CSOC20010

Introduction to Computational Social Science II

UCD School of Sociology

Spring, 2021-2022

Module Coordinator: Assoc Prof Taha Yasseri

Week 5 Assignment: Sentiment Analysis of Youtube comments

In this week's assignment, we are going to analyse the sentiment of the comments left on four YouTube videos of 4 K-pop bands. The four bands are "BLACKPINK", "EXO","BTS", and "TWICE". And the music videos are (in case you're interested!): BOOMBAYAH, CALL ME BABY, Blood Sweat & Tears, and SIGNAL.

1. Install sentistrenght using the following command in the terminal:

```
pip install sentistrength
```

2. Download W5.ipynb, SentiStrength.jar.zip, SentiStrenght_Data.zip, and the four .txt files. Each file contains the comments for one video.
3. Unzip SentiStrength.jar.zip and SentiStrenght_Data.zip in the same directory. It's important to keep all the files and the SentiStrenght_Data folder in the same directory.
4. Open W5.ipynb in Jupyter.
5. The ipynb file is self-explanatory. Run it once to make sure it runs without any issues and to understand its structure.
6. Change the file name in cell #4 to load different datasets for different music videos.
7. In cell #5 you can select a random sample of comments and calculate their sentiment.
8. For each band/music video, select a random sample of 200 comments and find one example of a comment with very positive/negative sentiment. Do you see comments whose sentiment id not accurately calculated based on your own judgment?
9. Now make a sample of 1000 comments per video and plot the histogram of the comments sentiments for each band separately.
10. Calculate the average sentiment of the four samples.
11. Write a report with the examples identified in 8, the histograms in 9, and discuss which video receives more positive/negative comments in general based on the results in 9 and 10.