추천 모델의 구축 방법을 설명한다. 제 4 장에서는 제안하는 최종 모델의 성능평가 및 추천 결과를 보여주며, 제 5 장에서는 결론을 서술한다.

2. 데이터 수집 및 전처리

인터넷 쇼핑몰 웹사이트 무신사[4]로부터 상의, 바지, 아우터에 해당하는 상품정보 데이터와 이를 구매한 사용자 데이터 총 76,357 개를 웹크롤러를 이용하여 수집하였다.

범주형 데이터의 경우 라벨 인코딩을 통해 숫자형 카테고리 값으로 변환하였고 결측값이나 이상치에 해당하는 값은 제거하였다. 댓글 리뷰의 경우 평점과 함께 고려하기 위해 감성분석을 통해 0 과 1 의 값을 부여하였다. 데이터의 속성으로 사용자의 아이디, 체형정보, 평점, 댓글 리뷰, 상품 이름, 색상, 이미지, 가격, 태그, 가이드(핏, 촉감, 신축성, 시즌)를 사용했다.

3. 하이브리드 추천 모델

3.1 딥러닝 모델

사용자의 취향 및 체형정보와 상품의 상세 정보를 모두 고려할 수 있는 딥러닝 모델을 구축하였다. 뉴럴 협업 필터링을 통해 사용자 벡터(키, 몸무게, 사이즈, 리뷰)와 상품 벡터(핏, 촉감, 신축성, 시즌, 해시태그, 색상)를 Latent 벡터로 축소하였다. 이후 사용자와 상품의 Latent 벡터를 합하여 뉴럴 네트워크에 입력하면 신경망 학습을 통해 사용자 평점 예측 결과를 출력하도록 구성하였다.

3.2 이미지 모델

VGG16 모델을 사용하여 각 이미지로부터 특징을 추출한 후 코사인 유사도 값을 계산하여 이미지 유사도 값으로 사용하였다. 이미지 비유사도 값은 다음 수식(1)을 이용하여 산출하였다.

$$\text{비유사도} = |1 - \text{코사인 유사도}| \quad (1)$$

이후 이미지 비유사도 값을 이용해 각 상품에 대한 사용자의 평점을 예측하였다.

3.3 하이브리드 추천 모델

딥러닝 모델의 평점 예측값과 이미지 모델의 평점 예측값에 0 부터 1 사이의 가중치를 부여하여 계산된 하이브리드 추천 점수와 사용자가 부여한 평점값과의 RMSE 값을 도출하였다. RMSE 값이 가장 작을 때의 가중치를 딥러닝 모델과 이미지 모델에 부여하여 두 모델을 혼합한 최적의 하이브리드 추천 모델을 구축하였다.

4. 성능 평가 및 예측 결과

하이브리드 추천 모델의 성능 평가지표로 추천 순서에 가중치를 두는 NDCG[5]를 사용하였다. 사용자가 상품과 얼마나 관련 있는지를 나타내는 Relevance 값은 사용자가 부여한 평점값으로 지정하였다.

$$NDCG = \frac{\sum_{i=1}^K (rel_i / \log_2(i + 1))}{\sum_{i=1}^K (rel_i^{opt} / \log_2(i + 1))} \quad (2)$$

이후 수식(2)를 통해 NDCG 값을 도출하였다.

모델 \ 품목	상의	바지	아우터
유사도 기반	0.9621	0.9504	0.9474
비유사도 기반	0.9794	0.9730	0.9875

<표 1> 품목별 하이브리드 추천 모델 성능평가 결과

각 품목별 유사도 기반 하이브리드 추천 모델과 비유사도 기반 하이브리드 추천 모델의 NDCG 값은 표 1과 같다. 딥러닝 모델과 비유사도를 이용한 이미지 모델을 혼합한 하이브리드 추천 모델은 모든 품목에서 유사도를 이용한 하이브리드 추천 모델보다 NDCG 값이 높아 성능이 더 좋다는 것을 확인하였다.

각각의 품목에 대하여 하이브리드 추천 모델을 적용한 개인 맞춤 추천 결과의 예는 다음 표 2와 같다.

품목 \ 상품	상의	바지	아우터
구매 상품			
추천 상품 1			
추천 상품 2			

<표 2> 품목별 하이브리드 추천 모델 적용 결과

5. 결론

본 연구에서는 사용자의 취향과 체형정보를 반영하는 딥러닝 모델과 상품의 이미지 비유사도를 반영하는 이미지 모델을 혼합하여 하이브리드 추천 모델을 구축하였다. 실험 결과, 이미지 비유사도를 이용한 하이브리드 추천 모델의 성능이 유사도를 이용한 하이브리드 추천 모델의 성능보다 높았다. 또한 제안하는 추천 모델을 통해 사용자의 평점을 예측함으로써 데이터 희소성 문제를 해결하였다. 결과적으로 사용자는 본인의 취향과 체형정보를 반영하여 개인 맞춤형 추천을 받는 동시에 기존에 구매한 상품과는 유사하지 않은 상품을 추천받을 수 있고, 이는 사용자의 만족도를 향상시킬 수 있다.

참고문헌

- [1] B. Smith and G. Linden, "Two Decades of Recommender Systems at Amazon.com" IEEE Internet Computing, Vol. 21, No. 3, pp. 12-18, 2017.
- [2] 이광한, "패션제품 추천을 위한 협업 딥러닝 시스템 개발" 국내석사학위논문 건국대학교 대학원, 2019.
- [3] Anelli, Vito Walter et al., "The importance of being dissimilar in recommendation.", The 34th ACM/SIGAPP Symposium on Applied Computing, Limassol, Cyprus, 2019, pp. 1-5.
- [4] 무신사, <https://www.musinsa.com/>
- [5] Järvelin, Kalervo and Jaana Kekäläinen, "IR evaluation methods for retrieving highly relevant documents.", ACM SIGIR Forum, Vol.51, No.2, pp. 243-250, 2017.