

Sentimental analysis of Song, lyrics and predict the Human emotions

Vivek Kumar¹, Ajay Kumar², Yogendra Narayan Prajapati³
MIET, Meerut¹²³

Article Info

Volume 82

Page Number: 15613 - 15617

Publication Issue:

January-February 2020

Article History

Article Received: 18 May 2019

Revised: 14 July 2019

Accepted: 22 December 2019

Publication: 28 February 2020

Abstract

This paper briefly describes the effects of music of an artist on a listener. With the choice of song of an individual, not only we can predict the emotions of the individual but also can assist to improve the mental state. Out of all the features of a song, lyrics play the most important role. The aim of this research is to predict the emotion of a song, either positive or negative. Texts in the lyrics can be used to check whether the song is appropriate or not for the listener by analyzing positive or negative content of the text. Sentimental analysis is a field of study which deals with classification of the polarity of emotions or features to find details of the subject. We proposed sentiment analysis method in which the linguistic analysis of the lyrics of songs of artists is done. The data is constructed using billboard Hot 100 artists. Top 10 songs of each artist are chosen using scribbles from last.fm website. The lyrics of each song are taken from A-Zlyrics.com and are refined. We will predict the sentiments of the songs belonging to an artist and whether it is safe to listen to the artist or not.

Keywords: Sentimental analysis, linguistic analysis, listener by analyzing positive or negative

I. INTRODUCTION

Sentiment analysis has been one of the top fields of research in the last few decades. It studies the opinions, thoughts, feelings or moods of the subject and on the basis of that study it determines the subject's attitude towards a particular topic is either positive, negative or neutral. Sentiment analysis also helps to determine and predict that effects of a product on an individual. A lot of work is going on in this field and here we are going to analyse one of its domain, music.

Music is an integral part of human's life. Listening to music is one of the many activities that an individual performs in his/her daily life. Music is the medium of expressing our feelings and we tend to choose that particular music that expresses our current thoughts and feelings. The feelings are best described by words and words in a song are called lyrics. The listener chooses a particular song and artist on the basis of the type of lyrics, beats, life story, genre etc. The type of artist we listen to expresses a lot about our character and feelings and that is what we are going to present in this paper. Till now a lot of exploration has been done on analysis of music on the

basis of lyrics, genre and audio features but the sentiment analysis of music on the basis of lyrics and artists still remains unexplored.

Here in this paper we are going to propose and algorithm that will perform to tasks. First, the lyrics of song and whether the content is positive, negative or neutral. Second, we will try to predict whether a particular artist puts a positive or negative effect on the listener. The latter task is performed on top 100 artists of the year, taken from Billboard top 100.

We have used SentiWordNet, which is a lexical resource for opinion mining. It gives three sentiment scores, negativity, positivity and objectivity. Direct score allocation to the words of lyrics would be less accurate. In order to increase the accuracy, we performed normalization on the lyrics of the song.

Previous work:

So far, the research has been done on finding and evaluating the emotions of a song on the basis of the music. This is implemented using various audio features like beats, genre, volume, lyrics, pitch etc. Mood of songs have been analyzed using lyrics and audio as a combination. SentiWordNet is used by

many researchers to allocate a number to the words, this number ranges between (-1,1) and tell the sentiment score of each word i.e. negativity and positivity. The lyrics are analysed using various algorithms like Naïve Bayes, Support Vector Machine and K-Nearest Neighbor (k-NN algorithm). Researchers have also used POS Tagger to classify the words of the lyrics.

Proposed Work

In this section we will describe the algorithm we conducted to analyze artists and their songs. Top 30 songs are selected for each artist of the TOP 100 artists of the year 2018, taken on the basis on scribbles from *last.fm*. The lyrics of these 30 songs are pre-processed, scored and analyzed.

1. Determining the vocabulary of songs

1.1 Normalization of lyrics

The lyrics of each song to be scored at once would lead to

Inconsistency, so they are first normalized i.e. we will perform text pre-processing on the lyrics. The normalization process included two steps: tokenization and lemmatization. Tokenization will perform the task of chopping off the lyrics (sequences of characters) into pieces, called tokens. The tokens generated will be free from some characters like punctuation. Lemmatization uses vocabulary and morphological analysis of words to eliminate the derivative form or inflectional endings of words and to return to base/root form. So, the word *saw* becomes *see* or *saw*, depending on whether it's a verb or noun.

For example, *I wannahold'em like they do in Texas plays*.

After tokenization becomes *I wanna hold 'em like they do in Texas plays*

After lemmatization becomes *be want hold like they do in Texas play*

1.2 Removal of stop words and POS Tagging

The stop words, like a, the, about, an etc. are removed from the lyrics document as they do not add to the emotion and sentiment of the song. The output after this process is passed to the POS Tagger which will

classify each word as noun, verb or adjective. POS tags will be useful at the time of scoring the words with the help of SentiWord Net.

2. Scoring using SentiWordNet

This step accounts for the score allocation to each word of the pre-processed lyrics. The total positive and negative score for each song is added up separately and dividing by the number of words in a song respectively, we calculated the percent positive and negative content in a song. This step is repeated for the 30 songs selected for each artist.

S.No	Song	Positive Score	Negative Score	Words
1	Problem	11.66724077	2.943691932	23
2	Break Free	2.244570871	6.471677918	34
3	One Last Time	9.517126097	4.315042417	30
4	Love Me Harder	23.64812016	3.581856258	61
5	Into You	1.479140562	18.25828843	65
6	Side to Side	3.64525857	5.976409801	56
7	No Tears Left to Cry	2.507280073	4.476879415	19
8	dangerous woman	1.596999275	5.552216775	69
9	Thank you next	20.09458135	8.661874558	53
10	Santa Tell Me	9.016466926	2.704205309	44
11	god is a woman	9.129456929	0.336805974	57
12	the way	17.86037077	2.761986397	110
13	Focus	8.0998573	1.583355464	54
14	7 rings	7.755569492	7.618973471	35
15	Breathin	2.8250321	4.902223919	20
16	Be Alright	7.911424553	6.884807927	54
17	Imagine	2.710552458	15.96221059	30
18	Everyday	11.95903044	3.454072296	38
19	Just a Little Bit of Your Heart	2.943923928	13.15250828	63

20	Best Mistake	8.856709017	7.776100538	77
21	Greedy	10.5879972	0.885187695	83
22	My Everything	6.03935857	6.428725923	28
23	Breakup with your girlfriend	-2.335598027	7.77908	39
24	Why Try	14.53962736	3.002757579	44
25	Baby I	9.56221858	1.297788727	39
26	Everytime	2.489075591	4.630269827	36
27	Let Me Love You	6.997818938	2.762091879	64
28	Honeymoon Avenue	5.200711759	10.57276393	50
29	Honeymoon Avenue	5.200711759	10.57276393	19
30	sweetener	4.429986043	2.689493875	94

S.No	Artist	Positive Score	Negative Score
1	Ariana Grande	8.786399134	5.421697088
2	Post Malone	5.499611593	9.596035633
3	Queen	7.245497419	5.959126998
4	Marshmello	5.332353799	5.767306769
5	Halsey	7.771275139	5.840449226
6	Maroon 5	6.908011859	7.640312169
7	Billie Eilish	7.123700132	7.842549714
8	Imagine Dragons	4.631276186	6.210930317
9	Cardi B	5.167737048	10.42757658
10	BTS	8.130091191	8.685409514
11	Travis Scott	5.198996889	6.14417098
12	Panic! at the Disco	7.09530951	7.667227484
13	Drake	6.785273751	6.368609988
14	Khalid	7.44699075	6.579541141
15	Lady Gaga	9.711111395	6.769220083
16	Luke Combs	4.879264372	6.213448888
17	XXXTENTACION	3.837035384	13.02844471
18	21 Savage	4.186015313	9.87279197
19	Lauren Daigle	6.612513194	4.764164222
20	Ed Sheeran	8.379375249	4.650796784
21	Luis Fonsi	8.641736018	6.730906006
22	Lil Baby	5.08885403	7.961830673
23	Breakup with your girlfriend	-2.335598027	7.77908
24	Why Try	14.53962736	3.002757579
25	Baby I	9.56221858	1.297788727
26	Everytime	2.489075591	4.630269827
27	Let Me Love You	6.997818938	2.762091879

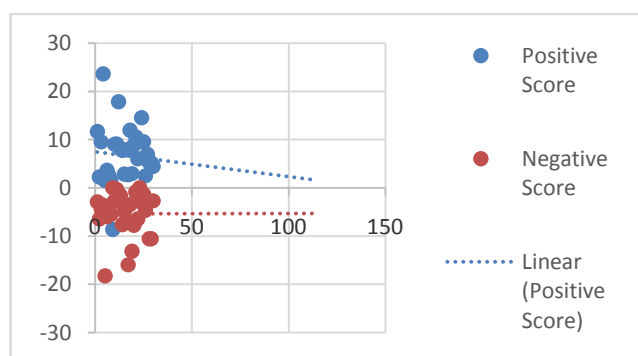


Figure for top 10 songs of an artist

This step is then repeated for 50 artists, and in total of 1500 songs.

The below chart shows the score of each artist, based on the top 30 songs. The score of all the 30 songs are added, and when divided by 30 would provide us with the average positive and negative content. The mean of these 50 artists or 1500 songs will give the average positive and negative content in a song. This means value can be used to determine the nature of the song.

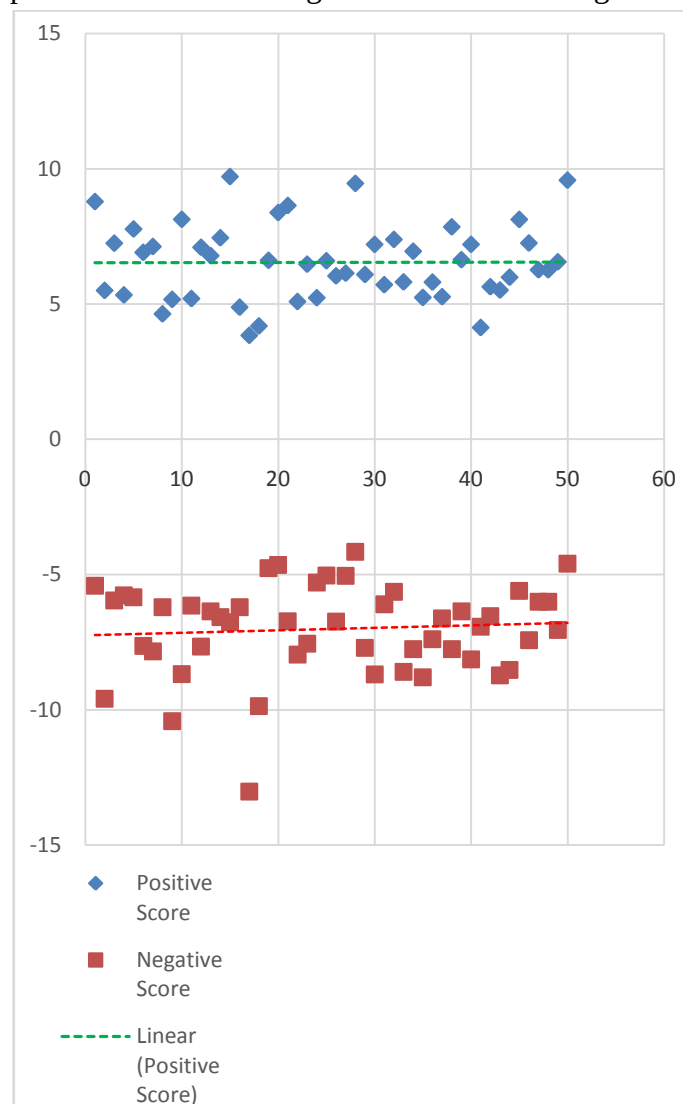
28	Honeymoon Avenue	5.200711759	10.57276393
29	Honeymoon Avenue	5.200711759	10.57276393
30	sweetener	4.429986043	2.689493875
31	Juice WRLD	7.202948168	8.700881216
32	Swae Lee	5.711213207	6.102661567
33	Bruno Mars	7.386414901	5.64414816
34	Eminem	5.813945375	8.596267303
35	Ella Mai	6.94930989	7.760101802
36	Mettalica	5.239722923	8.806410967
37	A BoogiewitdaHoodie	5.808131314	7.387587225
38	Kodak Black	5.266524307	6.623747745
39	Shawn Mendes	7.84960958	7.755077517
40	FloridaGeorgia Lines	6.636423506	6.366323597
41	Taylor Swift	7.203749379	8.14188247
42	Bastille	4.127090216	6.934295513
43	Meek Mill	5.512597224	8.728619583
44	Bad Bunny	5.987588542	8.531789874
45	Camila Cabello	8.126303128	5.60511179
46	Pink	7.252653058	7.42989819
47	Jason Aldean	6.260420408	6.007695467
48	Jason Aldean	6.260420408	6.007695467
49	Chain Smoker	6.556540031	7.051324075
50	Greta Van Fleet	9.577482266	4.595984906

RESULT

By comparing the mean values of each artist we can determine the neutrality of the artist at individual level. In the example taken here, we can state that for an artist the positive content is 8.79 and the negative content is 5.42. Thus, clearly the positive content in

the artist's song exceeds by 3.37 % making a positive impact on the listener. Hence, the artist is positive.

When repeated for 50 artists and their mean, the average negative content is 7.02 and average positive content is 6.54. The difference of these values, i.e., 0.48 gives us the threshold value. This threshold is a striking balance between the positive content and negative content in a song.



Conclusion and Future Work:

Inman previous researches, It has been observed that there is a significant impact of music on an individual's emotion and mental state. Researchers have found out that lyrics plays a major role in this phenomenon. This research helps us to get which artist makes trending positive or negative content. Hence can be used to built various applications like a locking an artist before reaching to the youth. Or

guessing the upcoming trend by plotting the previous years.

Other features like duration of songs, social tags, number of scrobbles can be added in further researches. These will help in getting more accuracy in threshold. We also intend to perform analysis on different languages including hindi songs also.

References

- [1] Robert Neumayer and Andreas Rauber. Integration of Text and Audio Features for Genre Classification in Music Information Retrieval. Vienna Institute of Technology, Institute of Software Technology and Interactive Systems
- [2] Will Archer Arentz, Magnus Lie Hetland, Bjorn Olstad. Retrieving musical information based on rhythm and pitch correlations
- [3] Byeong-jun Han, Seungmin Rho Roger B. Dannenberg Eenjun Hwang. SMERS: Music Emotion Recognition using Support Vector Regression. 10th International Society for Music Information Retrieval Conference (ISMIR)
- [4] Cyril Laurier, Perfecto Herrera. Audio Music Mood Classification Using Support Vector Machine. Music Technology Group, Universitat Pompeu Fabra
- [5] Patrik N. Juslin and John A. Sloboda. Music and Emotion: Theory and Research. Oxford, UK.
- [6] Robert E. Thayer. The Biopsychology of Mood and Arousal. Oxford University Press, 1989.
- [7] Vandana Korde, C Namrata Mahender. Text Classification and Classifiers Survey. International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.2, March 2012
- [8] Maite Tabode, Julian Brooke, Manfred Stede. Genre-based Classification for Sentiment Analysis. Proceedings of SIGDIAL 2009: the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue, 2009.
- [9] Ashis Kumar Mandal and Rikta Sen. Supervised Learning Methods for Bangla Web Document Categorization. International Journal of Artificial Intelligence & Applications (IJAIA), Vol.5, No.5, September 2014.
- [10] Menno Van Zaanen, Peiter Kanter. Automatic Mood Classification using TF*IDF based on Lyrics. International Society for Music Information Retrieval, 2010.
- [11] James A. Russell, Lisa Feldman Barrett. Core Affect, Prototypical Emotional Episodes, and Other things called Emotion: Dissecting the Elephant. Journal of Personality and Social Psychology.
- [12] David Rubin and Jennifer M. Talarico. A comparison of Dimensional Models of Emotion: Evidence from Emotions, Prototypical Events, Autobiographical Memories, and Words. US National Library of Medicine, National Institute of Health, PMC, Nov 2010.