EmoSense: Music mood detection and recommendation

Srishti Singh
srishti20409@iiitd.ac.in

Shreya Bhatia
shreya20542@iiitd.ac.in

Tarushi Gandhi
tarushi20579@iiitd.ac.in

Ashwin Sheoran
ashwin20288@iiitd.ac.in

Ashish Kamathi
ashish20364@iiitd.ac.in

Vedant Patil
vedant20348@iiitd.ac.in

Abstract

The use of smartwatches to obtain physiological data for music recommendation can overcome the challenges faced in music classification. Smartwatches are capable of monitoring physiological data such as heart rate, calories burned, steps taken etc. This data can be utilised to recommend appropriate music for a better workout experience. Music has been shown to have a significant impact on physical performance, improving motivation, focus, and endurance during exercise. By analysing the physiological data obtained from smartwatches, music recommendations can be personalised to the individual's fitness level and preferences, resulting in a more enjoyable and effective workout. This approach can also assist in injury prevention by recommending appropriate music to match the individual's pace and avoid overexertion.

***Keywords*—** I nformation Retrieval , Music Mood extraction , Sentiment Analysis, Recommender System , Physiological data , Smartwatches

1. Motivation

The increasing popularity of smartwatches, which are capable of monitoring fitness data, presents a unique opportunity to investigate the potential of AI and ML in music categorization and use of physiological data for better recommendation. This project aims to explore how listeners perceive and categorize music by utilizing machine learning methods to accurately categorize music based on its audio and lyrical content into different expressed emotions. Through machine learning, this project offers an exciting chance to better understand and appreciate the beauty of music and recommend them appropriately to user based on their physiological data like heart beat etc

2. Problem Statement

The objective of this project is to develop a mood detection system that can accurately predict the mood of a user based on the data collected from their Fitbit device (data like heart beats, pulse rate, SPO2, etc.) and recommend music accordingly. The aim is to create a personalised music experience for the user that can enhance their mood and overall wellbeing. The problem statement involves several challenges, such as processing large amounts of data collected from the Fitbit device, accurately identifying the user's mood based on the data, classifying the mood of the music based on lyrical and audio data, and recommending suitable music to match the users' mood. The proposed system needs to be accurate, efficient, and user-friendly to provide an enjoyable music experience to the user.

3. Related Work

3.1. Induced Emotion-Based Music Recommendation through Reinforcement Learning[1]

This paper introduces Moodify, a reinforcement learning (RL)-based music recommendation system that elicits emotions in the user to facilitate engagement in diverse contexts. The suggested RL approach assumes that the emotional state is dictated by a sequence of recently played music and learns how to choose music tracks that match a goal emotional state. As opposed to prior initiatives that suggested tunes for certain moods, Moodify creates an emotional state starting with a present emotion. The authors launched Moodify as a web application and carried out a pilot assessment study with 40 users and one million Spotify playlists. The findings point to excellent user satisfaction, responsiveness of the system, and suitability of the advice.

3.2. Artificial Neural Network (ANN) Enabled Internet of Things (IoT) Architecture for Music Therapy[6]

This paper proposes using music therapy as an alternative medicine technique to improve the well-being of patients suffering from pain, stress, and anxiety. An integrated system comprising IoT, BAN, and ANN is suggested to automate the music therapy process and provide immediate assistance to patients. The system involves monitoring patients' body parameters, categorizing their disease, and playing the most appropriate type of music over their handheld device. The ANN uses binary and categorical cross-entropy loss functions, Adam optimizer, and ReLU activation function to predict the patient's mood and suggest suitable music.

3.3. Music Emotion Classification based on Lyrics-Audio using Corpus based Emotion [5]

This article discusses how lyrics and audio can be used as features for music emotion classification. Corpus-based emotion (CBE) is employed, together with psycholinguistic and stylistic factors, to extract lyrical features. The extraction of energy, temporal, and spectral information is done for audio features. With an F-measure of 56.8 precent, Random Forest approach for both lyrics and audio elements produced the best result for classifying the emotions in music.

3.4. Multi-modal Music Emotion Classification based on audio and lyrics [3]

This paper uses multi-modal fusion emotion classification method based on audio and lyrics. Lyrics have been classified using Bert model, then LFSM based equalization was performed on the lyrics emotion classification results using the sentiment dictionary. For features in audio data they used Mel Frequency Cepstrum Coefficient, spectrum centroid and frequency-based energy distribution which are fed into LSTM model for music emotion classification. Their new fusion method achieved 5.77 percent and 4.03 percent improvement over the linear weighted multimodal and LASM fusion techniques.

3.5. Based on Improved Convolutional Neural Network [2]

The study describes a method for music emotion recognition that combines mel-frequency cepstral coefficient (MFCC) and residual phase (RP) to extract low-level audio features. Convolutional recurrent neural network (CRNN) is used to extract time-domain, frequency-domain, and sequence features of audio. Bidirectional long short-term memory (Bi-LSTM) network is used to obtain sequence information of audio features. The features are then fused and input into a softmax classification function with center loss function to recognize four music emotions. The proposed method achieved 92.06 percent accuracy, outperforming other methods. The method provides a new approach

3.6. Deep learning-based late fusion of multimodal information for emotion classification of music video [4]

The article discusses the creation of a diverse music video emotion dataset and its use in testing four unimodal and four multimodal convolutional neural networks (CNNs). The best unimodal classifier is integrated with corresponding music and video network features to create a multimodal classifier, which integrates whole music video features and uses a SoftMax classifier for final classification using a late feature fusion strategy. The multimodal structure achieved an accuracy of 88.56 percent, an f1-score of 0.88, and an AUC score of 0.987, demonstrating better performance than each unimodal emotion classifier.

4. Novelty

In our project, we aim to leverage the capabilities of smartwatches to collect physiological data and use it for music recommendation based on mood detection. This approach is novel, as it combines emotion detection from human body parameters with music recommendation, which has not been explored before. By utilizing smartwatches for this purpose, we can potentially enhance the overall health and fitness of individuals. Furthermore, the widespread availability and affordability of smartwatches make our project promising in terms of its potential impact on improving quality of life.

5. Methodology

Our system takes in the user's physiological data like heart beat rate and SPO2 levels (blood oxygen levels) to categorise their mood into three categories: Relaxed, Normal, and Stressed. The thresholds for heart rate and blood oxygen can be defined based on the literature reviewed so far. For example:

1. Relaxed: Heart rate below 60 bpm (beats per minute) and blood oxygen level above 95 percent.
2. Normal: Heart rate between 60 bpm and 100 bpm and blood oxygen level above 95 percent.
3. Stressed: Heart rate above 100 bpm or blood oxygen level below 95 percent.

To recommend songs to the user, we get them the top 10 songs that match their mood. For the sentiment analysis of a song, we use both the lyrics and audio features. We get songs from SpotifyAPI and use the VaderSentiment library for analysing the sentiment of the lyrics. TextBlob for natural language preprocessing and Lyricsgenius for accessing the Genius API, to get the sentiment scores of the lyrics. These sentiment scores have been assigned such that a score close to 1 corresponds to Happy songs, close to 0 is for neutral, and -1 for sad songs. After obtaining the audio features for each song in the dataset using the SpotifyAPI, we combine them with sentiment scores to create a comprehensive dataset. Then train a KMeans clustering model using this dataset, which groups songs based on their audio features and lyrical sentiment scores. While also attempting to make use of the listening history of the user through collaborative filtering, to get better recommendations. The purpose is to create clusters of similar songs so that when a user specifies their mood, the model can recommend songs from the cluster that best matches their mood. We then pick top 10 songs from the cluster representing the user's detected mood. We have made use of a vast and much newer dataset that pertains to the taste of present generations. For the evaluation purpose of the model, we selected five users and took their feedback as 'yes' or 'no' prompts to classify whether the songs played for them matched their mood or not, these are used to calculate precision, recall, and f1-score. Using the f1-score as the final metric, if it lies between 0 and 0.4, the recommendations aren't good enough for the user. Between 0.4 to 0.6 implies somewhat likeable recommendations, while anything above that implies excellent recommendations. Furthermore, if a user does not like a song, they can make this known using the two buttons in the GUI, and that song is taken out of possible future recommendations and replaced by the next best match. The GUI consists of an android native app that acts as a front end interface connected to backend and we run models and queries through REST API's posted on server with the help of Flask. Songs are now recommended after a thorough sentimental analysis of the lyrics AND the audio features from Spotify. The following audio features are used-['acousticness', 'danceability', 'energy', 'instrumentalness', 'liveness', 'loudness', 'speechiness', 'valence', 'sentiment'] After recommending songs, they are also played using the pygame library. After all are done playing, feedback is taken and f1-score is used as the main evaluation metric. In the event that the song isn't liked, feedback is taken and that song is replaced by the next best match.

Libraries and APIs used:

1. VaderSentiment: for analyzing the sentiment of lyrics
2. TextBlob: for natural language processing
3. Spotipy: for accessing the Spotify API
4. Lyricsgenius: for accessing the Genius API

External files used:

1. lyrics.csv, which contains the lyrics for all songs
2. spotifydata.json - contains audio factors obtained from the Spotify API

6. Database

We have used songs from Spotify. Here are some examples of what the Spotify API can return when called:

Search API: When you search for a keyword, the Spotify API returns a list of tracks, albums, artists, and playlists that match that keyword.

Top Tracks API: This API returns the most popular tracks for a specific artist based on their Spotify user data.

Recommendations API: This API returns a list of recommended tracks based on a variety of parameters, such as user listening history, mood, genre, and tempo.

Featured Playlists API: This API returns a list of featured playlists that are curated by Spotify based on different themes and genres.

The sample physiological data was gathered from an accessible fitbit user's watch.

7. Code

The code and README files can be found at our Github Repository.

8. Evaluation

Following techniques describe the state of art in evaluation metrics for recommendation systems: Personalization - assesses if a model recommends many of the same items to different users. A high personalization score indicates user's recommendations are different, meaning the model is offering a personalized experience to each user. Intra List Similarity - high intra list similarity indicates that the recommendation are of similar type. It is calculated as the average cosine similarity of all items in a list of recommendations. Precision@k is a fraction of top k recommended items that are relevant to the user P = (of top k recommendations that are relevant)/(of items that are recommended) Recall@k or HitRatio@k is a fraction of top k recommended items that are in a set of items relevant to the user. The larger the k, the higher the hit ratio since there is a higher chance that the correct answer is covered in recommendations. R = (of top k recommendations that are relevant)/(of all relevant items)

```
intra list similarity for  fin_happy = 0.64
intra list similarity for  mid_happy = 0.25
intra list similarity for  fin_sad = 0.79
intra list similarity for  mid_sad = 0.25
intra list similarity for  fin_nuetral = 0.91
intra list similarity for  mid_nuetral = 0.16
Personalization Score for final: 0.686721169655858
Personalization Score for mid: 0.6666666666666667
Precision@K for fin_happy (k=5) 0.4
Precision@K for mid_happy (k=5) 0.2
Precision@K for fin_sad (k=5) 0.6
Precision@K for mid_sad (k=5) 0.2
Precision@K for fin_nuetral (k=5) 0.4
Precision@K for mid_nuetral (k=5) 0.4
Precision@K for fin_happy (k=8) 0.5
Microsoft Windows [Version 10.0.22621.1555]
Recall@K for fin_sad (k=5) 0.6
Recall@K for mid_sad (k=5) 0.25
Recall@K for fin_nuetral (k=5) 0.4
Recall@K for mid_nuetral (k=5) 0.5
Recall@K for fin_happy (k=8) 0.6666666666666666
Recall@K for mid_happy (k=8) 0.6
Recall@K for fin_sad (k=8) 0.8
Recall@K for mid_sad (k=8) 0.75
Recall@K for fin_nuetral (k=8) 1.0
Recall@K for mid_nuetral (k=8) 1.0
```

9. Future Scope

Although the current smartwatches are capable of collecting and utilizing physiological data for music recommendation, they do not measure factors such as skin temperature, breathing, and skin conductance. By incorporating these additional metrics, it is possible to improve the accuracy of the emotion detection system, thus providing more accurate music recommendations. Furthermore, implementing a 5-scale emoji feedback system for music recommendations would enable users to provide more detailed emotional feedback on the recommended music. Combining this feedback data with music data can be used to develop a reinforcement learning model that would provide more personalized and accurate recommendations based on the user's emotional feedback. This approach would be more generalized and take into account the user's emotional response, leading to an even better overall music recommendation experience.

10. Acknowledgement

We are thankful to our instructor Dr. Rajiv Ratn Shah and the Teaching Assistants of our course Information Retrieval for guiding us through this project area and proposal.

References

[1] Roberto De Prisco, Alfonso Guarino, Delfina Malandrino, and Rocco Zaccagnino. Induced Emotion-Based Music Recommendation through Reinforcement Learning. *Applied Sciences*, 12(21):11209, Jan. 2022. Number: 21 Publisher: Multidisciplinary Digital Publishing Institute. 1

[2] Xiaosong Jia. A Music Emotion Classification Model Based on the Improved Convolutional Neural Network. *Computational Intelligence and Neuroscience*, 2022:6749622, 2022. 2

[3] Gaojun Liu and Zhiyuan Tan. Research on Multi-modal Music Emotion Classification Based on Audio and Lyirc. In *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, volume 1, pages 2331–2335, June 2020. 2

[4] Yagya Raj Pandeya and Joonwhoan Lee. Deep learning-based late fusion of multimodal information for emotion classification of music video. *Multimedia Tools and Applications*, 80(2):2887–2905, Jan. 2021. 2

[5] Fika Rachman, Riyanarto Sarno, and Chastine Fatichah. Music Emotion Classification based on Lyrics-Audio using Corpus based Emotion. *International Journal of Electrical and Computer Engineering (IJECE)*, 8:1720, June 2018. 2

[6] Shama Siddiqui, Rory Nesbitt, Muhammad Zeeshan Shakir, Anwar Ahmed Khan, Ausaf Ahmed Khan, Karima Karam Khan, and Naeem Ramzan. Artificial Neural Network (ANN) Enabled Internet of Things (IoT) Architecture for Music Therapy. *Electronics*, 9(12):2019, Dec. 2020. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute. 2