

Analysis of Social Trends through Lyrics

Kritik Mathur (111986679) and Radhika Dhawan (112074050)

Problem

It is a universally accepted fact that songs are a reflection of the society^[1]. Different music caters to different people across different eras, oftentimes defining an era as well. We analyse lyrics of music across 5 decades to get an understanding of social trends and how contemporary gets impacted by societal changes and major events.

This kind of analysis has its uses in:

- a. Marketing and advertising industry: It can be used in writing marketing slogan or copywriter specific to different target customers.
- b. Music Industry: It can be helpful to music producers to know how the trend of music is changing with time.
- c. Analysing changes in social behaviors of people and relating those to major events in the history.

Background

The techniques for text analysis we'll be applying are

1. Keyword Extraction

Since songs have a different structure than sentences, in the sense they generally don't have complete sentences and the order of the words is more flexible than in normal texts, common keyword extraction methods like RAKE^[2] won't be well suited for this. Hence, we'll be experimenting with a couple of lyrical topic detection techniques^{[3][4]}

2. Document to Vector

To find similarities in songs, genres, or artists, we need to convert text data into vectors. Some common approaches for document-to-vector that we will be trying are *tf-idf*, *bag-of-words*, *BERT*.

3. Sentiment Analysis

Different genres often have different sentiment associated with them and this information can be a useful characteristic in getting a hint of the general sentiment at the time that genre was popular. We'll use SentiWordNet^[5] for the aforementioned task.

Approach

Text Analysis:



2. ScatterPlot: We will be creating a scatter plot for visualizing the similarity of artists: the closer the vector of artists, the similar their lyrics should be.
3. LineChart:
 - a. For k-means clustering of artists, we would need to optimize the value of k, using an elbow chart.
 - b. We will also draw a line chart for visualizing the change in sentiments of lyrics in past decades.
4. BarChart: After sentiment and aggression analysis, we would draw barcharts to for the comparison of sentiments and male and female artists.

Dataset

362,237 lyrics from MetroLyrics.com^[7]: There are around 380,000+ lyrics in the data set from 18,230 different artists from 12 different genres arranged by year.

Various fields in the dataset:

1. Index: Index of the dataset
2. Song: Name of the song
3. Year: Year of the release
4. Artist: Name of the singer
5. Genre: Genre to which the song belongs
6. Lyrics: Complete lyrics of the song

Decade	Number of Songs
60s	1
70s	4496
80s	3874
90s	11285
00s	190847
10s	151718

Data Cleaning

Before we can use the data for the above mentioned tasks, we need to perform the following preprocessing:

1. Disregard rows with empty fields for tasks related to that particular field
2. Remove duplicate rows
3. Get rid of extra spaces

4. Remove stop words
5. Remove punctuations
6. Make all the words case-uniform
7. Normalize shorthands (like *ain't* to *is not*)

Future Work

1. Analysis of songs in different languages and culture.
2. Text analysis on other types of writing like novels, essays, news articles, public notices, etc.

References

- [1]. Longhurst, B., & Bogdanović, D. (2014). Popular music & society (3rd ed.). Cambridge: Polity Press
- [2]. Rose, Stuart & Engel, Dave & Cramer, Nick & Cowley, Wendy. (2010). Automatic Keyword Extraction from Individual Documents. 10.1002/9780470689646.ch1
- [3]. Kleedorfer, Florian, Peter Knees and Tim Pohle. "Oh Oh Oh Whoah! Towards Automatic Topic Detection In Song Lyrics." ISMIR (2008)
- [4]. Apoorva G., Mathur K., Mamidi R. (2018). What is this Song About?: Identification of Keywords in Bollywood Lyrics, 19th International Conference on Computational Linguistics and Intelligent Text Processing.
- [5]. Baccianella, Stefano & Esuli, Andrea & Sebastiani, Fabrizio. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.. Proceedings of LREC. 10
- [6]. Jonathan Yu. (2017). D3 Word Cloud.
- [7]. 362,237 lyrics from MetroLyrics.com (<https://www.kaggle.com/gyani95/380000-lyrics-from-metrolyrics>)