Music Recommendation System Using Audio

Rhythm Patel
rhythmkumar18083@iiitd.ac.in
IIIT Delhi
New Delhi, India

Mohnish Agrawal
mohnish18053@iiitd.ac.in
IIIT Delhi
New Delhi, India

Abhinav Suresh
abhinav18003@iiitd.ac.in
IIIT Delhi
New Delhi, India

Shaunak Pal
shaunak18098@iiitd.ac.in
IIIT Delhi
New Delhi, India

Ankit Mishra
ankit18019@iiitd.ac.in
IIIT Delhi
New Delhi, India

Rajat Prakash
rajat18078@iiitd.ac.in
IIIT Delhi
New Delhi, India

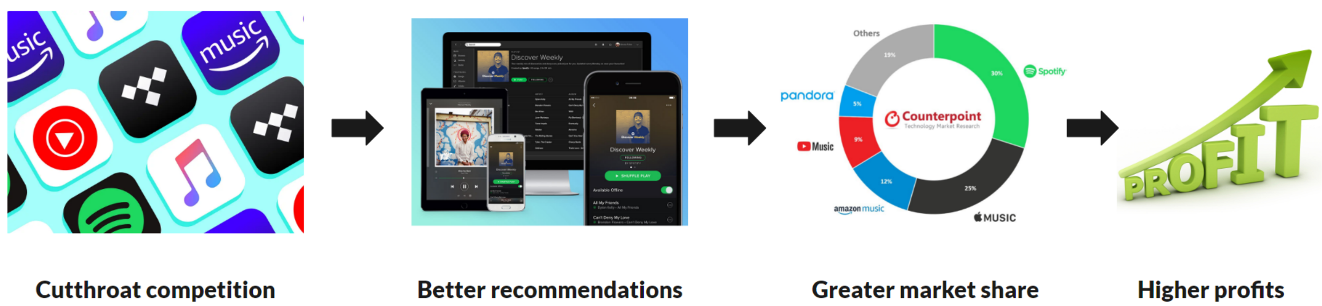

Figure 1: Problem Statement

KEYWORDS

information retrieval, music recommendation, deep learning, spectrogram

ACM Reference Format:
Rhythm Patel, Mohnish Agrawal, Abhinav Suresh, Shaunak Pal, Ankit Mishra, and Rajat Prakash. 2018. Music Recommendation System Using Audio. In *Proceedings of (Information Retrieval).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 PROBLEM STATEMENT

With the rapid growth of numerous digital music-streaming platforms like Spotify, Apple Music, and Amazon Music, the music content on the internet is greater than ever before, and so is the competition between them. It is impossible for any human to manually sort all this music and recommend the specific songs to users.

This is where Music Recommendation Systems come in. With various algorithms and a clean UI, these systems make it easier for customers to browse through millions of songs by music recommendations based on artist, genre, instrument, and reviews. The most popular type is genre: pop, rock, hip-hop, R&B, instrumental, indie, jazz etc.

We all know that there is a cutthroat competition between these music streaming platforms (Fig. 1. A better experience for the user enables them to get a greater market share. Consequently, one of the main features for the user is the recommendation system which recommends songs that a given user likes. Thus, such features drive higher profits for these for-profit companies. This is why music recommendation systems are important.

However, the perfect music recommendation system does not exist that has 100% accuracy. Another thing to note is that these are billion dollar companies, if not, trillion dollars, and have spent hundreds of millions of dollars on making a very very effective system. As a course project, we cannot have a better system than them. However, we can think outside the box and support creative & alternative ways of solving the problem.

The difference is: we will be recommending music using the audio content of the songs using spectrograms, along with/instead of other metadata which is the norm. This is described in the next section. We also have tested a baseline model. Also, one more thing to note is that this is an engineering project.

2 CURRENT SYSTEMS & LITERATURE REVIEW

"A Survey of Music Recommendation Systems with a Proposed Music Recommendation System" [3] is a review paper and provides a survey on different recommendation systems currently in use. They include content-based, collaborative, emotion-based, and other techniques. The paper also gives an in depth comparison of these systems, by exploring the strengths and weaknesses of each of them. Finally the paper also gives an overview of an improved hybrid recommendation system that claims to solve many of the problems that current systems face.

"Automatic Music Recommendation Systems: Do Demographic, Profiling, and Contextual Features Improve Their Performance?" [6] introduces a dataset of listening histories along with the listener's demographic informations and some other features that characterize aspects of listening behavior. This dataset is then used for the evaluation of accuracy of a music recommendation model. The paper's results show that using features like the listeners' age, country, and gender improve the recommendation accuracy by around 8 percent, and also when a new feature called "exploratoryness" was used, the accuracy increased by around 12 percent.

"Recommending music on Spotify with deep learning" [1] is written by Sander Dieleman who was interning at spotify at the time.he explains his work at spotify and some preliminary results. The article gives a brief overview of collaborative filtering, its virtues and its flaws. The article also describes content-based recommendation for when no usage data is available, and about predicting listening preferences with deep learning, which is music recommendation based on audio signals. Finally the article talks about what the covnets learn about music and some potential applications of his work.

"Why Spotify's music recommendations always seem so spot on" [2] describes about information about a user's listening history, music preferences, duration of play for certain songs, and how they respond to recommendations is fed into an algorithm (are they liking them, skipping them, replaying them, saving them). Spotify is attempting to model user behavior on the app by determining ways for projecting in-app actions into human features and emotion, as well as tying music experiences to mood and situational settings such as time of day, week, or season. CoSeRNN, a neural network that weighs specific variables such as prior listening history and current context to generate song choices that are appropriate for the moment, has been put to the test. Spotify groups music and users together based on shared features or qualities using a machine learning tool called the approximate nearest-neighbor search algorithm. MUSIG that learns "meaningful representations of tracks and users" based on individual features of songs (like genre, acoustics, danceability, the wordiness of lyrics) and how they relate to one another (like if they appear on the same playlist)

"How Spotify recommender system works" [4] shows how advanced NLP can also be executed on actual lyrics of the songs to indicate specific topics around which songs are composed. This might also be used for recommending songs with the same topics, but, possibly, from different genres. This feature helps with the cold-start problem. Raw audio has been processed, the CNN assigns some characteristics to each song. It can detect some of the key characteristics like mode, rhythm, loudness and sometimes even genre of song. Using these metrics CNN can then categorize songs in groups that have similar characteristics.

A paper published in 2012 [5] was one of the first papers to use Convolutional Neural Networks (CNN) to implement a music recommendation system suggesting the first content based system.

This paper from NUS Singapore scientists [7] improved upon Sander's method and proposed a novel algorithm on deep belief networks and probabilistic graphical methods that outperformed Collaborative Filtering in warm and cold start stages. Finally they combined Collaborative Filtering and their own model to outperform baseline Collaborative Filtering models with Root mean squared (RMS) reduced to 0.34 for warm start and 0.47 for cold start.

3 DATASET

3.1 Free Music Archive Dataset

Total tracks : 106574
Total genres : 163

Free Music Archive (FMA) is an open and easily available dataset which is capable of evaluating several tasks in Music Information Retrieval (MIR). MIR is a field concerned with browsing, searching, and organizing large music collections.

The full dataset size is 879 GB. But due to constraints, we'll probably be using the smaller version of this, which is 7.2 GB and has 8 genres (Pop, Rock, Instrumental, Hip-Hop, International, Electronic, Folk, Experimental)

3.2 CSV files with description

(1) **tracks.csv** - It contains metadata for all the tracks. There are 106574 tracks/rows and 6 columns : ID, title, artist, genres, tags and play counts.

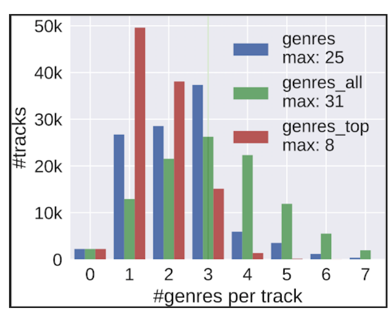

Figure 2: Genres per track

(2) **genres.csv** - It consists of names and parents of all the genres (163 genres) discussed. It is important for study on the basis of Hierarchy of genre and find the root genre.

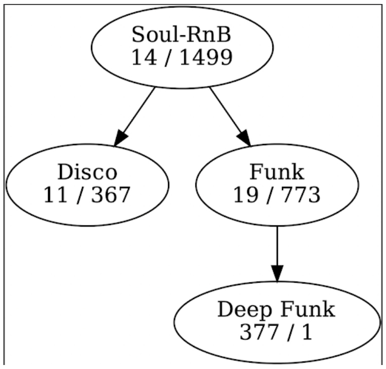

Figure 3: Genres per track

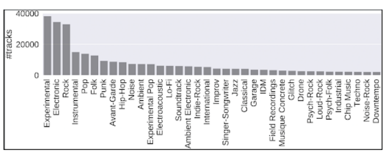

Figure 4: Genres per track

(3) **features.csv** - Some common features extracted using librosa.
(4) **echonest.csv** - Audio features by Spotify (previously known as Echonest) for 13129 tracks out of total tracks.

3.3 Multiple sizes of MP3-encoded audio data

(1) **fma_small.zip** - 8,000 tracks of 30s, 8 balanced genres (7.2 GB)
(2) **fma_medium.zip** - 25,000 tracks of 30s, 16 unbalanced genres (22 GB)
(3) **fma_large.zip** - 106,574 tracks of 30s, 161 unbalanced genres (93 GB)
(4) **fma_full.zip** - 106,574 untrimmed tracks, 161 unbalanced genres (879 GB)

4 BASELINE

Our baseline model is shown here in Figure 5. We finally got a training accuracy of 78% and validation accuracy of 67% which is pretty good for a dataset of just 8000 samples. The model was trained in 20 epochs with a learning rate of 0.0001. The graphs representing the training history is also shown here in Figures 6, 7, 8, 9.

Using this model we can get a latent representation of any song by taking a 30 second sample and then use it to find similar songs using cosine similarity.

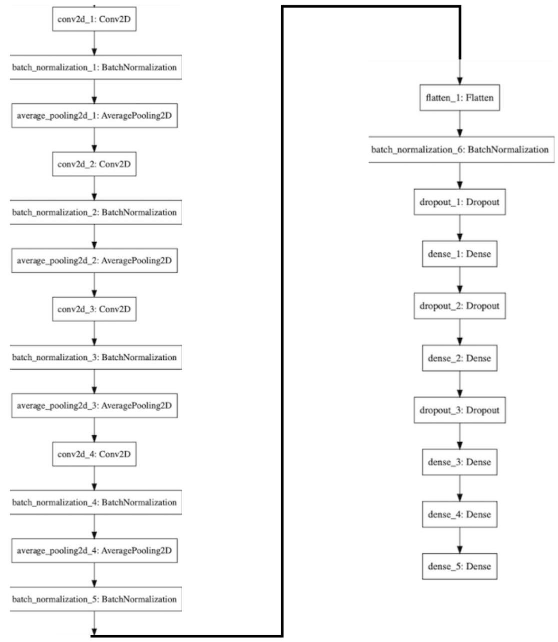

Figure 5: Baseline model structure

Drive link for the Keras weights: https://drive.google.com/drive/folders/1ch8F0XOf50H9PRuX3q0w2YcLaVnm9R_A

GitHub link for the project: https://github.com/AbhinavSE/IR2022_Project_8

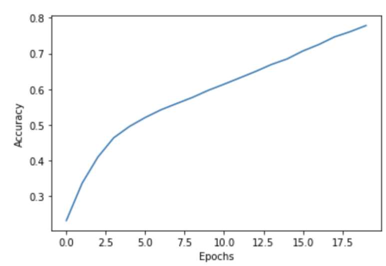

Figure 6: Accuracy v epochs

5 PROPOSED METHOD

Creating a model and creating a front end to manage following:

- User specific
- Search option
- Filtering music based genre
- Playlist creation
- Keeping track of history of music played
- Using model to recommend/filter music

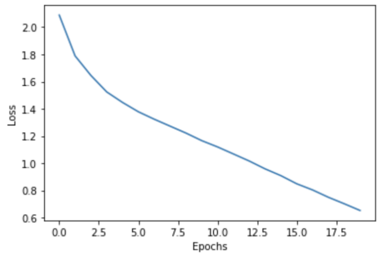

Figure 7: Loss v epochs

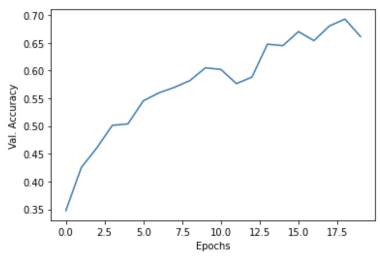

Figure 8: Validation Accuracy v epochs

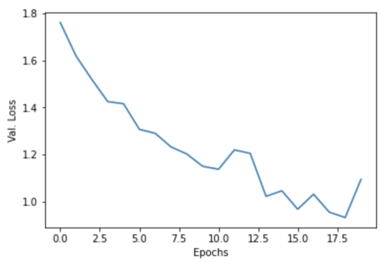

Figure 9: Validation Loss v epochs

We are using advanced methods KD-tree (data in each node is a K-Dimensional point in space) or Ball-tree to efficiently search our music.

6 PLAN OF WORK

Week 1 - 5
(Already done)
Choosing the exact problem statement and conducted existing literature review to get a sense of the problems faced in the current systems Chose the dataset that will be used for building the model

Week 6
(Already done)
The team is divided into sub-teams - A team that will build the model itself, the other will focus on building the frontend UI/UX. Both teams will start working on their respective micro-problem statements.

Week 7
(Already done)
For building the model, the focus would be shifted towards learning about various features that can be extracted from audio signals and then working on extracting that information from the dataset. Building the baseline model. The type of system, the architecture and the inner working of the system would then be decided.

Week 8 - 9
(Already done)
Building upon the baseline to make other models, evaluating them, and choosing the best one. Start working towards finalizing the recommendation model.

Week 10-12
Everyone shifts their focus and starts working on the frontend and integrating it with the backend.

Week 13-14
Adding finishing touches to the frontend. Evaluating the overall system. Going over the methodology followed, finding and fixing faults within the system. Writing up the report for the final presentation.

ACKNOWLEDGMENTS

We thank Dr Rajiv Ratn Shah, the professor of Information Retrieval course, for providing his constant help and support in the past, present, and the future. We also thank the Teaching Assistants for their guidance as well.

REFERENCES

[1] Sander Dieleman. 2014. Recommending music on Spotify with deep learning. https://benanne.github.io/2014/08/05/spotify-cnns.html
[2] Charlotte Hu. 2021. Why Spotify's music recommendations always seem so spot on. https://www.popsci.com/technology/spotify-audio-recommendation-research/
[3] Dip Paul and Subhradeep Kundu. 2020. A Survey of Music Recommendation Systems with a Proposed Music Recommendation System, Jyotsna Kumar Mandal and Debika Bhattacharya (Eds.). Springer Singapore, Singapore, 279–285.
[4] Daniel Roy. 2020. How Spotify recommender system works. https://www.linkedin.com/pulse/how-spotify-recommender-system-works-daniel-roy-cfa/
[5] Aaron van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep content-based music recommendation. In *Advances in Neural Information Processing Systems*, C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger (Eds.), Vol. 26. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2013/file/b3ba8f1bee1238a2f37603d90b58898d-Paper.pdf
[6] Gabriel Vigliensoni and Ichiro Fujinaga. 2016. Automatic Music Recommendation Systems: Do Demographic, Profiling, and Contextual Features Improve Their Performance?. In *ISMIR*.
[7] Xinxi Wang and Ye Wang. 2014. Improving Content-Based and Hybrid Music Recommendation Using Deep Learning. In *Proceedings of the 22nd ACM International Conference on Multimedia* (Orlando, Florida, USA) *(MM '14)*. Association for Computing Machinery, New York, NY, USA, 627–636. https://doi.org/10.1145/2647868.2654940