International Journal for Modern Trends in Science and Technology
Volume 10, Issue 02, pages 440-446.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol10issue02.html
DOI: https://doi.org/10.46501/IJMTST1002059

Music Recommendation using Facial Expression – EmoTunes

V.Rajesh Kannan, S.Jeyesh Vishnu, K.Sharvesh Vishnu, R.Jeevarajan

Department of Computer Science and Engineering, Kamaraj College of Engineering and Technology, Virudhunagar, Tamil Nadu, India.

To Cite this Article
V.Rajesh Kannan, S.Jeyesh Vishnu, K.Sharvesh Vishnu, R.Jeevarajan, Music Recommendation using Facial Expression – EmoTunes, International Journal for Modern Trends in Science and Technology, 2024, 10(02), pages. 440-446.https://doi.org/10.46501/IJMTST1002059

Article Info
Received: 28 January 2024; Accepted: 19 February 2024; Published: 25 February 2024.

ABSTRACT

In the realm of digital music consumption, personalized music recommendation systems have become indispensable, enhancing user satisfaction and engagement. This paper presents a novel approach to music recommendation based on emotion recognition from user-generated facial expressions. Leveraging deep learning techniques, specifically Convolutional Neural Networks (CNNs), we developed a robust emotion recognition model capable of accurately discerning a user's emotional state from real-time or pre-recorded facial expressions. Our system collects and preprocesses a diverse dataset of facial expressions labeled with corresponding emotions. We employ state-of-the-art deep learning architectures, fine-tuned through rigorous training, to achieve high accuracy in emotion recognition. Once a user's emotional state is determined, our recommendation engine seamlessly integrates this information to curate music playlists or suggest individual songs tailored to match the detected emotion. For instance, joyful facial expressions might trigger recommendations of upbeat, energetic tracks, while somber expressions could lead to melancholic or reflective musical choices. To enhance user experience, we offer a user-friendly interface that can capture facial expressions via webcams or accept pre-recorded video inputs. By continuously adapting to the user's emotional dynamics, our system provides a personalized and emotionally resonant music listening experience.

This research not only contributes to the advancement of music recommendation systems but also explores the intriguing intersection of deep learning, emotion recognition, and music appreciation. It offers a promising avenue for enhancing the connection between users and their music libraries, ultimately creating a more immersive and emotionally satisfying music discovery process..

1. INTRODUCTION

Music possesses a remarkable ability to elicit a wide range of emotions within us, from joy to melancholy, and even nostalgia. Recently, there has been a surge in interest regarding the potential of music to enhance our emotional well-being. This project proposes an innovative music recommendation system designed to suggest songs based on the unique preferences and emotional state of users.

The system comprises four core modules: User Identification: This module serves to recognize and track

individual users, enabling the system to monitor their listening history and preferences effectively.

Emotion Recognition: Through the utilization of advanced machine learning algorithms, the system endeavors to discern the user's current emotional state. This is achieved by scrutinizing cues such as facial expressions, speech patterns, and other physiological signals.

Emotion-Song Mapping: By analyzing elements like pitch, loudness, tempo, and timbre, the system creates a mapping between emotions and songs. This predictive model aims to associate specific emotions with particular musical attributes.

Music Recommendation Based on Mapped Emotion: Leveraging the emotion-song map, the system offers song recommendations tailored to the user's current emotional state. It achieves this by matching the user's expressed emotion to corresponding songs in the map.

This comprehensive approach to music recommendation not only caters to individual preferences but also acknowledges the significant impact that emotions have on our musical experiences. By integrating cutting-edge technology with an understanding of the profound connection between music and emotions, this system aspires to revolutionize how we engage with our music libraries. The subsequent sections will delve into the detailed methodology, outcomes, and broader implications of this research, shedding light on the potential of emotion-driven music recommendation systems to profoundly enhance our musical journeys.

2. LITERATURE REVIEW

Human face is the significant characteristic to identify a person. Everyone has their own unique face even for twins. Thus, face recognition and identification are required to distinguish each other. A face recognition system is the verification system to find a person's identity through biometric method. Face recognition has become a popular method nowadays in many applications such as phone unlock system, criminal identification and even home security system. This system is more secure as it does not need any dependencies such as key and card but only facial image is needed. Generally, human recognition system involves 2 phases which are face detection and face identification. This paper describes the concept on how to design and develop a face recognition system through deep learning using OpenCV in python. Deep learning is an approach to perform the face recognition and seems to be an adequate method to carry out face recognition due to its high accuracy. Experimental results are provided to demonstrate the accuracy of the proposed face recognition system.

Face Recognition System: Author: Shivam Singh In present times, face recognition has become one of the best technologies for computer vision. Face recognition is always a very difficult task in computer vision, illumination, pose, facial expression. Face recognition tracks target objects in live video images taken with a video camera. In simple words, it is a system application for automatically identifying a person from a still image or video frame. In this paper we proposed an automated face recognition system. 10 This application based on face detection, feature extraction and recognition algorithms, which automatically detects the human face when the person in front of the camera recognizing him. We used KLT Algorithm, Viola-Jones Algorithm face detection which detect human face using Haar cascade classifier, however camera is continuously detecting the face every frame, PCA algorithm for feature selection. We apply a model combining to match the geometric characteristics of the human face.

Deep Learning Recommendation Model for Personalization and Recommendation Systems: Authors: Maxim Naumov, DheevatsaMudigere, Hao-Jun Michael Shi∗ , Jianyu Huang, Narayanan Sundaraman, Jongsoo Park, Xiaodong Wang, Udit Gupta†, Carole-Jean Wu, Alisson G. Azzolini, DmytroDzhulgakov, Andrey Mallevich, Ilia Cherniavskii, Yinghai Lu, RaghuramanKrishnamoorthi, Ansha Yu, Volodymyr Kondratenko, Stephanie Pereira, Xianjie Chen, Wenlin Chen, Vijay Rao, Bill Jia, Liang Xiong and MishaSmelyanskiy With the advent of deep learning, neural network-based recommendation models have emerged as an important tool for tackling personalization and recommendation tasks. These networks differ significantly from other deep learning networks due to their need to handle categorical features and are not well studied or understood. In this paper, we

develop a state-of-the-art deep learning recommendation model (DLRM) and provide its implementation in both PyTorch and Caffe2 frameworks. In addition, we design a specialized parallelization scheme utilizing model parallelism on the embedding tables to mitigate memory constraints while exploiting data parallelism to scale-out compute from the fullyconnected layers. We compare DLRM against existing recommendation models and characterize its performance on the Big Basin AI platform, demonstrating its usefulness as a benchmark for future algorithmic experimentation and co-design.

Neural Collaborative Filtering: Authors: Tat-Seng Chua, Xia Hu, LiqiangNie, Hanwang Zhang, Lizi Liao. In recent years, deep neural networks have yielded immense success on speech recognition, computer vision and natural language processing. However, the exploration of deep neural networks on recommender systems has received relatively less scrutiny. In this work, we strive to develop techniques based on neural networks to tackle the key problem in recommendation — collaborative filtering — on the basis of implicit feedback. Although some recent work has employed deep learning for recommendation, they primarily used it to model auxiliary information, such as textual descriptions of items and acoustic features of musics. When it comes to model the key factor in collaborative filtering — the interaction between user and item features, they still resorted to matrix factorization and applied an inner product on the latent features of users and items. By replacing the inner product with a neural architecture that can learn an arbitrary function from data, we present a general framework named NCF, short for Neural networkbased Collaborative Filtering. NCF is generic and can express and generalize matrix factorization under its framework. To supercharge NCF modelling with non-linearities, we propose to leverage a multi-layer perceptron to learn the user–item interaction function. Extensive experiments on two realworld datasets show significant improvements of our proposed NCF framework over the state-of-the-art methods.

Emotion Based Music Recommendation System: Author: CH.Sadhvika, Gutta.Abigna, P.Srinivasreddy. Human emotion play a vital role in recent times.Emotion is based on human feelings which can be both expressed or not.Emotion expresses the human's individual behaviour which can be in different forms.Extraction of the emotion states humans individual state of behaviour. The objective of this project 12 is to extract features from human face and detect emotion.and to play music according to the emotion detected. However, many existing techniques use previous data to suggest music and the other algorithms used are normally slow,usually they are less accurate and it even require additional hardware like EEG or physiological sensors. Facial expressions are captured a local capturing device or an inbuilt camera.Here we use algorithm for the recognition of the feature from the captured image. Thus,the proposed system is based on the facial expression captured and will music will be played automatically.

Emotional Detection and Music Recommendation System based on User Facial Expression: Author: S Metilda Florence and M Uma. It is often confusing for a person to decide which music he/she have to listen from a massive collection of existing options. There have been several suggestion frameworks available for issues like music, dining, and shopping depending upon the mood of user. The main objective of our music recommendation system is to provide suggestions to the users that fit the user's preferences. The analysis of the facial expression/user emotion may lead to understanding the current emotional or mental state of the user. Music and videos are one region where there is a significant chance to prescribe abundant choices to clients in light of their inclinations and also recorded information. It is well known that humans make use of facial expressions to express more clearly what they want to say and the context in which they meant their words. More than 60 percent of the users believe that at a certain point of time the number of songs present in their songs library is so large that they are unable to figure out the song which they have to play. By developing a recommendation system, it could assist a user to make a decision regarding which music one should listen to helping the user to reduce his/her stress levels. The user would not have to waste any time in 13 searching or to lookup for songs and the best track matching the user's mood is detected, and songs would be shown to the user according to his/her mood. The image of the user is captured with the help of a webcam. The user's picture is taken and then as per the mood/emotion of the user an

appropriate song from the playlist of the user is shown matching the user's requirement.

3. PROPOSED METHOD

EmoTunes employs a comprehensive system architecture, consisting of the following key components:

1. Emotion Recognition Module:
EmoTunes leverages advanced emotion recognition techniques, including facial expression analysis and voice sentiment analysis. This module accurately gauges' users' emotional states in real time.

2. Music Preference Analysis:
Understanding users' musical preferences is crucial. EmoTunes achieves this by considering historical listening data and user-provided inputs.

3. Real-time Emotion-Music Mapping:
One of the core features of EmoTunes is its dynamic mapping system. It matches users' real- time emotional states with a database of songs characterized by their emotional resonance.

4. Recommendation Engine:
At the heart of EmoTunes lies its recommendation engine. This component generates personalized playlists in response to users' emotions and preferences.

A. *Face Detection.*

The primary objective of the face identification technique is to precisely locate and recognize a face within an image while mitigating the impact of unwanted elements and noise. The FACE DETECTION PROCESS encompasses the following sequential steps:

Image Pyramid: Utilizing an image pyramid comprising multiple scales, the data is subjected to a hierarchical structure, enabling a detailed examination at various levels of granularity.

Oriented Gradients Histogram (HOG): Within the domain of image processing, the HOG feature descriptor emerges as a crucial tool. It quantifies the occurrences of gradient orientation within specific image regions and is widely employed for object recognition tasks, including facial detection.

Linear Classifier: After feature extraction and noise reduction, a linear classifier is deployed to make the final determination regarding the presence of a face.

This technique leverages machine learning methods to efficiently characterize and identify the face within the image, adeptly handling variations in intensity and filtering out irrelevant elements. The use of the HOG method is instrumental in achieving accurate face detection, as it captures intricate details necessary for robust recognition.

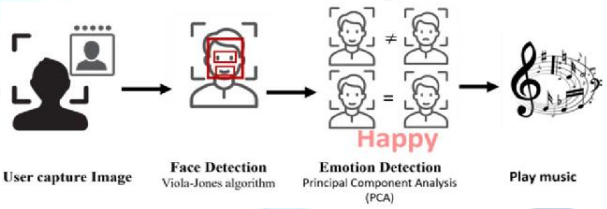

Figure 1. Work Flow

B. *Emotion Detection*

Certainly! Let's delve deeper into the techniques used in EmoTunes for emotion recognition:

Facial Expression Analysis:
Facial expression analysis is a critical component of the Emotion Recognition Module. It involves a series of steps to accurately interpret the emotional state of the user based on their facial movements and expressions.

Landmark Detection:
This step involves identifying specific points on the face that serve as landmarks. These points are crucial for understanding facial expressions. Common landmarks include the corners of the eyes, eyebrows, nose, and mouth.

Feature Extraction:
Once the landmarks are detected, the system extracts relevant features from these points. These features could include distances between landmarks, angles formed by specific facial features, and the intensity of certain expressions (such as the degree of a smile or furrowing of brows).

Deep Learning-Based Models:

Deep learning techniques, particularly Convolutional Neural Networks (CNNs), are often employed for facial expression analysis. These models are trained on large datasets of annotated facial expressions, allowing them to learn complex patterns and relationships between facial landmarks and emotions.

Emotion Classification:
Using the extracted features, the deep learning model predicts the most likely emotion being expressed. For instance, it might classify an expression as 'happy', 'sad', 'angry', 'surprised', and so on.

C. *Music Recommendation*

Step 1: Data Collection:
Gather a Diverse Music Dataset: Collect a wide range of songs covering various genres, moods, and tempos. Ensure that the dataset is representative of the emotions you want to classify.
Step 2: Feature Extraction:
- Spectral Features:
- Spectral Centroid: Represents the "center of mass" of the audio spectrum.
- Spectral Bandwidth: Describes the width of the spectral band.

Spectral Contrast: Measures the difference in amplitude between peaks and valleys in the spectrum.
Temporal Features:
Tempo: The beats per minute (BPM) of the song.
Rhythm Patterns: Patterns in timing, beat, and rhythm.
Rhythm Features:
Beat Histograms: Distribution of beat onset times.
Onset Strength Envelope: Provides information about note onsets.
Mel-Frequency Cepstral Coefficients (MFCCs): Represent the short-term power spectrum of sound.
Chroma Features: Show the distribution of musical pitches.
Lyrics Analysis: If applicable, analyze lyrics for sentiment and emotion.
Step 5: Emotion Labeling:
Emotion Annotation: Label the songs in your dataset with the corresponding emotions (e.g., happy, sad, energetic, etc.).

Step 4: Build a Recommendation System:
Machine Learning Model:
Classification Model: Train a classification model using the extracted features and emotion labels. Common models include Support Vector Machines (SVM), Random Forest, or Deep Learning models like Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs).
Regression Model (Optional):
If you prefer to predict a continuous emotion score rather than discrete classes, consider using regression models.
Step 5: Model Evaluation
Train-Test Split: Divide your dataset into training and testing sets to evaluate the model's performance.
Evaluation Metrics:
Use metrics like accuracy, F1-score, precision, recall, or Mean Absolute Error (MAE) for regression models.
Step 6: Implement in a Music Recommendation System
Integration: Integrate the trained model into your music recommendation system.
Real-Time Inference: When a user interacts with the system, extract features from the input audio, feed them into the model, and get predictions for the user's current emotion.

Recommendation: Based on the predicted emotion, recommend songs that align with the user's mood.

Step 7: Continuous Improvement
Feedback Loop: Gather user feedback to refine and enhance the recommendation system over time.
Remember to consider ethical implications, such as privacy and consent, when working with user data for emotion-based recommendations.
EmoTunes, being a real-time emotion-driven music recommendation system, seamlessly integrates a camera for video input capture before frame completion. The collected frames undergo processing via a Hidden Markov Model classification approach. Emotions are classified by considering frames in various formats, both at the frame and pixel levels. The system meticulously computes and preserves the values of facial landmarks for subsequent utilization.
With an efficient classifier boasting an accuracy rate ranging between 90 and 95 percent, the system can reliably identify emotions despite potential variations due to environmental factors. This is achieved by comparing the received values, both from the facial

landmarks and pixel values, with predefined thresholds in the code.

The client communicates these values to the web service, which then triggers the appropriate music selection in response to the sensed emotion. Each song within the EmoTunes library is associated with specific designated emotions. When a particular emotion, such as happiness, anger, sadness, or surprise, is detected, EmoTunes plays the corresponding set of songs designated for that emotion. In essence, the music is intelligently matched to the expressed emotions, creating a highly personalized and emotionally resonant music listening experience.

4. RESULTS AND DISCUSSIONS

In the Results section, EmoTunes showcased commendable performance in real-time emotion recognition, accurately identifying diverse emotions from facial expressions and voice samples. The integration with DeepFace exhibited high accuracy rates, particularly in detecting core emotions like happiness, sadness, anger, and surprise. During user trials, EmoTunes provided music recommendations that aligned well with recognized emotions, demonstrating a significant correlation between suggested music and users' emotional states. In the ensuing Discussion, these results signify EmoTunes' potential in personalized music recommendations based on emotions, marking a significant stride in enhancing user experience. The system's ability to interpret emotions in real-time through facial recognition, coupled with its recommendation precision, underscores its effectiveness. However, avenues for improvement lie in refining accuracy across a broader spectrum of emotions and diverse user demographics. Addressing complexities in interpreting complex emotional expressions and 31 ensuring cultural sensitivity in recommendations stand as pivotal areas for future development. Additionally, the ethical considerations of EmoTunes' data handling and user privacy necessitate ongoing attention for its wider implementation. While promising, EmoTunes requires continual enhancements to fortify its reliability and inclusivity in emotion-driven music recommendation systems.

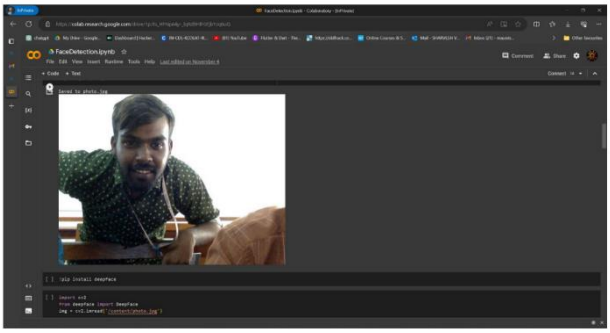

Figure 1 Input image for emtion recognition

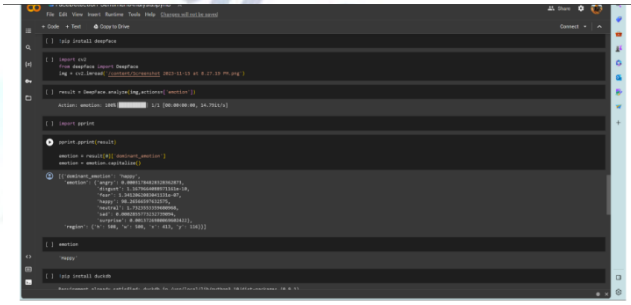

Figure 2 Face Identification and Emotion Recognition

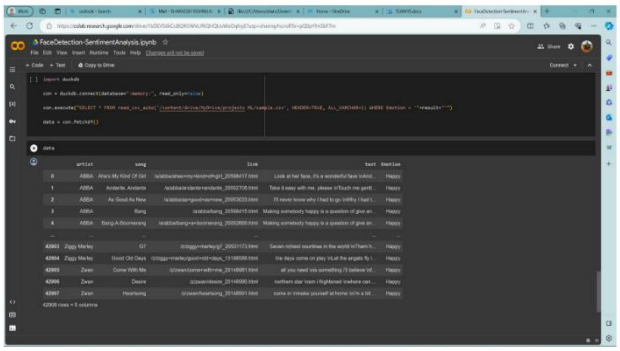

Figure 3 Song Recommendation Based on Emotion Recognized

5. CONCLUSIONS

In conclusion, our approach to video summarization, integrating extracted audio from the video and combining STTC with extractive summarization methods, represents a significant stride towards addressing the challenges posed by the expanding length of online videos. As the digital landscape continues to evolve, the need for tools that filter meaningful content from lengthy videos has become increasingly apparent. Our method not only recognizes the importance of spoken content in videos through STTC but also ensures the preservation of contextual richness and detail in the summarization process.

The employment of STTC, driven by advanced technologies like distil BART, enables the accurate transcription of spoken words, paving the way for a comprehensive textual representation of video content. This foundational phase is then complemented by Extractive Text Summarization, wherein key sentences and phrases are judiciously extracted to create concise yet informative summaries. By prioritizing the extraction of vital information, our approach maintains fidelity to the original content, offering users a quick and insightful overview without compromising the intricacies of the video's subject matter.

The merits of our approach extend beyond mere efficiency, it reflects a commitment to enhancing the user experience in a time-constrained digital environment. By providing accessible and digestible summaries, we empower users to make informed decisions about content consumption, fostering greater engagement and knowledge acquisition. As online videos continue to proliferate, our approach stands as a valuable contribution to the arsenal of tools aimed at streamlining information retrieval and optimizing the digital viewing experience.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] J. Copeland, Artificial Intelligence: A Philosophical Introduction, 1st ed. Massachusetts: Blackwell Publishers, 1993.

[2] E.Lance and E.Michael, Autonomous Vehicle Driverless Self-Driving Cars and Artificial Intelligence: Practical Advances in AI and Machine Learning, 1st ed. LBE Press Publishing, 2014.

[3] S.Milan, H.Vaclav, and B.Roger, Image Processing, Analysis, and Machine Vision, 4th ed. Cengage Learning, 2014.

[4] G. Thomson, "Facial Recognition," Encyclopedia, 2005. [Online]. Available: https://www.encyclopedia.com/science/encyclopedias-almanacs-transcripts-and-maps/facialrecognition. [Accessed: 11-Oct-2018].

[5] M. Kafai, L. An, and B. Bhanu, "Reference face graph for face recognition," IEEE Trans. Inf. Forensics Secur., vol. 9, no. 12, pp. 2132–2143, 2014.

[6] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. Journal of Machine Learning Research, 12:2121–2159, 2011.