

SENTIMENT ANALYSIS AND CLASSIFICATION OF MUSIC LYRICS

GitHub: <https://github.com/xzhangfox/ANALYSE-AND-CLASSIFY-MUSIC-BY-LYRICS>

Introduction

Music is an important part of human society. It can pump us up when we work out, comfort us when we are depressed. Because of the variety of music, it can always satisfy our needs. On the other hand, music also has a strong market value. In total, the American music industry's 2019 retail revenue was about \$11 billion, up from \$10 billion in 2018 and \$9 billion in 2017. As a rapidly growing industry, music has great potential. And as one of the most important parts of music, lyrics offer NLP direction to dig deeper. The sentiment and genre of lyrics can be analyzed by NLP and machine learning models. Music classification has potential value in the market.

Background

Based on the online articles, the researchers are seeing the value of the increasing amount of musical information that is being published daily on platforms like Social Networks, Digital Libraries, and Web Pages. One of them is from the Pompeu Fabra University. The music professionals designed the tutorial aims to provide the audience with potential applications of Natural Language Processing to MIR and Computational Musicology. It gives us a handful of sources to learn the linguistic, semantic, and statistical-based approaches to extract and formalize knowledge about music from naturally occurring text.

Literature Review

This project is guided by research papers and projects about how to do lyrics data collection, what kind of sentiment the lyrics can have, and how to annotate the lyrics. And also, we spent most of the time selecting the models for best fitting our data.

Motivation and Scope

We wanted to use what we learned in the NLP course to make the lyric data truly marketable. To achieve this goal, we fixed the research framework on the goal of how to make the lyrics data needed by the market. You will see in our project how we preprocess the data, adjust and select the optimal model, and how the final result is applied to the actual example.

Dataset

1) Data collection and description

The data is from the million song dataset website and we randomly selected 1000 songs from it. But the dataset doesn't include lyrics. Therefore, we wrote the script by python to download lyrics from Lyric Wiki by traversing artists and music titles. Table 1 shows the links of the datasets.

Every single observation should be the whole lyrics. But our raw data contains too many distractions, like Articles, Pronouns, Conjunctions. Especially for lyrics data, we found there are a lot of musical terms like Verse and Chorus.

2) Exploratory Data Analysis

After we removed such stopwords, there are still many words that have no sentiment. You can see our bigrams and trigrams here. The "oh oh" or "la la la" can't express any emotion without rhythms. So we must figure out how to target the data we need.

3) Data Preprocessing

The basic lexicon we used for filtering of lyrics is ANEW (Affective Norms for English Words) which provides a set of normative emotional ratings for 1034 unique English

words. Since our dataset wasn't big enough, we incorporated another sentiment lexicon, which contained both positive(well, love, like, won, right, best, top, beautiful, etc.) and negative(dead, lost, attack, killed, problems, issues, etc.) emotions, to filter out more words that contained emotions. After filtering, five songs without emotional lyrics were eliminated, and 37,351 emotional words were selected. This data sets the stage for future modeling.

Modeling and Result

1) Modeling training

We decided to use the Latent Dirichlet Allocation model(LDA). LDA is a generative probabilistic model that assumes each topic is a mixture over an underlying set of words, and each document is a mixture of over a set of topic probabilities. Compared to the other topic models like Latent semantic analysis(LSA) and Hierarchical Dirichlet Process(HDP), the LDA model gives us the best Topic Coherence, which measures the degree of semantic similarity between high scoring words in the topic.

As Figure 1 shows, the model shows the best performance measured by the coherence score. Thus, six topics built by lda model are shown in the report.

Figure 2 shows the wordcloud of each topic and the label of each topic which is given manually. As we can see, some topics seem similar, such as topic 1, 3 and 5, which both contain the word 'love' but there are some differences. As Figure 3 shows, topic 0 represents a very clear angry emotion as it contains many relative words like 'war', 'burn' and 'fight'. Topic 1 suggests some sad love story as 'love' has a very high frequency but words like 'pain' and 'hurt' also appear. Topic 2 can be hardly defined since it seems that its wordcloud contains complex emotions with not only 'happy' but also 'hate'. Topic 3 looks much clearer which contains all positive words. Topic 4 seems to represent a more peaceful and comfortable emotion as 'home', 'joy' and 'bird'. Topic

5 also suggests a complex emotion including both love emotion with 'love', 'like' and sad emotion with 'sick', 'lie'.

To better understand the model, we have also used the popular visualization package called pyLDAvislization. In Figure 5, It includes the Intertropical Distance Plot and the top words associated with each topic.

2) Model Testing

In this part, we choose two new songs which are not included in our training dataset to test our model. The first song is Eminem's 《One Shot, two shot》 (Song1), and the second song is Bob Marley's 《Three Little Birds》 (Song2). Table 2 shows parts of the lyrics of the two songs.

The lyrics of Song1 contains many positive words like 'shot', 'hit', 'killed'. And the model result (Figure 4) looks very accurate. It shows that topic 0 has the largest possibility with over 60%, which we defined as an angry topic.

The lyrics of Song 2 contain many positive words like 'happy', 'love'. And the model result(Figure 5) shows that topic 2 has the largest possibly with over 60%. And 21% for topic 5 and 14% for topic 3. In this case, it seems really hard to decide as each one seems reasonable. It may result from the fact that these three topics have some similarities, and the lyrics of this song are so simple that not too many features can be captured by the model.

Future Work

For the future work, we would like to conduct the classification model like Naive Bayes to train and predict the song's mood based on the available features. We believe this kind of project can be applied to the music industry and helps to build the music suggestion engine.

References

- Kaggle:
- <https://www.kaggle.com/mousehead/songlyrics>
- Oramas, S.(2018), Natural Language Processing for Music Knowledge Discovery,
•<https://arxiv.org/abs/1807.02200>
- Oramas, S.(2018), Natural Language Processing for Music Information Retrieval,
•<https://www.upf.edu/web/mdm-dtic/tutorial-natural-language-processing-for-music-information-retrieval>

Appendix

```
(0, '0.057*war" + 0.051*burn" + 0.035*sun" + 0.027*sin" + 0.027*fight" + 0.025*easi'))
(1, '0.360*love" + 0.032*god" + 0.031*like" + 0.025*babi" + 0.023*heart" + 0.014*pain'))
(2, '0.060*babi" + 0.057*fire" + 0.049*time" + 0.030*happi" + 0.027*heart" + 0.026*love'))
(3, '0.075*girl" + 0.061*christma" + 0.056*love" + 0.053*lone" + 0.032*merri" + 0.027*babi'))
(4, '0.097*home" + 0.047*joy" + 0.043*good" + 0.023*babi" + 0.018*time" + 0.017*love'))
(5, '0.044*like" + 0.042*life" + 0.025*lie" + 0.019*love" + 0.017*sick" + 0.017*slow'))
```

Figure 1: LDA Model Results



Figure 2: Wordclouds of Each Topic

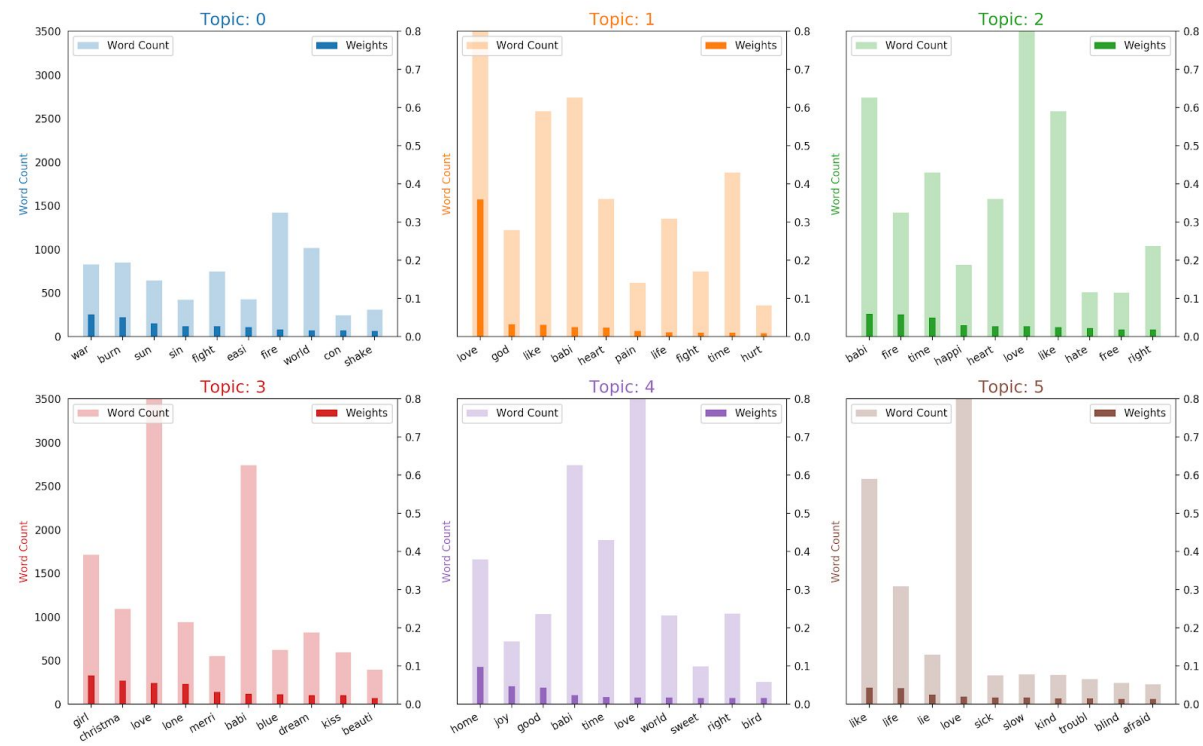


Figure 3: Most relevant words of each topic

Eminem | One shot, two shot - Violence, Anger

Sentiment Words

{'like', 'bar', 'right', 'life', 'odd', 'music', 'hand', 'money', 'mess', 'damn', 'kill', 'believ', 'truck', 'human', 'gun', 'bodi', 'thank', 'peopl', 'fight', 'face', 'drunk', 'collaps', 'vest', 'god', 'wife', 'black', 'floor', 'car', 'tire', 'plenti', 'dumb', 'use', 'miss', 'hit', 'bitch', 'hug', 'argument', 'aliv', 'clear', 'man', 'swift', 'cat', 'fun', 'hope', 'bizzarr', 'sound', 'cliqu', 'fast', 'strut', 'hous', 'fuck', 'desert', 'door', 'fire', 'spray', 'shit', 'outsid', 'diss', 'save', 'parti', 'bullet', 'reveng'}

[(0, 0.61559546), (1, 0.024770392), (2, 0.00114846), (3, 0.001139392), (4, 0.0011462743), (5, 0.3562001)]

Figure 4: LDA Model Result of Song 1

Bob Marley | Three Little Birds - Love Happy

Sentiment Words

{'blue', 'happi', 'babi', 'right', 'mind', 'life', 'girl', 'world', 'eas', 'sky', 'invest', 'fine', 'imagin', 'love'}

[(0, 0.0050705294), (1, 0.0051040943), (2, 0.6410427), (3, 0.13757965), (4, 0.0050775856), (5, 0.2061254)]

Figure 5: LDA Model Result of Song 2

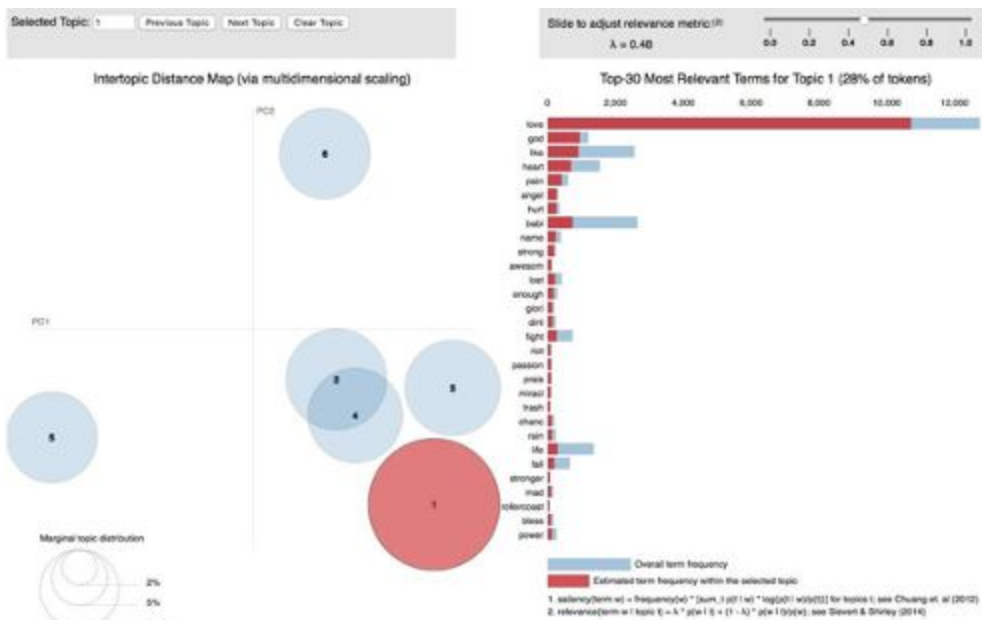


Figure 6: Topic Modeling Visualization by Using pyLDAvis

Million song dataset	http://millionsongdataset.com/pages/getting-dataset/
Lyric Wiki	https://lyrics.fandom.com/wiki/LyricWiki

Table1: Links of datasets

Song1- 《One Shot, two shot》	Song2- 《Three Little Birds》
I'm tryna slip through the exit and get to where my car is at Bitches screamin' everywhere and niggas is wildin' Two minutes ago we was all jokin' and smilin' This chick is clingin' onto me sobbin' and sighin' Sayin' she didn't mean to diss me earlier and she cryin'	Imagine me and you, I do I think about you day and night, it's only right To think about the girl you love and hold her tight So happy together If I should call you up, invest a dime

<p>But its real and cats is gettin' killed</p> <p>So I hugged her and used her body as a human shield</p> <p>And she got hit now she yellin'</p> <p>(Don't leave me!)</p> <p>I told her I'd be right back and the dumb bitch believed me</p> <p>I squeezed through the back door and made my escape</p> <p>I ran and got my 38 I hope its not to late</p> <p>One shot two shot three shot four shots</p> <p>.....</p> <p>God damn there just went another damn bullet I'm hit</p> <p>My vest is barely able to handle it, its to thin</p> <p>If I get hit again I can't do it, I scoop deep</p> <p>.....</p> <p>I ain't come in this bitch to party I came in this bitch to fight</p> <p>Although I can't stay here to fight 'cause I'm poppin' niggas tonight</p> <p>That's right bitches I'm drunk with revenge</p> <p>Shot a bouncer in the neck for tryna check when I get in</p> <p>Swift told me to meet him here so its clear that this fucker</p> <p>Shoot out the back of his truck goes up in this motherfucker</p> <p>.....</p>	<p>And you say you belong to me and ease my mind</p> <p>Imagine how the world could be, so very fine</p> <p>So happy together</p> <p>I can't see me lovin' nobody but you</p> <p>For all my life</p> <p>When you're with me, baby the skies'll be blue</p> <p>.....</p> <p>So happy together</p> <p>So happy together</p> <p>How is the weather</p> <p>So happy together</p> <p>We're happy together</p> <p>So happy together</p> <p>Happy together</p> <p>So happy together</p> <p>So happy together (ba-ba-ba-ba ba-ba-ba-ba)</p>
--	---

Table 2: Parts of Lyrics of Song 1 and Song 2