oii oii oii oxford internet institute university of oxford

ÅBO AKADEMI

Information Studies

Using and modifying SentiStrength

Mike Thelwall
University of Wolverhampton, UK

Statistical Cybermetrics Research Group

CYBEREMOTIONS

UNIVERSITY OF WOLVERHAMPTON

Contents

- Using SentiStrength in English
- Adapting SentiStrength to Russian
- Evaluating the results

Using SentiStrength in English

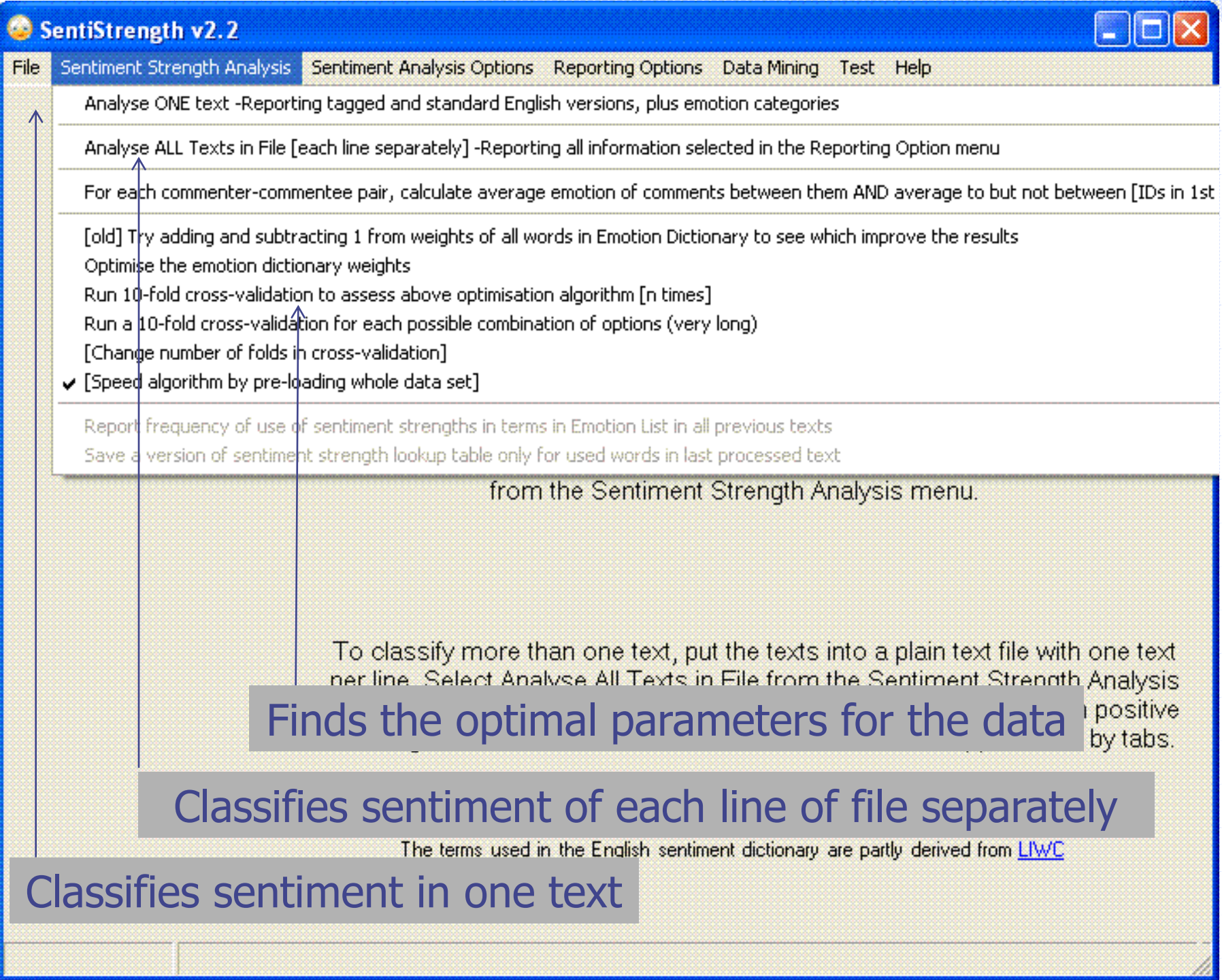
- Windows version:
  - Download program and zipfile SentiStrength_Data.zip from http://sentistrength.wlv.ac.uk/
- Unzip SentiStrength_Data.zip, then start SentiStrength.exe and point to the unzipped SentiStrength_Data folder
- Ready to go!

SentiStrength Input files

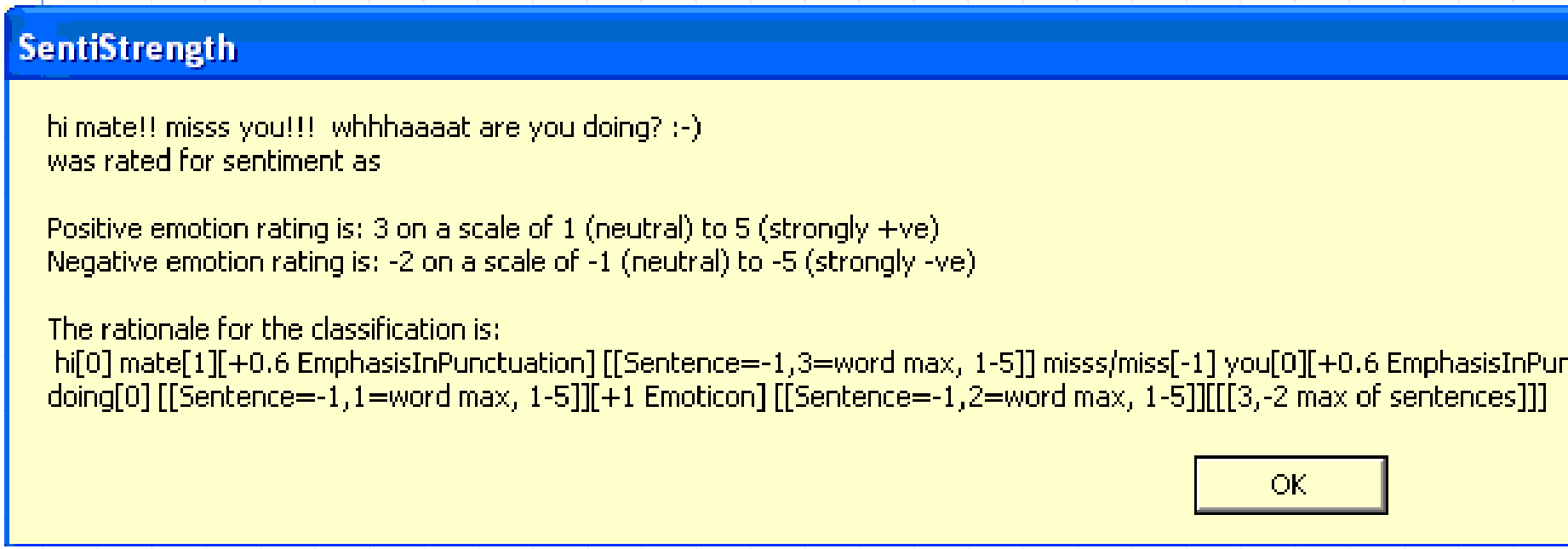
- EmotionLookUpTable.txt - a list of emotion-bearing words with a strength 1 to 5 or -1 to -5.
- Emo**ticon**LookUpTable.txt - as above but for a list of emoticons. :)
- EnglishWordList.txt - a list of English words – used for spelling corrections.
- IdiomLookupTable.txt – idiomatic phrases and sentiment strengths

SentiStrength Input files

- NegatingWordList.txt – negating words –e.g., not, don't.
- BoosterWordList.txt - sentiment intensity modifiers -e.g., very, extremely, quite, some.
- SlangLookupTable.txt – slang translations


SentiStrength v2.2
File
Sentiment Strength Analysis
Sentiment Analysis Options
Reporting Options
Data Mining
Test
Help
Analyse ONE text -Reporting tagged and standard English versions, plus emotion categories
Analyse ALL Texts in File [each line separately] -Reporting all information selected in the Reporting Option menu
For each commenter-commentee pair, calculate average emotion of comments between them AND average to but not between [IDs in 1st
[old] Try adding and subtracting 1 from weights of all words in Emotion Dictionary to see which improve the results
Optimise the emotion dictionary weights
Run 10-fold cross-validation to assess above optimisation algorithm [n times]
Run a 10-fold cross-validation for each possible combination of options (very long)
[Change number of folds in cross-validation]
[Speed algorithm by pre-loading whole data set]
Report frequency of use of sentiment strengths in terms in Emotion List in all previous texts
Save a version of sentiment strength lookup table only for used words in last processed text
from the Sentiment Strength Analysis menu.
To classify more than one text, put the texts into a plain text file with one text
Finds the optimal parameters for the data
positive
by tabs.
Classifies sentiment of each line of file separately
The terms used in the English sentiment dictionary are partly derived from LIWC
Classifies sentiment in one text

One text


SentiStrength
hi mate!! misss you!!! whhhaaaat are you doing? :-)
was rated for sentiment as
Positive emotion rating is: 3 on a scale of 1 (neutral) to 5 (strongly +ve)
Negative emotion rating is: -2 on a scale of -1 (neutral) to -5 (strongly -ve)
The rationale for the classification is:
hi[0] mate[1][+0.6 EmphasisInPunctuation] [[Sentence=-1,3=word max, 1-5]] misss/miss[-1] you[0][+0.6 EmphasisInPun
doing[0] [[Sentence=-1,1=word max, 1-5]][+1 Emoticon] [[Sentence=-1,2=word max, 1-5]][[[3,-2 max of sentences]]]
OK

Multiple texts

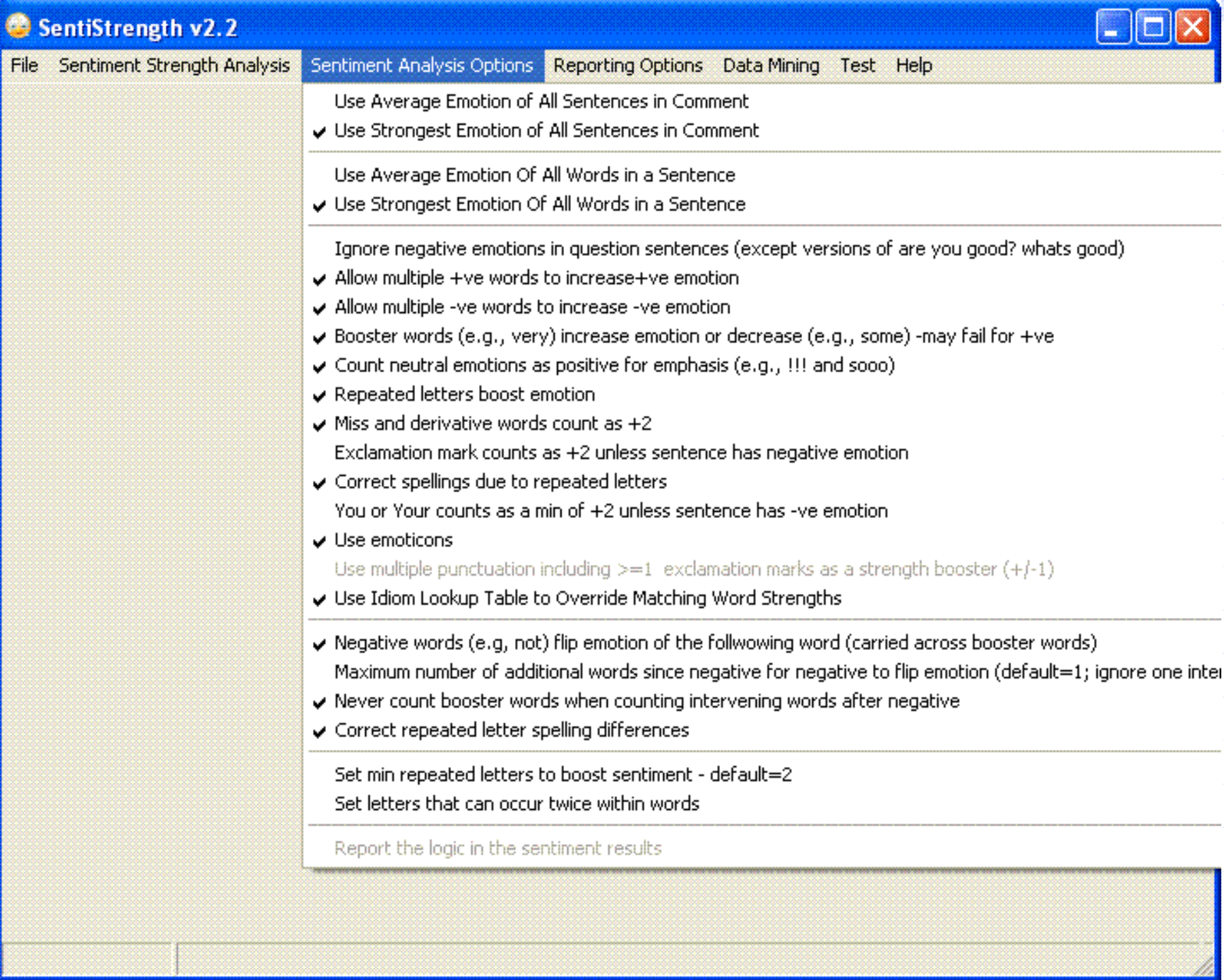
- Input file is list of texts, one per line
- Output file is a copy of the texts, plus the classifications

I just thought that I would say HI... ----- Love you
After the series it looked like shit!!
Damn its been a good while that i don’t see u

4 1 I just thought that I would say HI... ----- Love you
1 4 After the series it looked like shit!!
3 2 Damn its been a good while that i don’t see u

Optimisation and validation

- For the optimisation and cross-validation options the input must be a Gold Standard.
- Positive – tab – Negative – tab – text
- Accuracy statistics can be calculated
- The optimisation step alters the sentiment dictionary term weights to fit the data better
  - E.g.., love (+4) -> love (+3)


SentiStrength v2.2
File
Sentiment Strength Analysis
Sentiment Analysis Options
Reporting Options
Data Mining
Test
Help
Use Average Emotion of All Sentences in Comment
Use Strongest Emotion of All Sentences in Comment
Use Average Emotion Of All Words in a Sentence
Use Strongest Emotion Of All Words in a Sentence
Ignore negative emotions in question sentences (except versions of are you good? whats good)
Allow multiple +ve words to increase+ve emotion
Allow multiple -ve words to increase -ve emotion
Booster words (e.g., very) increase emotion or decrease (e.g., some) -may fail for +ve
Count neutral emotions as positive for emphasis (e.g., !!! and sooo)
Repeated letters boost emotion
Miss and derivative words count as +2
Exclamation mark counts as +2 unless sentence has negative emotion
Correct spellings due to repeated letters
You or Your counts as a min of +2 unless sentence has -ve emotion
Use emoticons
Use multiple punctuation including >=1 exclamation marks as a strength booster (+/-1)
Use Idiom Lookup Table to Override Matching Word Strengths
Negative words (e.g, not) flip emotion of the follwowing word (carried across booster words)
Maximum number of additional words since negative for negative to flip emotion (default=1; ignore one inte
Never count booster words when counting intervening words after negative
Correct repeated letter spelling differences
Set min repeated letters to boost sentiment - default=2
Set letters that can occur twice within words
Report the logic in the sentiment results

Java version

- Ask Mike for location
- Commercial version
- Quicker and more options than the Windows version
- Need to also download and unzip the Windows version SentiStrength_Data folder
- Runs on any computer with Java runtime installed

Using the Java version

- Process one text (must be escaped text):
  - java -jar SentiStrength.jar sentidata C:/SentStrength_Data/ text i+don't+hate+you.
- Process all texts in file
- java -jar SentiStrength.jar sentidata C:/SentStrength_Data/ input C:/test.txt

Java version options

- As for Windows version but can also:
  - Listen at IP number
  - Process stdin -> stdout
  - Run interactively from command line
- Has some linguistic options
  - E.g., can allow negation after sentiment terms (happy not)
- Can do binary/trinary/scale classifications instead of default

Modifying SentiStrength for a different domain

- Create a gold standard for that domain
- Use the optimise option to optimise the sentiment word strengths in EmotionLookUpTable.txt.
- Use SentiStrength with the new EmotionLookUpTable.txt.

Modifying SentiStrength for a different language

- Translate all the input files in SentiStrength_Data
- Pay particular attention to making the list of terms in EmotionLookUpTable.txt as complete as possible.
- Create a gold standard for appropriate text in that language
- Use the optimise option to optimise the sentiment word strengths in EmotionLookUpTable.txt & to evaluate the result
- Use SentiStrength with the new EmotionLookUpTable.txt.

Example – Russian/ French

- амортизация ?
- ампутировать ?
- анархия ?
- аннулирование ?
- банальный ?
- бандит ?
- банкрот ?

- atroce ?
- atrophie?
- attaque ?
- attenter ?
- atterré ?
- audacieux ?
- austère ?

What sentiment score should each word have? (1-5 or -1 - -5)

Wildcard/Kleene star

- absence -2
- absent* -2
- absurd* -2
- abuse* -4
- abusi* -4
- accepta* 2
- abyss -2

Allows groups of words to match
In SentiStrength's sentiment dictionary

Summary

- SentiStrength has Windows and Java versions
- Can be modified for new languages or domains
- Needs linguistic work, not programming work, to modify

Bibliography

- 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.
- http://sentistrength.wlv.ac.uk – see user documentation on this site, including Java documentation