San Diego State University

Generating Music Recommendations with Sonic Features and Collaborative Filtering

Brenden Church

Nahla Elshafey

Alejandro Gomez

Van Luu

Alex Ray

Computer Science 582: Introduction to Speech Processing

Dr. Chuck Konopka

6 May 2023

**Abstract**

**Introduction**

Since the emergence of music streaming services there are many services on the market that offer on-demand streaming of their music library, such as Spotify, Soundcloud, and Apple Music, to name a few. These platforms compete with each other to offer the best streaming platform, including having larger or more diverse libraries to choose from and offering stable service. Another way these services compete are through the music recommendation algorithms that are integrated into the services. In the case of Spotify, this is done using metadata and collaborative filtering based on the current user's listening history compared with every user in Spotify’s database.

One method of Collaborative Filtering is the K-means algorithm, which is an unsupervised clustering method. This algorithm begins by initializing a number of centroids equal to a target value. These centroids represent the statistical average of the position of all the data points associated with it. The algorithm then assigns each data point in the data set to the nearest centroid, then updates the position of the centroid given the additional data point. The end result is K unique clusters that are grouped by having similar features.

Another method of Collaborative Filtering is the K-Nearest-Neighbors algorithm, which is a supervised learning model. The K-Nearest-Neighbors algorithm attempts to find the closest related items to a given item by calculating the distances between the items and finding the shortest options. Once the objects are grouped together, the model assigns a label to the cluster.

Both of these algorithms can be used to group data into clusters, and it generally comes down to whether the classification features of a supervised machine learning algorithm are needed when choosing between the K-means and the K-nearest neighbors algorithm.

This is an effective algorithm for generating user music recommendations–it is used in one of the most popular music streaming platforms in the world, but many implementations fail to account for any of the acoustic data or sonic features found in each song. To adapt this model to not just excel at finding music that people will enjoy, but to recommend similar sounding songs as well, the model needs to be updated to support the new data. In this essay, the Collaborative Filtering process will be demonstrated with audio metadata as well as acoustical features. Then, the viability of using audio features with Collaborative Filtering to generate music recommendations will be analyzed. The goal of this project is to explore how to make an improved music recommendation algorithm using machine learning concepts.

**Review of Related Literature**

The application of Collaborative Filtering for Music Recommendation Algorithms is explored in the research essay “Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations” by authors Seok Kee Lee, Yoon Ho Cho, and Soung Hie Kim. There are two types of Collaborative filtering: user-based, which attempts to place similar users into groups, and item-based, which attempts to group similar items, in this case, songs available in the dataset. Collaborative filtering is accomplished by calculating a similarity score for each user or item, and finding the “nearest neighbors,” or most similar items or users to a given item or user. In the 2010 research essay by Lee, Cho, and Kim, they find a way to implicitly define user interests in certain songs by tracking page visits, preview listens, and song purchases. This data can then be applied to the collaborative filtering algorithm in order to generate music recommendations.

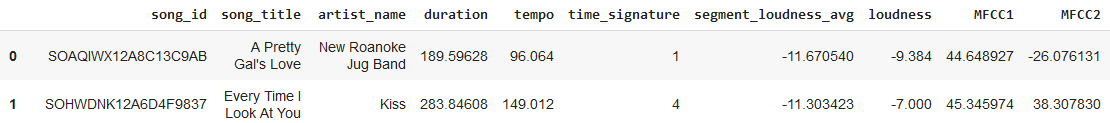
While this system has worked well for generating music recommendations, it does not factor in the actual sound of the music, and may have other drawbacks as well. In the research paper “Artist and Style Exposure Bias in Collaborative Filtering Based Music Recommendations” by authors Andres Ferraro, Dmitry Bogdanov, and Xavier Serra, some of the drawbacks of this technique are explored. In this paper, the authors develop a simulation to illustrate how these filtering algorithms can narrow the listening interests of users and eventually be susceptible to feedback loops. They also demonstrate how this algorithm may concentrate the interests of users to fewer items in the catalog, giving less opportunities for the vast majority of artists on the platform.

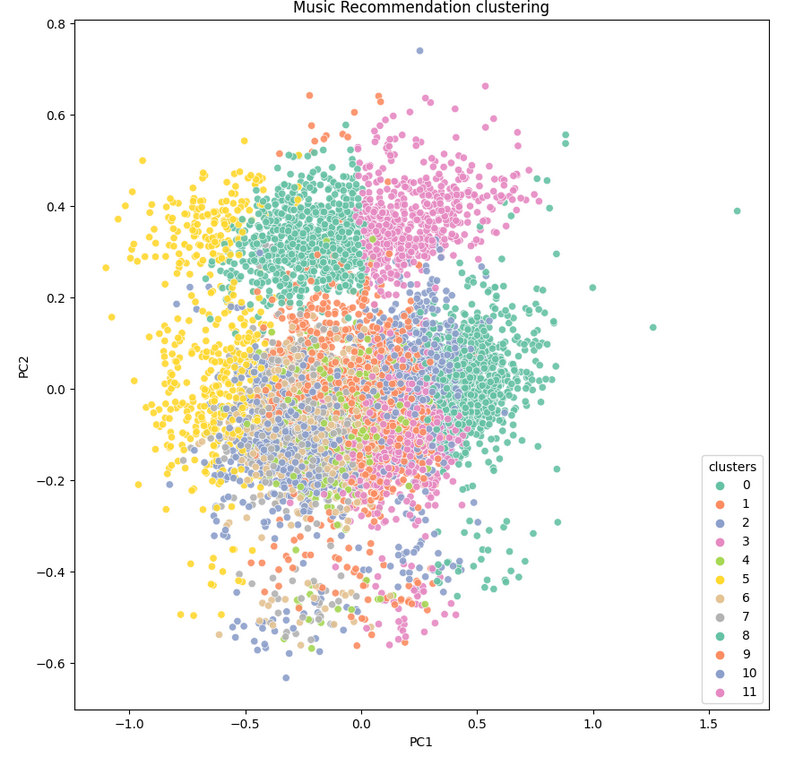
By adding more information for the model to learn off of, the algorithm can be enhanced to provide better results and a better understanding of which features in the data influence successful music recommendations. In the research paper “A Music Recommender Based on Audio Features,” authors Qing Li, Byeong Man Kim, Dong Hai Guan, and Duk Whan Oh analyze and explore adding additional features such as genre and audio features into Collaborative filtering-based music recommendation systems. In their tests, they incorporate audio features including Mel frequency cepstral coefficients and the sum of scale factor into the dataset. The results show that incorporating audio features into the recommendation system can give more ideal results.

**Presentation**

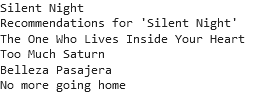
In order to explore the implications of using audio features with collaborative filtering-based music recommendation systems a dataset that includes these features must be chosen first. The Million Song Dataset was chosen because it offers features such as timbre, chroma, and loudness data over the course of the song as well as metadata for the song and a very large sample size.

The audio features used to classify and analyze the dataset were timbre, chroma, loudness, tempo, and the time signature. Timbre was represented through Mel-frequency cepstral coefficients calculated in equal length segments throughout the audio track. Mel-frequency cepstral coefficients are a logarithmic representation of the audio signal. They are calculated with the inverse Fourier transform of the log of the short term power spectrum of the original signal. The Mel-scale binning and weighting of the audio signal allows for feature extraction with the same inclinations as the human ear. The ear does not react linearly to the pitch of a given audio signal, so the binning and weighting act to model this phenomenon through a logarithmic representation. The chroma values are representations of the 12 main pitches in the western equal tempered scale. The values are highly sensitive to small changes in pitch and timbre of the signal. This allows for a very unique representation of a song, when averaged over the whole song. Loudness is the amplitude of the input signal. There are 2 values we analyze, the first is the average loudness over the whole song, and the second is the max amplitude sampled at regular intervals and averaged for the whole song. Tempo is the speed at which a song is played, and is measured in beats per minute. Time signature is how the beats are divided within the song structure.

In order to use learning algorithms on this dataset, first the useful data needs to be extracted from the sample. For each song in the sample, a 28 dimension feature vector is generated. The first entry is the general loudness of the track, followed by the tempo and the time signature. Then, the average segmented loudness is computed. The next 24 features in the vector are dedicated to the timbre and chroma features for each song. Over the course of the song, the average output of the timbre and chroma throughout the 12 frequency band is averaged, giving an estimate on the usage of each band of frequencies over the course of the song. In Figure 1, the first nine features of the vector are shown for two songs, “Every Time I Look at You” and “A Pretty Gal’s Love.”

After processing the features with the K-means algorithm, the 10,000 songs provided in the Million Song Dataset subset are clustered into 11 unique groups. A graph demonstrating the results of the clustering process is shown in Figure 2 provides a visual representation of the dataset. 

In order to find recommendations for a song, the K-nearest-neighbors algorithm is processed on an item. A demonstration of finding the nearest 4 neighbors to the song “Silent Night” is shown in Figure 3.



**Conclusion**

Overall our music recommendation algorithm provides a relatively simple application of using embedded audio features as a basis for grouping and recommending songs to users. As with many real-world applications of machine learning models, there are some limitations to the model. Model validation is a large challenge since there are no supervised labels that allow for an easy verification of model output. A weak supervision method could consist of using known metadata around the song’s genre and artist, and comparing whether the output of the model is within a genre or by an artist who creates music in a neighboring song. In addition, this same problem extends into the use of other models, as more powerful models require a sizable set of supervised labels for very accurate use. Real-world applications may be able to solve this problem with user feedback, where users can accept or reject certain recommendations, which can be used as another form of weak labels for which to fit the underlying models upon. In spite of these limitations, the music recommendation algorithm provides a strong example of grouping songs through their distinct audio attributes for more robust recommendations to the user.

**Works Cited**

Ferraro, Andres, et al. “Artist and Style Exposure Bias In Collaborative Filtering Based Music Recommendations.” *International Society for Music Information Retrieval.*  2019.

Lee, Seok Kee, et al. “Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations.” *Information Sciences*, Volume 180, Issue 11, 2010.

Li, Qing, et al. “A Music Recommender Based on Audio Features.” *Special Interest Group on Information Retrieval*. 2004.