Sentiment Analysis Project On Songs Lyrics

Janet Victoria, Tahmina Choudhury

Lehman College, City College of New York

Abstract

Sentiment analysis is the interpretation and classification of emotions of text. It is a process of determining the emotional tone from a series of words, which allows us to understand the tone and the emotion expressed in the text. In this project, lyrics of songs over the years are analyzed using an emotion lexicon to detect the trends of emotions.

Introduction

Sentiment analysis classifies emotions contained within texts by using different text analysis techniques. Some text analysis techniques are clustering, visualization, categorizing, information extraction, etc. NLP is a field itself and it has been around more than 50 years and grew out of the linguistics field due to the rise of computers (Brownlee). There have been many sentiment analysis projects on tweets, books, reviews, articles, song lyrics, etc. Lyrics from different genre songs have a variety of emotions in their words.

In our project, we will use a lyrics dataset and analyze it by using different text analysis techniques. The dataset is imported into a Google Collaboratory(Colab) notebook using a programming language called Python and displayed and eventually manipulated as a Pandas Dataframe. In addition to studying the emotion associations of the songs in the lyrics dataset, we also performed some simple analyses on the dataset to understand it better.

Table 1. Lyrics dataset displayed as a data frame

	song	year	artist	genre	lyrics
0	when-you-were-with-me	2009	a	Hip-Hop	I stopped by the house we called our home\nIt ...
1	careless-whisper	2009	a	Hip-Hop	I feel so unsure\nAs I take your hand and lead...
2	2-59	2007	a	Hip-Hop	Mark:] Sunday football I got boot off the pitc...
3	power-of-desire	2007	a	Hip-Hop	[Chris:] Fallin' for a fantasy\nI threw away m...
4	you-re-not-in-love	2007	a	Hip-Hop	something in the way we touch\nyou hold my han...

Methods and Materials

We used a python library, NLTK, to remove stop words in the lyrics. Stop words that are commonly used words like 'the', 'in', 'a', 'in', etc. Removing the stop words from a text, in this project's case the text being the lyrics, cleans up the text and leaves the more relevant words in the text, making it easier to detect emotions with the lexicon. The emotion lexicon we used is called the NRC Emotion Lexicon and it is a list of English words and their associations with eight basic emotions (anger, fear, trust, joy, anticipation, surprise, sadness, and disgust) and additionally, two sentiments (positive and negative) (Mohammad). In this lexicon, it uses the number 1 which means that there exists an association between the word. It also uses number 0 meaning no association with the word. For each song, the string of words representing the lyrics of a song was passed through the NRC Emotion Lexicon using a for loop and we keep count and add up the times each emotion associated with the words in the lyrics appears in each song. We count up if the emotion has the number one. Then, the emotion that appears the most in the words of each song is added to an array. A new column is added to the dataset which has an array of all the emotions for each song. The newly modified dataset with the new 'emotion' column is saved in a new CSV as "lyrics_emotion_v1.csv".

Results

The new modified dataset was saved assigned to a variable named "lyrics_emotion_dataset.". This new 'emotion' column was used to find the number of emotion combinations based on the new dataset. We chose the top 10 emotion combinations of the dataset because there was many combinations (in fact, 371 in total), and we were interested in the most commonly occurring emotion combinations. We wanted to see the partial percentages of the emotion combinations and how it was divided up. Figure 1 stated below shows that positive emotion has the most combinations in the lyric's dataset. Figure 2 shows that throughout the years positive emotion is the most prominent emotion.

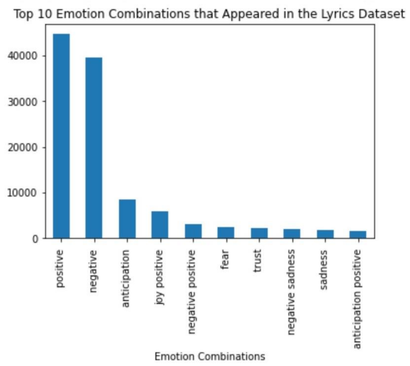


Figure1. Top ten emotion combinations found in the Lyrics dataset

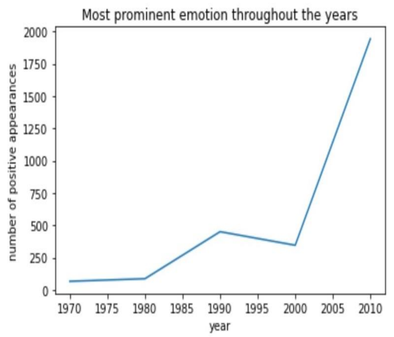
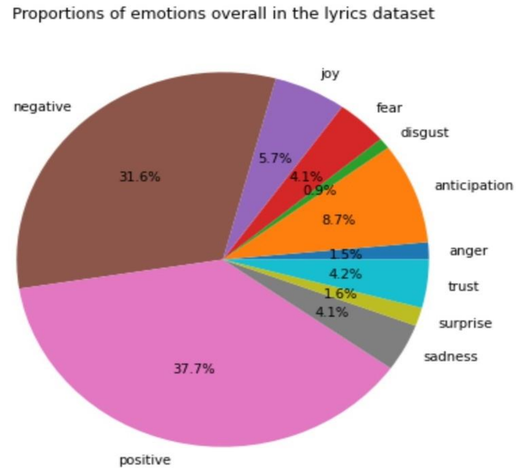


Figure 2. Most prominent emotion throughout the years found in the new modified dataset.



Discussion

Based on the figures, the emotion that was shown the most in the lyrics dataset is positive. Positive emotion seems to have about 44729 combinations overall in the lyrics modified dataset. About 37.7% of positive emotion represents the overall emotions' proportions. This emotion also is the most prominent emotion throughout the years. However, anticipation and positive had the least combinations. Along with disgust having the lowest percentage of emotions' proportions.

Throughout the process of analysis, we've noticed that the majority of the songs have single or multiple emotions. Also, positive was shown mostly throughout the years which affected our analysis. Our future improvements would perhaps be to find a dataset with more variety in genres and includes the recent year. And if a dataset doesn't exist, we can create a dataset by combining smaller datasets.

Conclusions

Sentiment analysis is extremely useful in our everyday life. In social media, through sentiment analysis, it allows us to gain an overview of the wide public opinion on certain topics. For instance, in our project, we used sentiment analysis to analyze the emotions from a set of lyrics dataset. Through the process of analyzing the dataset, it allows us to visualize the emotions one song implies. Overall, from our data, it is shown that the emotion of 'positive' and 'negative' are the most popular once. Most of the songs portrayed the emotions of either positive or negative. Moreover, we also noticed that there are very few songs that were written throughout the years of 1970s to 1990s with positive emotions. However, from 2000 to 2010, the songs with the most positive emotions have been increasing throughout.

In future work, it would be interesting to explore a dataset of songs of a different language because we can figure out when words from a different language translated in English have the same emotion as the English lyrics dataset. We also want to explore more genres and use another lexicon that detects word associations and labels each word as a color.

Acknowledgements:

Many thanks to Helen Jin and the PTS³ team

References:

1. Brownlee, Jason. "What Is Natural Language Processing?" *Machine Learning Mastery*, 7 Aug. 2019, machinelearningmastery.com/natural-language-processing/.
2. Mohammad, Saif. "NRC Word-Emotion Association Lexicon (Aka EmoLex)." *NRC Emotion Lexicon*, saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm.