Audio content-based Music Recommendation System

Akanksha Pandey
MT20048
IIIT Delhi
akanksha20048
@iiitd.ac.in

Atul
2017032
IIIT Delhi
atul17032
@iiitd.ac.in

Prashant Jain
2018253
IIIT Delhi
prashant18253
@iiitd.ac.in

Shivani Mishra
MT20062
IIIT Delhi
shivani20062
@iiitd.ac.in

ABSTRACT

As online music streaming becomes the dominant medium for people to listen to their favorite songs, music streaming services are now able to collect large amounts of data on the listening habits of their customers. These streaming services, like Spotify, Apple Music or Pandora, are using this data to provide recommendations to their listeners.
This project compares the music data in their ability to automatically recommend songs into their musical understanding depending on the features extracted from the music data.

1. PROJECT INTRODUCTION

"Music is a moral law. It gives soul to the universe, wings to the mind, flight to the imagination, and charm and gaiety to life and to everything" - Plato. The words of Plato rightly describe the importance of music in the world. As the field of music is evolving, large number of songs are being published every day.

A recommender (or recommendation) system (or engine) is a filtering system which aim is to predict a rating or preference a user would give to an item, i.e. a song. Recommendations done using content-based recommenders can be seen as a user-specific classification problem. This classifier learns the user's likes and dislikes from the features of the song. We are using as of now the most straightforward approach that is keyword matching. The idea behind is to extract meaningful keywords present in a song description a user likes, search for the keywords in other song descriptions to estimate similarities among them, and based on that, recommend those songs to the user.

2. LITERATURE SURVEY

Recommendation System of Music can be attempted in a variety of ways. A typical approach involves processing a dataset of audio files, extracting features from them, and then using a dataset of these extracted features to train a machine learning classifier.

In [1] B. McFee et al. propose a method for optimizing content-based similarity by learning from a sample of collaborative filter data. The optimized content-based similarity metric can then be applied to answer queries on novel and unpopular items, while still maintaining high recommendation accuracy. The proposed system yields accurate and efficient representations of audio content, and experimental results show significant improvements in accuracy over competing content-based recommendation techniques.

In [2] D.Kim et al. suggests a method for personalized services. They extract the properties of music from music's sound wave. They use STFT (Shortest Time Fourier Form) to analyze music's property. And they infer user's preferences from user's music list. To analyze users' preferences they propose a dynamic K-means clustering algorithm. The dynamic K-means clustering algorithm clusters the pieces in the music list dynamically adapting the number of clusters. They recommend pieces of music based on the clusters.
By using our K-means clustering algorithm, they can recommend pieces of music which are close to user's preference even though he likes several genres. They perform experiments with one hundred pieces of music. In this paper, they present and evaluate algorithms to recommend music.

M. Soleymani et al. [3] proposed a music recommender-system based on psychological study [4] done by P.J Rentflow et al. to describe music preferences. The study provided five set of attributes namely Melow, Unpretentious, Sophisticated, Intense and Contemporary (MUSIC) to narrate the preference of music.
For dataset, they used 249 audio files collected from 5 substudies.The dataset contained nine points user rating system, some metadata like artist, genre and title, and scores on five factors of the music preference model (MUSIC).
Timbral features which shows the perceived quality of sound or musical tone and auditory temporal features which helps in revealing the small and sudden stimuli in audio were extracted for genre recognition.
while training the model, PCA was used for dimensionality reduction and MLR, SVR, and RSS were used. For performance measure they used RMSE and the coefficient determination of r^2.
The authors concluded that the best performance for at-

tribute detection was found when we combine the audio modulation feature with sparse representation. They also showed that the model did not have any cold start problem.

In [5], S.D.Teh Chao Ying et al. presented an approach for lyrics based genre classification which useed mood information. The authors selected 10 genres(pop,blue,country, etc) and 10 mood categories(happy, sad, angry etc). For dataset, they collected 1000 English songs.The study assumed that genre and mood are complementary with each other.
For pre-processing the data, they manually cleaned the lyrics removing phrases like "back to intro" and replacing them with words itself to get complete text. Authors used variants of tf-idf including wf-idf. For training purpose they used kNN, Naive Bayes and SVM with ten-fold cross-validation which was further averaged over five repeated runs.
They noted that with Lwf-idf, Pop hadAuthors noted highest accuracy of 76.94 for pop using Lwf-idf and with Nlwf-idf, 75.71. They concluded that Lwf-idf and NLwf-idf weighting equations indicated thar additional weights in this domain is a promising approch and it can greatly help in classifying Genre using a lyrics based system.

[6] proposes a Shazam system audio identification system which computes a spectrogram from an audio using STFT (short term fourier transform). Peak picking strategy is used to extra all the local maxima in the magnitude spectrogram (these time frequency points represent closest search to the audio). The graph is further reduced to a constellation map, a low-dimensional sparse representation of the original signal by means of a small set of time-frequency points.The peaks are highly characteristic, reproducible, and robust against many, even significant distortions of the signal which facilitates its high identification rate, while scaling to large databases.

In [?] Tags form a basis for many music recommendation systems, which uses information about moods, musical key, etc to recommend audio content. However, such tags tend to be less accurate, subjective, and rather noisy. Crowd (or social) tagging, one popular strategy in this context, employs voting and filtering strategies based on large social networks of users for "cleaning"the tags. However, this approach is that it relies on a large crowd of users for creating reliable annotations. While mainstream pop/rock music is typically covered by such annotations, less popular genres are often scarcely tagged. This phenomenon is also known as the "long-tail" problem. The accuracy of these system is very less 30-40% and is entirely dependent on the crowd's choice. To overcome these problems, content-based retrieval strategies are preferred as they do not rely on any manually created metadata but are exclusively based on the audio content and cover the entire audio material in an objective and reproducible way.

3. DATASET AND DATA ANALYSIS

We are using song.csv dataset for music recommendation which is freely available. This dataset contains name, artist, and lyrics for 57650 songs in English. The data has been acquired from LyricsFreak through scraping. We are working with text and words so, Term Frequency-Inverse Document Frequency (TF-IDF) can be used for this matching process. And because of the dataset being so big, we are going to resample only initial 10000 songs. Later we will include audio features from FMA dataset.

4. BASELINE RESULT

If we use only lyrics and find tf-idf for lyrics and use similarity score for recommending the songs then we are getting some recommendations but it is just using similarity score and is not very accurate. Similarity 'scores of recommended song is low [0.01 - 0.3]

5. PROPOSED METHOD

We are using 10000 initial samples of the songs. We use TF-IDF vectorizer that calculates the TF-IDF score for each song lyric, word-by-word. Then we have calculated the similarity of one lyric to another. We have used the cosine similarity. Once we get the similarities, we have stored it in a dictionary the names of the 100 most similar songs for each song in our dataset. We have then used that similarity scores to access the most similar items and output a recommendation based the number of recommendation user want.
The next part is feature extraction in which some number of audio features will be extracted from each track. Features can be broadly classified as time domain and frequency domain features. The feature extraction can be done using libROSA, a Python library.
In the last part of the exercise we will create a ML personalized song recommend-er system by leveraging the audio features and content based recommendation system. We can here propose different-different machine learning models to classify songs in more correct manner and we are going to try our best to achieve maximum accuracy to fulfil users need.

6. PROJECT STATUS

We have completed some of the steps of the project for example first is to collect dataset. Then we have done some preprocessing on the dataset in order to prepare it for the recommendation. We are also done with the recommendation of the songs in small level like we have used similarity score based on tf-idf values of lyrics and asked user to input the number of songs he wants to be recommended, and our model will give the output accordingly. Right now our system is facing cold start problem. Later we will include some audio features also to overcome this problem.

7. REFERENCES

[1] B. McFee, L. Barrington, and G. Lanckriet, “Learning content similarity for music recommendation,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 8, pp. 2207–2218, 2012.

[2] D. Kim, K. Kim, K. Park, J. Lee, and K. M. Lee, “A music recommendation system with a dynamic k-means clustering algorithm,” in *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*, 2007, pp. 399–403.

[3] M. Soleymani, A. Aljanaki, F. Wiering, and R. C. Veltkamp, “Content-based music recommendation using underlying music preference structure,” in *2015 IEEE International Conference on Multimedia and Expo (ICME)*, 2015, pp. 1–6.

[4] P. J. Rentfrow, L. R. Goldberg, and D. J. Levitin, “The structure of musical preferences: A five-factor model.” *Journal of Personality and Social Psychology*, vol. 100, no. 6, pp. 1139–1157, 2011. [Online]. Available: https://doi.org/10.1037/a0022406

[5] S. D. Teh Chao Ying and L. N. Abdullah, “Lyrics-based genre classification using variant tf-idf weighting schemes,” *Journal of Applied Sciences*, vol. 15, pp. 289–294, 2015. [Online]. Available: https://scialert.net/abstract/?doi=jas.2015.289.294

[6] J. V. Balen, “Automatic recognition of samples in musical audio,” 2011.

[7] D. Bogdanov, M. Haro, F. Fuhrmann, A. Xambo, E. Gomez, and P. Herrera, “Semantic audio content-based music recommendation and visualization based on user preference examples,” pp. 13–33, 2012.