MUSIC RECOMMENDATION SYSTEM

INTRODUCTION-

Finding new songs that suit our own interests in the vast and ever-expanding world of music can be both thrilling and stressful. By utilizing data and algorithms, the "Personalized Music Recommendation System" initiative seeks to make the task of selecting the ideal soundtrack easier. This project makes sure that every user is greeted with melodies that suit their tastes by providing a customized music recommendation experience based on lyric recognition.

Keywords:

NumPy, Pandas, Cosine similarity, CountVectorizer.

Plan Outline:

1. **Data Collection**

   From Spotify, kaggle dataset and many more. Database containing name of songs with their artists and lyrics.

2. **User Interface**

   Users can search songs by entering specific keywords. Users will be given several options too, like top 5 songs based on their similarity with other songs.

3. **Data Processing**

   After data collection, data would be cleaned and processed. Remove all the duplicates and spelling errors. Make a standardized system.

   - Acoustic characteristics are extracted
   - A feature vector for the composition is built
   - Vector similarity is estimated
   - A list of recommendations is formed.

Literature Survey

A great amount of work in recent years have been done in music perception, psychology, neuroscience and sport which study the relationship between music and the impact of human behaviour.

People have considered that music is an important aspect of their lives and they listen to music, an activity they engage infrequently. People sometimes feel it is difficult to choose from millions of songs. With commercial music streaming services which can be accessed from mobile devices, the availability of digital music currently is abundant compared to the previous era. Music service providers need an efficient way to manage songs and help their customers to discover music by giving quality recommendations.

A music recommender system is a tool that music providers may use to predict and give consumers with appropriate songs based on the characteristics of previously heard music. Sorting through all of the digital music available now is more time-consuming and leads to information fatigue than it did in the past.[3]

Because of the Internet's explosive expansion, sentiment analysis has emerged as one of the most important areas of natural language processing in cloud computing (NLP). Sentiment analysis may be effectively used to extract latent emotion in text, helping businesses or organizations make better decisions.[4]

Compared in the past, the availability of digital music has risen due to internet music streaming services accessible via mobile phones. Sorting through all of the songs gets difficult and leads in information overload. Many individuals regard music as an important aspect of their life and place a high value on it. When a person is happy, sad, or emotional, he wants to unwind by listening to music. Users typically utilize search engines to identify music of interest to them, but as technology has improved, other methods of searching have emerged. [1]

As a consequence, developing a music recommendation system that can automatically search song albums and propose acceptable songs to consumers is highly helpful. Using such a mechanism, it may predict and then deliver the right tunes to its consumers based on their mood. The designed recommender system is unique in that it is dependent on the user's mood. There are two well-known approaches to developing content-based music recommendation systems. The first technique, which employs a powerful classification algorithm, is widely used, but the second employs deep learning methods to improve the effectiveness of the recommender system.[1]

A good music recommender system should be able to automatically detect preferences and generate playlists accordingly. Meanwhile, the development of recommender systems provides a great opportunity for industry to aggregate the users who are interested in music. More importantly, it raises challenges for us to better understand and model users' preferences in music.

The article covers the process of developing a music recommendation system, in addition to various approaches, potential workarounds, and upcoming advancements. The first step in data pre-processing is data cleansing. Since much of the data we work with nowadays is dirty, extensive data cleansing is required. Some include garbage data, while others lack numbers. If these unaccounted-for factors and anomalies were not appropriately addressed, our method would not yield reliable findings. We've created a new feature by combining the name of the artist with the song title. We are going to examine a subset of this dataset, specifically the first 10,000 songs. The song and artist name are then combined into a single column, which is then totaled by the number of times a song has been listened to by all visitors. Data transformation facilitates data compression and further enhances interpretability. [3]

We investigate the viability of embedding songs using triplet neural networks according to content-based musical similarity. Our network is trained using triplets of songs, meaning that a third song by a different artist is implanted more distantly from the first two songs by the same artist. We compare two models that are trained with distinct methods of selecting this third song: based on shared genre labels or randomly. We employ common audio characteristics and run our tests with tracks from the Free Music Archive. According to the first findings, music may be embedded into shallow Siamese networks for a straightforward artist retrieval job. Although our approach of producing training data with artists and genre labels is fairly low-cost, it might not be enough to build a strong embedding model for content-based music recommendation. Since many musicians have a wide range of acoustic styles, their songs are not good indicators of acoustic similarity. We hope to evaluate the model in the future using similarity data from a listening research.Rather of producing complete training data from a listening research, several strategies for choosing hard triplet. [2]

divided into overlapping frames with a specific number of samples (the sampling method determines the number of samples).,the signal's frequency). By using the Fourier conversion to every frame, we are able to extract the signal's spectrum. Descriptive statistics are computed based on the spectrum shape. Add other traits, then combine them over the whole quantity of frames (we calculate the mean, lowest, maximum, divergence from the norm, etc.). [5]

Python is increasingly being used as a scientific language. Matrix and vector manipulations are extremely important for scientific computations. Both NumPy and Pandas have emerged to be essential libraries for any scientific computation, including machine learning, in python due to their intuitive syntax and high-performance matrix computation capabilities.

CBRS recommends products based on their features and the similarity of their elements to those of other items. Assuming a user has already viewed a move from the omedy genre, BRS will recommend moves that also belong to the omedy genre. A content-based recommender works on data provided by the user, either explicitly (rating) or implicitly (clicking on a link). Based on the data, a user profile is created, which is subsequently utilised to make recommendations to the user. As the user provides more input or takes actions on the recommendations, the engine becomes more and more accurate.Python is becoming increasingly popular as a scientific language. Matrix and vector manipulations are critical for scientific computations.Due to their intuitive syntax and high-performance matrix computation capabilities, both Numy and pandas have emerged to be esential libraries for any scientific computation. NumPy stands for ‘Numerical Python’ or ‘Numeric Python’. It is an open-source module of Python which provides fast mathematical computation on arrays and matrices. Since arrays and matrices are an essential part of the Machine Learning ecosystem, NumPy along with Machine Learning modules like Scikit-learn, Pandas, Matplotlib, TensorFlow, etc. complete the Python Machine Learning Ecosystem [6]

Modules used:

1. Panda (helps to load the data frame in a 2D array format and has multiple functions to perform analysis tasks in one go.)
2. Nltk ( )
3. Sklearn (contains multiple libraries having pre-implemented functions to perform tasks from data preprocessing to model development and evaluation.)

Evaluation Metrics:

Various metrics are used to evaluate the effectiveness of content-based recommendation systems, including precision, recall, F1-score, Mean Average Precision (MAP), and Area Under the Receiver Operating Characteristic (ROC) Curve (AUC-ROC).

Similarity Measurement:

A critical aspect involves defining similarity measures between songs based on their extracted features. Techniques like cosine similarity, Euclidean distance, or more complex similarity metrics are employed.

Conclusion:

The music recommendation system is a comprehensive and evolving tool designed to simplify the daunting task of discovering new songs in the vast world of music. By leveraging data, algorithms, and techniques like cosine similarity and CountVectorizer, this system aims to provide users with personalized music recommendations based on lyrics.

The literature survey highlights the significance of music in people's lives and the challenges associated with sorting through an abundance of digital music. Various approaches, such as sentiment analysis and content-based recommendation systems, are discussed, emphasizing the importance of understanding user preferences and emotions to enhance recommendation accuracy.

Key components, including data collection, user interface, data processing, and the use of Python libraries like NumPy and Pandas, are outlined in the plan. Additionally, methods such as embedding songs using triplet neural networks and signal processing techniques for feature extraction are explored as ways to enhance music recommendation accuracy.

Explores future directions and trends that could shape the next generation of recommender systems.
Discusses the significance of User Generated Content (UGC), the evolving nature of vocabularies, and the challenge of providing users with novel and engaging recommendations they might not have otherwise encountered.

Overall, the paper covers the foundational aspects of content-based recommender systems, reviews current techniques and their applications, and explores potential trends and challenges that could influence the evolution of recommender systems in the future.

Reference

1] Music Recommendation System using Python

-Mr. Chirag Desai1 , Shubham Bhadra2 , Mehul Parekh3 K J Somaiya College of Engineering, Mumbai, Maharashtra, India -year of publication-2023

2] Content-based Music Similarity with Triplet Networks

- Joseph Cleveland , Derek Cheng , Michael Zhou , Thorsten Joachims ,Douglas Turnbull
- year of publication- 2022

3] Recommendation System for music based on content and popularity ratings -

-Timanshi Bhardwaj , Aastha Jain ,Karan Choudhary Manav Rachna of International Institute of Research and Studies, Faridabad, India - year of publication- 2021

4] Emotion Based Music Recommendation System By Using Different Ml Approach

-Bagde Sarika Vidyasagar, Asst.Prof. V. Karwande - year of publication-2021

5] Content-based Music Recommendation System

- Aldiyar Niyazov St.Petersburg State University, Elena Mikhailova ITMO University, St.Petersburg State University, Olga Egorova ITMO University St.Petersburg, Russia - year of publication- 2021

6] Music Recommendation System Using Machine Learning

-Varsha Verma, Ninad Marathe, Parth Sanghavi, Dr. Prashant Nitnaware [Department of Information Technology, PCE, Navi Mumbai, Maharashtra, India] - year of publication-2021

SUBMITTED BY -
REETI AGARWAL - 23070122177
RIYA AGRAWAL - 23070122184
SAUMYA KUMAR - 23070122195
NIMITA JESTIN - 23070122259