

콘텐츠 기반 추천을 위한 이미지 특징 추출 방법 비교 연구

김진수[◇], 이상락[●], 채동규[●]

[◇]한양대학교 인공지능학과

[●]한양대학교 컴퓨터소프트웨어학과

{josjkim, sangrak, dongkyu}@hanyang.ac.kr

A Comparative Research on Image Feature Extraction Methods for Content-based Recommendation

Jinsoo Kim[◇], Sang Rak Lee[●], and Dong-Kyu Chae[●]

[◇]Dept. of Artificial Intelligence, Hanyang University

[●]Dept. of Computer Science, Hanyang University

요 약

콘텐츠 기반 추천은 아이템의 특징(feature)을 이용하여 사용자가 과거에 소비했던 아이템과 유사한 아이템을 추천한다. 아이템 간 유사도 산출을 위한 적절한 아이템 정보 선택과 특징 추출은 추천 과정에 있어 중요한 요소로 작용한다. 시각적 정보가 중요한 아이템은 이미지 특징 추출 방법이 추천 성능에 영향을 준다. 이미지로부터 특징을 추출하는 방법에는 JPEG와 같은 알고리즘적 방법, CNN 기반의 딥러닝 모델을 이용한 방법, 인코더-디코더 구조를 갖는 모델의 인코더를 이용한 방법 등이 있다. 본 논문에서는 각각의 특징 추출 방법에 대해 서술하고 이를 비교한다. 또한 실험을 통해 각 이미지 특징 추출 방법에 따른 추천 결과를 확인하고 정성적 평가를 진행한다.

1. 연구 배경

콘텐츠 기반 추천은 아이템의 특징(feature)을 이용한다. 추천 연구 분야에서는 주로 협업 필터링 (콘텐츠가 아닌 사용자-상품 행렬을 기반으로 하는 추천 기법)이 주로 연구되었으며, 콘텐츠 기반 추천에 대한 연구는 상대적으로 활발하지 않았다. 그러나 최근 딥러닝 및 표현 학습의 발전으로 인해 아이템으로부터 양질의 특징 정보를 추출하는 것이 가능해졌다. 이에 따라 콘텐츠 기반 추천의 성능 또한 기존에 비해 월등히 높아질 수 있는 가능성이 생겼다고 볼 수 있다.

본 논문은 이미지를 활용한 콘텐츠 기반 추천에 대해서 다룬다. 아이템의 특징 중 시각적 정보가 중요한 분야로는 패션, 공예품 등이 있다. 이러한 도메인에 속하는 상품들의 이미지 데이터에 대해 효과적으로 사용될 수 있는 몇 가지의 특징 추출 (feature extraction) 방법들을 정리하고, 각각의 장단점에 대해 논의해 보고자 한다. 특히 최근의 딥러닝 기반 생성 모델의 이미지 표현 방법을 소개하고 콘텐츠 기반 추천에서의 사용 가능성에 대해 논의한다. 본 논문에서 논의된 내용들을 통해 관련 연구자들이 효과적으로 자신의 연구에 맞는 방법을 찾고 적용하는 것에 도움을 줄 수 있을 것으로 기대된다.

2. 이미지 특징 추출 방법

2.1 JPEG Coefficient

참고문헌 [1]에서는 이미지 특징 추출 방법으로 JPEG Coefficient를 제안한다. JPEG Coefficient는 JPEG 손실 압축 과정에서 이산 코사인 변환 (discrete cosine transform)을 통해 나온 압축된 이미지 정보이다. JPEG

Coefficient는 이미지 특징 추출을 위한 모델이 아닌 알고리즘 기반의 방법이다. 특히 이미지를 압축해서 표현하는 방식이기 때문에 원본 이미지의 크기를 줄여 추천 과정에서 속도를 높일 수 있다. 그러나 JPEG는 손실 압축 알고리즘이기 때문에 압축 정도에 따라 원본 이미지의 정보가 많이 손실될 수 있다는 단점이 있다. 이는 모델의 정확도에 안좋은 영향으로 이어질 수 있다.

2.2 CNN의 피쳐 맵 (Feature Map)

참고문헌 [2]에서는 이미지로부터 특징 정보를 추출하고 이를 추천에 활용하기 위해 이미지넷 데이터로 사전 훈련된 VGG 모델을 사용하였다. VGG 모델은 여러 합성곱 층 (Convolution layer) 들로 구성되어 있으며, 각 층은 필터 (혹은, 커널) 들의 연산을 통해 피쳐 맵이 생성된다. 위 논문은 가장 마지막에 산출된 피쳐 맵을 이미지의 특징 정보 벡터로써 활용하였다.

피쳐 맵은 이미지의 다양한 특징을 보존하고 있기 때문에 이미지 데이터를 분석하는 과정에서 활용가치가 높다. 하지만 추출된 특징 정보가 사용된 딥러닝 모델이 사전 학습한 데이터나 성능 등에 의존적이라는 한계가 있다. 또한 모델에 따라 피쳐 맵의 크기가 커지면 특징 벡터의 크기가 늘어나게 되고, 이는 추천 모델의 속도 저하를 야기한다. 또한 CNN의 층이 깊어질수록 피쳐 맵을 직관적으로 해석하기 어려운 문제가 있다.

2.3. 스타일 벡터

최근 참고문헌 [3]에서는 이미지로부터 스타일 벡터를 추출하는 방법이 제안되었다. 스타일 벡터는 pSp

(Pixel2Style2Pixel) 프레임워크 하에 있는 인코더를 통해 생성된다. pSp 인코더는 ResNet backbone의 Standard Feature Pyramid를 통해 이미지로부터 피쳐 맵을 추출한다. 작은 매핑 네트워크는 각각에 대응하는 피쳐 맵으로부터 18개의 목표 스타일을 추출하도록 훈련된다. 스타일 벡터의 1번부터 3번 스타일은 작은 피쳐 맵으로부터 생성되고 이후 스타일도 마찬가지로 중간 피쳐 맵 (4번부터 7번까지), 가장 큰 피쳐 맵 (8번부터 18번까지) 으로부터 생성된다. 매핑 네트워크인 Map2Style은 작은 fully Convolutional 네트워크로 LeakyReLU 활성화 함수로 이어지는 2-stride Convolution 세트를 사용하여 공간의 크기를 점진적으로 줄여나간다. 이렇게 생성된 512차원의 벡터는 아핀 변환 (Affine transformation) 을 통해 생성자 (Generator) 의 이미지 생성을 위한 입력 값으로 사용된다.

스타일 벡터는 아직 콘텐츠 기반 추천에 활용된 바는 없으나, 기존의 이미지 추출 방법에 비해 몇 가지 장점을 지니고 있기 때문에 추천을 위한 새로운 수단으로써 주목할 가치가 있다. 스타일 벡터의 가장 큰 장점은 벡터가 Coarse, Medium, Fine 세 종류로 구성되어 있어 이미지 데이터의 다양한 특징 보존이 가능하고, 이에 따라 추천 모델 훈련 및 추천 수행 시 보다 다양한 전략을 수립할 수 있다. 여기서 Coarse, Medium, Fine은 각각 $4^2 \sim 8^2$, $16^2 \sim 32^2$, $64^2 \sim 1024^2$ 해상도의 이미지로부터 추출되는 특징 정보를 의미하며 각자 스케일이 다르므로 서로 다른 종류의 특징을 잡는다.

반면 단점으로는 다양한 도메인에 대한 훈련이 힘들다는 문제가 있다. 데이터가 포함하는 도메인 간의 차이가 크다면 알맞은 스타일 벡터를 만들어 내기 어렵다. 또한 많은 양의 데이터와 양질의 backbone 모델 성능에 의존적인 단점이 있다.

3. 비교 실험

본 장에서는 위에서 설명한 3가지 이미지 특징 추출 방법으로 콘텐츠 기반 추천을 수행하고 결과에 대해 논의한다. 전체적인 추천 과정은 다음과 같다. 각 이미지 특징 추출 방법을 통해 특징 벡터를 추출한다. 그 후 어떤 가상의 유저가 시스템에 접속해서 어떠한 상품을 클릭했다고 가정하자. 해당 상품의 이미지가 타겟으로 주어지면, 타겟 이미지 (그림에서 selected 로 표기) 와 추출된 특징 벡터들 간의 유사도를 계산한다. 유사도를 기준으로 상위 5개의 아이템을 선정해서 추천한다. 실험을 위해 Amazon Review Data (2018) [4]의 신발 이미지 데이터를 사용했다.

먼저 JPEG Coefficient에 대해서 두 가지 실험을 진행하였다. 첫 번째 실험은 손실 압축 정도에 따른 추천 결과의 변화이다. [그림 1]에서 볼 수 있듯이 전체적으로 큰 차이는 없었으며, 60% 품질에서 약간의 변화가 있었다. 다음으로는 95% 이미지 품질에 대해서 더 많은 타겟 상품으로 추천을 수행하였다. 결과는 [그림 2]에서 확인할 수 있으며, 추천 결과를 살펴보면 랭킹이 높은 구간에서는 비교적 비슷한 상품을 잘 추천하지만 낮은 구간에서는 다소 유사성이 없어 보이는 추천 결과를 보여주

기도 한다. 또한 같은 신발이라도 그 종류 (운동화, 구두, 샌들, 등산화 등) 에 따라 추천 성능과 편차가 존재하는 것을 확인하였다.



[그림 1] JPEG 60/80/95% 품질별 추천 결과.



[그림 2] JPEG 품질 95% Top-5 추천 결과.

다음으로 VGG모델을 활용한 피쳐 맵 기반의 추천에 대한 실험을 수행하였다. 실험 결과는 [그림 3]에 있다. JPEG를 이용한 추천에 비해 대체로 더 합리적인 추천 결과를 확인할 수 있었다. 특히 신발 항목 내에 존재하는 소분류에 대해서도 알맞은 추천 결과가 나온 것을 확인할 수 있었다.

마지막 실험으로 스타일 벡터를 이용하여 두 가지의 추천 실험을 진행하였다. 첫 번째 실험은 스타일 벡터의 각 스타일 별 (Coarse, Medium, Fine) Top-5 추천을 수행하였다. [그림 4] 를 통해 스타일마다 서로 상이한 추천 결과가 나오는 것을 확인하였다. 특히 Fine 스타일에서 같은 신발이지만 방향이 서로 반대인 신발 이미지 간에도 유사도가 높게 나왔다는 점에서, 모델이 방향에 관계 없이 신발 자체의 특징에 잘 집중하는 모습을 보였다.



[그림 3] VGG-16 피쳐 맵 Top-5 추천 결과



[그림 4] 스타일 벡터 종류 별 Top-5 추천 결과

두 번째 실험은 모든 스타일을 고려하여 추천을 수행하였다. 결과는 [그림 5]에서 확인할 수 있으며, 한 종류의 스타일이 지배적인 경우 해당 스타일 내에서 추천이 되는 경향과 모든 스타일을 고려하여 총합이 보다 유사한 아이템이 추천되는 경향을 모두 확인할 수 있었다. 예를 들면, [그림 5] 첫 번째 행의 추천 결과는 한 종류의 스타일이 지배적인 경우를 보여준다. 회색 운동화를 보면 Medium 스타일 벡터의 유사도가 다른 스타일 벡터에 비해 높기 때문에 동일한 가중치를 부여했을 때, 모든 스타일을 고려한다고 해도 Medium 스타일 벡터의 유사도가 높은 아이템이 최종 추천 후보로 선택된다. [그림 5] 마지막 행의 추천 결과는 모든 스타일 벡터 유사도의 총합이 높은 아이템이 추천되는 결과를 보여준다. 한 종류의 스타일 벡터만을 고려했을 때 추천된 아이템 목록이 아닌 새로운 아이템이 추천된 것을 확인할 수 있다. 이를 응용한다면 스타일 별 가중치를 달리하여 추천 결과를 더 다양하게 만들거나 개별 사용자가 원하는 추천 결과 위주로 화면에 표시되도록 만들 수 있을 것으로 기대된다.

4. 결론

본 논문은 콘텐츠 기반 추천에 사용될 수 있는 여러 가지 이미지 특징 추출 방법에 대해 살펴보았다. 뿐만 아니라 실제 데이터를 이용해서 각 방법을 적용해보고 추천 결과를 도출해서 비교 분석 해보았다. JPEG Coefficient 방법은 경우에 따라 주어진 타겟 이미지와 관련이 없는 추천결과를 제공하기도 하였다. CNN기반 모델의 피쳐 맵을 이용한 방법은 이미지 특징 보존에 이점이 있고 그에 따라 좋은 추천 결과를 보여주었다. 마지막으로 스타일 벡터는 다양성이라는 측면에서 더 좋은 추천 결과와 향후 응용 가능성을 보여주었다.



[그림 5] 개별 스타일과 모든 스타일 비교

감사의 글

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 ((1) No.2022-0-00819, (2) No.2020-0-01373,인공지능대학원지원(한양대학교))

참고문헌

- [1] Ullah, Farhan, Bofeng Zhang, and Rehan Ullah Khan. "Image-based service recommendation system: A JPEG-coefficient RFs approach." IEEE Access 8: 3308-3318. 2019.
- [2] Vasudevan, Smrithi et al. "Image-based recommendation engine using VGG model." Advances in Communication and Computational Technology. Springer, Singapore, 257-265. 2021.
- [3] Richardson, Elad et al. "Encoding in style: a stylegan encoder for image-to-image translation." in CVPR, 2021.
- [4] Ni, Jianmo, Jiacheng Li, and Julian McAuley. "Justifying recommendations using distantly-labeled reviews and fine-grained aspects." in EMNLP-IJCNLP, 2019.