Text Analysis and Visualization with Python

Arieda Muço[1]

Central European University

[1]Based on: https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f

Sentiment Analysis

If you want to understand people, especially your customers...then you have to be able to possess a strong capability to analyze text.

–Paul Hoffman, CTO Space-Time Insight

What is Sentiment Analysis?

- Sentiment Analysis, or Opinion Mining, is a sub-field of Natural Language Processing (NLP) that identifies and extracts opinions and sentiments from given text

Why is sentiment analysis so important?

- Majority of this data we have today is unstructured text
  - Sources like emails, chats, social media, surveys, images, videos, articles, and documents.
- It's challenging
  - Huge amount of data involved (difficult and time consuming)
  - The kind of language used in them to express sentiments, i.e., short forms, memes and emoticons


talkwalker

Why is sentiment analysis so important?

- Sentiment Analysis allows us to make sense out of text by being able to automate this entire process

Why is Sentiment Analysis a Hard to perform Task?

- Understanding emotions through text are not always easy. Sometimes even humans can get misled, so expecting a 100% accuracy from a computer (today) is like asking for the Moon!
  - A text may contain multiple sentiments all at once. For instance: *The intent behind the final season of Game of Thrones was great, but it could have been way better.*
  - Two polarities, i.e., Positive as well as Negative. How do we conclude whether the review was Positive or Negative?
- Computers can't comprehend Figurative Speech (yet). Can't understand use of similes, metaphors, hyperboles etc qualify for a figurative speech.
  - *The best I can say about the season finale is that it was interesting.*
  - The word 'interesting' does not necessarily convey positive sentiment and can be confusing for algorithms

Python Libraries

We are mainly going to use the following libraries

- pip install nltk or conda install nltk (if you haven't installed it yet)
- VADER (Sentiment Analysis)
  - Valence Aware Dictionary and sEntiment Reasoner (VADER) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media
  - VADER uses a combination of A sentiment lexicon is a list of lexical features (e.g., words) which are generally labelled according to their semantic orientation as either positive or negative
  - For more info: https://www.nltk.org/_modules/nltk/sentiment/vader.html

VADER

- VADER has been found to be quite successful when dealing with social media texts, NY Times editorials, movie reviews, and product reviews.
- VADER not only tells about the Positivity and Negativity score but also tells us about how positive or negative a sentiment is
- The developers of VADER have used Amazon's Mechanical Turk to get most of their ratings, You can find complete details on their GitHub
- Link to paper http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf

Advantages of using Vader

- It works exceedingly well on social media type text, yet readily generalizes to multiple domains
- It doesn't require any training data but is constructed from a generalizable, valence-based, human-curated gold standard sentiment lexicon
- It is fast enough to be used online with streaming data
- It does not severely suffer from a speed-performance tradeoff

Let's try it out!!!