
노래 추천 시스템

김준화 / 손정용

01 데이터 수집 / 전처리

님의 뮤직

홈

좋아요

플레이리스트

DJ플레이리스트

최근 들은

많이 들은

팬맷은 아티스트

친구

곡

앨범

영상

플레이리스트

총 3,901개

나랑 같은 좋아요

최신순 | 좋아요순 | 가나다순

전체 선택

듣기

다운

담기

선물

<input type="checkbox"/>	NO	곡명	아티스트	앨범	좋아요	뮤비	다운
<input type="checkbox"/>	1	Love Love Love (Feat. Yoong Jin Of Casker)	에픽하이 (EPIK HIGH)	Remapping The Human S...	68,229 ▾		
<input type="checkbox"/>	2	첫 사랑니 (Rum Pum Pum Pum)	f(x)	'Pink Tape' f(x) The 2nd Alb...	78,867 ▾		
<input type="checkbox"/>	3	Airplane	f(x)	'Pink Tape' f(x) The 2nd Alb...	37,555 ▾		
<input type="checkbox"/>	4	SUMMERIDE	박재범	SUMMERIDE	17,577 ▾		
<input type="checkbox"/>	5	Lovers On The Moon	AJ Mitchell	SKYVIEW	16,219 ▾		
<input type="checkbox"/>	6	Holiday	Little Mix	Holiday	12,570 ▾		
<input type="checkbox"/>	7	Oops (Feat. Charlie Puth)	Little Mix	Glory Days (Expanded Edit...	58,506 ▾		
<input type="checkbox"/>	8	ATHLETIC GIRL	H1-KEY	Athletic Girl	839 ▾		

유저들이 좋아요 누른
노래 목록을 수집!

01 데이터 수집 / 전처리

멜론 차트에 있는 곡들의 댓글 창에서
유저마다 주어지는 고유한
memberkey 수집



수집한 memberkey 에서 탈퇴한 회원,
좋아요 누른 곡 개수가 10개 이하
1500개 이상인 유저 제거



남은 유저 중에 1000명의
좋아요 목록 수집

01 데이터 수집 / 전처리

	user_id	title	artist	album	quantity
0	38795083	STAY	The Kid LAROI	Stay	1
1	38795083	abcdefu	GAYLE	abcdefu	1
2	38795083	폰서트	10CM	4.0	1
3	38795083	pet	10CM	4.0	1
4	38795083	그라데이션	10CM	5.3	1
...
336623	10566406	세 단어	젝스키스	세 단어	1
336624	10566406	너의 모든 순간	성시경	별에서 온 그대 OST Part.7	1
336625	10566406	공감 (Feat. 서영은)	정준영	공감	1
336626	10566406	기억해 줄래	젝스키스	학원별곡	1
336627	10566406	말하는 대로	처진 달팽이 (유재석 & 이적)	무한도전 서해안 고속도로 가요제	1

336628 rows × 5 columns

좋아요 누른 사람이
2명 이하인 노래 제거

최종 336628개 데이터
유저 1000명, 23546가지 노래

02 Model

ALS (Alternative Least Squares)

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	X		X		X	
User 2		X	X			
User 3				X		X
User 4					X	
User 5	X	X		X		X
User 6			X	X		
User 7	X	X	X		X	X
User 8		X		X		
User 9			X			

R

\approx

	UF1	UF2
User 1		
User 2		
User 3		
User 4		
User 5		
User 6		
User 7		
User 8		
User 9		

U

X

V

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
IF1						
IF2						

- 먼저, U 를 랜덤하게 초기화하고 V 의 해를 구합니다.
- 다시 돌아가 V 의 해를 이용해 U 의 해를 구합니다.
- R 을 최대한 근사시키는 수렴 값을 얻을 때까지 반복합니다.

02 Model

Implicit 라이브러리

Fast Python Collaborative Filtering for Implicit Datasets.

- 암묵적(implicit) dataset을 사용하는 다양한 모델을 굉장히 빠르게 학습할 수 있는 패키지
- implicit의 als(AlternatingLeastSquares) 모델을 사용

```
from implicit.als import AlternatingLeastSquares

# Implicit AlternatingLeastSquares 모델의 선언
# 1. factors : 유저와 아이템의 벡터를 몇 차원으로 할 것인지
# 2. regularization : 과적합을 방지하기 위해 정규화 값을 얼마나 사용할 것인지
# 3. use_gpu : GPU를 사용할 것인지
# 4. iterations : epochs와 같은 의미입니다.
als_model = AlternatingLeastSquares(factors=100, regularization=0.1, use_gpu=False, iterations=15, dtype=np.float32)
```

03 결과

1. 비슷한 노래 추천 .similar_items()

```
1 # 비슷한 노래 찾기
2 def get_similar_music(favorite_music):
3     title_id = title_to_idx[favorite_music]
4     ids, scores = als_model.similar_items(title_id, N=20)
5     idx_to_title = {v:k for k,v in title_to_idx.items()}
6     for pair in zip(ids, scores):
7         print(pair, idx_to_title[pair[0]])
```

```
1 favorite_music = 'Next Level'
2 get_similar_music(favorite_music)
```

```
(620, 0.9999999) Next Level
(3602, 0.7933783) Savage
(4161, 0.7116712) Black Mamba
(602, 0.6586029) ASAP
(380, 0.60926235) LOVE DIVE
(3593, 0.59806633) Dreams Come True
(1017, 0.5876398) ELEVEN
(333, 0.58617747) TOMBOY
(539, 0.57642347) Queendom
(332, 0.5508086) INYU
(312, 0.5341946) 도깨비불 (Illusion)
(3588, 0.5320269) Step Back
(1813, 0.52837694) Weekend
(3228, 0.518613) SMILEY (Feat. BIBI)
(405, 0.51813024) Feel My Rhythm
(27, 0.5161157) Lovesick Girls
(3654, 0.5115826) On The Ground
(708, 0.5084981) 라일락
(1123, 0.50479186) Psycho
(3335, 0.4984385) 롤린 (Rollin')
```

```
1 favorite_music = '거리에서'
2 get_similar_music(favorite_music)
```

```
(163, 0.99999994) 거리에서
(234, 0.59763736) 또 다시 사랑
(59, 0.54047) 다시 사랑한다 말할까
(5029, 0.51866883) 사랑합니다...
(143, 0.5167015) 소주 한 잔
(1670, 0.5143186) 이별택시
(11893, 0.5123641) 차마...
(14287, 0.51056576) 어제보다 오늘 더
(33, 0.50914973) 벌써 일년
(18268, 0.5068631) 늑 (조관우)
(982, 0.5059282) 천년의 사랑
(5624, 0.5054064) 중독된 사랑
(1695, 0.5013237) 못해 (Feat. 美)
(123, 0.49958152) My Love (And)
(17099, 0.49860647) 헤어지는 중 입니다 (불광동 휘발유)
(161, 0.4960256) 삭제
(5607, 0.49137312) 너를 보내고
(242, 0.49085617) 내가 저지른 사랑
(141, 0.48898047) 내게 오는 길
(20, 0.48750106) 슬픔활용법
```

03 결과

2. 나에게 노래 추천 .recommend()

```
# 노래 추천
def get_recommendations(your_name):
    user = user_to_idx[your_name]
    ids, scores = als_model.recommend(user, csr_data[user], N=20)
    idx_to_title = {v:k for k,v in title_to_idx.items()}
    for pair in zip(ids, scores):
        print(pair, idx_to_title[pair[0]])
```

```
1 your_name = '김준화'
2 my_music_list('한계')
3 my_music_list('운이 좋았지')
4 my_music_list('도깨비불 (Illusion)')
5 my_music_list('시차')
6 my_music_list('사건의 지평선')
7 my_music_list('꿈과 책과 힘과 벽')
8 my_music_list('Feelings')
9 my_music_list('aenergy')
10 my_music_list('Strawberries & Cigarettes')
11 my_music_list('Never Not')
12 my_music_list('서울의 잠 못 이루는 밤 (Feat. 이수현)')
13 my_music_list('볼꽃놀이 (Remember Me)')
14 my_music_list('Our love is great')
15 my_music_list('Opening Sequence')
16 my_music_list('하룻길')
17 my_music_list('Attention')
18 my_music_list('Hype boy')
19 my_music_list('for him.')
20 my_music_list('Angel Baby')
21 my_music_list('난춘 (亂春)')
```

```
1 your_name = '김준화'
2 get_recommendations(your_name)

(672, 0.0637255) Square (2017)
(380, 0.055552848) LOVE DIVE
(1953, 0.051733978) That's Hilarious
(3602, 0.05077818) Savage
(332, 0.04815747) INYU
(4015, 0.04642431) Life's Too Short (English Ver.)
(4161, 0.044485644) Black Mamba
(4833, 0.04326672) 지켜줄게
(405, 0.04311622) Feel My Rhythm
(2013, 0.042626206) 자각몽 (Lucid Dream)
(2181, 0.04217812) YOUTH
(602, 0.041507505) ASAP
(1017, 0.041425016) ELEVEN
(612, 0.03876651) Lost
(1070, 0.038752995) 그건 아마 우리의 잘못은 아닐 거야
(3593, 0.03864543) Dreams Come True
(333, 0.038606156) TOMBOY
(4146, 0.038547087) Forever (약속)
(7477, 0.03837939) Drive
(4099, 0.038114198) YEPPI YEPPI
```