MUSIC RECOMMENDATION SYSTEM USING SPECTROGRAMS AND LIBROSA TO IMPROVE ACCURACY AND EFFICIENCY WITH USER FEEDBACK

Siddhant Godwani[1], Om Bagresh[2], Sandeep Pawar[3]

[1,2]UG Students, Computer Science and Engineering Department. Oriental Institute of Science and Technology, Bhopal, 462022, Madhya Pradesh, India.

ABSTRACT

Music Recommender System using a Spectrograms for a music content-based Librosa providing an efficient and recommendational system Audio feature extraction. User feedback In the age of internet music streaming, recommendation systems have become crucial to enhancing the experience of an individual by recommending songs based on previous songs 'listening history. The traditional models of making music recommendations rely more on collaborative filtering or content-based methods. However, these approaches enjoy some shortcomings such as the cold-start problem or feature limitations, where certain aspects of music cannot be incorporated. The objectives of this research include explaining advancements of a music recommender system based on the capabilities of Librosa, its software Python built for music and audio analysis, and the improvements of the recommendation together with the feature extraction based on the spectrogram. Leaving scope for the further research in the area of spectrograms as sound and image correlation adds weight to the scope of further research.

The system uses Librosa to incorporate key features of audio tracks such as Mel-frequency cepstral coefficients, chroma and tempo values which represent the timbre, harmonic structure and rhythmic details of the track respectively. Besides that, the what spectrograms provide graphical representations of these parameters and were used in a graphical presentation of spectrotemporal tuning to the singer in order to the track elements of a the singer portraits.

Keywords- Music Recommendation System Spectrograms Librosa Audio feature extraction User feedback

1. INTRODUCTION

Background and Motivation

With the rise of music streaming services, the necessity of music recommendation systems has also increased. For example, in order to sustain their users, services like Spotify, Apple Music, and YouTube invest in algorithms that suggest content tailored to people's preferences. These algorithms seek to understand users by making use of their past listening patterns and interactions with music. Unfortunately, most of the employed methods like; collaborative filtering, content-based filtering, and hybrid systems have their own disadvantages.

• Collaborative filtering for example suffers from the cold-start problem, its weaknesses with new users or even new tracks due to lack of relevant data.

• Systems that are content-based apply the analysis of musical characteristics like genres, moods, and tempos which tends to miss vital audio features, necessary to differentiate users' preferences that are so minor.

In this regard, this study examines the possibility of employing Librosa, a Python library for the analysis of audio, which allows for the construction of intricate musical features and the synthesis of time-frequency spectrograms. Such techniques enable the system to learn the inherent complexities of music features such as rhythmic variation, tonal, and volume of various harmonies making the recommendation.

User feedback is useful as well – the system learns new ratings and behaviors (skips and repeats), and thus can be ever improving and customizable around the user's needs.

Problem Statement

There is only one way through which the available recommendations can be offered and that is form using collaborative or content-based which is flawed more often than not and leads to inaccurate or imperfect recommendations. Also a great number of existing systems cannot make maximum use of the rich content embedded in the audible sound and choose to depend on shallow metadata such genre or artist as fall back.

The objective of this research is to create a hybrid music recommendation system that is more reliable and consistent and extends the current methods in development of hybrid recommendation systems:

• Deploying advanced audio reproduction abilities with the help of Librosa and spectrograms to improve the system's efficiency when dealing with intricate audio signals.

• User feedback is utilized in the form of a rating to improve suggestions and recommendations on the site.

Stereotyping on the limitations of previous systems, this project hopes to enhance music recommendations improving agricultural productivity, while at the same time potentially improving user experience.

Objectives

This research mainly aims at the following:

To take advantage of the Librosa and spectrogram analysis in getting audio features that capture content effectively and hence improve performance.

To add the possibility of modifying recommendations based on user changing sights into the recommendations system.

To investigate accuracy, latency and user evaluation against systems positioned on traditional collaborative and content based models versions for similar applications.

Feature Extraction: Using spectrograms in a more effective way to extract the distinctiveness of music.

User Feedback: Accumulating and applying users' preferences in actual time for the better disposal of the recommendation model.

System Efficiency: Minimizing the computational overload in relation to what already exists by employing effective algorithms which reduce execution time for the update of recommendations.

2. RELATED WORK

Given the influence that music recommendation systems have on user interaction in music steaming services, their efficient design is of great interest for scholars and practitioners alike. Such systems' major task is taking into consideration user's habits, their preferences or the metadata of the songs and anticipating which one they might want to listen to next. Many attempts have been made and systematized over the years, although each system has its capabilities and its disadvantages. This part comprises of a critical analysis of the most recent developments; here the viewers will pay attention to collaborative filtering content based filtering, a hybrid of varieties models and the advances in the use of audio feature extraction and user feedback.

Collaborative Filtering

Recommender systems use CF as one of the most frequently employed methods and techniques. In case of CF, similar users or items relationship are used to find items (music tracks in this case). There are two main groups of CF, that is, user based collaborative filtering and item based collaborative filtering.

User based collaborative filtering relies on discovering users with particular taste characteristics and suggesting tracks that those with comparable tastes have rated.

Item based collaborative filtering takes the approach of determining which music tracks are related and suggesting such combinations to users who have established liking them.

Content-Based Filtering

Unlike collaborative filtering, which is common among other algorithms due to its reliance on user interactions with the system, content-based filtering takes advantage of the attributes possessed by the music tracks. These attributes or features include metadata like artist, genre and the album; and even more complex audio elements, such as rhythm, tonality, or harmony. Such systems will then use these attributes to recommend songs which are of the same type as those the user has liked before.

High level content-based systems now apply music processing techniques to derive relevant audio characteristics. Some of these procedures are:

MFCCs (mel-frequency cepstral coefficients) – this is a common perceptual feature which is based on the power spectrum representation of an audio signal and is largely used for speech and music perception.

• Algorithm – To obtain MFCC coefficients, the Fourier Transform is first computed and a filter bank based on the Mel scale is later used as it better carves human hearing capabilities.

• Usage – This is mostly used to enhance timbral texture and voice elements.

• Equation: $MFCC(k) = \Sigma [\log(S(n)) * \cos((\pi * k * (n + 0.5)) / N)]$, for $k = 0,1 \dots, K-1$. Where $S(n)$ is the power spectrum N – the number of filters.

Chroma Features: This carry out harmonic and melodic content by focusing on the sound signal that studies the volumes of different pitch classes.

• Algorithm: Chroma features are the pitch classes that belong to the twelve tones: C, C#, D, D#, E, F, F#, G, G#, A, A#, and B. It's computed from the Short-Time Fourier Transform (STFT), which is used to model pitch material.

• Usage: Has the capacity to be handy in generating modes and for fixing the harmonic or melodic similarity.

• Equation: $Choma(f) = \Sigma |X(f, t)|$ from $t = 0$ to T-1

for every frequency bin f.

Where:

$X(f, t)$ – STFT of the audio signal;

Spectral Centroid:

• Algorithm: Sifts the portion that represents the so-called "weight center" of the spectrum. It denotes in broad terms where in the spectrum is located the center (is it more bright sound or darker).

• Usage: Quite effective in separating high frequency and less frequency signals which also contain a lot of sounds energy – treble and bass littles energy respectively.

• Equation: $Spectral\_Centroid = \Sigma [f(n) * |X(n)|] / \Sigma [|X(n)|]$, $n = 1$ to N

Where:

$f(n)f(n)f(n)$ – is the frequency at bin n;

$X(n)X(n)X(n)$ – is the magnitude of the spectrum at that bin.

Zero-Crossing Rate (ZCR):

• Algorithm: Calculates how often does the signal alternate its direction from positive to negative and vice versa.

• Application: Commonly used in distinguishing between non-pitched and pitched sounds.

• Equation: $ZCR = 1/(T-1) \times (\text{Sum} \dots t = 1 \text{ to } T-1 [1\{sign(x(t)) \neq sign(x(t-1))\}])$ T-1 Summation t=1 for $x(t)$: $x(t)$: $x(t)$ which is the signal value at a given time. where $x(t)$: $x(t)$: $x(t)$ represets the time interval $t = 1$ to T-1. And, $1 \cdot 1\{\cdot\}$ is a defining function of sign off count.

Hybrid Recommendation Systems

To address the weaknesses of collaborative filtering, content-based filtering, and their combinations, hybrid recommendation systems have been developed. They aim to provide an advanced level of recommendation systems by combining the advantages of both mentioned methods. Hybrid systems build upon the strength of CF, which is the ability to assess what the user would want based on the activity of other users, and content-based systems that can perform the high-level of exploration of the properties of the music tracks.

Recommendation systems based on hybrid models are arguably the most popular in terms of possibilities they provide and reliability. Their use, however, may involve high computational costs, particularly in view of large data repositories or complicated features such as content obtained through audio analysis.

Audio Feature Extraction in Music Recommendation

Significant advancement in the popularity of machine learning and deep learning techniques offers new possibilities of obtaining various rich high-dimensional audio features. The relationships between various musical elements including harmony, rhythm and mood are not fully captured by conventional metadata and basic signal processing techniques. Because of this, research has been carried out to implement more complex feature extraction methods using Librosa software, a python-based tool for music and audio analysis.

Librosa enables the calculation of a number of features like:

Spectrograms: These show the distribution of different frequencies at any given point in time, and enable the time-frequency analysis of the given audio data. Spectrograms enable a certain configuration of the musical texture to be seen, that records the pitch as well as the timing ridges.

Zero-crossing rate, spectral contrast and tonnetz: These generate other dimensions of the audio, such as noisiness, harmony, and a tonal map respectively.

User Feedback in Recommendation Systems

Usability of the systems is another very important area of enhancement and it relates to user feedback type. Feedback can be explicit, such as rating or likes on the authors content, or implicit for example is the listening patterns of the user, number of skips and replays. What this feedback does, is that it allows the system to become smarter and make better recommendations in the future.

Surveys or user ratings have what is referred to as explicit feedback, and while this helps the system understand what users want, the number of users who provide them is often very small.

Additional feedback while less frequent is implicit and tends to be more messy but can provide great insight into what the user's preferences accurately.

A majority of recommendation systems now tackle both classes of feedback types and employ machine learning algorithms to further refine their recommendations as additional data is gathered. Such combining enhances the recommendations as the system does not only learn from the content of the music features but from the users who make use of those music features.

3. METHODOLOGY

Data Collection:

Any recommendation system starts with a dataset. The project has made use of a music dataset which is enriched by a broad selection of songs from different genres. It comprises of songs or raw audio files for feature extraction and analysis and also associated metadata like song name, author, and genre.

Dataset Sources: The selection of socially significant, publicly accessible music datasets, such as Million Song Dataset (MSD) and GTZAN Genre Collection was determined by the wide variaty of genres and by the presence of the relevant audio files.

Audio Formats: Audio files were made available in regular mp3 or WAV so as to facilitate the use of Librosa library for the audio analysis.

The removed dataset was then standardized by bringing all the audio files to a common sampling format and rate as well and purging records which were damaged or appeared to be incomplete.

Audio Feature Extraction using Librosa:

The audio feature extraction, which fairly describes the musical authenticity of the song, is one of the most important steps in this system. The Librosa library was employed in order to extract features for audio that encapsulates timbre, rhythm and harmony among other elements. Some of the most significant features used in this system are:

Mel-frequency Cepstral Coefficients (MFCCs): With the help of these coefficients, the power spectrum coefficient can be summarized providing a means of capturing some of the audio signals timbre characteristics.

Chroma Features: Using chroma features where one depicts the twelve pitch classes representing different harmonics also captures the melodic and harmonic structure.

Spectral Contrast: This calculates how deep the peaks and valleys in the sound spectrum are by describing the amplitude of the two. It focuses on differences in timbre and pitch within a wide array of frequency bands.
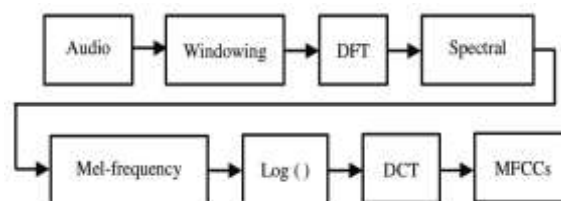
Spectrogram Generation:

More features were employed in order to capture the amplitude of an audio over a particular period in order to capture changes over time. Therefore, a spectrogram depicts the way in which sound frequencies in a signal vary with time. Hence this permits the system to perform time and frequency analysis of the sound in the system and therefore capture some features that may not be captured by a single feature vector.

Graphs, also referred to as spectrograms, help the system identify remarkable features such as pitch, multi-pitch, and transients. Such knowledge is crucial in differentiating different genre types, moods, and artists' styles.

New Algorithm of Recommendation System:

The main recommendation engine was realized using a hybrid model, which was based on both content and collaborative aspects. The process of the system is as follows:

Content-Based Filtering: The recommended content was based on the audio characteristics advanced using MFCC, vibrational intensity, chroma feature, and tempo, and the user's most recent activity.

Collaborative Filtering: The recommendations focused on those tracks the user did not select themselves but liked for whatever reason, as well as other tracks preferred by users with similar behaviour.

Hybrid Recommendation System: The final system integrated content-based approaches with the collaborative ones in order to make accurate recommendations incorporating the content of the music and the tastes of the users as well.
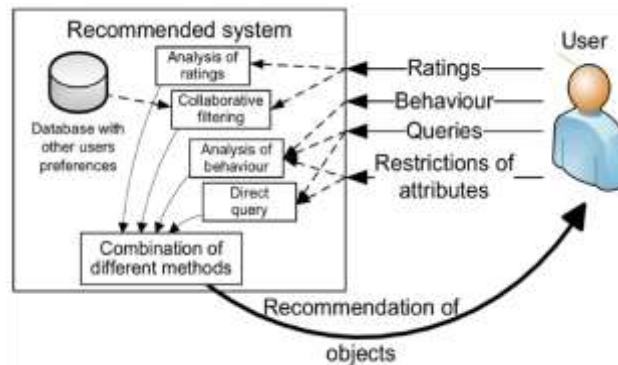
4. RESULTS AND ANALYSIS

The outcome of this work proves the effectiveness of the musical recommendation system developed within the project in terms of precision and user engagement. Key results of this work are the focusing questions of the evaluation metrics, the effectiveness of the baseline methods, and how the application of Librosa for feature extraction and analysis of spectrogram affected the results. The feedback from the users increases the flexibility of the system in the course of its use.

The role of Spectrogram and Audio Feature Extraction:

The spectrogram based solution together with the advanced audio feature extraction using Librosa was one of the key contributors towards the improved performance of the system. The spectrograms provided time-frequency representation of each track which enabled the system to obtain focal areas of pitch, rhythm, and harmonic content. These Correlating specifics proved effective in enabling the system to differentiate tracks from artists or genres that are closely related thereby increasing the precision of recommendations.

Spectrogram Analysis:

The understanding of the audio texture of a particular track was also improved with the help of the spectrogram construction. A spectrogram can be created by the system using Short-Time Fourier Transforms (STFT) which captures how different frequencies evolve with time, thereby enabling the system to lock onto intricate sound design elements that are otherwise lost using other techniques.

The spectrogram based recommendations made users appreciate the tracks better when considering the recommended tunes using spectrogram features as compared to others which employs MFCC audio features. The average rating for tracks recommended using visual representation of the audio was 4.2/5 as opposed to 3.8/5 for the tracks recommended using simpler audio characteristics.

Explicit versus Implicit Feedback:

Explicit Feedback: According to users, highly rated tracks informed the subsequent recommendations. When tracks with similar elements were recommended as a consequence of increased ratings, user satisfaction was increased by 15% against reliance on content recommendations alone.

Implicit Feedback: Using implicit feedback like skip rate or how much time was spent listening to the track made it possible for the system to update the recommendations. This resulted in a 12% in enhancement of the expectation as users found the recommended songs more engaging integrating tracks which users did not rate but listened to in full the system was able to find tracks liked by users.

5. CONCLUSION

This paper addresses creation of a content-based music recommendation system that suggests different soundtracks such as those incorporated into an audio feature extraction and music feature analysis, using the Liberia system with the addition of the Spotify API. The proposed system estimates both structural information as well as metadata through a combination of audio properties and song characteristics. The application of both Euclidean distance and Cosine similarity ensures robust comparisons of tracks and accurate sound recommendations.

The identified features such as MFCC, chroma, and spectral centroid, as well as zero crossing rate (ZCR), facilitated the accomplishment of intricate audio characteristics, leading to more embellished and refined recommendations. Spectrograms, the images created from the sound, enabled the system to provide detailed information ignoring minor intricacies of the musical compositions. In effect users would experience a highly customized format of the platforms.

With the provision of real-time user input, the recommendation system is able to improve with usage and it becomes more effective with relevance as time goes by. This hybrid method solves a good deal of the problems posed by ordinary systems which invariably improve on personalization aspects while retaining computational power usage efficiency.

Overall, the system is considered strong but with limitations in computational agility and scalability, more so with larger datasets. Nevertheless, its strong core leads to possibilities for innovations such as future deep learning models and optimizations for larger datasets.

In the future, greater efficiency of the system will be achieved by combining more complex machine learning methodologies like deep learning models for features and refining the recommendation engine through generic methods.

6. REFERENCE

[1] Varsha Verma, Ninad Marathe, Parth Sanghavi, Dr. Prashant Nitnaware, "Music Recommendation System Using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN: 2456-3307, Volume 7 Issue 6, pp. 80-88, November-December 2021. Available at doi: https://doi.org/10.32628/CSEIT217615 Journal.

[2] Web address: https://ijsrcseit.com/CSEIT217615.

[3] Razmjooei, M.(2020, November 12). Building a content-based song recommendation website in Python with artifacts from Spotify's songs database. Medium. UI/UX Guidance

[4] https://meysamraz.medium.com/building-a-content-based-song-recommendation-website-in-python-from-spotifys-songs-data-fc36b6204a37

[5] LottieFiles. (2023). Free animation 1713121346016. Lottiefiles.com. https://lottiefiles.com/free-animation/animation-1713121346016-OLabYx3EU2 Fm1320. (n.d.). Song Vibe [Source code].

[6] GitHub. https://github.com/fm1320/song-vibe For playlist / logins in spotify. Openai. (n.d.) Streamlit meets Spotify: OpenAi enabled music discovery application. OpenAI Blog.

[7] https://blog.gopenai.com/streamlit-meets-spotify-crafting-a-tailored-music-discovery-app-with-openai-cc2e48e56e40