An empirical study on Music recommendation Using Deep learning

Pranav Shinde, Rutik Jangam, Hritik Dhende, Aditya Thorat, Dr. Shalaka Deore.

[1] *Pranav Shinde, Computer Dept., Modern Education Society's College of Engineering Pune, Maharashtra, India*
[2] *Rutik Jangam ,Computer Dept., Modern Education Society's College of Engineering Pune, Maharashtra, India*

[3] *Hritik Dhende, Computer Dept., Modern Education Society's College of Engineering Pune, Maharashtra, India*

[4] *Aditya Thorat, Computer Dept., Modern Education Society's College of Engineering Pune, Maharashtra, India*

[5] *Dr. Shalaka Deore, Computer Dept.,Modern Education Society's College of Engineering Pune, Maharashtra, India*

ABSTRACT

Abstract — *Sentiment has performed an essential role in research on music recommender systems as being one of the primary elements that impacts human music listening behavior. Music is a fundamental sensation that almost everyone enjoys. However, listeners are frequently confused owing to individual preferences and the vast number of music options available, with new material being released on a daily basis. One of the most important aspects in determining what music to listen to is an individual's personal mood. Multiple forms of studies have been conducted on this method for the objective of determining an individual's mood through the usage of face photographs. For the purposes of applying our methodology, these approaches were examined for their relevance in image processing and machine learning implementations. This examination of various methodologies was crucial in the development of our strategy, which will be broadened in the future.*

Keyword : - *Convolutional Neural Networks, Fuzzy Classification.*

I. INTRODUCTION

Expression is the result of a complicated interplay between joints and muscle fibers in the person's face. Multiple medical terminologies and other associated techniques can be used to describe this physically. However, it is even more complicated and is largely used by people for interaction. Linguistics and expressions are used extensively in human meetings and dialogue. The majority of facial expressions and mannerisms offer a great deal of knowledge about what the person is saying. As a result, it is an essential aspect of language and interactions, both of which have a significant effect in an individual's general well-being.

Homo sapiens are extremely sociable beings, which distinguishes from those other species inhabiting this earth. Individuals have intricate relationships that are extremely beneficial to the psychological and physiological wellbeing of individuals as well as groups of persons. People are exceptionally skilled at detecting human faces thanks to their brains' built-in ability to recognize expressions quickly. Expressions are made up of the bulk of human gestures that are employed for successful and perceptive dialogue and collaboration.

Face expressions may be quite important in identifying the conversation's intent and interpretation on their own. Computer systems are incapable to distinguish a face or a facial picture from every other image, hence this technique is not as developed. This is because the computer treats the picture as a compilation of pixel values instead of recognizing the image's features. This is a very different psychological approach from how individuals analyze and develop knowledge about the outside world and society.

As a result, there is a requirement for a solution that can assess the attitude as well as what it represents based on facial characteristics. Many face expressions in humans are linked to whatever the subject wishes to communicate or how they are feeling at the moment. This provides a reference point for other people to recognize and engage with this person by utilizing their facial expressions as an indicator for the feelings they are experiencing. As a result, facial expressions must be discovered and employed efficiently for the aim of assessing a person's emotions with extraordinary precision. Music is considered one of the most prominent services that plays an important role in people’s enjoyment activities since technology and science have progressed and people’s lifestyle standards have improved. In the presence of huge music collections, assisting individuals in selecting the appropriate music for various situations is an important duty.

Traditional songs recommender systems instances, such as soundtrack and radio suggestions, can only really analyze users' tastes based on constrained interactions and ambiguous feedbacks, which appears to work well for portraying users' long-term musical tastes but is tricky to handle with in immensely changing circumstances. Increasingly, academics have emphasized on user sentiment in music suggestion, demonstrating the relevance of the user's feelings in boosting the suggestions' performance. Therefore, this survey paper utilizes a number of different approaches and analyzes them in this publication to reach our approach which will be elaborated in the upcoming editions of this research.

This literature survey paper segregates the section 2 for the evaluation of the past work in the configuration of a literature survey, and finally, section 3 provides the conclusion and the future work.

II RELATED WORKS

H. Zhang et al. [1] explain a cross-cultural study to examine the emotional preferences of music in daily activities for context-aware music selection. The authors proposed EmoMusic, a unique emotion-based music recommendation service for everyday activities that provides explanation and control of music emotion through an interactive interface, based on the emotional map between activity and music obtained from the research. User research was presented to evaluate the EmoMusic application, and participants provided positive comments. The results demonstrate that the service is usable and that visualization and interaction in a music emotion space can increase acceptance of music recommendations.

M. Bakhshizadeh et al. presents a system for deriving listening moods from users' listening history by grouping music tracks. A unique strategy of displaying these clusters to illustrate the listening moods in a user-friendly manner is shown in this research. The grouping depends on the audio qualities of the music songs, which include acoustics, danceability, energy, mode, speechiness, and valence [2]. This grouping results in the creation of customized music playlists. Furthermore, the case study supports the idea that it is fair to limit the suggestions to songs with the most comparable audio attributes since users tend to stay in the same listening mood rather than change it quickly. However, a large number of users must be investigated to get a general and reliable judgment on the efficacy of the given assumption.

C. Zhou et al. innovatively include music suggestions into the dialogue system and employ an interactive "user ask, system react" approach [3]. A clever recommender will elicit more interactions and so perform better. In the suggested scenario, the recommender is tasked with satisfying the aforementioned interactions while also providing contextual and real-time results. As a result, the bandit-based method is used in this study to tackle the real-time interactive recommendation issue. Bandit is a type of online recommendation system that has exceptional performance and is used in a variety of extremely dynamic shifting circumstances.

C. Dhahri et al. present a mood-based music recommendation system that selects songs that fit the user's mood without knowing the user's preferences in advance. The authors first anticipate the user's mood depending on their public social profiles. The user's mood is distinct into three categories: good, negative, and neutral. Then, using data obtained from online blogging services, they integrate the user's mood and song's emotion in a latent space to learn a generic function [4]. To conquer the previous drawback the authors use a merger of a reinforcement learning (RL) framework and a statistical test termed as Page-Hinkley (PH) test, in which the recommendation process is iterative and adaptive utilizing a change point detection approach. The algorithm changes the song's selection function based on implicit input at each trial and permits a change in the song's position over the song map when a change in the user's implicit feedback is noticed. Using the user's implicit feedback, the computer gradually creates a customized map per mood for each user from a generic song map.

K. Sakurai et al. propose a novel music playlist generating approach based on reinforcement learning and acoustic feature mapping. Using a dimensionality reduction approach, potential song features are mapped to a space of a specific reinforcement learning environment in the suggested method. Then, based on the user's preferences, the authors place an agent in the environment [5]. They train the agent to identify the best behavior in the feature map environment, taking into account all of the potential songs. Finally, they construct the playlist depending on the behavior of the trained agent. Because the agent is trained based on the distribution of all songs, the suggested technique may build a playlist with a high degree of diversity and seamless track transitions.

Z. Yu et al. proposed a convolutional neural network-based facial microexpression detection model (CNN). After model training using the FER2013 data set, the scientists attained an identification rate of 62.1 percent. The content-based recommendation system was utilized to automatically recommend music to users depending on the spotting of both facial expression and mood [6]. Unlike prior algorithms that just propose music depending on the users' historical listening preferences, the algorithm presented in this research increases the user's emotion recognition, allowing the recommended music to better meet the users' listening expectations.

B. Kostek presents some concepts for a music recommendation system based on social collaborative filtering. Music information retrieval (MIR), Query-by-Example (QBE), Query-by-Category (QBC), music content, music annotation, bridging semantic gaps in the music domain, and other words pertinent to music recommendation systems are defined [7]. The history of music recommender systems is briefly addressed, and the role of machine learning vs statistics in recommender system functioning is briefly studied.

K. Kittimathaveenan et al. introduce the Color-to-Music technique that provides an alternate method for selecting songs based on a color selection. This research was divided into three stages: the initial stage was the construction of a music library for the correlation of color and emotion, as well as the association of music and emotion [8]. The library data used to create the Hue, Saturation, and Value color model were as follows: Hue represented musical instruments, Saturation alluded to pace, and Value was critical. The second stage was to design two different types of graphical user interfaces for color choosing. The final stage was to collect data from 120 participants.

H. Han et al. present the content similarity approach which is utilized for music recommendation, and the content is expressed by the MFCC feature value. The recommended method is essentially split into three steps: first, the music feature vector is computed; second, the distance between the vectors is calculated to reflect the song's similarity; and third, the suggested result is generated [9]. Multiple recommendation outcomes have an average accuracy of 87 percent. Furthermore, due to the restricted volume of data, it is unrealistic to cover all of the songs. As a consequence, when the same song belongs to numerous categories and the collected music only belongs to one, the accuracy of the recommendation results suffers.

G. Yamaguchi proposes an IGA approach for recommending music pieces that are related to the user's unclear image. As previously said, the IGA melody is extremely likely to fit the user's sensitivity and desire, thus the music pieces chosen based on that melody can likewise have the same impact [10]. As a result, it is envisaged that the approach suggested in this research would be able to recommend preferred music compositions that the user is unfamiliar with. From a different perspective, the suggested solution employs the user's tacit knowledge preference pattern in the retrieval process.

S. Gilda et al a multimedia materials recommendation system that employs AV values of mood tags rather than tags themselves was proposed by C. B. Moon et al. to partially tackle the problem of synonyms in Folksonomy in SNS service [11]. In this research, before analyzing recommendation performance, the ANOVA test is used to assess the AV values distribution of multimedia materials and tags based on 12 moods. The ANOVA test and the homogeneity test of dispersion using the AV table of multimedia items revealed that there are distribution disparities in AV values amongst 12 moods. According to the results of the investigation, the AV values of multimedia items with a mood tag and its synonyms are located in the Thayor model's mood tag region.

A. Budhrani et al. [12] suggested a system that employs a deep learning technique to do genre categorization on the music and then recommends the song using word2vec. For categorization, it is necessary to get a big sample of songs and index them according to their genre, then, using the skip-gram model, it will locate comparable context songs for the suggestion. Therefore, the suggested system functions as a full music recommendation system for providing a fantastic user experience.

S. Deepak presents a system in which music was evaluated by utilizing audio signal processing techniques to extract elements that demarcate it based on genre. The Mel Frequency Cepstral Coefficient (MFCC) characteristics dominate audio signal processing because they are similar to how the human ear works. The most crucial step in this approach is to remove the silent parts from the audio stream [13]. The model's performance is due to the non-silent audio tracks. Music is time-series data, and the LSTM network can analyze it to uncover patterns across time that are specific to each data type. This implementation demonstrates that both the LSTM model and the LSTM+SVM model perform well on the GTZAN dataset.

Using a triplet loss network and a collection of reference rankings, L. Pretet et al. offer a technique for learning a similarity ranking. The authors show that learning the ranking using the triplet loss produces better results than learning the tags utilized by the ground truth similarity function. This is a consistent outcome across all of their measurements. Finally, they demonstrate that auto pooling, which was developed in the context of audio event detection, also provides for improvement in the case of music similarity [14].

D. Wang et al. [15] describe a content- and context-aware music recommendation model capable of performing exact music recommendations using heterogeneous inputs. The authors propose a model to learn latent real-value low-dimension feature representations from both interactive/context data and textual content data, inspired by the collaborative filtering (CF) technique based on matrix factorization. This work is distinct from earlier work in three ways: First, the proposed method incorporates and leverages heterogeneous information to alleviate the data sparsity problem, Second it can cope with various aspects of music when interacting with different neighbors adaptively, and third, it is capable of precisely capturing music pieces' dynamic features and learning the structure and text embedding to further improve recommendation performance.

III. CONCLUSION AND FUTURE SCOPE

This survey article examines related studies on a successful and beneficial approach for music suggestion based on a user's mood as expressed through facial expressions. Face expression is among the most challenging and mysterious techniques ever attempted in the image processing framework. People exhibit the majority of their feelings via their expressions, which may be utilized for various reasons like as identifying a person's state of mind. Recognition of a person's mood is amongst the most beneficial deployments since it can be utilized in a variety of ways to enhance an individual's quality of life. As a result, in this research study, there has been a thorough examination of similar works in order to arrive at our method to music suggestion based on a person's mood assessment. For successful music suggestions, we consider employing convolutional neural networks in combination with Fuzzy Classification. In future research studies on this subject, this technique will be explored.

REFERENCES

[1] H. Zhang, K. Zhang and N. Bryan-Kinns, "Exploiting the emotional preference of music for music recommendation in daily activities," 2020 13th International Symposium on Computational Intelligence and Design (ISCID), 2020, pp. 350-353, DOI: 10.1109/ISCID51228.2020.00085.

[2] M. Bakhshizadeh, A. Moeini, M. Latifi, and M. T. Mahmoudi, "Automated Mood Based Music Playlist Generation By Clustering The Audio Features," 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE), 2019, pp. 231-237, DOI: 10.1109/ICCKE48569.2019.8965190.

[3] C. Zhou, Y. Jin, X. Wang and Y. Zhang, "Conversational Music Recommendation based on Bandits," 2020 IEEE International Conference on Knowledge Graph (ICKG), 2020, pp. 41-48, DOI: 10.1109/ICBK50248.2020.00016.

[4] C. Dhahri, K. Matsumoto and K. Hoashi, "Mood-Aware Music Recommendation via Adaptive Song Embedding," 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2018, pp. 135-138, DOI: 10.1109/ASONAM.2018.8508569.

[5] K. Sakurai, R. Togo, T. Ogawa, and M. Haseyama, "Music Playlist Generation Based on Reinforcement Learning Using Acoustic Feature Map," 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE), 2020, pp. 942-943, DOI: 10.1109/GCCE50665.2020.9291748.

[6] Z. Yu, M. Zhao, Y. Wu, P. Liu, and H. Chen, "Research on Automatic Music Recommendation Algorithm Based on Facial Micro-expression Recognition," 2020 39th Chinese Control Conference (CCC), 2020, pp. 7257-7263, DOI: 10.23919/CCC50068.2020.9189600.

[7] B. Kostek, "Listening to Live Music: Life Beyond Music Recommendation Systems," 2018 Joint Conference - Acoustics, 2018, pp. 1-5, DOI: 10.1109/ACOUSTICS.2018.8502385.

[8] K. Kittimathaveenan, C. Pongskul, and S. Mahatanarat, "Music Recommendation Based on Color," 2020 6th International Conference on Engineering, Applied Sciences and Technology (ICEAST), 2020, pp. 1-4, DOI: 10.1109/ICEAST50382.2020.9165455.

[9] H. Han, X. Luo, T. Yang and Y. Shi, "Music Recommendation Based on Feature Similarity," 2018 IEEE International Conference of Safety Produce Informatization (IICSPI), 2018, pp. 650-654, DOI: 10.1109/IICSPI.2018.8690510.

[10] G. Yamaguchi and M. Fukumoto, "A Music Recommendation System based on Melody Creation by Interactive GA," 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2019, pp. 286-290, DOI: 10.1109/SNPD.2019.8935654.

[11] C. B. Moon, J. Y. Lee, D. Kim, and B. M. Kim, "Analysis of Mood Tags for Multimedia Content Recommendation in Social Networks," 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), 2019, pp. 452-454, DOI: 10.1109/ICUFN.2019.8806025.

[12] A. Budhrani, A. Patel and S. Ribadiya, "Music2Vec: Music Genre Classification and Recommendation System," 2020 4th International Conference on Electronics, Communication, and Aerospace Technology (ICECA), 2020, pp. 1406-1411, DOI: 10.1109/ICECA49313.2020.9297559.

[13] S. Deepak and B. G. Prasad, "Music Classification based on Genre using LSTM," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 985-991, DOI: 10.1109/ICIRCA48905.2020.9182850.

[14] L. Prétet, G. Richard and G. Peeters, "Learning to Rank Music Tracks Using Triplet Loss," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2020, pp. 511-515, DOI: 10.1109/ICASSP40776.2020.9053135.

[15] D. Wang, X. Zhang, D. Yu, G. Xu, and S. Deng, "CAME: Content- and Context-Aware Music Embedding for Recommendation," in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 3, pp. 1375-1388, March 2021, DOI: 10.1109/TNNLS.2020.2984665.