

SHOW ME THE MONEY - SENTIMENT ANALYSIS

Van Anh HOANG

Introduction to Digital Humanities: Final Report

hvananh@toki.waseda.jp

1. Introduction

As an active fan of the Korean rap show Show Me The Money, I wanted to analyze Reddit, the platform that I often use to discuss with the online fan community.

Often recognized on the thread as smtm, it appears under the subreddit r/khiphop, which gathers international fans who use English to communicate and share information on the latest music information of Korean hip-hop and R&B artists.

Show Me The Money came out with season 10 in late 2021. They release 1 episode every week, and the thread also contains shows with translated English subtitles done by fans and members of this community.

There is quite a few drama in the show, like this year, the winner of Season 4, Basick came back to compete and garner fame again after an unsuccessful career post winning. The runner-up of that season, Mino, is actually one of the producers (judge/coach) this year, so there's a lot of drama surrounding that. There are also controversial contestants, like a rookie rapper who's known for his super fast rapping. Some people find that incredible, but he also has quite a number of haters.

I set the date of data retrieval to be the day before the show starts and 2 weeks after it has ended, since it normally takes some time to translate and edit the subtitled videos.

Since there are good and bad seasons, I want to see how the audience finds this season based on the emotions shown in their comments. Since I have been watching the show, I found this year's season went quite well. I want to confirm whether other members of the community share the same feelings as me.

2. Background

I'm planning to use NLTK and lexicon-based NRC that we learned in week 12 to follow the fans' emotions and thoughts on this show.

To quote from the website, "the NRC Emotion Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive)". I will focus more on the sentiments.

3. Data and Method

I extracted the comments on reddit and tried to clean the text using nltk. Since there were many irrelevant columns on reddit, I only saved the comments, under the body column, and tokenized and lemmatized the data and changed it to a csv file.

Then, by using the NRC lexicon, I analyzed the emotions in the words used. I did not choose the other existing applications over NRC lexicon because they were less exact compared to NRC.

4. Results

Since I struggled with the pre-processing part, I don't have the right data to implement the data analysis part on. I will consequently write a hypothesis of the emotions.

I would guess there are mixed emotions, however, positive emotions prevail because after all, fans are following the show and found a community to share their thoughts.

5. Conclusion and discussion

These are also hypothetical. Since the keyword is quite broad because it's the shortened version of the name of the show, it can't include all the comments that belong to the discussion for season 10. Additionally, since I only used smtm as a keyword, there might be comments on older seasons where fans are watching, which affects the results of my findings on season 10. However, I think the mistake is minimal because the majority are watching the ongoing season.

Additionally, to be even more detailed, I thought I could have worked on sentiment analysis on the name of the two potential winners of the show and see if it could have guessed who the winner would be from the fans' comments.

I continuously struggled with the pre-processing part because most examples I found analyzed a short text that they wrote directly on the code. I couldn't figure out how to do it with my dataframe, so I struggled with getting an accurate result.

6. List of tools and methods used

- Tools
Google Collab

NLTK

NRC lexicon

- Code

```
import nltk
from nltk import word_tokenize
nltk.download('punkt')
nltk.download('wordnet')
```

References

Mohammad, S. (n.d.). *NRC Word-Emotion Association Lexicon (aka EmoLex)*. Saif Mohammad. Retrieved February 6, 2022, from <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>