Perceiving Personality Through Sentiment Analysis

by Sidharth Dhawan
Class of 2016
with Dr. Christiane Fellbaum

Motivation and Goal

- Goal of the project:
    - Create algorithms and statistics to characterize a person’s outlook / disposition in a useful way.
        - Starting from their written documents.
    - Useful Metrics:
        - positive/ negative mood

- Motivation:
    - Could be a useful tool for psychiatrists
    - A useful tool in the workplace to understand how to work with different people
    - A useful tool for personalized social media applications

Problem Background & Related Work

- Understanding personality through social media (recent)
  - “Predicting Personality from Twitter”
    - J. Golbeck, C. Robles, M. Edmondson, K. Turner
    - uses profile information to assess personality (namely the “big five” scale)
- Understanding personality purely from language
  - “Linguistic Styles: Language Use as an Individual Difference”
    - J. Pennebaker, L. King
    - Investigates several linguistic features of writing and how they are related to personality.
  - “From Ace to Zombie: Some Explorations in the Language of Personality”
    - L. Goldberg

Approach

- Uses “longitudinal” approach
  - 5,000 documents of 100 or more words.
  - These documents span several years in the lifespan
  - Can measure the way different events of a person’s life affect their writing.

- Uses the latest techniques from Sentiment Analysis
  - may be able to augment early personality detection techniques with most recent investigations in detecting sentiment from text.
  - an area that investigates the way sentiment is embedded in text

- Data is that of private correspondences
  - makes it more honest, unfiltered, and thorough than many other sources.

Implementation & Progress to Date

Data Collection

- Crawling Presidential Letters:
    - the University of Virginia Press contains written documents of several presidents
        - http://rotunda.upress.virginia.edu/founders/
    - Built a web crawler in python.

Data Collection

- The data:
  - Adams, Jefferson and Washington
  - 5,000 letters per president
  - On the website, they were stored as a large set of links
  - I have stored each letter as a separate .txt file in my file system
    - chronologically ordered

Most Sentiment-Indicative Words, cont’d

- score(word) = senti(word) * f(word)

- senti(word): typical level of Sentiment
  - Taken from Sentiwordnet
- f(word): Frequency score
  - either the total occurrences of each word
  - or the log of occurrences of each word

- Results for different metrics
  - log-occurrences produces very similar lists for both presidents
    - want to know idiosyncrasies

Sorted Lists of (Positive) Sentimental Words:

Recall, wordscore = sentiwordnet(w) * freq(w)

Adams:

1. happy
2. proper
3. honourable
4. worthy
5. sincere
6. virtue
7. affection

Washington:

1. good
2. proper
3. best
4. well
5. new
6. hope
7. better

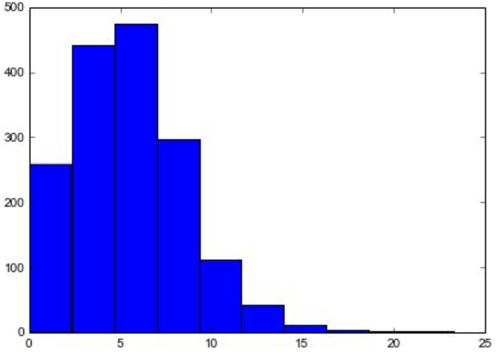
Present and Future Uses

- Present
  - Evaluate the overall sentiment in a letter by counting occurrences of these words.
  - See next slides

- Future
  - Can mine this list to understand preferences
    - Ex. Does Washington often express a desire for privacy?

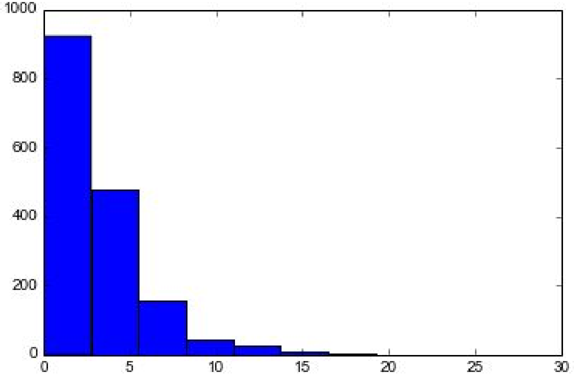
Measuring Aggregate Sentiment Data (Adams)

negativity of letter



< Adams: negative data

(note the high mode)



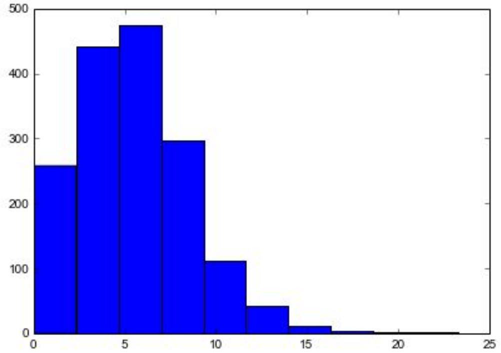
< Adams: positive data

Above are histograms of letter positivity and negativity

Positivity = positive words / wordcount.

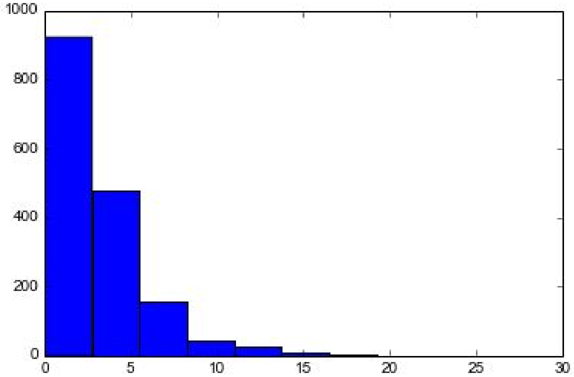
Positive and negative affect (mood).

Histograms of Positivity / Negativity levels

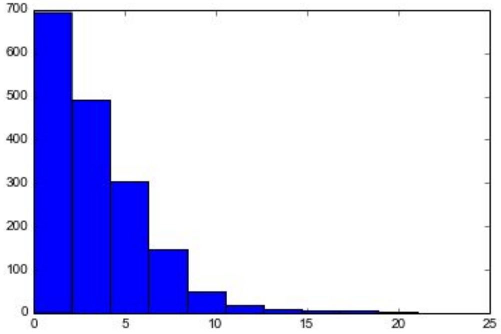


< Adams:
neg data

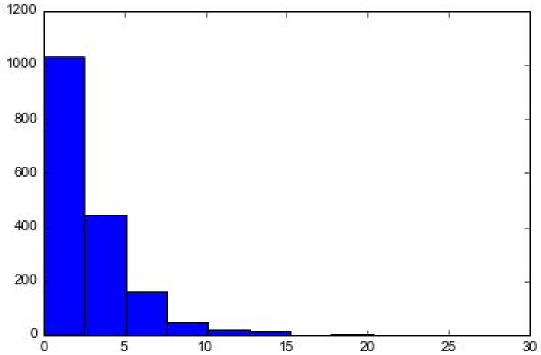
(note the
high
mode)



< Adams:
pos data



< Wash:
neg data



< Wash:
pos data

Descriptive Stats from this Data

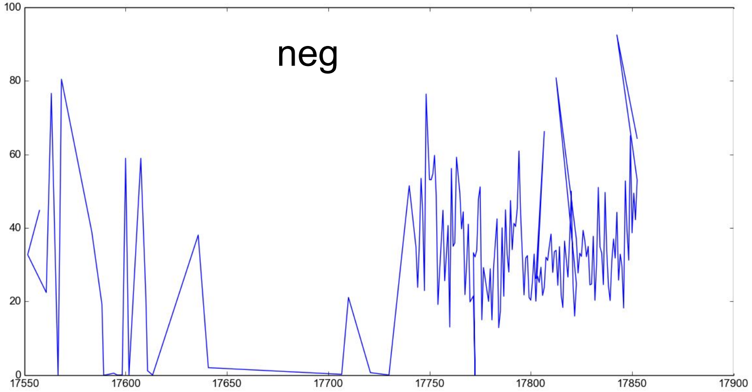
- Terseness:
    - % of letters with a combined positive and negative score of less than 5:
    - Adams: 19.5%
    - Washington: 42.7%

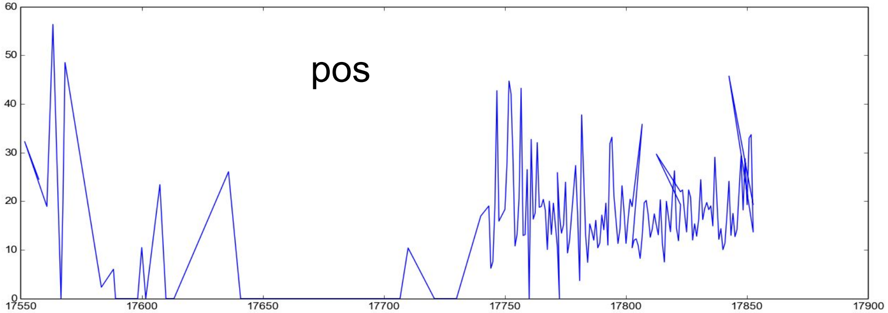
- Positive Affect:
    - median positivity score:
    - Adams: 2.4
    - Washington: 2

- Negative Affect:
    - median negativity score:
    - Adams: 5.8
    - Washington: 3

- In the future, it will be useful to collect this data for more people and compute percentile scores

Measuring Changes in Sentiment Over time





Graphs show Adams’ positivity and negativity plotted over time. Each point = positivity / negativity for one month.

Future: Map this against significant events or changes in Adams’ life

Results

- Evaluation Tool: Amazon Mechanical Turk
  - Several workers will be allowed to access and complete the task; each of whom is paid at an hourly rate.
  - Gives objective measure of sentiment of letters, so I can assess accuracy of metrics

- Evaluation Method: Evaluate 100 word- samples of letters
  - Ex. To evaluate positive or negative affect, score the sample from 1-5 based on its level of optimism or pessimism, where:
    - 1: strongly / overtly negative
    - 2: guardedly negative
    - 3: no emotion whatsoever
    - 4: guardedly positive
    - 5: strongly / overtly positive

Results

A guardedly positive sentence (sentiment is subtle / implied):

- “Now that a new king has been named, the construction of the railroad track will be accomplished more quickly.”

An overtly positive sentence:

- “The ascension of the new king is the best news to reach us in months, as it will expedite the railroad construction process.”

Results

The average positive and negative affect level for each president will then be computed, and compared with the results of my statistics to check accuracy.

Questions?

Extra Slides

Most Sentiment-Laden Words [SKIP]

Goal: find the words that each president uses most often to express emotions.
(First step: remove stopwords: like “the”, “and”, etc.)

- **PMI (word w)**: how often is word “w” used *near* positive words, like “good”, “excellent”, etc. ?
  - PMI = Pointwise Mutual Information
  - unique for each dataset

- **Sentiwordnet (word w)**: In the english language, what is the average positive sentiment of “w”, over all the different ways it is used?
  - sentiment connotations predetermined