

Twitter and song lyrics Sentiment analysis

Fredrik Wilhelm Butler Wang

Informatikk / 2020

`fredriwb@stud.ntnu.no`

Abstract

This report explores the correlation between song lyrics and tweets written by the same artists by using sentiment analysis, to give an indication on whether what artists say in their songs can relate to how they express them self on social media.

In this report two methods of sentiment analysis as been used on song lyrics as well as the performing artists or band tweets from their own twitter account. The first being a Naive Bayes classifier from NLTK that defines each tweet or lyrics line as positive or negative. The second method makes use of a VADER analysis tool that classifies text as neutral as well as positive and negative.

1 Introduction

Sentiment analysis is the use of various text analysis and processing to study and interpret emotions that exist within the text that is analyzed. The usage of sentiment analysis on social media, and also reviews and feedback of products is already quite popular and is a service that many companies provide, for instance social media monitoring to provide a indication of a product or a company's popularity among users. Because sentiment analysis classifies the emotion of the text that is analyzed, the motivation for this report is to see if it exists a correlation between an artists two main ways of communicating and expressing them self to the world, that is their songs and their own social media accounts. Also a more overall objective of this report is to perform and study the process of an sentiment analysis.

2 Background

This report uses two types of models to perform sentiment analysis. One being a Naive Bayes Classifier. How the model works is described in more detail in the 4 section but it is based on Bayes' theorem, or Bayes' rule, that calculates conditional probabilities (Joyce, 2003).

The VADER model, developed by Hutto and Gilbert (2014) is the other model being used and instead of having training data it uses a sort of gold standard lexicon created by humans to analyze the data. More details on how the model work can be found in the 4 section

For accessing data this report uses genius.com for song lyrics, Genius is a website that contains almost any song lyrics, and it also has an API that is open for the public that allows for fast downloading of large amount of lyrical data. This report did not rely heavily on the Genius API, but is definitely recommended if one wants to do something familiar in a larger scale.

The other type of data that has been used is tweets from Twitter.com, these tweets were downloaded using the Twitter API which is also open for public use.

3 Related Work

A description of some of the work that has been done relating to sentiment analysis on tweets and song lyrics is found here.

In general there are exists a lot of papers about sentiment analysis, for instance Pang and Lee (2008) takes a look at the performance of various techniques using Machine Learning, one of them being the Naive Bayes classifier which has been used in this report.

There are many papers written on sentiment analysis based on tweets from Twitter, or social media posts in general. For instance Kouloumpis et al. (2011) that looked at the effects of part-of-speech features, existing sentiment lexicons and the presence of intensifiers, for instance emoticons when doing a sentiment analysis on tweets.

As for sentiment analysis on song lyrics the research field appears to be a bit smaller, although Napier and Shamir (2018) gives an insight on how analyzing song lyrics, especially from different time periods can tell us a lot about the way language being used in songs change over the years.

4 Architecture

Here we find a description of the different models used, also a discussion on the data that is being used is included in this section.

4.1 Naive Bayes Classifier

The Naive Bayes classifier from nltk is a classifier that is based on the Naive Bayes algorithm, which in short means that between each pair of features the classifier makes an naive assumption that their is independence between the pairs. This means that the classifier is a very simple model. It still has a lot of advantages as it is very scalable because the model uses counting for seeing whether a text, in this case a tweet or lyrics line contains a word. Therefore training on large data sets is easy and efficient. In this report a classifier was trained on twitter samples from the nltk corpus. This corpus contains tweets that already have been classified as positive or negative. After the training the accuracy of the classifier was calculated to 99 percent correctness which shows that the model indeed is quite accurate for classifying tweets

4.2 VADER

The VADER (for Valence Aware Dictionary for sEntiment Reasoning) model used is a Vader model in nltk that is designed for the language and sentiment used in social media context. The model has a standard of lexical features that it considers important and as a sort of "gold standard". These features are used by the model to indicate the intensity of sentiment of, in this case, tweets. It can therefore give the grade of how positive, negative and neutral a tweets is. In simpler terms one might say that the models understand that one sentence is for instance more positive than another, due to the content of a certain lexical feature. This also means that the model does not have to be trained on any data, which makes it quite nice for a project like this where computer processing power is limited. In this report the VADER model calculated the sentiment score of each lyrics line or tweet and then, based on the values categorized them as either neutral, positive or negative.

Figure 1 shows the architecture of the VADER model, where firstly the twitter feed of the artist or their song lyrics is retrieved, then a process of cleaning up the data is done, mainly for the tweets as the lyrics are, as previously described already handled. Then the model is applied the data and eventually calculate a score for each line of text it is fed. More details on the code is given on the Github repo which can be found at 4.4.

4.3 Data being used

The data used in the report come mainly from two different places, being Genius for song lyrics and Twitter for tweets. Both methods use both the song lyrics and the tweets. 5 artists was chosen at somewhat random although with a goal of selecting artists from different genres and periods, while also making sure that they do have a somewhat active Twitter account. For each artist/band 5 songs have been downloaded from genius.com, with the removal of some parts, mainly the removal of section headers as these don't have anything to do with the actual lyrics. For each artist there exists a json file containing all the lyrics. For tweets 500 tweets from each account was downloaded using the Twitter API into an json file, with the removal of links, some stop words and mentions of other users.

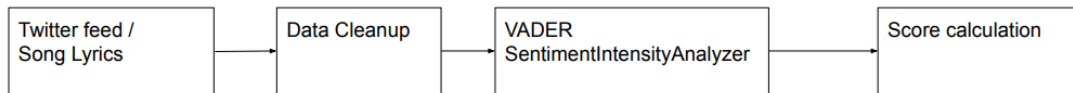


Figure 1: The architecture of the VADER Model

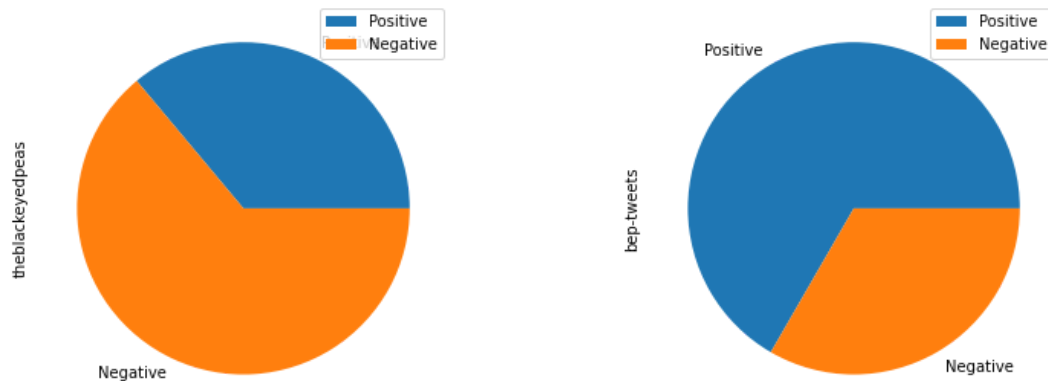


Figure 2: Results from the NB classifier on The Black Eyed Peas. Lyrics on the left and tweets on the right

4.4 Github repo

A link to the github repo can be found here: <https://github.com/fredrikbw/sentimentArtists>

5 Experiments and Results

The setup used for this experiment, along with the results of the experiments is described in this section. The full tables and diagrams can be found in the Github repo for this project.

The experiment in this report consists of performing sentiment analysis using the two models previously described on each of the artists tweets and song lyrics. The results from the the analysis are then to be compared using tables and diagrams.

5.1 Experimental Setup

To repeat the experiments all data is available in the linked Github repo. In the repo one will also find parameters that one can change with comments explaining what the parameter does.

The repo also includes a requirements.txt file containing the needed Python modules. The code has been written in Python 3.7.

5.2 Experimental Results

In figure 2 the result from the Naive Bayes classifier is rather interesting as it has found that for the band The Black Eyed Peas their lyrics are more negative than positive, but for their tweets it's about the opposite.

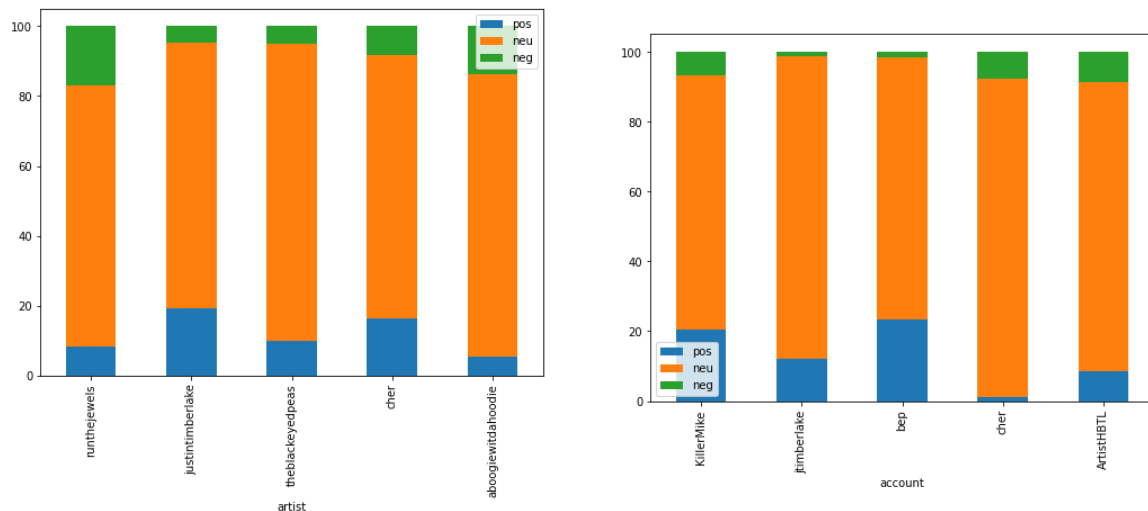


Figure 3: Results from the VADER model, song lyrics on the left. Tweets on the right

As seen in Figure 3 The VADER model produced two diagrams, categorizes both tweets and line of lyrics into positive, neutral and negative. One can clearly see from the differences in the graph that there is not a direct correlation between song lyrics and tweets, neither was this expected. One of the most interesting things about the results is that for the music group RunTheJewels, and it's member Killer Mike, the positive and negative values have switched places, giving the indication that the artist has a negative language emotion vise in their songs, but are mostly positive on social media.

Another interesting result is the extreme lack of positiveness on the artist Cher's Twitter account. This could be a result of the fact that this particular artist is known for both rants as well as twitter wars with other celebrities (cher/twitter, 2015)

All of the results can bee seen in the Github repo.

6 Evaluation and Discussion

Here is an evaluation regarding the learning outcomes from the conducted experiments and the process in general, as well as discussion on the limitation and validity of the work.

From a learning perspective it has become more clearer on a more general basis how powerful the usage of sentiment analysis can be. For the type of usage done in this report the categorization of whether or not the data is positive, negative or neutral may not give such a clear indication of what emotion the person singing or tweeting has. Though it does give some insight, and a pointer to the differences between musicians.

This work is obviously limited by both a small data sample as well as its usage of fairly simple sentiment analysis tools. For more accurate measurements the use of larger data samples would give better and more precise results to compare tweets and lyrics. Another factor consider is the cleaning of data, both tweets and lyrics, for instance the removal of more unnecessary words in song lyrics like lines consisting of "La la la", as one might think such a line does not say that much about emotions. What also could make a difference is a more advanced analysis of what each emoji in a tweet means in a sentiment analysis, so that the analysis tool used has a deeper understanding than just the most basic emojis and emoticons. The usage of more complex models, would also give a more detailed result. By having more models one might also be able to easier detect flaws of the different models and inspect text that got particularly different results depending on the model being used.

7 Conclusion and Future Work

Here is a conclusion on what as been achieved in this report, as well as an discussion regarding future work.

7.1 Conclusion

The results of this experiment does not really say that much about the correlation. Partly because their is not enough data, but also because the classifiers used are mainly created for handling social media posts and not also lyrics. Regarding the goal of performing sentiment analysis on text it has been rather interesting to see how fast one can perform such analysis, meaning the work that has been done can without to much hassle be scaled up to a much larger scale to hopefully achieve more interesting and precise results.

7.2 Future Work

A way of extending the work could be to not only look at positiveness and negativeness, but also the degree of excitement in the text, as well as exploring what words that are typical for each artist to see if certain ways of expressing themselves can be identified in both music as well as social media. Napier and Shamir (2018) Did show that their is a lot of features to extract from song lyrics and therefore it could be interesting to see whether other types of emotions and usage of words can be related to the way artists tweets. Another thing that is interesting is the analysis of whether an artist is currently active or not. To see if there exists some relation between an artists who currently does not use their social media to promote their own music, but instead uses it to express their thoughts on society, politics, etc. As one might think that tweeting opinions might often contain more negative loaded words.

For any future work it would also be interesting to use more models, who for instance rely on neural networks, and thus getting a more broad approach to sentiment analysis.

References

- cher/twitter. Cher tweet 17.06.2015, 2015. URL <https://twitter.com/cher/status/611054781809848320>.
- Clayton J Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*, 2014.
- James Joyce. Bayes' theorem. 2003.
- Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore. Twitter sentiment analysis: The good the bad and the omg! 2011. URL <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2857/3251>.
- Kathleen Napier and Lior Shamir. Quantitative sentiment analysis of lyrics in popular music. *Journal of Popular Music Studies*, 30(4):161–176, 2018.
- Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.