

jaccard_type7_testFirstOne

November 26, 2021

```
[2]: import os
import json
import pandas as pd
from collections import defaultdict
from tqdm.notebook import tqdm
```

0.1 Preprocessing

```
[3]: # isolate type6 and type7 testing
with open('challenge_set.json', 'r') as file:
    challenge_set = json.load(file)

    type_6 = challenge_set['playlists'][5000:6000]
    type_7 = challenge_set['playlists'][6000:7000]

    with open('challenge_type_6.json', 'w') as f:
        json.dump(type_6, f)

    with open('challenge_type_7.json', 'w') as f:
        json.dump(type_7, f)
```

0.2 Jaccard

```
[4]: import pandas as pd

def jaccard(s1, s2):
    numer = len(s1 & s2)
    denom = len(s1 | s2)
    return numer / denom if denom else 0

def pre_calc_jaccard(dataset):
    jaccard_dict = {}

    unique_track = set()
    for sample in tqdm(type_7):
```

```

        playlist = [t["track_uri"] for t in sample["tracks"]]
        unique_track.update(playlist)

    for track in tqdm(all_tracks):
        for ref_track in unique_track:
            jaccard_dict[(track, ref_track)] = 0
    → jaccard(playlist_per_track[track], playlist_per_track[ref_track])
    return jaccard_dict

def apc(dataset):
    predictions = []
    for sample in tqdm(dataset):
        k = sample["num_holdouts"]
        playlist = [t["track_uri"] for t in sample["tracks"]]

        tracks_scored = []
        for track in tqdm(all_tracks):
            if track in playlist: # skip tracks thats already in `playlist`
                continue

            jaccard_sum = 0
            for ref_track in playlist:
                jaccard_sum += jaccard(playlist_per_track[track], 0
    → playlist_per_track[ref_track])

            tracks_scored.append((jaccard_sum, track))

        tracks_scored.sort(reverse=True)
        recommend_tracks = [t[1] for t in tracks_scored[:k]]

        pred = {
            "pid": sample['pid'],
            "tracks": playlist + recommend_tracks
        }
        predictions.append(pred)

    df = pd.DataFrame(predictions)
    return df

```

```

[5]: # map pid to playlist name
playlist_name = {}

track_per_playlist = defaultdict(set)
playlist_per_track = defaultdict(set)

path_home = os.path.expanduser('~')

```

```

path_data = os.path.join(path_home, 'spotify_mpd/data')

for filename in tqdm(os.listdir(path_data)[:100]):
    full_path = os.path.join(path_data, filename)
    with open(full_path, 'r') as file:
        mpd_slice = json.load(file)

        for playlist in mpd_slice['playlists']:
            pid = playlist['pid']

            playlist_name[pid] = playlist['name']

            for track in playlist['tracks']:
                track_uri = track['track_uri']

                track_per_playlist[pid].add(track_uri)
                playlist_per_track[track_uri].add(pid)

```

```
0%|          | 0/100 [00:00<?, ?it/s]
```

```
[6]: all_tracks = list(playlist_per_track.keys())
```

0.3 Precalculated

```
[ ]: with open('challenge_type_6.json', 'r') as f:
      type_6 = json.load(f)

jaccard_dict = pre_calc_jaccard(type_6)
```

0.4 Non-precalculated

```
[7]: with open('challenge_type_7.json', 'r') as f:
      type_7 = json.load(f)

df_pred_type7 = apc(type_7[:1])
```

```
0%|          | 0/1 [00:00<?, ?it/s]
```

```
0%|          | 0/688034 [00:00<?, ?it/s]
```

```
[15]: with pd.option_context('display.max_colwidth', None):
      display(df_pred_type7)
```

```
pid \
```

0 1007147

0 [spotify:track:4wCmqSrbyCgxEXROQE6vtV, spotify:track:2bKYRpXeCquRfWYjYqHU4k, spotify:track:

[16]: type_7[0]

```
[16]: {'name': 'Oldies',
      'num_holdouts': 106,
      'pid': 1007147,
      'num_tracks': 131,
      'tracks': [{'pos': 7,
                   'artist_name': 'Gotye',
                   'track_uri': 'spotify:track:4wCmqSrbyCgxEXROQE6vtV',
                   'artist_uri': 'spotify:artist:2AsusXITU8P25d1RNhcAbG',
                   'track_name': 'Somebody That I Used To Know',
                   'album_uri': 'spotify:album:1HjSyGjmLNjRAKgT9t1cna',
                   'duration_ms': 244973,
                   'album_name': 'Making Mirrors'},
                 {'pos': 14,
                   'artist_name': 'Player',
                   'track_uri': 'spotify:track:2bKYRpXeCquRfWYjYqHU4k',
                   'artist_uri': 'spotify:artist:0fgtHVpOPfXb07S8Jx443z',
                   'track_name': 'Baby Come Back',
                   'album_uri': 'spotify:album:5vMR67pULOuK1BpEKw0qlld',
                   'duration_ms': 253204,
                   'album_name': 'Player'},
                 {'pos': 18,
                   'artist_name': 'Christina Aguilera',
                   'track_uri': 'spotify:track:2EVv4Vxu3HGPr32CQecgHx',
                   'artist_uri': 'spotify:artist:1l7ZsJRRS8w1W3WfJfPfNS',
                   'track_name': 'Genie in a Bottle',
                   'album_uri': 'spotify:album:2alhZwaLZ5wXZ0oJAo38aS',
                   'duration_ms': 217573,
                   'album_name': 'Christina Aguilera'},
                 {'pos': 22,
                   'artist_name': 'Chris de Burgh',
                   'track_uri': 'spotify:track:1523Yp80GgISzfWgBisj1l',
                   'artist_uri': 'spotify:artist:2RpHsR0rX075xfIwHn6B2U',
                   'track_name': 'The Lady In Red',
                   'album_uri': 'spotify:album:4HLgcxU5r4ZAZWUBg7ycEq',
                   'duration_ms': 256173,
                   'album_name': 'Double Package: Into The Light / High On Emotion'},
                 {'pos': 26,
                   'artist_name': 'Bee Gees',
                   'track_uri': 'spotify:track:4t2euPkyLU5n5j1HDZ2Tr',
                   'artist_uri': 'spotify:artist:1LZEQNv7sE11VDY3SdxQeN',
```

```

    'track_name': 'Night Fever',
    'album_uri': 'spotify:album:684Fi6YqIP9xU9JeboAgVM',
    'duration_ms': 212253,
    'album_name': 'Greatest'},
  {'pos': 28,
    'artist_name': 'Frank Sinatra',
    'track_uri': 'spotify:track:40tAOP3DPqmVD6L1h45Jp6',
    'artist_uri': 'spotify:artist:1Mxqyy3pSjf8kZZL4QVxS0',
    'track_name': 'My Way',
    'album_uri': 'spotify:album:0Cx7IYEdG2T70z9xFknhDT',
    'duration_ms': 274933,
    'album_name': 'Ultimate Sinatra'},
  {'pos': 35,
    'artist_name': 'Bonnie Tyler',
    'track_uri': 'spotify:track:4McRlwqJQIERlJFiJEgbP0',
    'artist_uri': 'spotify:artist:0SD4eZCN4KrOwQk56hCdh2',
    'track_name': 'Total Eclipse of The Heart',
    'album_uri': 'spotify:album:0e18vg5n45L6CYr6L6xxh',
    'duration_ms': 267786,
    'album_name': 'The Very Best Of'},
  {'pos': 48,
    'artist_name': 'Lipps Inc.',
    'track_uri': 'spotify:track:0N5Du3CNQilIBTxkwDQadv',
    'artist_uri': 'spotify:artist:0lwRI7lvmlRY5DiA5Xa6wQ',
    'track_name': 'Funkytown - Single Version',
    'album_uri': 'spotify:album:0yS1j89glNzVI07nF4pxMs',
    'duration_ms': 238240,
    'album_name': 'Mes Soirées Disco'},
  {'pos': 52,
    'artist_name': 'Roxette',
    'track_uri': 'spotify:track:6qB7YcFpeBEQa0D6Q0482y',
    'artist_uri': 'spotify:artist:2SHhfs4BiDxGQ3oxqf0UHY',
    'track_name': 'It Must Have Been Love',
    'album_uri': 'spotify:album:6Zd10LqFX5geleqvJ9xtAL',
    'duration_ms': 258786,
    'album_name': "Don't Bore Us - Get To The Chorus! Roxette's Greatest Hits"},
  {'pos': 55,
    'artist_name': 'Cher',
    'track_uri': 'spotify:track:7sJMvQ0kYs6u3vMQUYvVoK',
    'artist_uri': 'spotify:artist:720aDtakiy6yFqkt4TsiFt',
    'track_name': 'After All (Love Theme From Chances Are)',
    'album_uri': 'spotify:album:60Q5H90JaMs0sFzImKDJ8s',
    'duration_ms': 244360,
    'album_name': 'Heart Of Stone'},
  {'pos': 66,
    'artist_name': 'The Police',
    'track_uri': 'spotify:track:5COLFQARavkPpn7JgA4sLk',

```

```

    'artist_uri': 'spotify:artist:5NG030tJxFlKixkPSgXcFE',
    'track_name': 'Every Breath You Take - Remastered 2003',
    'album_uri': 'spotify:album:28e0riEfl7IGbQDNvWIWXX',
    'duration_ms': 253250,
    'album_name': 'Synchronicity'},
  {'pos': 70,
    'artist_name': 'Mr. Mister',
    'track_uri': 'spotify:track:63nVsCAdxu90deRhe58wiE',
    'artist_uri': 'spotify:artist:7Bah8E0kCETqEpAHI6CPzQ',
    'track_name': 'Broken Wings',
    'album_uri': 'spotify:album:1vmioa1hsabaxkYnYKROT1',
    'duration_ms': 345840,
    'album_name': 'Welcome To The Real World'},
  {'pos': 72,
    'artist_name': 'REO Speedwagon',
    'track_uri': 'spotify:track:5WwqdeavrQrbeAMDxGawse',
    'artist_uri': 'spotify:artist:55vs7NT1KxcFjbMC4y202E',
    'track_name': "Can't Fight This Feeling",
    'album_uri': 'spotify:album:35KafpmKh0nDLzBLV75MpR',
    'duration_ms': 294773,
    'album_name': "Wheels Are Turnin'"},
  {'pos': 74,
    'artist_name': 'Journey',
    'track_uri': 'spotify:track:2w4EpqGasrz9qdTwocx54t',
    'artist_uri': 'spotify:artist:0rvjqX7ttXeg3mTy8Xscbt',
    'track_name': 'Open Arms',
    'album_uri': 'spotify:album:43wpzak90mQfrjyksuGwp0',
    'duration_ms': 202173,
    'album_name': 'Escape'},
  {'pos': 80,
    'artist_name': 'Daryl Hall & John Oates',
    'track_uri': 'spotify:track:41dDygr3r7e926oGUXfrLt',
    'artist_uri': 'spotify:artist:77tT1kLj6mCWtFNqi0mP9H',
    'track_name': "I Can't Go for That (No Can Do) - Remastered",
    'album_uri': 'spotify:album:7rfpaXxmQG7dnFycZjLae0',
    'duration_ms': 307306,
    'album_name': 'Private Eyes'},
  {'pos': 86,
    'artist_name': 'Kenny Loggins',
    'track_uri': 'spotify:track:4YR6Dextuoc3I8nJOXgzKI',
    'artist_uri': 'spotify:artist:3Y3xIwWyq5wnNHPP5gPj0W',
    'track_name': 'Footloose',
    'album_uri': 'spotify:album:68ZonGoyvksZu4sqqNXect',
    'duration_ms': 226293,
    'album_name': 'The Essential Kenny Loggins'},
  {'pos': 88,
    'artist_name': 'Debbie Gibson',

```

```

    'track_uri': 'spotify:track:6PMDILZiYo4YzMBbE0dGK9',
    'artist_uri': 'spotify:artist:18jZvCsW1PJ4FDQ5gEXuKp',
    'track_name': 'Lost In Your Eyes',
    'album_uri': 'spotify:album:5PnwSSRbLcCLpNMe2LFh0C',
    'duration_ms': 213000,
    'album_name': 'Electric Youth'},
  {'pos': 95,
    'artist_name': 'Scorpions',
    'track_uri': 'spotify:track:6wjSN06uJYlgbbfauyWhiw',
    'artist_uri': 'spotify:artist:27T030eWyCQRmDyuvr1kxY',
    'track_name': 'Wind Of Change',
    'album_uri': 'spotify:album:4mGV3pZqcrUk00SC8HBT96',
    'duration_ms': 313226,
    'album_name': 'Crazy World'},
  {'pos': 98,
    'artist_name': 'Nazareth',
    'track_uri': 'spotify:track:082KyGrUIooHcyyElugozU',
    'artist_uri': 'spotify:artist:6fvN9GmMCVKb5LYOWsnjFP',
    'track_name': 'Where Are You Now',
    'album_uri': 'spotify:album:5rtW1DEsBzYQp9kpB4t1W7',
    'duration_ms': 233853,
    'album_name': 'Sound Elixir'},
  {'pos': 101,
    'artist_name': 'Whitney Houston',
    'track_uri': 'spotify:track:4eHbdreAnS0rDDsFfc4Fpm',
    'artist_uri': 'spotify:artist:6XpaIBNiVzIetEPCWDvAFP',
    'track_name': 'I Will Always Love You',
    'album_uri': 'spotify:album:7JVJlkNNobS0GSoY4tCS96',
    'duration_ms': 271093,
    'album_name': 'The Bodyguard - Original Soundtrack Album'},
  {'pos': 102,
    'artist_name': 'Cyndi Lauper',
    'track_uri': 'spotify:track:2A6yzRGMgSQCUpR2ptm6A',
    'artist_uri': 'spotify:artist:2BTZlqwOntH9MvilQ3ewNY',
    'track_name': 'True Colors',
    'album_uri': 'spotify:album:31TRqoVBTQi0lzlPLtvINn',
    'duration_ms': 226066,
    'album_name': 'True Colors'},
  {'pos': 104,
    'artist_name': 'Roxette',
    'track_uri': 'spotify:track:18eqtLHIuDSnx54o0IRUy0',
    'artist_uri': 'spotify:artist:2SHhfs4BiDxGQ3oxqf0UHY',
    'track_name': 'Spending My Time',
    'album_uri': 'spotify:album:5Q4kTNheoI86WVkl1YQKt4x',
    'duration_ms': 275666,
    'album_name': 'Joyride'},
  {'pos': 114,

```

```

'artist_name': 'Atlantic Starr and The Jets',
'track_uri': 'spotify:track:5A0WdpiZuxn02NKx1L8o7e',
'artist_uri': 'spotify:artist:4KhV561Hcvr0HpS3s3vT0x',
'track_name': 'You Got It All',
'album_uri': 'spotify:album:2MwgUMtXUv76duKJdU4yHb',
'duration_ms': 261720,
'album_name': 'Back To Back - Atlantic Starr and The Jets'},
{'pos': 116,
 'artist_name': 'Smokie',
 'track_uri': 'spotify:track:5RDuitr7B8gv9lQXgaaW43',
 'artist_uri': 'spotify:artist:5rIq0JspxDq89aBBCUda1X',
 'track_name': 'Have You Ever Seen the Rain',
 'album_uri': 'spotify:album:10mpyIo02CEuG6gtX9wrKQ',
 'duration_ms': 255906,
 'album_name': 'The World and Elsewhere'},
{'pos': 117,
 'artist_name': 'Boney M.',
 'track_uri': 'spotify:track:78His8pbKjbDQF7aX5asgv',
 'artist_uri': 'spotify:artist:54R6Y0I7jGUCveDTtI21nb',
 'track_name': 'Rivers of Babylon',
 'album_uri': 'spotify:album:0txzXbDfTn3vAdx77iCaXd',
 'duration_ms': 261266,
 'album_name': 'Nightflight to Venus'}],
'num_samples': 25}

```

[]: