Melon Playlist Continuation

CUAI 4기 메로나팀
권송아(소프트웨어학), 김소은(응용통계학), 김태윤(소프트웨어학), 김효민(소프트웨어학), 최은서(소프트웨어학), 홍지호(기계공학)

[요약]
이 프로젝트는 카카오 아레나의 Melon Playlist Continuation 참가를 가정하고 진행하였으며, 프로젝트의 목표는 음악 플레이리스트에 수록된 곡과 태그의 절반 또는 전부가 숨겨져 있을 때, 주어지지 않은 곡들과 태그를 예측하는 것이다. 숨겨진 곡과 태그를 예측할 수 있는 모델을 만든다면, 이를 주어진 플레이리스트에 대해 그 플레이리스트와 어울리는 곡을 추천해주는 용도로 사용할 수 있다. 멜론 플레이리스트 데이터 EDA를 통해 장르, 곡, 태그에 대한 데이터의 특징을 파악하였다. 이후 오토인코더를 사용한 협업 필터링을 기반으로 모델을 설계하여 평가 지표로서 nDCG와 Recall을 사용하였다. 또한 사전 훈련된 모델을 통해 코사인 유사도를 기반으로 플레이리스트 제목과 유사한 태그를 추천하는 시스템을 설계하였다.

1. 서 론

국내 음원 스트리밍 시장에서 독보적인 점유율을 유지하던 '멜론'은 '지니', '플로' 등 국내 음악 서비스는 물론이고 '애플 뮤직', '스포티파이', 유튜브에서 음악 관련 기능만 집중한 '유튜브 뮤직'이 주요 경쟁자로 부상함에 따라 그 지위가 흔들리고 있다.

이처럼 음원 시장에서의 각 플랫폼의 생존 경쟁도 치열하다. 사용자의 취향이나 분위기에 맞는 곡을 제안하는 추천시스템이 중요하게 부각되고 있다. 음원 스트리밍 플랫폼에는 수천만 개가 넘는 곡을 서비스 하는데, 이 수많은 곡들 중 자신의 취향에 맞는 음악을 사용자가 일일이 찾는 것은 시간이 많이 걸리는 작업이다. 따라서 사용자의 취향에 맞는 곡을 최대한 많이 효율적으로 탐색하는 것을 도와주는 시스템을 구축해야 한다.

이에 본 참가팀은 협업 필터링(Collaborative filtering) 기법을 적용하여 플레이리스트의 태그와 곡을 추천하는 모델을 설계하고자 한다. 예측 곡과 예측 태그를 nDCG의 가중평균값으로 점수를 매겨 모델의 성능을 점검하고자 한다. 감추어진 곡과 태그를 정확히 예측하는 모델을 완성하게 된다면, 음원 스트리밍 플랫폼이 사용자의 플레이리스트에 어울리는 곡을 효율적으로 추천하는 것에 도움을 줄 수 있을 것으로 예상된다.

2. 본 론

1) EDA

장르 데이터인 genre_gn_all.json을 통해 총 254개의 장르코드가 존재함을 확인하였다. 곡 별 메타 데이터인 song_meta.json을 통해 대부분의 곡들은 한 개의 대분류 장르와 매핑되어 있고 2014~2019년도에 발매된 곡의 비중이 높았다. 플레이리스트의 태그는 1개부터 11개까지 존재하고, 노래는 1개부터 200개까지 존재한다.

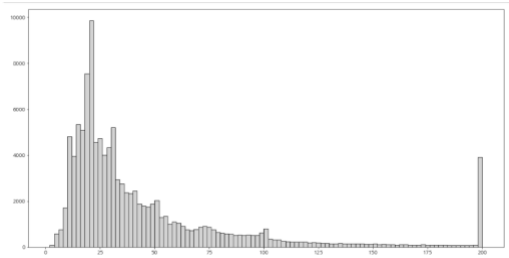


그림 1. 플레이리스트에 수록된 곡

	min	25%	50%	mean	75%	max	std
songs	1.0	19.0	30.0	45.94	54.0	200.0	43.95

표 1. 플레이리스트에 수록된 곡 통계

train.json을 통해 플레이리스트에는 평균 약 46개의 곡이 수록되어 있고 약 4.1개의 태그가 수록되어 있다. 수록곡의 약 51%, 태그의 약 40.2%는 는 두개 이상의 플레이리스트에 중복 수록되어 있다. 플레이리스트 아이디는 다르지만 같은 이름을 갖는 플레이리스트도 약 2% 존재했고 대부분 곡 수와 태그 수도 비슷한 편이었다. 기분전환, 계절, 드라이브 등의 태그가 매핑 기준 상위권에 있었으며 다른 태그와 조합이 없는 장르 관련 태그가 상위권을 차지하고 있었다.

2) 노래 추천

2-1) 데이터 세트

학습에 이용할 수 있는 전체 데이터 세트에는 총 115,071개의 플레이리스트와 707,989개의 노래가 존재했다. 이 중, 10개 미만의 노래가 수록된 플레이리스트는 학습에서 제외시켰다. 그 이후 검증 용도로 30,000개, 테스트 용도로 10,000개의 플레이리스트를 분리시켰다.

2-2) MultiDAE를 이용한 협업필터링 추천 시스템

노래 파트는 협업 필터링을 이용해 플레이리스트에

어울리는 새로운 노래를 예측하였다. 플레이리스트에 수록된 노래들이 비슷하게 수록된 다른 플레이리스트들을 찾아 새로운 노래를 추천하는 방식이다.

데이터의 차원의 수가 크기 때문에 데이터를 잠재공간으로 압축해 학습을 진행하는 오토인코더(AutoEncoder)를 사용해 협업 필터링을 구현했다.

오토인코더 모델에는 DAE(Denoising AutoEncdoer)와 VAE(Variational AutoEncoder)가 있다. DAE는 기본 오토인코더 모델보다 Robust한 결과를 내기 위해 입력 데이터에 노이즈를 추가해 학습을 진행한다. 그리고 VAE는 은닉층이 입력 데이터와 최대한 같은 분포를 가지도록 학습한다. 베이지안 추론(Bayesian Inference)을 이용해 입력 데이터과 가까운 분포를 가지는 은닉층에서 원본 데이터와 가까운 값을 출력하는 것이 VAE의 목표이다.

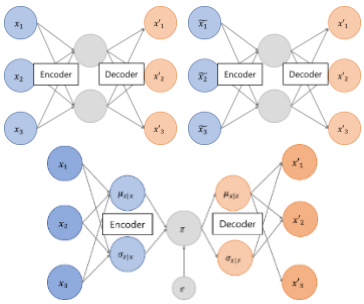


그림 3. AE, DAE, VAE

후보 모델로 MultiDAE(Multinomial DAE)와 MultiVAE(Multinomial VAE) [1]를 고려했다. 추천시스템에서의 ranking loss는 직접적으로 최적화되기 어렵고 보통 근사값을 이용한다. 참고한 논문에서 다항 가능도 함수(Multinomial Likelihood)가 ranking loss를 감소시키는데 적합하다는 것과 implicit한 feedback에 적합하다는 것을 알 수 있었다.

두 모델을 MovieLens-20M 데이터 세트를 이용한 baseline 코드를 돌려보았을 때, MultiVAE가 근소하게 더 높은 성능을 보였지만 연산량이 더 많았다. 우리가 사용한 데이터 세트는 이보다 더 큰 데이터 세트였기 때문에 성능을 조금 포기하는 대신 MultiDAE를 택했다.

입력값으로는 플레이리스트에 들어있는 노래를 1로, 들어있지 않은 노래를 0으로 표현한 행렬을 사용한다. MultiDAE의 layer들의 차원은 [노래 수, 200, 노래 수]이다. 참고한 논문에서 가장 높은 성능을 보였던 구조를 택했다.

학습 과정에서는 인코더에 해당하는 가중치와 디코더에 해당하는 가중치를 학습한다. 각 layer를 통과할 때마다 입력 값과 가중치를 곱한 값에 bias가 더해진다. 활성화 함수로는 tanh를 사용했고 l2 정규화를 적용했다. 또, 가중치는 xavier 초기화를 사용했고, bias는 잘린 정규분포에서 샘플링을 해 초기화를 했다.

DAE의 objective 함수는 아래와 같다. 인코더 네트워크의 결과 값이 주어졌을 때, 입력 값이 나올 확률에 log를 취한 것이다.

$$\mathcal{L}_u(\theta, \phi) = \log p_\theta(\mathbf{x}_u | g_\phi(\mathbf{x}_u))$$

학습을 최적화하기 위해 adam optimizer를 사용했고, Learning rate는 1e-3, regularization rate는 1e-2, epoch 6회로 학습을 진행했다.

2-3) 결과

그림 4. 노래 학습 데이터와 예측값

이별을 담은 플레이리스트를 모델에 넣었을 때, 이별에 관련한 '떠나갈래', '좋을텐데'라는 노래가 담긴 노래를 예측한 것을 볼 수 있었다.

```
Test NDCG@3000=0.00521 (0.00016)
Test Recall@3000=0.01782 (0.00038)

Test NDCG@1000=0.00365 (0.00015)
Test Recall@1000=0.00999 (0.00027)

Test NDCG@500=0.00285 (0.00014)
Test Recall@500=0.00636 (0.00022)

Test NDCG@100=0.00192 (0.00014)
Test Recall@100=0.00275 (0.00016)
```

그림 5. 평가지표 nDCG, Recall

평가지표로는 관련성이 높은 아이템을 순서대로 예측했는지 평가하는 nDCG와, 실제 추천되어야 하는 것이 얼마나 포함되어 있는지 나타내는 지표인 Recall을 사용했다. @뒤의 값은 평가의 대상이 상위 top-N개를 뜻한다.

이 모델은 보통 추천시스템에서 수많은 데이터 중 후보군을 추려내는데 사용하는 모델이다. 따라서 절대적인 수치가 작아 보이지만 의미가 없는 값이 아니다. Recall 3000 값을 보면 1000개 중 17개가 바르게 추천된 것을 볼 수 있는데, 3000개 중에서는 51개가 바르게 추천이 되었음을 의미한다. 큰 데이터 세트에서 후보군을 만든 다음 랭킹 모델에 넣어서 바르게 추천되어야하는 51개의 순서가 앞으로 오도록 학습을 진행하면 더 나은 성능의 추천시스템을 만들 수 있을 것이다.

3) 태그 추천

3-1) 데이터 세트

데이터 전처리는 비슷한 분야의 태그를 합치는 방향으로 진행되었다. 먼저 태그 데이터에서 한글이 아닌 영어, 숫자로 이루어진 데이터를 삭제하였고 이를 바탕으로 태그간 유사도 분석을 진행하였다. 유사도 분석은 FastText[2]에서 사전 훈련된 모델을 통해 이루어졌다. 비슷한 태그끼리 매칭이 이루어 진 후 가장 많은 태그와 매칭이 된 태그를 상위 태그로 선정하였고 상위 태그와 비슷한 태그를 같은 라벨로 라벨링하였다. 그 이후 결측치를 제거 하였고 태그, 플레이리스트, 노래로 이루어진 새로운 데이터프레임으로 바꾸는 작업을 거쳤다.

데이터 전처리를 거쳐 만들어진 111,133개의 플레이리스트와 16,451개의 태그 중 20,000개의 플레이리스트는 학습에 10,000가 테스트에 이용되었다.

3-2) 플레이리스트 제목을 이용한 태그 추천시스템

태그의 경우 텍스트 기반제목 벡터를 이용한 추천

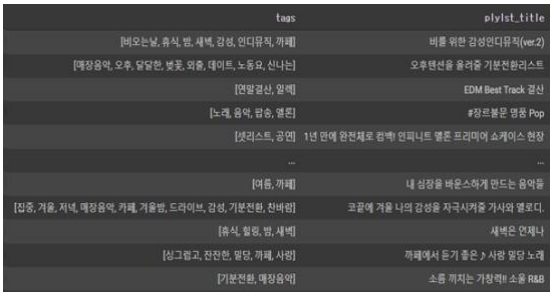
시스템은 플레이리스트 제목과 코사인 유사도가 큰 플레이리스트들을 찾아 새로운 태그를 추천하는 방식이다.

학습 과정에서 우선, 학습 데이터 세트 중 학습에 의미가 없는 플레이리스트는 제거한다. 플레이리스트 이름이 영어로만 구성되어 있거나, 토크닝 결과 의미가 없는 단어만을 가진 플레이리스트가 5,312개가 제거되었다. 그 다음 플레이리스트 이름의 벡터값을 구한다. 플레이리스트 이름의 벡터값은 플레이리스트 이름의 벡터의 평균값이다. 이 값은 플레이리스트 이름을 토크닝한 결과 생성되는 단어들의 벡터값을 모두 더한 후 문서 길이로 나누는 연산의 결과이다.

마지막으로, 코사인 유사도를 이용하여 테스트 세트에 있는 태그를 입력하면, 해당 플레이리스트 제목과 유사한 다른 플레이리스트 중 상위 5개를 찾아(코사인 유사도가 큰 플레이리스트) 테스트 세트가 가진 태그와 중복없이 태그를 추천해 준다.

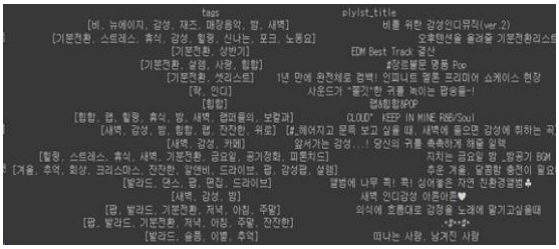
3-3) 결과

특정 태그([그림6])를 입력으로 하면 해당 플레이리스트와 유사한 다른 플레이리스트 5개의 태그들을 취합해 입력 태그의 갯수만큼 추천([그림7])한 것을 볼 수 있다.



tags	plylst_title
[비오는날, 휴식, 밤, 새벽, 감성, 인디뮤직, 까페]	비를 위한 감성인디뮤직(ver.2)
[매장음악, 오후, 달달한, 벚꽃, 외출, 데이트, 노동요, 신나는]	오후텐션을 올려줄 기분전환리스트
[연말결산, 일렉]	EDM Best Track 결산
[노래, 음악, 팝송, 멜론]	#장르불문 명품 Pop
[셋리스트, 공연]	1년 만에 완전체로 컴백! 인피니트 멜론 프리미어 쇼케이스 현장
...	...
[여름, 까페]	내 심장을 바운스하게 만드는 음악들
[집중, 겨울, 저녁, 매장음악, 카페, 겨울밤, 드라이브, 감성, 기분전환, 찬바람]	코끝에 겨울 나의 감성을 자극시켜줄 가사와 멜로디.
[휴식, 힐링, 밤, 새벽]	새벽은 언제나
[싱그럽고, 잔잔한, 달달, 까페, 사랑]	까페에서 듣기 좋은♪ 사랑 달달 노래
[기분전환, 매장음악]	소름 끼치는 가창력!! 소울 R&B

그림 6. 테스트 세트로 주어진 태그-플레이리스트



tags	plylst_title
[비, 뉴에이지, 감성, 재즈, 매장음악, 밤, 새벽]	비를 위한 감성인디뮤직(ver.2)
[기분전환, 스트레스, 휴식, 감성, 힐링, 신나는, 포크, 노동요]	오후텐션을 올려줄 기분전환리스트
[기분전환, 상반기]	EDM Best Track 결산
[기분전환, 설렘, 사랑, 힐링]	#장르불문 명품 Pop
[기분전환, 셋리스트]	1년 만에 완전체로 컴백! 인피니트 멜론 프리미어 쇼케이스 현장
[락, 인디]	사운드가 "꽉 "한 곡를 놓이는 감성들~!
[힐링]	댄스힐링POP
[힐링, 캠, 힐링, 휴식, 밤, 새벽, 캠미들의, 보컬과]	CLOUD" KEEP IN MIME R&B/Soul
[새벽, 감성, 밤, 힐링, 캠, 잔잔한, 위로]	[#,헤어지고 문득 보고 싶을 때, 새벽에 들으면 감성에 취하는 곡
[새벽, 감성, 커페]	앞서가는 감성...! 당신의 귀를 촉촉하게 해줄 일렉
[힐링, 스트레스, 휴식, 새벽, 기분전환, 금요일, 공기정화, 피톤치드]	지치는 금요일 밤 _밤공기 BGM
[겨울, 추억, 회상, 크리스마스, 잔잔한, 달앤비, 드라이브, 팝, 감성팝, 설렘]	추운 겨울, 달콤함 충전이 필요
[발라드, 댄스, 팝, 댄칩, 드라이브]	클럽에 나무 혹! 혹! 싱어송은 자연 친환경댄뱁♣
[새벽, 감성, 밤]	새벽 인디감성 아른아른♥
[팝, 발라드, 기분전환, 저녁, 아침, 주말]	의식에 흐름대로 감정을 노래에 맡기고싶을때
[팝, 발라드, 기분전환, 저녁, 아침, 주말, 잔잔한]	·*·*·*
[발라드, 슬픔, 이별, 추억]	떠나는 사람, 남겨진 사람

그림 7. 플레이리스트별 추천된 태그목록

Recall : 0.0549132179783643

그림 8. 평가지표 Recall 값

평가지표로는 실제 추천되어야 하는 것이 얼마나 포함되어 있는지 나타내는 지표인 Recall을 사용하였는데, 그 이유는 태그가 추천된 순서보다는 특정 태그가 추천되었는지 되지 않았는지가 중요했기 때문이다.

이 모델도 노래 추천과 마찬가지로 수많은 데이터 중 후보군을 추려내는 데 사용하는 모델이다. Recall 값([그림 8])을 보면 1,000개 중 54개가 바르게 추천된 것을 볼 수 있다. 태그 출현 횟수를 단순히 세어 추천하기보다는 제목별 유사도를 태그에 가중치로 적용하여 추천한다면 더 나은 성능의 추천시스템을 만들 수 있을 것이다.

3. 결 론

본 연구에서는 카카오 아레나에서 제공하는 멜론 플레이리스트 데이터를 이용해 플레이리스트의 숨겨진 곡과 태그를 예측하는 모델을 설계하였다. 그 후, nDCG, Recall 지표를 통해 결과의 정확도를 확인하였다. 이를 통해 주어진 플레이리스트에 대해 비슷한 분위기를 가진 곡을 추천함으로써 음원 스트리밍 서비스 이용자로 하여금 자신의 취향에 맞는 노래를 추천 받을 수 있다.

본 연구에서는 협업 필터링 기반의 모델에만 의존하였다. 그러나 추후 오디오 웨이브 데이터를 이용한 콘텐츠 기반 필터링 기법을 함께 사용한다면 음원 스트리밍 서비스 이용자에게 더 정확한 추천 시스템을 제공할 수 있을 것이라 기대한다.

참고 문헌

[1] Dawen Liang, Rahul G. Krishnan, Matthew D. Hoffman, and Tony Jebara, "*Variational autoencoders for collaborative filtering*", The Web Conference (aka WWW), 2018.02
[2] fasttext
https://fasttext.cc/docs/en/crawl-vectors.html
[3] 단어 벡터를 이용한 추천시스템
https://wikidocs.net/book/2155