ABSTRACT

In this paper, we present a personalized music recommendation system based on the KNN and machine learning algorithms.In personalized music recommendation system, we propose a collaborative filtering and content filtering recommendation algorithm to combine the output of the network with the log files to recommend music to the user. The proposed system contains the log files which stores the previous history of playlist of music by the user. The proposed music recommendation system extracts the user's history from the log file and recommends music under each recommendation.Content-based methods gives recommendations based on the similarity of two song contents or attributes while collaborative methods make a prediction on possible preferences using a matrix with ratings on different songs. The plagiarism system extracts the music from input and finds music that are close to the query music which the query has plagiarized. We use the million song dataset to evaluate the personalized music recommendation system. The data cleaning is done by the data science algorithms. The plagiarism detection is done by finding the similar music genre which minimizes the issue of copyrights.

Index Terms— collaborative filtering algorithm. KNN, cosine similarity, tf-idf, CSR Matrix,SVM.

Table of Contents

1. INTRODUCTION

With the explosion of network in the past decades, internet has become the major source of retrieving multimedia information such as video, books, and music etc. People have considered that music is an important aspect of their lives and they listen to music, an activity they engaged in frequently. However, the problem now is to organize and manage the millions of music titles produced by society. A good music recommender system should be able to automatically detect preferences and generate playlists accordingly. The proposed system is to detect music plagiarism based on music similarity. The plagiarism system extracts the music from input and finds music that are close to the query music which the query has plagiarized. Meanwhile, the development of recommender systems provides a great opportunity for industry to aggregate the users who are interested in music. We need to generate the best music recommendation system which is need to predict based on customization, by using KNN, Machine Learning.

Everyones taste in music is unique which means that no matter what music you make, someone is bound to enjoy listening to it. While the Music industry may favor certain types of music more than others, it is important to understand that there isnt a single human culture on earth that has existed without music. Music is of great benefit to us, regardless of whether we are renowned recording artists, karaoke singers or merely fans of music. The number of songs available exceeds the listening capacity of single individual.

According to the MarsBands.com there are at least 97 million songs. These are only the songs officially released. If we included songs everyone knows or the incredibly old Celtic songs with no names, we would reach 200 million songs since the website most likely does not include Happy Birthday or a nameless song from 1400 AC. This is when we take only the artists who had their name officially on Music charts. Starting there, let's say that there are currently around 1 million songwriters alive that we know about. If we use the same percentage as above, we can guess that there have been about 15.3 million songwriters ever.

To get an idea, there are 4 million songs on Spotify that have never been played. In total, there must be billions just there and Spotify itself is by no means the limit of music. What about all the CDs and records made over the past century which have not been digitized? What, indeed, about song passed down the generations in small African communities? There are trillions and trillions of songs in the world, so many that an estimate is impossible, and the potential more an infinitely greater number which have not yet been made, a world of music for us to enjoy.

Currently, based on users' listening behaviour and historical ratings, collaborative filtering algorithm has been found to perform well. Since 2005, an annual evaluation event called Music Information Retrieval Evaluation exchange (MIREX) is held to facilitate the development of MIR algorithms.In content-based method similarity between the songs are checked and songs are recommended based on the similarity score.The content -based method can also be considered for plagiarism detection, plagiarism can be detected based on the similarity score.

By using mood prediction we can easily recommend songs to the user according to their interest and mood.The mood can be predicted in many ways such as by using lyrics,face emotion detection etc.Here we used lyrics based mood prediction and it also calculates the similarity score and recommends the songs.

The music recommender systems are double edged swords. The are of valuable use both to the user as well as the provider. They keep the user engaged by finding interesting music in the form of recommendations, lessening the burden on the user by reducing the set of choices to choose from. They give the scope for exploration and discovery of music that the user may not know exists. Because it is a music recommender there is never less entertainment.

1.1. Music Plagiarism

The second component of recommender systems is music plagiarism. Music often characterized by its melody, harmony, rhythm and timbre.Music plagiarism can be suspected under the following cases: when two music have similar successive music notes, when two music shares the unique parts of music which is rarely uses in others, when music has similar music progress with different key and instruments, plagiarism can be suspected. This system is composed of four modules:(1)Music Extraction

Module: music of the query music is extracted,(2)Similarity Calculation Modules: the similarity between the sequence of the input polyphonic music and those of music in the database is calculated,(3) The similar section of the music in the database is detected.This can be done by using content-based filtering,The model only learns to recommend items of the same type that the user is already using or, in our case, listening to. Even though this could be helpful, the value of that recommendation is significantly less because it lacks the surprise component of discovering something completely new.

1.2. Motivation for work

Different recommendations primarily need to work for the satisfication the users.Identifying user grievances thereby resolving them leads to customer satisfaction as well as trustworthiness.The today's world many people are busy and suffering a lot in their life so to overcome that problem for atleast sometime and the best solution is listening to the music.So we dediced to project on a music recommendation system so that users can listen a song based on their interests , can get the recommendation based on the similarity between the lyrics and also based on their mood songs can be recommended.

1.3. Problem statement

The basic task in music recommendation system with plagiarism detection is to generate the best music recommendation system by predicting based on customization and detecting the similar music genre to avoid copyrights issue, by using Collaborative filtering, Content based, Machine Learning, Data Analysis.

2. LITERATURE SURVEY

2.1. Introduction

An ideal music recommender system should be able to automatically recommend personalised music to human listeners.So far, many music discovery websites such as Last.fm, All music, Pandora, Audio baba Mog, Spotify, Apple Genius, have aggregated millions of users, and the development is explosive . In this section, we present the most popular approaches, metadata information retrieval, collaborative filtering, content-based information retrieval, emotion-based model , context-based information retrieval and hybrid models .

The proposed system contains the plagiarism in addition to the recommendation system which acts as a great advantage to resolve the copyright problems, the plagiarism module deals the check of similar music genre and detects the songs with similar musical notes.

The collaborative filtering technique is used to recommend songs to the users of similar groups. Involving collaborative filtering technique is an advantage to recommend songs by achieving the customization. The three types of collaborative filtering involved are memory based collaborative filtering, model based filtering, hybrid based filtering.

The existing recommender systems using collaborative filtering algorithms have gained a great success. Netflix opened a challenge for the best collaborative filtering algorithm , and the winning algorithm using latent factor models could make 10.09% improvements over the algorithm used by Netflix at that time. Amazon uses user-user based and item-item based collaborative filtering , which greatly contributes to the success of the business. Recently a newer algorithm using neural network, neural collaborative filtering was proposed.

The content-based filtering technique is used to predict the song by analyzing the song track. It is rooted in information retrieval and information filtering that recommends a song which is similar to those the user has listened to in the past rather than what the user have rated ‘like’ Lots of research have been paid attention on extracting and comparing the acoustic features in finding perceptual similar tracks . The most representative ones so far are timbre, rhythm. Based on the extracted features,

the distance between songs are measured. Three typical similarity measurements are K-means clustering with Earth-Mover's Distance, Expectation-Maximization with Monte Carlo Sampling, Average Feature Vectors with Euclidean Distance.

For content based algorithm, a lot of researchers have proposed different methods using Machine Learning technique, such as Decision Tree based , Support Vector Machine based , and even logistic regression . We can fully utilize the knowledge we learnt from the class to implement these algorithms.

Digitization of music has led to easier access to different forms music across the globe. Increasing work pressure denies the necessary time to listen and evaluate music for a creation of a personal music library. One solution might be developing a music search engine or recommendation system based on different moods. Develop a mood classification system from lyrics as well by combining a wide range of semantic and stylistic features extracted from textual lyrics.

2.2. Existing Systems

2.2.1. Collaboration Filtering

Collaborative filtering (CF) uses the numerical reviews given by the user and is mainly based upon the historical data of the user available to the system . The historical data available helps to build the user profile and the data available about the item is used to make the item profile. Both the user profile and the item profile are used to make a recommendation system. The Netflix Competition has given much popularity to collaborative filtering .

Collaborative filtering is considered the most basic and the easiest method to find recommendations and make predictions regarding the sales of a product. It does have some disadvantages which has led to the development of new methods and techniques.

2.2.2. Memory-Based Collaborative Filtering(Neighbourhood based)

People with similar interests are combined to form a group and every user is a part of that . User –based CF and Itemimplement and scales well with correlated items. There is no need items being recommended. There are many limitationsof memory problem, sparsity and their dependencies on human ratings .

2.2.3. Model-based CF

Complex patters which are based on training data, are the models (such as data mining algorithms, machine learning) and then intelligent predictions are made for CF tasks for the real world data which arebased on learnt models. It intuitive rationale for recommendations. Model disadvantage of model-based CF is that it loses useful information for dimensionality reduction techniques .

2.2.4. Hybrid Collaborative Filtering Techniques

In hybrid recommender system, different techniques of collaborative approaches and other recommender techniques (usually content based approaches), combined to get better results.

Various problems like cold-start, data sparsity and scalability can be avoided by using hybrid approach [40].There are different ways of combining CF with other recommender techniques which are following:

- Hybrid Recommenders Incorporating CF and Content-Based Features
- Hybrid Recommenders Combining CF and Other Recommender Systems
- Hybrid Recommenders Combining CF Algorithms .

2.2.5. Content Based Recommender System

Content based systems focus on the features of the products and aim at creating a user profile depending on the previous reviews and also a profile of the item in accordance with the features it provides and the reviews it has received .It is observed that reviews usually contain product feature and user opinion in pairs . It is observed that users' reviews contain a feature of the product followed by his/her opinion about the product. Content based recommendation systems help overcome sparsity problem that is faced in collaborative filtering based recommendation system.

Content based collaborative filtering is more widely used to compare pure CF and pure Content-base. In CF the problem of sparsity is overcome (converting sparse user filled matrix into full user rating matrix) by using content-based prediction. Relevant entities of an item and relations are kept together as input.

1. Fully observable: Complete structure and values of contextual factors are known explicitly, at the time when recommendations are made.
2. Partially observable: Some of the information is known explicitly about the contextual factors.
3. Unobservable: There is no information of contextual factors explicitly available in It.

2.2.7. A Recommender System Based on Genetic Algorithm

The recommender system by Hyun-Tae Kim, Eungyeong Kim, Jong-Hyun Lee and Chang Wook Ahn is a hybrid approach of Collaborative Filtering (CF) and Genetic Algorithm (GA). The proposed system aims to effectively adapt and respond to immediate changes in users' preferences. The experiments conducted in an objective manner exhibit that our system is able to recommend items suitable with the subjective favorite of each individual user.

The content-based filtering technique is applied to generate the initial population of GA. The recommender system is divided into three phases:

1.Feature extraction

2.Feature evaluation

3.Interactive phase

2.2.7.1. Advantages

The experimental results exhibited that the average scores, which are objectively collected by means of user evaluations, increases by degrees as the generation grows.

2.2.7.2. Limitations

It's really hard for people to come up with a good heuristic which actually reflects what we want the algorithm to do. It might not find the most optimal solution to the defined problem in all cases.