VADER-Sentiment-Analysis Introduction

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is *specifically attuned to sentiments expressed in social media*. It is fully open-sourced under the [MIT License] (we sincerely appreciate all attributions and readily accept most contributions, but please don't hold us liable).

It describes the dataset of the paper:

> **VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text**
> (by C.J. Hutto and Eric Gilbert)
> Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

For questions, please contact:

C.J. Hutto
Georgia Institute of Technology, Atlanta, GA 30032
cjhutto [at] gatech [dot] edu

Python Code Example

For a **more complete demo**, point your terminal to vader's install directory (e.g., if you installed using pip, it might be `\Python3x\lib\site-packages\vaderSentiment`), and then run `python vaderSentiment.py`.

The demo has more examples of tricky sentences that confuse other sentiment analysis tools. It also demonstrates how VADER can work in conjunction with NLTK to do sentiment analysis on longer texts...i.e., decomposing paragraphs, articles/reports/publications, or novels into sentence-level analysis. It also demonstrates a concept for assessing the sentiment of images, video, or other tagged multimedia content.

If you have access to the Internet, the demo will also show how VADER can work with analyzing sentiment of non-English text sentences.

```
    from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
    #note: depending on how you installed (e.g., using source code download versus pip
install), you may need to import like this:
    #from vaderSentiment import SentimentIntensityAnalyzer

# --- examples -------
sentences = ["VADER is smart, handsome, and funny.",      # positive sentence example
             "VADER is not smart, handsome, nor funny.",   # negation sentence example
             "VADER is smart, handsome, and funny!",       # punctuation emphasis handled
correctly (sentiment intensity adjusted)
             "VADER is very smart, handsome, and funny.",  # booster words handled
correctly (sentiment intensity adjusted)
             "VADER is VERY SMART, handsome, and FUNNY.",  # emphasis for ALLCAPS handled
             "VADER is VERY SMART, handsome, and FUNNY!!!",# combination of signals - VADER
appropriately adjusts intensity
             "VADER is VERY SMART, uber handsome, and FRIGGIN FUNNY!!!",# booster words &
punctuation make this close to ceiling for score
             "The book was good.",                                      # positive sentence
             "The book was kind of good.",                 # qualified positive sentence is
handled correctly (intensity adjusted)
             "The plot was good, but the characters are uncompelling and the dialog is not
great.", # mixed negation sentence
             "At least it isn't a horrible book.",         # negated negative sentence with
contraction
             "Make sure you :) or :D today!",              # emoticons handled
             "Today SUX!",                                 # negative slang with
capitalization emphasis
             "Today only kinda sux! But I'll get by, lol"  # mixed sentiment example with
slang and constrastive conjunction "but"
              ]

analyzer = SentimentIntensityAnalyzer()
for sentence in sentences:
    vs = analyzer.polarity_scores(sentence)
    print("{:-<65} {}".format(sentence, str(vs)))
```

For a **more complete demo**, go to the install directory and run `python vaderSentiment.py`. (Be sure you are set to handle UTF-8 encoding in your terminal or IDE.)

Output for the above example code

```
VADER is smart, handsome, and funny.---------------------------- {'neg': 0.0, 'neu':
0.254, 'pos': 0.746, 'compound': 0.8316}
VADER is not smart, handsome, nor funny.----------------------- {'neg': 0.646, 'neu':
0.354, 'pos': 0.0, 'compound': -0.7424}
VADER is smart, handsome, and funny!---------------------------- {'neg': 0.0, 'neu':
0.248, 'pos': 0.752, 'compound': 0.8439}
VADER is very smart, handsome, and funny.----------------------- {'neg': 0.0, 'neu':
0.299, 'pos': 0.701, 'compound': 0.8545}
VADER is VERY SMART, handsome, and FUNNY.----------------------- {'neg': 0.0, 'neu':
0.246, 'pos': 0.754, 'compound': 0.9227}
VADER is VERY SMART, handsome, and FUNNY!!!---------------------- {'neg': 0.0, 'neu':
0.233, 'pos': 0.767, 'compound': 0.9342}
VADER is VERY SMART, uber handsome, and FRIGGIN FUNNY!!!--------- {'neg': 0.0, 'neu':
0.294, 'pos': 0.706, 'compound': 0.9469}
The book was good.---------------------------------------------- {'neg': 0.0, 'neu':
0.508, 'pos': 0.492, 'compound': 0.4404}
The book was kind of good.-------------------------------------- {'neg': 0.0, 'neu':
0.657, 'pos': 0.343, 'compound': 0.3832}
The plot was good, but the characters are uncompelling and the dialog is not great.
{'neg': 0.327, 'neu': 0.579, 'pos': 0.094, 'compound': -0.7042}
At least it isn't a horrible book.------------------------------- {'neg': 0.0, 'neu':
0.637, 'pos': 0.363, 'compound': 0.431}
Make sure you :) or :D today!----------------------------------- {'neg': 0.0, 'neu':
0.294, 'pos': 0.706, 'compound': 0.8633}
Today SUX!------------------------------------------------------ {'neg': 0.779, 'neu':
0.221, 'pos': 0.0, 'compound': -0.5461}
Today only kinda sux! But I'll get by, lol----------------------- {'neg': 0.179, 'neu':
0.569, 'pos': 0.251, 'compound': 0.2228}
```

About the Scoring

- The `compound` score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). This is the most useful metric if you want a single unidimensional measure of sentiment for a given sentence. Calling it a 'normalized, weighted composite score' is accurate.

  It is also useful for researchers who would like to set standardized thresholds for classifying sentences as either positive, neutral, or negative. Typical threshold values (used in the literature cited on this page) are:

  1. **positive sentiment**: `compound` score >= 0.5
  2. **neutral sentiment**: (`compound` score > -0.5) and (`compound` score < 0.5)
  3. **negative sentiment**: `compound` score <= -0.5

- The `pos`, `neu`, and `neg` scores are ratios for proportions of text that fall in each category (so these should all add up to be 1... or close to it with float operation). These are the most useful metrics if you want multidimensional measures of sentiment for a given sentence.