Sentiment Analysis on Songs Between 1958-2019

Shashank Bengaluru Srinivasa
State University of New York at Binghamton
Binghamton, USA
sbengal2@binghamton.edu

Naveen Chowdary Poda
State University of New York at Binghamton
Binghamton, USA
npoda1@binghamton.edu

***Abstract*—You always hear people complain how the music in older times was better and meaningful than the music in the current times. For obvious reasons the kind of music people listen to changes with time. In this paper we try to analyse how the music and music preference has changed over the years.**

***Index Terms*—songs,sentiment analysis,music,billboard**

I. PROBLEM

If you listen to the 70s,80s music you can't help but appreciate how beautiful and soulful those songs are compared to today's songs.The present day music seems way too negative when put against the music from a few decades earlier.

To check if this is in-fact true, in this paper, we analyse a data-set of songs that were produced between 1958 - 2019. The data-set is extended to provide us with the required information. We obtain the lyrics of each song, and assign a polarity score to it by running it through textblob, a python library for textual analysis.

Using the polarity score for each song we try to observe how the music has changed from the 60s to today, in parallel to this we create an user- interface to the database to retrieve information for a set of queries.

II. SOFTWARE DESIGN AND IMPLEMENTATION

A. Data-set

We use the billboard hot 100 data-set that has all the songs that appeared in the billboard hot 100 charts between 1958-2019. 'Billboard hot 100' has long been a reliable, industry standard ranking system for songs, ranking them weekly.

The basic data-set contains only song, performer, date it first appeared on the charts, number of weeks it stayed on the charts and its peak-position. To perform a sentiment analysis this data-set is insufficient.Therefore we try to extend this data-set using the genius and spotify API. Lyrics for each song is obtained using the genius API and genre for the same is obtained using the spotify API.

Further the WeekID(date) and peak-position fields of the songs in the basic data-set are in string format. We convert these fields to their query-relevant formats by running through a python script over the mongo database.

B. Software Design and NoSQL-Databse and Tools Used

We use **mongo DataBase** as the NoSQL-Database considering the ease and number of functions it offers for data manipulation. The data-set is prepared by extending it by adding lyrics and genres to each song, converting the fields into required, relevant formats, dropping the useless fields. Each song is assigned a polarity by running its lyrics through textblob library using a python script.

Pymongo is used to connect and perform operations over the mongo database.

Matplotlib library is used to plot the graphs using python script.

C. Supported Queries

The application supports the following queries:

- **Search for a song**: retrieves the information of song that includes song name, artist, genre. If there are more than one song with the same name, the query is made more specific by asking the user to enter the artist's name.
- **Search for songs by an artist**: retrieves the information of all songs by a particular artist.
- **Search for songs by a genre**: retrieves the information of all songs of a particular genre.
- **Get the hot charts for a year**: retrieves the information of all songs present in the hot charts for a particular year.
- **Get all the top songs**: retrieves the information of all top-ten songs for each week in every year between 1958-2019
- **Get all the top artists**: retrieves the top-ten artists for each week in every year between 1958-2019
- **Get all the top genres**: retrieves the top-ten genres for each week in every year between 1958-2019
- **Plot sentiment for all songs**: plots a line-graph for polarity of each song against the date it appeared on the billboard charts .

III. SENTIMENT ANALYSIS

The sentiment analysis is performed to understand how the lyrics and music preference of people has changed over the years.

The article "Sentiment analysis of all billboard Hot 100 songs over time (1958–2019)" on medium.com provided us with a basic idea on how to perform a sentiment analysis [1].

A. Song Sentiment Analysis

The average polarity is calculated for each year and a line-graph is plotted against the time. The fig 1 shows the varying polarity of song lyrics from 1958 to 2019 taken as an average of all songs against the year of their appearance on the Billboard charts.

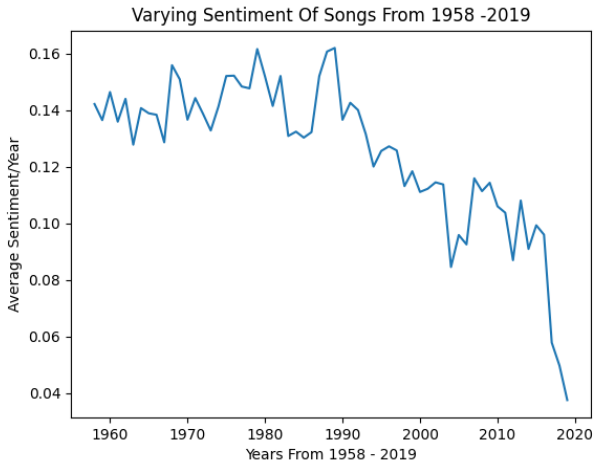

Fig. 1. Average song polarity per year

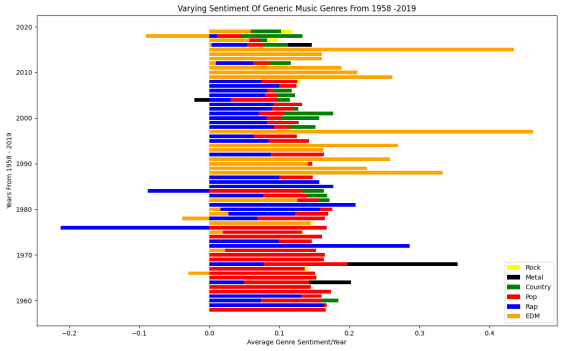

Fig. 2. Average sentiment per genre per year

From the plot it can be observed that the lyrics of songs have become increasingly negative over the years. Further,lyrics were most positive in the year 1979 and least positive in the year 2019.

B. Genre Sentiment Analysis

The average polarity is calculated for each genre in each year and a bar-graph is plotted against the time.

The figure 2 shows the distribution of popular genres per year plotted against the polarity of their songs.

From the genre sentiment analysis it can be observed that the ”EDM” genre has had the most postive lyrics whereas ”Rap” genre has had the most negative lyrics in the past 60 years.

IV. Project Outcome

The source code is on github: click here

Acknowledgment

We thank our Professor, Kyoung-Don Kang for all his assistance for this project. We offer our sincere appreciation for all the learning opportunities.

References

[1] Salim Zubair, “Sentiment Analysis of all Billboard Hot 100 Songs Over Time (1958–2019),” https://towardsdatascience.com/sentiment-analysis-of-all-billboard-hot-100-songs-over-time-1958-2019-3329439e7c1a