[1] S L Medha*
[2] Prerana M
[3] Prema R

Emotion-aware Music Information Retrieval System

JES Journal of Electrical Systems

Abstract: - This study introduces an innovative system designed to refine song recommendations based on the user's expressed emotions. By leveraging a comprehensive database of songs spanning from 1950 to 2019, the system ranks and retrieves top songs, integrating a feedback loop to enhance personalized experiences. It also incorporates the Streamlit platform, enabling interactive input for personalized music recommendations. The objective is to elevate user engagement by prioritizing music selections finely attuned to the user's emotional state, offering an enriched and tailored music recommendation service.

Keywords: Music information retrieval, emotion-based model, metadata-based model, cosine similarity, content-based recommendation, feedback loop.

I. Introduction

In the digital age, music streaming platforms have transformed music discovery. At the heart of these platforms lie recommendation systems that enhance user experience by suggesting relevant content and act as digital curators, guiding listeners through vast musical landscapes and introducing them to new songs that would align to their tastes. However, a significant challenge remains as many traditional recommendation algorithms [1] struggle to capture and respond to users' nuanced, ever-changing emotional states, a key aspect of music selection and enjoyment.

This research addresses a critical gap in current music recommendation technology: The absence of real-time emotional feedback. While existing systems [2] [3] mostly rely on static data points such as listening history, genre preferences, or collaborative filtering, they often overlook the dynamic and profound influence of emotions on music choices. This oversight frequently results in recommendations that, although aligned with the user's general taste, may fail to resonate with their current emotional state.

To address this gap, this study introduces an innovative approach that incorporates real-time emotional feedback into the recommendation process. The aim is to demonstrate how emotional attunement in algorithms can significantly enhance the quality and relevance of song suggestions, thus contributing to the evolution of music recommendation systems. This method seeks to create a more responsive and intuitive recommendation process by prioritizing the user's current emotional state in song selection.

II. Literature Review

The paper [4] classifies emotions into four categories (“happiness”, “sadness”, “anger” and “fear”) based on three features (relative tempo, the mean and standard deviation of average silence ratio) using a BP neural network. The results showed that the mood in music is computable. However, mood detection in music is subjective, as different listeners may perceive the same piece of music differently. The limited feature set and the small test corpus also may not be representative of the wide variety of musical styles and moods, affecting the generalizability of the findings.

The increasing digitization of music has made automatic music recommendations relevant in today’s world. Since collaborative filtering suffers with the cold-start problem, this paper [5] focuses on the advantages of content-based methods which can recommend niche items that doesn’t require user interaction data. New and unpopular songs can still be recommended based on the song content, making it a better choice for accurate music recommendations.

The core idea in [6] is to rank music tracks based on their emotional content. Instead of assigning absolute emotion values (like in traditional rating systems), this approach compares music pieces to each other and determines their relative positions on the Cartesian space. Annotators were asked to compare two songs and determine which one is more emotionally intense or positive. So, a set of songs is compared in a series of pairwise

[1] Student, Department of Computer Science Engineering, PES University, Bengaluru, India.
[2] Student, Department of Computer Science Engineering, PES University, Bengaluru, India. preru1809@gmail.com
[3] Associate Professor, Department of Computer Science Engineering, PES University, Bengaluru, India. prema.r@pes.edu
* Corresponding Author Email: medhadev2111@gmail.com

matches. This ranking is done using an RBF-ListNet algorithm. However, this algorithm can be computationally expensive, especially when dealing with large datasets. It also requires careful tuning of its parameters, such as the value of the RBF kernel. Setting the RBF kernel parameter too small can result in overfitting, while setting it too large can lead to underfitting. Thus, balancing these two is very challenging.

[7] also explores the differences in content-based and collaborative-based recommendations, along with other approaches as well. Content-based models measure similarity between songs to provide recommendations without needing human ratings, whereas collaborative-based models provides recommendations on the collective behaviour of users, such as their listening histories and ratings. It identifies patterns among users to suggest new music. But, this method leads to many downfalls such as the popularity bias and the cold-start problem. Popularity bias leads to recommendations of popular music with more ratings, neglecting less popular music that might suit the user's taste better. The cold-start problem results in poor prediction accuracy due to the limited ratings in the early stages of a user's interaction with the system.

Many emotion-based systems [8] [9] [10] focus on detecting the user's mood through a live video feed before recommending songs. But, these systems face challenges due to the complexity of processing video streams and the latency introduced while waiting for a few seconds to detect a face, which adds significant time overhead.

III. Dataset And Methodology

3.1 Dataset:

The preprocessed dataset [11] consists of music metadata information, with 28372 entries and 5 major features, which include the track name, artist name, genre, a list of lyrics and the release date which ranges from the year 1950 to 2019.

3.2 Proposed Methodology:

Preprocessing and Feature Engineering: To determine the sentiment of each track, the track name and lyrics are combined into a unified feature labeled 'text'. This feature undergoes several preprocessing steps, including tokenization, stop word removal, and lemmatization, to produce the processed lyrics. Subsequently, a comprehensive dictionary of keywords is established for each sentiment, encompassing synonyms associated with each sentiment. Sentiment computation involves counting the occurrences of sentiment-related keywords within the processed lyrics. The sentiment with the highest count of these keywords is identified as the predominant sentiment for the track. As a result, an additional column is introduced into the dataset, indicating the sentiment of each track.

Data Loading and Interface Setup: The system loads the dataset comprising of the songs, which includes the metadata. An interface is set up using Streamlit, allowing users to interact with the system.

User Input and Sentiment Detection: Users input keywords related to their current mood. The system matches these keywords to predefined sentiments (happy, sad, angry, romantic, relaxed).

Song Recommendations: Based on the detected sentiment, the system finds and displays the top 10 songs that match the user's mood. Users have the option to explore more songs by selecting a specific genre and release year range. The system filters the dataset to show 5 songs that match the chosen genre, sentiment, and year range. The system also recommends 5 additional songs from a genre that fits the detected sentiment. Techniques like Term Frequency-Inverse Document Frequency (TF-IDF) and cosine similarity are used to rank songs based on how well they match the user's keywords. The TF-IDF vectorizer processes the song lyrics, and the cosine similarity calculator determines the relevance of each song to the input keywords.

TF-IDF (Term Frequency-Inverse Document Frequency) is a numerical statistic used to measure how important a word is to a document in a collection of documents (corpus). It combines two metrics: term frequency (TF) and inverse document frequency (IDF).

$$TF(t,d) = \frac{Number\ of\ terms\ t\ in\ document\ d}{Total\ number\ of\ terms\ in\ document\ d}$$

$$IDF(t) = \log\left(\frac{N}{DF(t)}\right)$$

$$TF - IDF(t,d) = TF(t.d)\ \times IDF(t)$$

Where:

- N is the total number of documents in the corpus.
- DF(t) is the document frequency of term t, i.e., the number of documents that contain the term t.

Cosine similarity is a metric used to measure how similar two documents are irrespective of their size. It calculates the cosine of the angle between two vectors.

$$Cosine\ Similarity = \frac{A\,.\,B}{||A||.||B||}$$

Preference Management: Users can save their preferences for each sentiment, which the system will store in a user preferences database for future recommendations. If users are not satisfied with the recommendations, they can clear their preferences and start the process over.

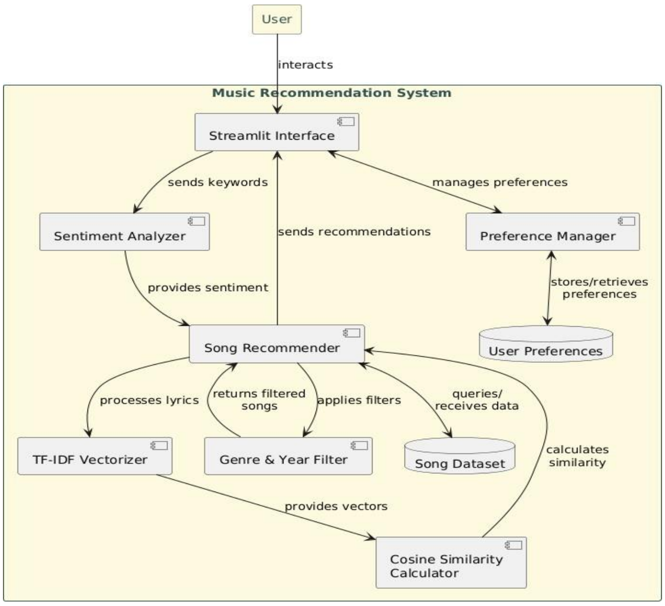

Figure 1. System Architecture

Performance Metrics Summary: The evaluation of the system, as shown in Section 4.1 of the paper, demonstrates exceptional performance. Without a clear set of correct answers (ground truth), it's not possible to use traditional metrics like precision, recall, and NDCG on their own. These metrics usually require a known set of relevant items to be effective. Instead, relevance is measured using Precision@10, Recall@10, and NDCG@10. These metrics look at the top 10 recommended songs, providing a clear picture of how well the system performs in recommending the best items. This approach helps to understand the system's effectiveness even when a complete set of correct answers is not available.

Precision at k: It is the proportion of relevant items among the top k items retrieved.

$$Precision@k = \frac{1}{k}\sum_{i=1}^{k} rel(i)$$

Where:

- rel(i) indicates whether the i-th item in the list is relevant.

Recall at k: It is the proportion of relevant items retrieved in the top k items compared to the total number of relevant items.

$$Recall@k = \frac{\sum_{i=1}^{k} rel(i)}{Total\ number\ of\ relevant\ items}$$

Normalized Discounted Cumulative Gain (NDCG) at k: It is the quality of the top k items retrieved, considering the position and relevance of items, normalized against the ideal ranking.

$$DCG@k = \sum_{i=1}^{k} \frac{2^{rel(i)} - 1}{\log_2(i+1)}$$

$$IDGC@k = \sum_{i=1}^{\min(m,k)} \frac{2^{i} - 1}{\log_2(i+1)}$$

Where:

- m is the number of relevant items

$$NDCG@k = \frac{DCG@k}{IDCG@k}$$

IV. Results And Discussion

4.1 System Evaluation:

From Table 2, Precision@10 value of 1 across all sentiments indicates that all top 10 recommended items are relevant for each sentiment. Similarly, NDCG@10 value of 1 signifies that the top 10 items are perfectly ranked, with the most relevant items appearing at the top.

The Recall@10 values are comparatively low. Recall measures the proportion of relevant items retrieved in the top 10 compared to the total number of relevant items available. So, given the high number of songs per sentiment shown in Table 1, retrieving only 10 relevant songs constitutes a small fraction of the total relevant songs. Thus, despite the low Recall, the system performs well overall, as it maintains perfect precision and ranking among the top recommendations.

Table 1: Distribution of songs by sentiment

Sentiment	Count
Happy	12413
Sad	6315
Angry	3039
Romantic	5341
Relaxed	1264

Table 2: Performance Metrics of the System

Sentiment	Precision@k	Recall@k	NDCG@k
Happy	1	0.02	1
Sad	1	0.08	1
Angry	1	0.05	1
Romantic	1	0.08	1
Relaxed	1	0.56	1

4.2 Sentiment Analysis by Genre:

The sentiment scores from Figure 2(a) demonstrate the connection between the various emotions and musical genres. Happiness is most strongly associated with the Blues genre (0.495), while a romantic mood is closely linked to Country music (0.237). Anger shows a high correlation with Hip Hop (0.409), just as happiness is significantly linked to Reggae (0.467). Similarly, sadness is also allied with Country music (0.259). These statistics highlight the unique emotional landscapes inherent to each genre. This insight can enhance the effectiveness of song recommendations, ensuring they resonate with the user's current moods.

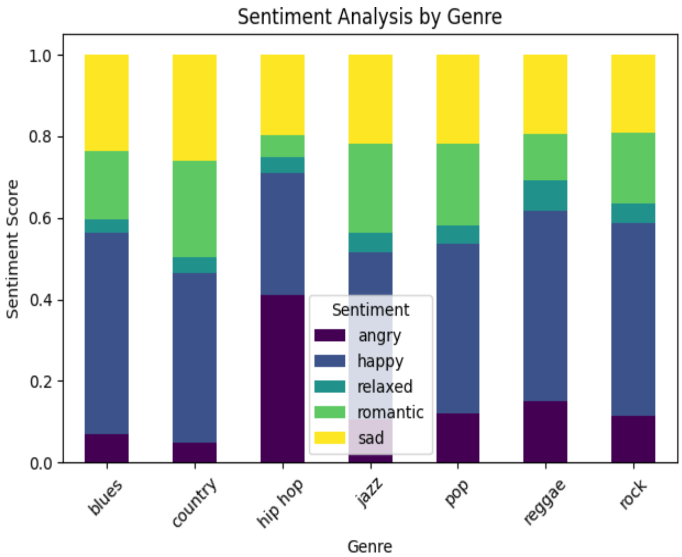

Figure 2(a). Sentiment Analysis by Genre

4.3 Genre Distribution over Release Years:

The genres from Figure 2(b) reflect evolving cultural and musical trends over the decades. Blues was steady, peaking in the 1980s and again in 2018. Country grew steadily from the 1950s on with occasional fluctuation. Hip-hop has been strong since the late 1970s, flourished in the late 1990s, and remains popular since 2015. Jazz was very popular in the 1950s, declined a bit in the 1970s, and has remained stable with periodic shifts. Pop peaked consistently in 1980s and 1990s. Reggae took firm hold in the 1970s, retreated in the ensuing decades and has been on the rise since 2015. Rock dominated its nascent years, peaked in the 1960s and 1970s.

These trends provide valuable context for the recommendation system, allowing it to suggest songs that align with both historical patterns and user preferences.

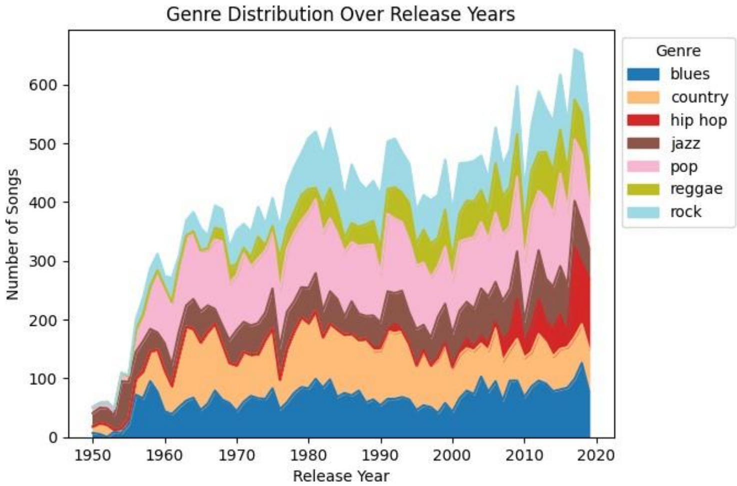

Figure 2(b). Genre Distribution over Release Years

4.4 User Feedback and System Performance:

The music recommendation system leverages both explicit and implicit user feedback to continuously improve its performance. Explicit feedback includes saved sentiment preferences and the option to clear them, while implicit feedback is gathered through user interactions with genre selections, year ranges, and recommended songs. Users directly input their mood keywords, select genres and year ranges, and save their preferences. The system logs these interactions to understand user behavior and preferences over time. This feedback collection allows for analysis of patterns in genre preferences and sentiment associations. The analysis directly impacts recommendations, personalizing them based on saved preferences and adapting to user choices over time. The system iteratively refines its genre-sentiment associations and allows users to reset their profiles, providing insights into evolving music tastes. Performance is evaluated through engagement metrics like search frequency and preference-saving rates. By offering a mix of mood-matched and genre-exploratory recommendations, the system aims to balance familiarity with discovery, enhancing user satisfaction and engagement in their music exploration journey.

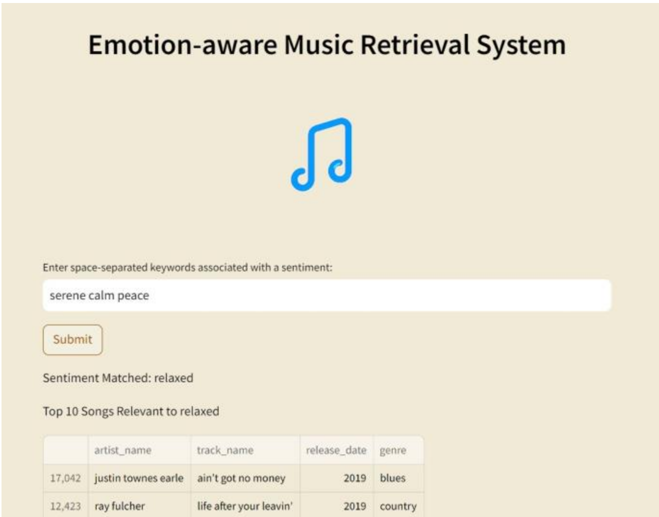

Figure 2(c). Music Recommendation Page

V. Conclusion And Future Scope

The Emotion-aware Music Information Retrieval system enhances personalized music recommendations by prioritizing user emotions and their preferences. By leveraging a comprehensive song database spanning several decades, the system offers a diverse range of music options tailored to the user's emotional state.

The ranking mechanism, which is driven by user-expressed moods, intelligently prioritizes song recommendations, enhancing the music discovery experience. Feedback based on genre and a range of release years, combined with the system's recommendations tailored to the user's mood and time frame, further refines the suggestions.

The system currently only incorporates the metadata information, which does not work for tracks that contain solely instrumental music. In the future, user engagement and satisfaction can be further elevated by incorporating audio features, making the system more robust and capable of handling a wider variety of music.

References

[1] Y. Qin, A historical survey of music recommendation systems: Towards evaluation. McGill University (Canada), 2013.

[2] D. Jannach, I. Kamehkhosh, and G. Bonnin, Music Recommendations: Algorithms, Practical Challenges and Applications, Nov. 2018, pp. 481–518.

[3] E. Shakirova, "Collaborative filtering for music recommender system," in 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 2017, pp. 548–550.

[4] Y. Feng, Y. Zhuang, and Y. Pan, "Popular music retrieval by detecting mood," in Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, 2003, pp. 375–376.

[5] A. Van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," Advances in Neural Information Processing Systems, vol. 26, 2013.

[6] Y.-H. Yang and H. H. Chen, "Ranking-based emotion recognition for music organization and retrieval," IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 4, pp. 762–774, 2010.

[7] Y. Song, S. Dixon, and M. Pearce, "A survey of music recommendation systems and future perspectives," in 9th International Symposium on Computer Music Modeling and Retrieval, vol. 4, Citeseer, 2012, pp. 395–410.

[8] A. Mahadik, S. Milgir, J. Patel, V. B. Jagan, and V. Kavathekar, "Mood-based music recommendation system," International Journal of Engineering Research & Technology (IJERT), vol. 10, no. 06, 2021.

[9] H. I. James, J. J. A. Arnold, J. M. M. Ruban, M. Tamilarasan, and R. Saranya, "Emotion-based music recommendation system," Emotion, vol. 6, no. 3, pp. 2096–2101, 2019.

[10] V. P. Sharma, A. S. Gaded, D. Chaudhary, S. Kumar, and S. Sharma, "Emotion-based music recommendation system," in 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), IEEE, 2021, pp. 1–5.

[11] S. Shahane, "Music dataset: 1950 to 2019," 2020. [Online]. Available: https://www.kaggle.com/datasets/saurabhshahane/music-dataset-1950-to-2019