

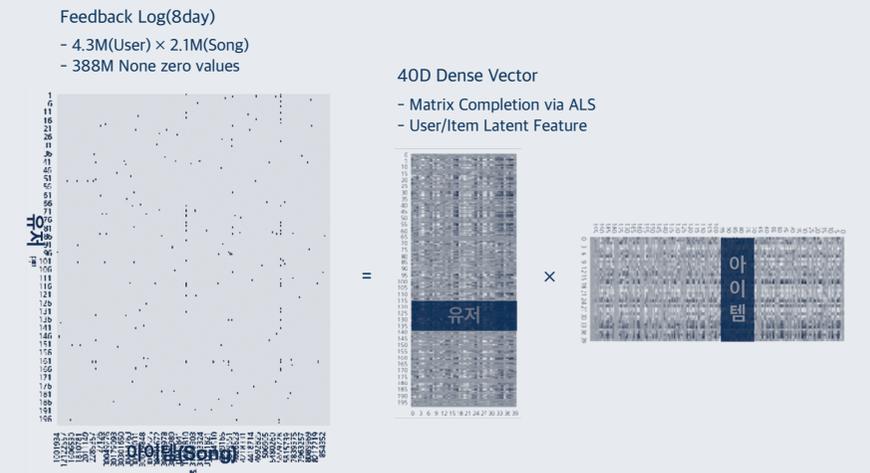
# 눈으로 듣는 음악 추천 시스템

이번 글은 음악 추천 시스템에서 임베딩(embedding)한 음원 벡터(song vector)를 장르와 인기도, 발매 연도의 관점에서 고찰해보는 발표 내용을 담았다. 실제 추천팀에서는 하루 1억 5000만 건의 스트리밍(streaming) 데이터, 5600만 개의 사용자 플레이리스트 데이터, 18만 개의 디제이(DJ) 플레이리스트 데이터를 협업 필터링(collaborative filtering, CF) 방식으로, 3000만 곡의 음원 및 메타 정보는 콘텐츠 기반 접근 방식(contents based filtering, CBF)을 통해 다양한 음원 벡터를 모델링하고 있다.

## 음악 추천 시스템에 담긴 추천 알고리즘: CF와 CBF

[그림 1] 협업 필터링 방법 중 대표적으로 쓰이는 행렬분해(matrix factorization) 방법

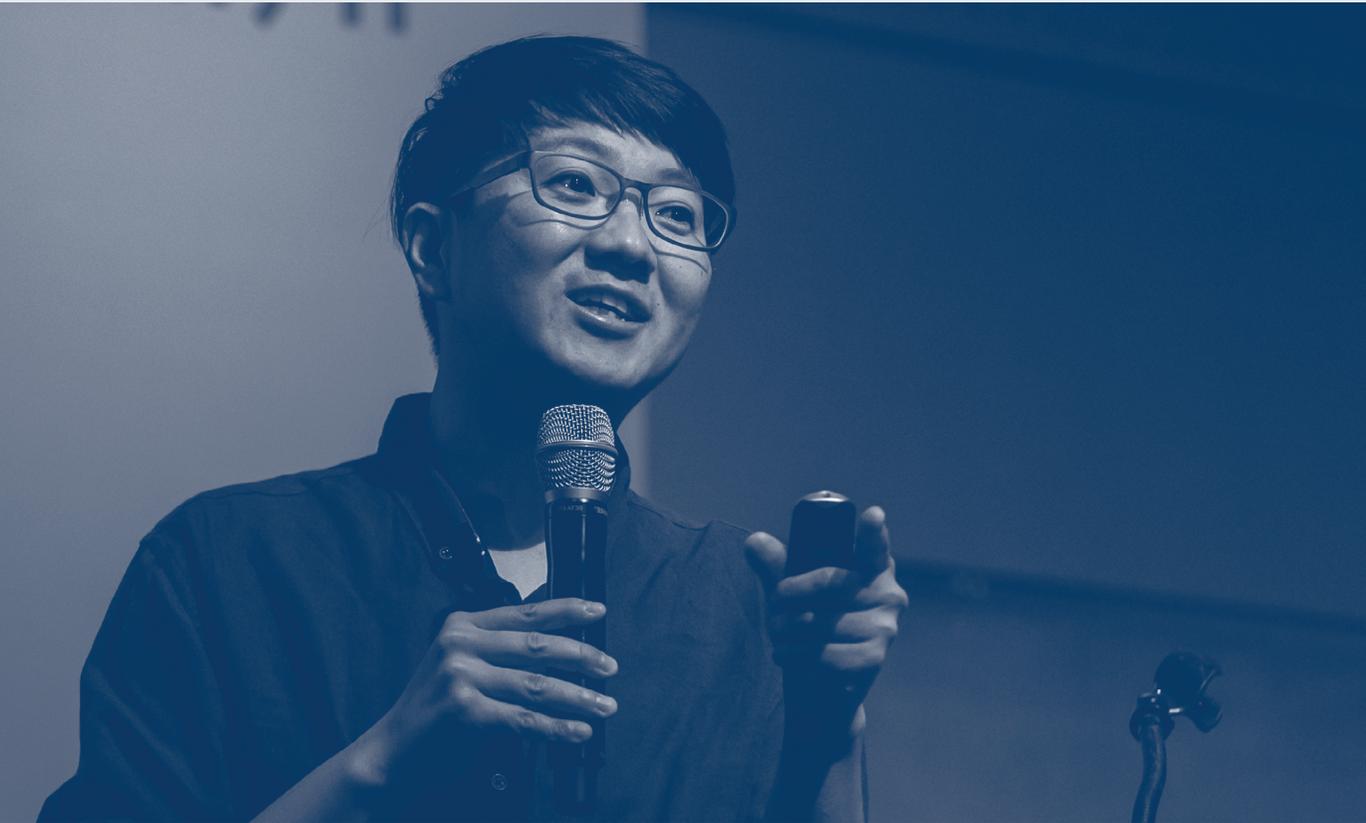
CF의 대표 기술인 Matrix Factorization



[그림 1]과 같이 협업 필터링 방식 중 성능이 뛰어나다고 알려진 행렬분해(matrix factorization, MF)를 이용해 유저(user)와 아이템(song)의 잠재 특징(latent feature)들을 원본보다 밀도 있는 특정 차원의 벡터들로 뽑아낼 수 있는데, 카카오에서는 40차원의 벡터를 사용한다. 또한 계산상의 이점을 얻기 위해 잘 알려진 ALS(alternating least square) 기법을 사용해 매트릭스 컴플리션(matrix completion)을 진행한다. 협업 필터링을 통해 얻은 음원 벡터는 여기서 바로 아이템에 대해 얻은 잠재 특징(latent feature) 벡터이고, 각 벡터들은 사용자들의 소비 특성에 기반해 유사한 속성을 가진 음원들의 벡터 간 거리를 가깝게 표현했을 것이라 기대된다.

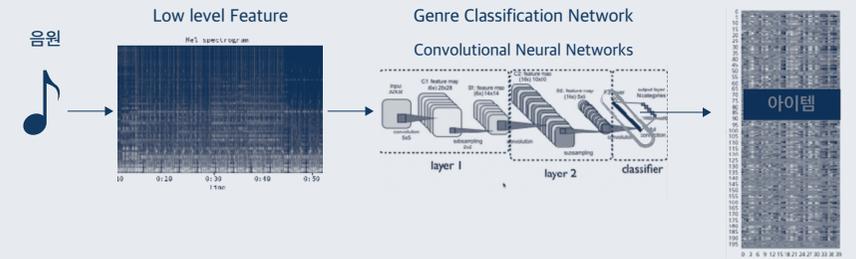
컨퍼런스 발표 | 최규민 pi.314@kakaocorp.com

글 | 김태훈 th.kim@snu.ac.kr 서울대학교 음악오디오연구실에서 음악을 비롯한 콘텐츠 추천 시스템을 연구하고 있습니다. 추천 시스템의 발전은 창작하는 사람과 향유하는 사람 사이의 피드백 선순환을 일으켜 양질의 콘텐츠 생산에 기여할 수 있다고 믿고, 기술로써 그에 조금이나마 보탬이 되기 위해 노력 중입니다.



[그림 2] 콘텐츠 기반 음원 벡터를 뽑는 과정

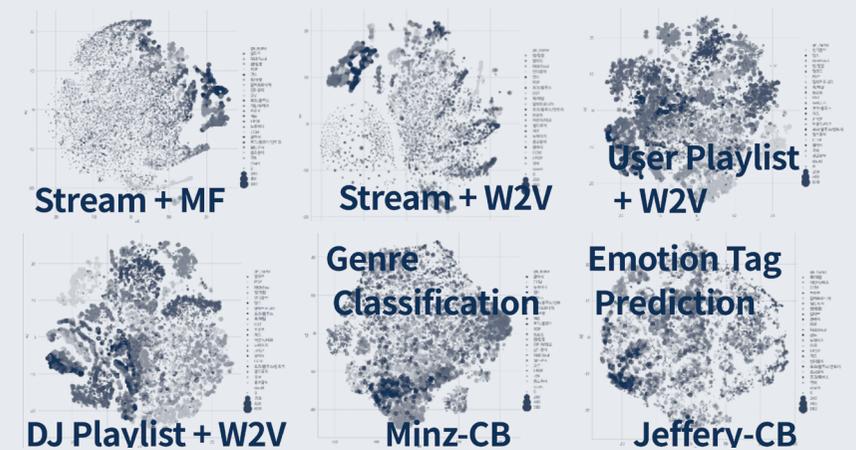
CBF: 장르 Classification Model



음원의 오디오 특성을 이용해 음원 벡터를 구성할 수도 있는데, 카카오에서는 음원의 멜-스펙트로그램(Mel-Spectrogram)과 같은 낮은 수준 특성(low-level feature)을 뽑아 입력으로 사용하고, 가지고 있는 메타 정보를 이용해 분류 태스크(장르 혹은 태그)를 진행하는 딥 뉴럴 네트워크(deep neural network, DNN)를 구성한다. 해당 네트워크로 학습을 시킨 후 특정 레이어(layer)를 가져오면 각 음원의 콘텐츠 기반 특성으로부터 클래스를 분류하는 데 사용된 벡터를 얻을 수 있다. 이것을 콘텐츠 기반 접근을 통해 얻은 음원 벡터로 사용하는 것이다([그림 2] 참조).

음악 추천을 위해 사용하는 음원 벡터와 그 특징

[그림 3] 카카오 추천팀에서 사용하는 6가지 음원 벡터



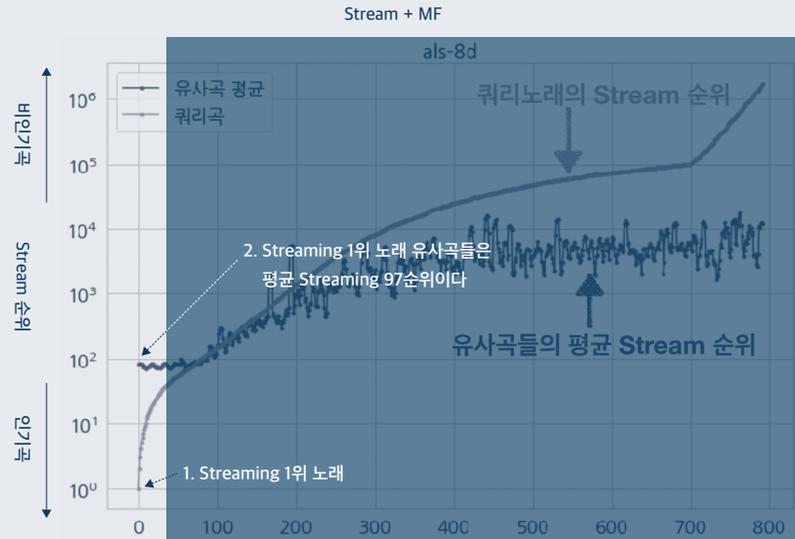
카카오는 크게 6가지의 음원 벡터를 가지고 있다. 각각은 [그림 3]에서 볼 수 있는 것처럼 스트리밍 데이터 + 행렬분해, 스트리밍 데이터 및 사용자가 직접 만든 플레이리스트를 Word2Vec(word to vector, W2V)으로 처리한 것, 전문 DJ의 플레이리스트를 Word2Vec으로 처리한 것, 그리고 장르 분류 모델을 이용한 콘텐츠 기반의 음원 벡터, 감정 태그를 분류하는 모델로부터 얻을 수 있는 음원 벡터이다. 각기 다른 데이터와 방법론을 거쳐 만들어진 음원 벡터이다 보니 특정 쿼리(query) 곡에 대한 유사곡 추천 결과 양상도 조금씩 다르게 나타난다.



쿼리곡 장르와 유사곡으로 추천된 곡들의 장르가 같은 비율은 'Stream + MF'가 45.2%, 'Stream + W2V'가 60.1%, 'Playlist + W2V'가 59.9%로 조금씩 다른 비율을 보여주었다.

[ 그림 6 ] Stream + MF 음원 벡터로 유사곡을 추천한 경우 인기도 분포

유사곡들의 Popularity 분포는?



또한 'Stream + MF' 음원 벡터의 경우는 쿼리곡이 스트리밍 순위로 몇 위의 곡이든 유사곡들의 평균 순위를 내보면 대부분 100위~1만 위 사이 곡들로 추천된다(그림 6) 참조).

[ 그림 7 ] 각 음원 벡터 방법으로 유사곡을 추천한 결과, 추천된 곡들의 평균 인기도 분포 비교

유사곡들의 Popularity 분포 Compare?



이에 반해 'Stream + W2V'와 'Playlist + W2V'의 경우에는 쿼리곡의 순위에 비해 유사곡으로 추천되는 곡들의 평균 순위가 더 낮은 경향을 보여주었다(그림 7) 참조).

[ 그림 8 ] 각 음원 벡터 방법으로 유사곡을 추천한 결과, 추천된 곡들의 평균 발매 연도 분포 비교

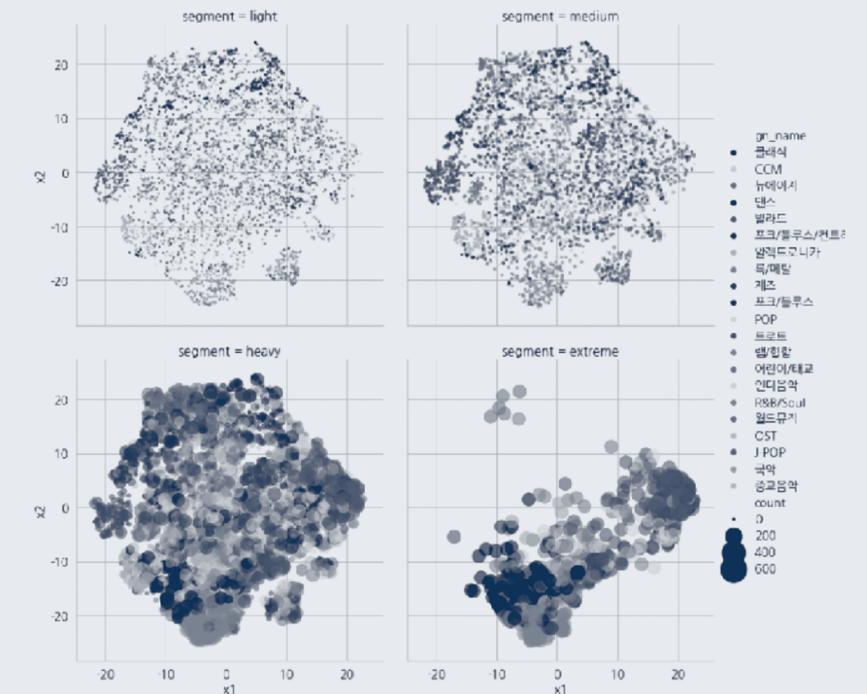
유사곡들의 발매 연도는?



또한 비슷한 방식으로 쿼리곡에 대해 유사곡으로 추천되는 곡들의 평균 발매 연도를 살펴본 경우, [그림 8]과 같이 'Stream + MF'는 쿼리곡보다 최신곡들이 추천되었다면, 'Playlist + W2V'는 쿼리곡과 같은 시대의 곡들이 유사곡으로 추천되었고, 'Stream + W2V'는 그 중간 정도 성격을 나타내었다.

[ 그림 9 ] 콘텐츠 기반 방법론(장르 분류 모델)으로 뽑아낸 음원 벡터의 시각화

Audio + Genre Classification DL



- Song Feature가 장르를 적절히 표현하는 것으로 보임
- 장르 간 글로벌한 특성까지 적절히 표현되어 보임
- Song Popularity와 상관없이 고르게 분포되어 있음

마지막으로 콘텐츠 기반 방법론으로 뽑아낸 음원 벡터의 경우 장르끼리 적절히 잘 군집화되는 특징을 보였으며, 접근 방법의 특색대로 인기도와 상관없이 고르게 분포되어 있는 모습을 보여 앞의 경우들과는 달리 상위 1000곡에 속하는 노래들도 음악적 특성에 의해 군집화되는 경향을 확인할 수 있었다(그림 9) 참조).

또한 장르별로 나누어 그림을 그려보면 OST, 인디음악 등 ‘음악적 특성으로 구분되기보다는 일종의 구분에 의한 장르’의 경우, 다른 장르들의 경우와는 달리 별도의 군집 특성 없이 전체적으로 흩어져서 분포되는 것을 확인할 수 있었으며, 발매 연도를 살펴봐도 딱히 상관성을 발견할 수는 없었다. 앞에서 살펴본 바와 같이 음원 벡터마다 다양한 특성을 갖고 있는데, 실제 음악 추천 환경(멜론, 카카오뮤직)에서 어떠한 음원 벡터가 좋은지는 내년 콘퍼런스 행사를 기약하며 세션이 마무리되었다.

‘눈으로 듣는 음악 추천 시스템’ 세션은 ‘음악’이라는 대상에 접근하는 방법에 따라 ‘유사하다’는 개념이 얼마나 다를 수 있는지를 보여주었다고 생각한다. 인기도에 따른 바이어스(popularity bias)를 피해 유사곡을 추천하려면 콘텐츠 기반으로 접근해볼 수 있다. 같은 스트리밍 데이터를 가지고도 음원 벡터를 만들어내는 방법에 따라 추천된 유사곡들의 평균 인기도나 발매 연도와 같은 특성이 달라졌다. 각 방법들의 장점을 청취 맥락이나 유저의 취향에 맞게 잘 살린다면 보다 다양한 유사곡 추천 성능을 보일 수 있을 것이다.



〈눈으로 듣는 음악 추천 시스템〉 브랜치로 연결되는 QR 코드입니다.