Semantic Drift Velocity: Change in meaning over time

Isaac Barrett, Wesley Dahar

Intro

Research Questions

1. Have the meanings of words changed over time?
2. Is language less information-dense now, or in the past?
3. Has the average vocabulary grown or decreased?

Many words shift meaning over time, which leads to a decrease in information density and average vocabulary.

Scope

- American English
- 1600 to present
- 1 unit of time is 1 decade

Methodology

1. Train with each decade to generate word vector data
    a.   Continuous model, with cached vector-field on each epoch
    b.   Discrete models, trained from-empty each epoch
2. Compare vector fields across decades
    a.   Note the words that experience change
        i.   Relative to closed-class words (discrete)
        ii.   Relative to observed consistent words (discrete)
        iii. Relative to former positions (continuous)
    b.   Observe trends in word drift
3. Word count from each decade to measure vocabulary

Data and Tools

- Gutenberg Archives by century
- gensim
- numpy

Why?

1. Predict velocity and acceleration of words to determine words that are changing
2. Prepare ML algorithms to deal with changing words over time
3. Observe the shift in human use of language

Questions?