Music Recomendations using Nearest Neighbors

Simpler Techniques

Korede Adegboye
STATS 780 course project
05/04/2022

Outline

Motivation

Data

Methods

Result

Discussion

Motivation

- **Domain:** When the end of a playlist approaches, users alter their attention to choosing the next song.
  - Distracting or time-consuming when the user is driving or at work.
  - To increase the user experience by personalization
  - Seamless continuation of a users playlist — curated playlists
- **Statistical:** Light-weight solution
  - Published literature
  - Acquire a foundational understanding of modern issues and methods

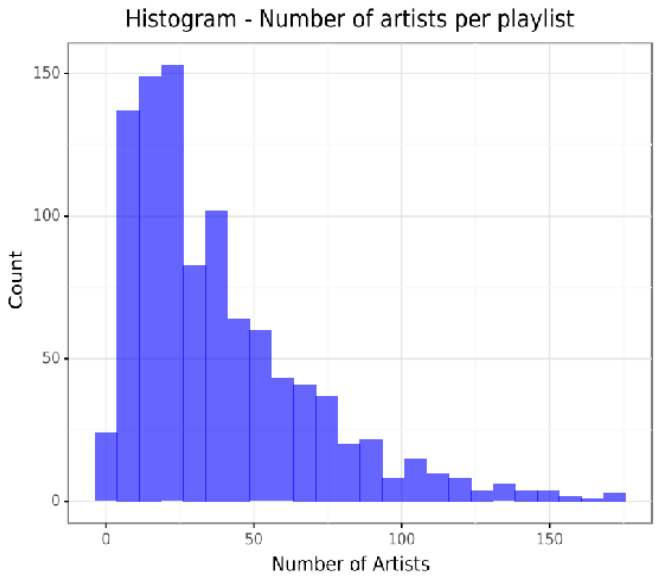
Data

Table 1: Playlists – Collaborative-based filtering. $n < p$, $k = \{\text{no: } 0, \text{ yes:}1\}$

	Song_1	Song_2	$\cdots$	Song_p
Playlist_1	k_1	k_{12}	$\cdots$	k_{1p}
Playlist_2	k_{2p}	k_{2p}	$\cdots$	k_{2p}
$\vdots$	$\vdots$			
Playlist_i	k_{ip}	k_{ip}	$\cdots$	k_{ip}
$\vdots$	$\vdots$		$\cdots$	
Playlist_n	k_{np}	k_{np}	$\cdots$	k_{np}

Table 2: Audio Features – Content-based filtering. $n > p = 13$, $K \in \mathbb{R}$

	playlist ID	loudness	$\cdots$	danceability	energy	track ID
1	1	K_{12}	$\cdots$	$\cdots$	K_{1p}	track_1
1	1	K_{12}	$\cdots$	$\cdots$	K_{1p}	track_2
$\vdots$	$\vdots$					
2	2	K_{22}	$\cdots$	$\cdots$	K_{2p}	track_1
$\vdots$	$\vdots$		$\cdots$			
n	n	K_{n2}	$\cdots$	$\cdots$	K_{np}	track_1



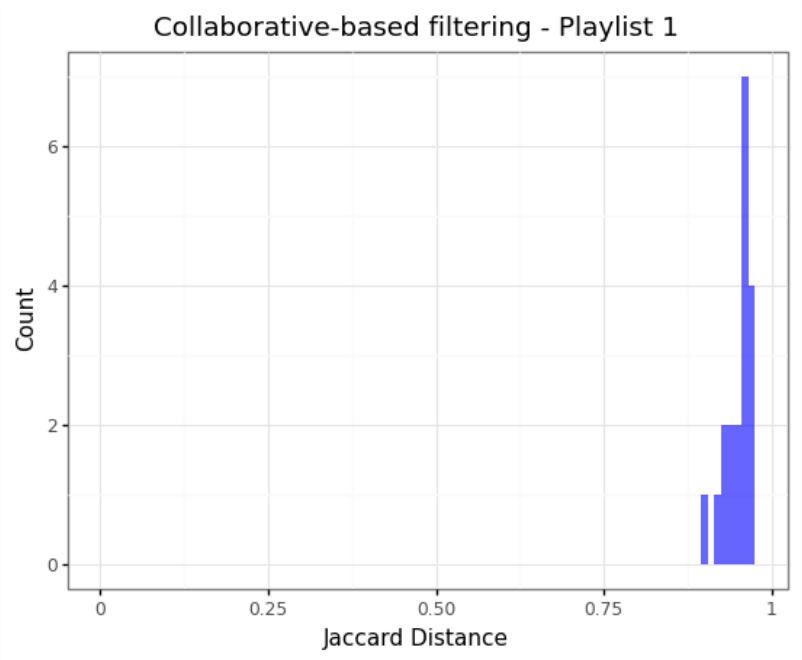
Source: 1000 random playlists from *Spotify Million Playlist Dataset Challenge*.

- Spotify's API

Methods

- ~~K~~NN for distance/similarity metrics
  - Collaborative-based Filtering — considers **multiple viewpoints** (playlists). Assumes that those "determined" to have closely related behaviours, also would like the same songs.
  - Content-based filtering — Uses an items (tracks) feature to **extract information** about the users and therefore, suggest a song.
- Similarity metrics: jaccard and euclidean
- Standardization for audio features
- To reduce dimensionality, Principal Component analysis (PCA) suggested 8 features

Results



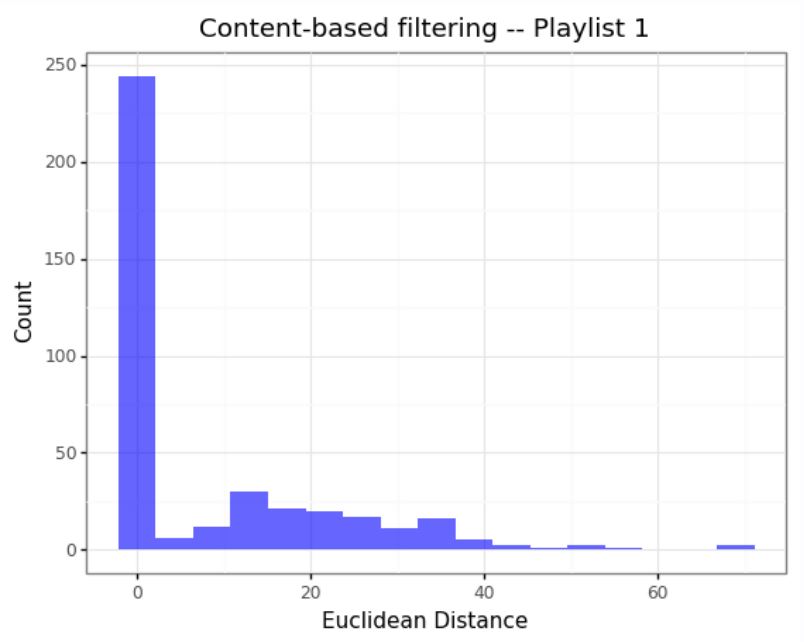


Table 3: Evaluation

	MAP	NDCG
Collaborative	0.0	0.0
Content	0.0	0.0

- Both approaches give Mean Average Precision (MAP): 0.0
  - **R-precision** is defined as the proportion of the top-R retrieved documents that are relevant, where R is the number of relevant documents for the current query

Discussion

- The methods considered give a baseline for cases (clusters of user behaviors)
  - When a user does not have many songs, a collaborative approach is useful
  - When there are many songs, then use content-based filtering
  - **Hybrid**
- Slow learning rate. Large amount of data is needed.
- Interpretation of "similarity" can be misleading. Songs can be deemed to be "close" but upon listening to the recommended songs, it is much different.
- Simpler techniques relies **algorithmic approaches**
  - Further filtering needed to get the recommended songs
- More features — artist popularity, genre
- Feature Selection on K-Nearest Neighbor Algorithm Using Similarity Measure

Thank you!

References

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani.2021. An Introduction to Statistical Learning V2: With Applications in r. Springer.

Ludewig, Malte, Iman Kamehkhosh, Nick Landia, and Dietmar Jannach. 2018. "Effective Nearest-Neighbor Music Recommendations." In, 1–6. https://doi.org/10.1145/3267471.3267474.

Spotify. 2018. Alcrowd | Spotify Million Playlist Dataset Challenge | Challenges. Alcrowd. https://www.aicrowd.com/challenges/spotify-million-playlist-dataset-challenge.

R. Puspadini, H. Mawengkang and S. Efendi, "Feature Selection on K-Nearest Neighbor Algorithm Using Similarity Measure," 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT), 2020, pp. 226-231, doi: 10.1109/MECnIT48290.2020.9166612.