oxford internet institute university of oxford

ÅBO AKADEMI

Information Studies

Sentiment Analysis: The Emotionality of Discourse

Mike Thelwall
Statistical Cybermetrics Research Group
University of Wolverhampton, UK

Statistical Cybermetrics Research Group

CYBEREMOTIONS

UNIVERSITY OF WOLVERHAMPTON

Sentiment Strength Detection in the Social Web with *SentiStrength*

Detect positive and negative sentiment *strength* in short informal text

- Does not rely on standard grammar and spelling
- Uses nonstandard emotion expression forms from the social web (e.g., :-) or haaappppyyy!!!)
- Classifies positive 1 to 5 AND negative -1 to -5 sentiment

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social Web. *Journal of the American Society for Information Science and Technology* , 63(1), 163-173

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558.

SentiStrength Algorithm - Core

- List of 2,489 positive and negative sentiment terms and strengths (1 to 5), e.g.
  - ache = -2, dislike = -3, hate=-4, excruciating -5
  - encourage = 2, coolest = 3, lover = 4
- Sentiment strength is highest in sentence; or highest sentence if multiple sentences

	positive, negative
My legs $\overset{-2}{\text{ache}}$.	1, -2
You are the $\overset{3}{\text{coolest}}$.	3, -1
I $\underset{-4}{\text{hate}}$ Paul but $\underset{2}{\text{encourage}}$ him.	2, -4

Extra rules (total: about 20)

spelling correction	nicce -> nice
booster words alter strength	**very** happy
negating words flip sentiments	**not** nice
repeated letters boost sentiment/+ve	niiiice
emoticon list	:) =+2
exclamation marks count as +2 unless –ve	hi!
repeated punctuation boosts sentiment	good!!!
Negativity ignored in questions	u h8 me?
sentiment idiom list	shock horror = -2

Online as http://sentistrength.wlv.ac.uk/

Tests against human coders on data sets of >1000 texts

Data set	Positive scores - correlation with humans	Negative scores - correlation with humans
YouTube	0.589	0.521
MySpace	0.647	0.599
Twitter	0.541	0.499
Sports forum	0.567	0.541
Digg.com news	0.352	0.552
BBC forums	0.296	0.591
All 6 data sets	0.556	0.565

SentiStrength agrees with typical humans as much as they agree with each other

1 is perfect agreement, 0 is random agreement

Why the bad results for BBC? (and Digg)

Irony, sarcasm and expressive language e.g.,

- David Cameron must be very happy that I have lost my job.
- It is really interesting that David Cameron and most of his ministers are millionaires.
- Your argument is a joke.

Other SentiStrength Languages

- **OK**: Spanish, Finnish, German, Dutch, Russian, Turkish, Italian
- **Untested**: French, Polish, Greek, Swedish, Portuguese, Persian, Arabic, Welsh, Irish
- **Basic**: Chinese, Filipino, Hausa, Indonesian, Japanese, Korean, Shona, Swahili

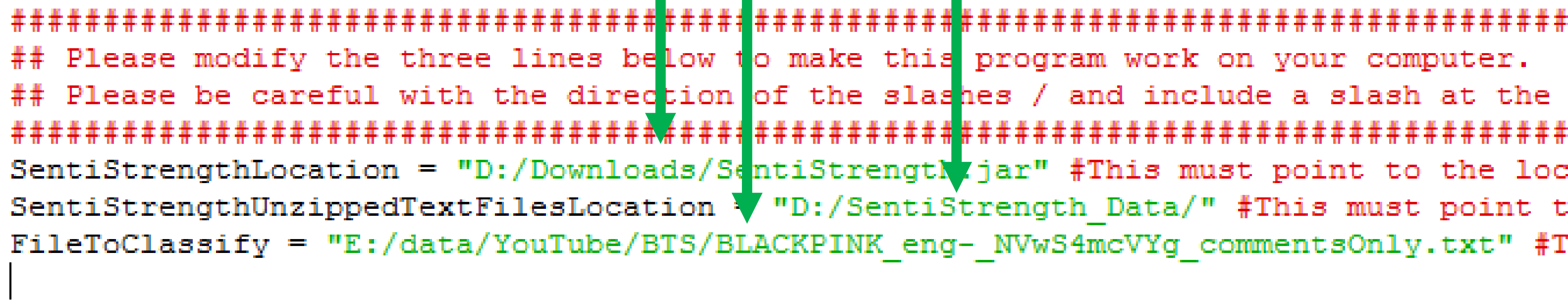
Workshop task

- Classify YouTube comment sentiment with SentiStrength
- Read the very strongly positive or negative comments to discover
  - What strong sentiment is expressed about and
  - How the magnitude and topic of sentiment differs between groups or videos.

Python code

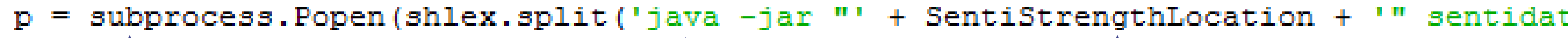
Download Python and SentiStrength programs and other files to your computer

Enter file locations on your computer here



```
######################################################################################
## Please modify the three lines below to make this program work on your computer.
## Please be careful with the direction of the slashes / and include a slash at the
######################################################################################
SentiStrengthLocation = "D:/Downloads/SentiStrength.jar" #This must point to the loc
SentiStrengthUnzippedTextFilesLocation = "D:/SentiStrength_Data/" #This must point t
FileToClassify = "E:/data/YouTube/BTS/BLACKPINK_eng-_NVwS4mcVYg_commentsOnly.txt" #T
```

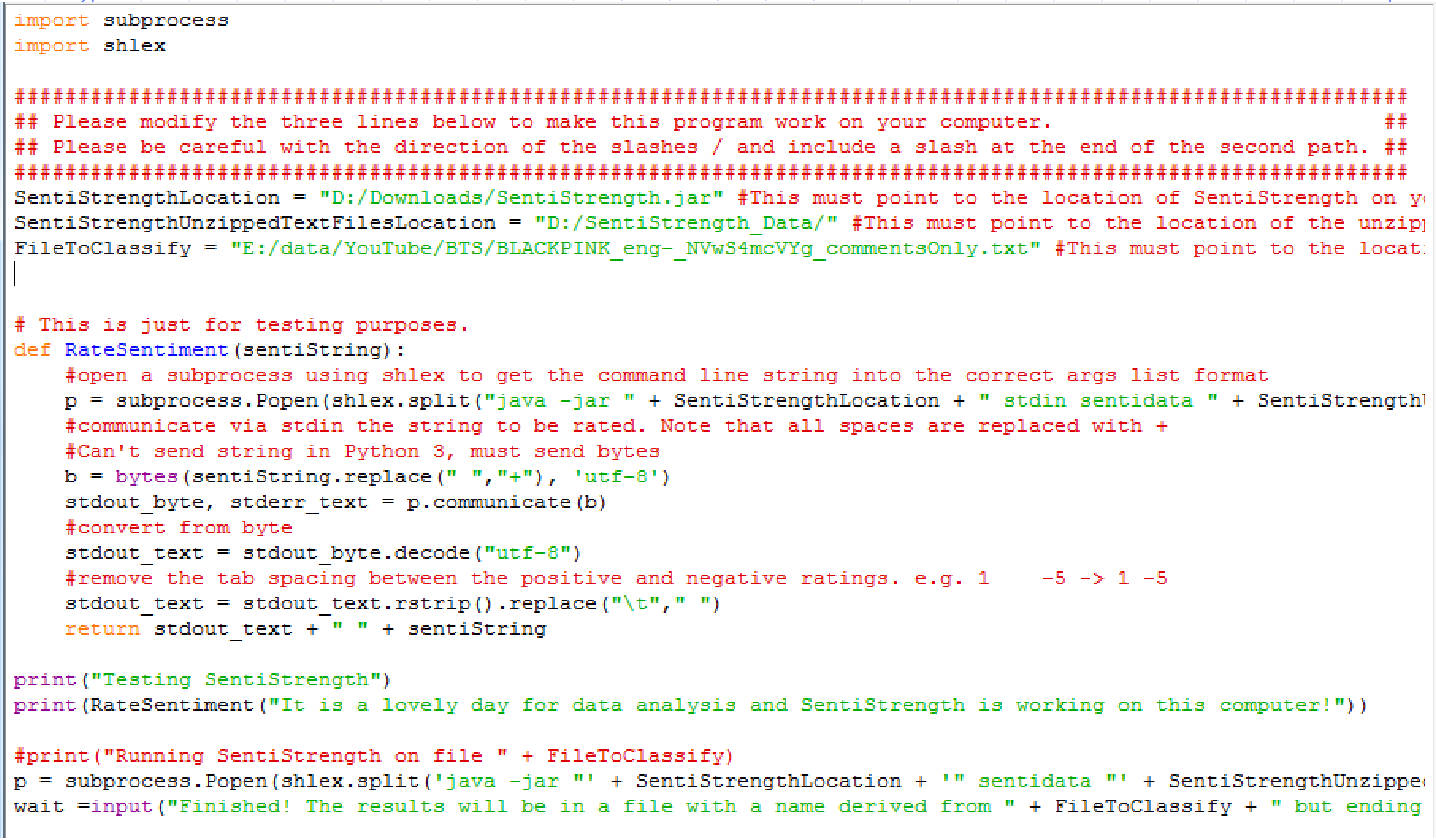
Other code….



```
p = subprocess.Popen(shlex.split('java -jar "' + SentiStrengthLocation + '" sentidat
```

Calls the SentiStrength java program & sends the file to process

All code



```
import subprocess
import shlex

#####################################################################################
## Please modify the three lines below to make this program work on your computer.     ##
## Please be careful with the direction of the slashes / and include a slash at the end of the second path. ##
#####################################################################################
SentiStrengthLocation = "D:/Downloads/SentiStrength.jar" #This must point to the location of SentiStrength on y
SentiStrengthUnzippedTextFilesLocation = "D:/SentiStrength_Data/" #This must point to the location of the unzip
FileToClassify = "E:/data/YouTube/BTS/BLACKPINK_eng-_NVwS4mcVYg_commentsOnly.txt" #This must point to the locat


# This is just for testing purposes.
def RateSentiment(sentiString):
    #open a subprocess using shlex to get the command line string into the correct args list format
    p = subprocess.Popen(shlex.split("java -jar " + SentiStrengthLocation + " stdin sentidata " + SentiStrength
    #communicate via stdin the string to be rated. Note that all spaces are replaced with +
    #Can't send string in Python 3, must send bytes
    b = bytes(sentiString.replace(" ","+"), 'utf-8')
    stdout_byte, stderr_text = p.communicate(b)
    #convert from byte
    stdout_text = stdout_byte.decode("utf-8")
    #remove the tab spacing between the positive and negative ratings. e.g. 1    -5 -> 1 -5
    stdout_text = stdout_text.rstrip().replace("\t"," ")
    return stdout_text + " " + sentiString

print("Testing SentiStrength")
print(RateSentiment("It is a lovely day for data analysis and SentiStrength is working on this computer!"))

#print("Running SentiStrength on file " + FileToClassify)
p = subprocess.Popen(shlex.split('java -jar "' + SentiStrengthLocation + '" sentidata "' + SentiStrengthUnzippe
wait =input("Finished! The results will be in a file with a name derived from " + FileToClassify + " but ending
```