



Content Based Music Recommender System

D. Krishna Teja

Advisor : Dr. Carol Romanowski
Rochester Institute of Technology



Abstract

With the digitization of music, it has paved path for many changes, some of the significant ones being no CD's, online streaming services, and too many options. With many options, the users are confused to select what they want to hear thus the need for a recommender system had arisen. Since manual annotation to a large set of music collections is arduous, Topic Modelling is preferred to automatically identify the underlying topics present in the text sources. Lyrics are sometimes way to confusing and poetic to find the latent topics. In this project, the interpretations of songs were used to get more accurate results.

Introduction

• One of the main interest of music listeners is to understand the subject of the song, but finding the subject of a song is a very hard task when compared to finding the metadata such as title, artist, genres and lyrics.

• Many researchers have attempted to extract subject information using several supervised and unsupervised techniques on the song lyrics. While they produced quite good results, but the lyrics contain many words which are ambiguous in nature and made it difficult for the machine to understand the context of it.

• The analyzation of the interpretations of songs is an active research topic, Hence decided to perform my algorithm on the interpretation of the songs.

Technologies Used



Data

Searched a lot for an API/dataset which had interpretations of songs but could not find one. So finally decided to create one. Implemented a program with the help of python's beautiful soup package which parsed the websites and gathered lyrics, interpretations and tags of several albums. The websites used for these purposes are www.allmusic.com, and www.songmeanings.com. Due to the restriction of time, we created a small dataset which consisted about 250 songs. Even though its small its good enough to run the selected algorithms and get good results.

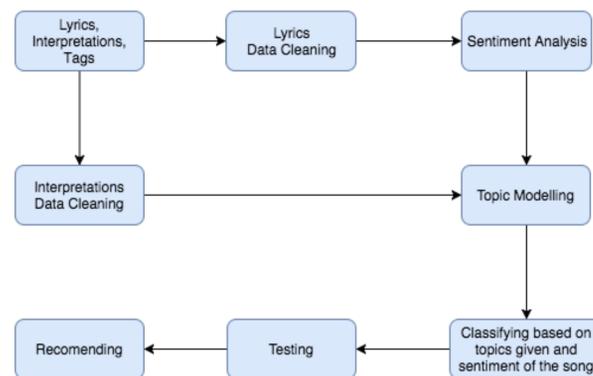


Fig 1: Methodology

Data Cleaning

Data collection and cleaning was the toughest part of the project. After collecting the data, we have performed word clouds on the data set to get a much better understanding. After looking at the word clouds decided to implement several data cleaning techniques such as Tokenization, Lower case conversion, Type checker, Stemming, Stop word removal, Punctuation removal, User Tag names removal, Digits removal.

Tokenization : Large corpus broken down into single words.

Type Checker : Used a dictionary to check if all the words are valid.

Stemming : Find the originate / base word ex: lovely -> love

Stop word removal : Added some additional words to the common English stop word list to better the accuracy of the system. Ex: Lyrics, song (These words do not add weight to the topics).

User Tag removal : Each comment has a user tag which is not required for the system hence wrote a regex to remove all such tags.

Results

General Commentok this song is part of the album the wall to fully understand this song u first must understand the album and watch the movie it helps. in the movie pink sits locked in a hotel room, somewhere in la. too many shows to much dope, too much applause: a burned out case. on the tv, an all too familiar war film flickers on the screen. we shuffle time and place, reality and nightmare as we venture into pink's painful memories, each one a "brick" in the wall he has gradually built around his feelings. Slowly he withdraws from the real world and slips further into his nightmare as he imagines himself as an unfeeling demagogue, for whom all that is left is the demonstration of power over his unthinking audience, the culmination of the odious excess of his own world and the world around him. his internal self trial follows as the witnesses of his past life, the very people who have contributed to the building of the wall, come forward to testify against him. thats the story of the wall.

comfortably numb is a song about how pink has finally burned out he can no longer tell the real world from the one he has created. his wall is built and he can no longer make it out. He has lived in pain and anguish for so long that he has finally given into the pain and the drugs. He now sits in his room realizing the only people that care anyone are the people only trying to make money off him like his manager. Instead of getting ready for his shows he now just lets them drug him up get him dressed and get him going hoping the drugs last long enough for him to get on stage perform then crash.

while most of the artists from pink floyd did experiment with many drugs this song is mainly about syd barrett and how he quit the band early on when things started getting good it wasn't what he wanted and soon gave up quit the band.

Fig2: Interpretation from Pink Floyd's The Wall



Fig3: Word Cloud of that album

Topics in LDA model:
Topic #0:
ice depth dragging trying evils suddenly skating theres parents soft people aware waters album fact finally hard way silent beginning
Topic #1:
link silent crack way arent slip hard generations beginning soft tear suddenly aware memories trying love album mother dont fact
Topic #2:
naive modern pinks parents thats stained understanding life theres love evils eyes cynical going generations disapproval ice roger fact means
Topic #3:
people tear stained reproach album song ice slip skating voice generations modern bad going suddenly trying parents life roger arent
Topic #4:
think look disapproval incident oooh love understanding really baby babe song parents fact way soft link million going album slip
Topic #5:
means memories mother says going silent slip evils crack modern maybe finally beginning pink tear fact generations just hard waters
Topic #6:
pink blue roger babe going dont things million understanding eyes dragging arent fact love soft says maybe mother voice waters
Topic #7:
life million just waters fact maybe baby cynical link blue really slip dragging memories things love crack modern song understanding
Topic #8:
eyes world things finally innocent cynical look generations million roger dragging parents mother really suddenly pinks means tear song aware
Topic #9:
aware bad voice thats dont oooh link look memories generations evils just dragging slip things roger finally theres going million

Happy: 0.21381
sad : 0.21614
Anger: 0.21226
Grief: 0.17889
Romantic: 0.17889

Sentiment : Negative
The Album's Tag is Sad

Fig 4: Results of the topic modelling

Implementation

Topic Modelling: Devised a statistical model using LDA which allows the set of observations to be explained by unobserved groups that explain why there is semantic connection between the topics. LDA is an uni-gram method[2], so as to tweak the accuracy of the system n-grams technique was also implemented. There was a slight increase in accuracy but due to the tradeoff between space, time and accuracy, LDA was chosen as the primary algorithm.

Sentiment Analysis: The lyrics of the song is used to find the polarity of the song[1]. The Naive Bayes classifier was used to train on a list of positive and negative words list.

Conclusion & Future Work

- When tested with the tags from another site, got an accuracy of 89.4%.
- Some albums had a contrasting Sentiment and Tag.
- Implement an application which when integrated automatically goes through the user playlists and gets similar tracks
- Also implement a sarcastic factor to check how sarcastic the song are.
- Integrate with audio component of songs for more extensive and accurate classification and check for contrast with audio mood.

References

- [1]Fell, Michael, and Caroline Sporleder. "Lyrics-based Analysis and Classification of Music." COLING. 2014.
- [2]Topic Extraction papers: LDA and N-gram Algorithms.

Contact : kxd8041@g.rit.edu