Recommending Books Based On Emotional Arcs

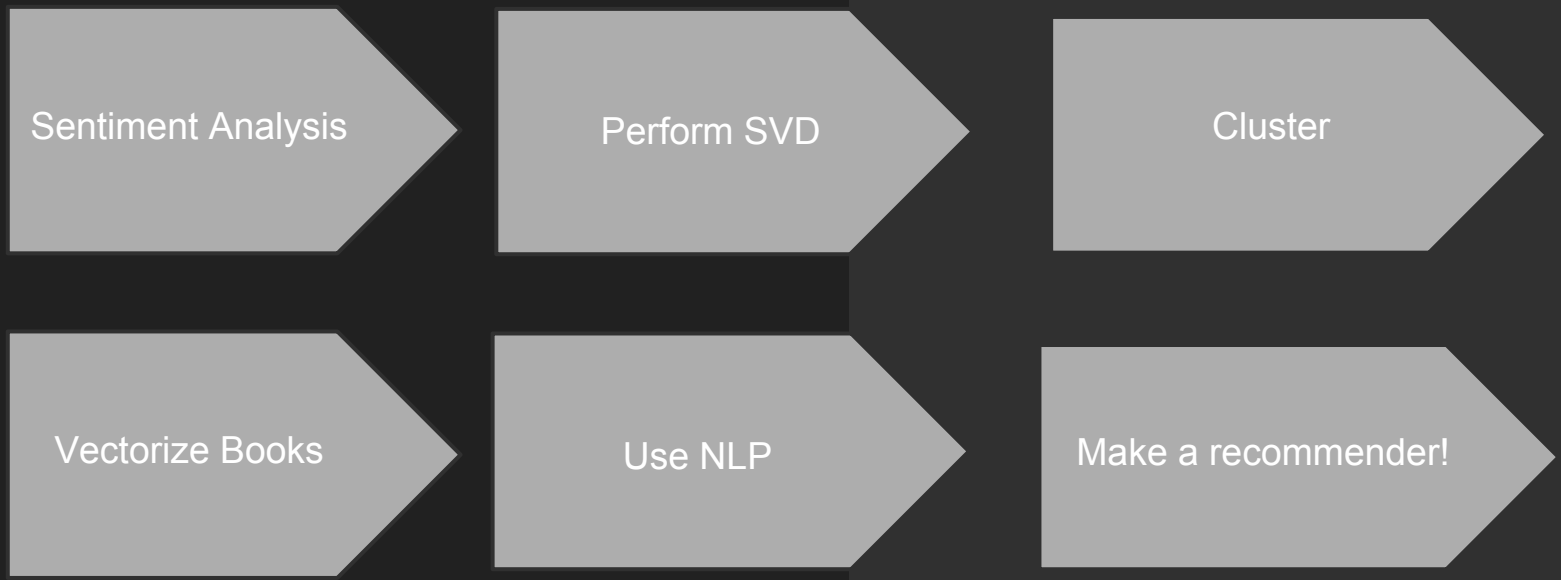
Christian Hansen

March 10, 2017

Goal:

Use sentiment analysis and nlp features to recommend books.

Pipeline:



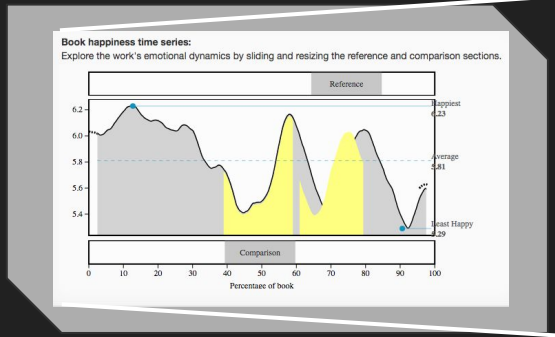
Emotional Acs != Plot Arcs

Emotional Arc

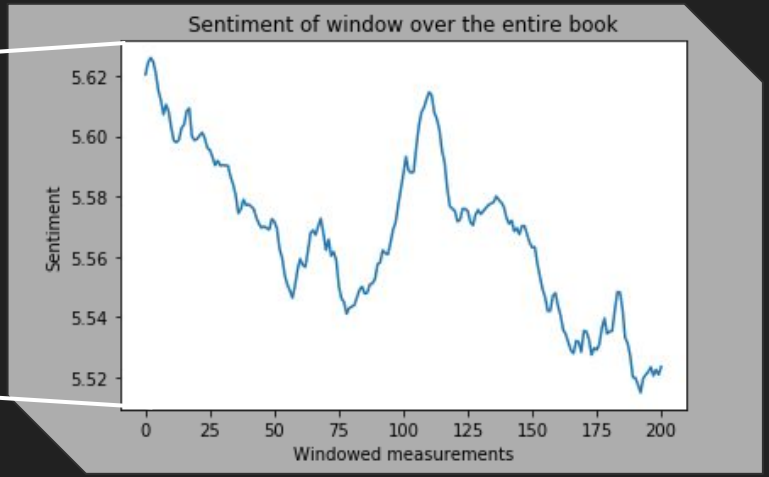
- Sentiment of words over length of book.
- Can be categorized, or clustered.

Plot Arc

- Highly Conceptual.
- Classically hard to define.



Hedonometer.org



A Journey To The Centre Of The Earth, by Jules Verne

Tools

Python
MongoDB
labMT Sentiment Library

Data

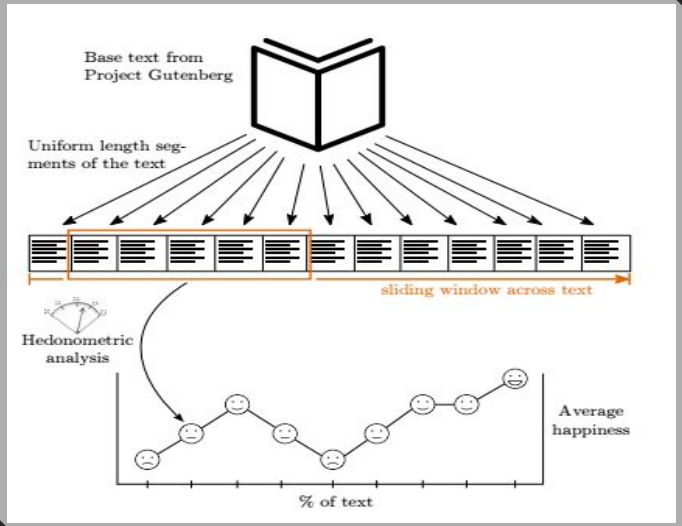
Project Gutenberg
~1300 Full text books

Emotional Arc

- Using sentiment analysis to measure the “happiness” score (0-9) of a word over an entire book.

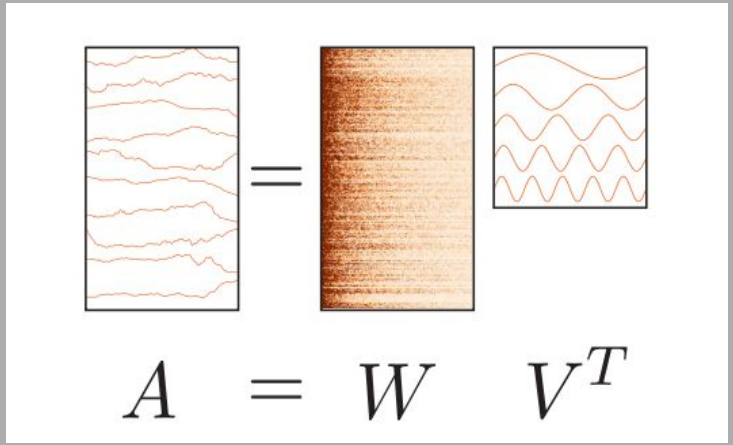
Window length = (length of book - window size + 1) / Resolution

- Each book will have the same number of windows.

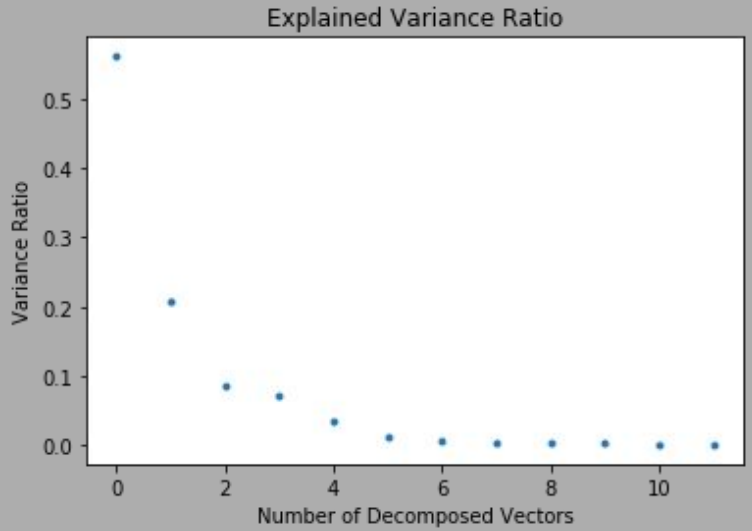


https://arxiv.org/pdf/1606.07772.pdf

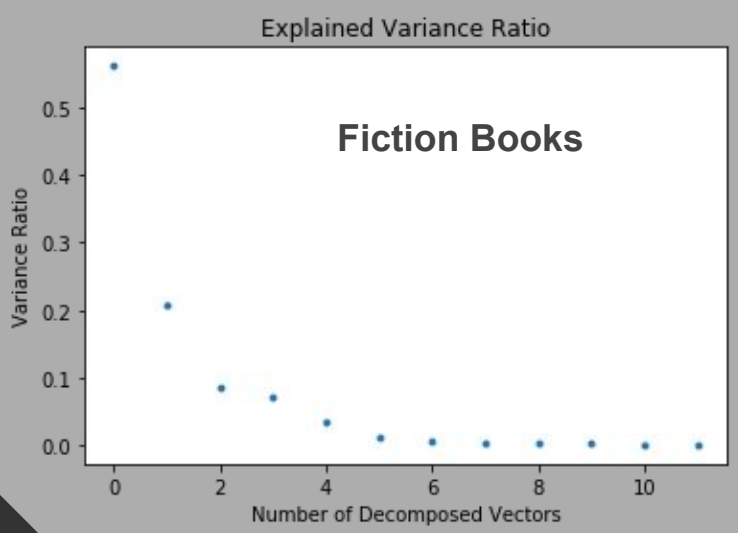
- Use SVD to decompose the sentiment arrays.
- Examine top components based on explained variance.

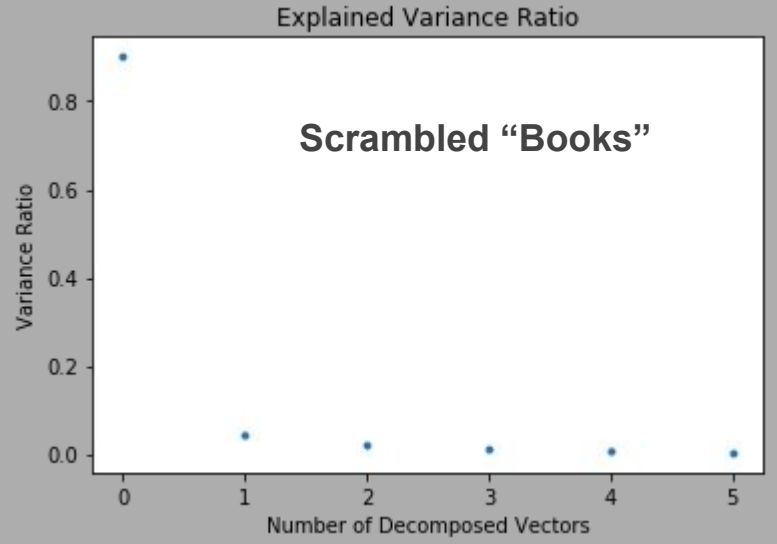


https://arxiv.org/pdf/1606.07772.pdf



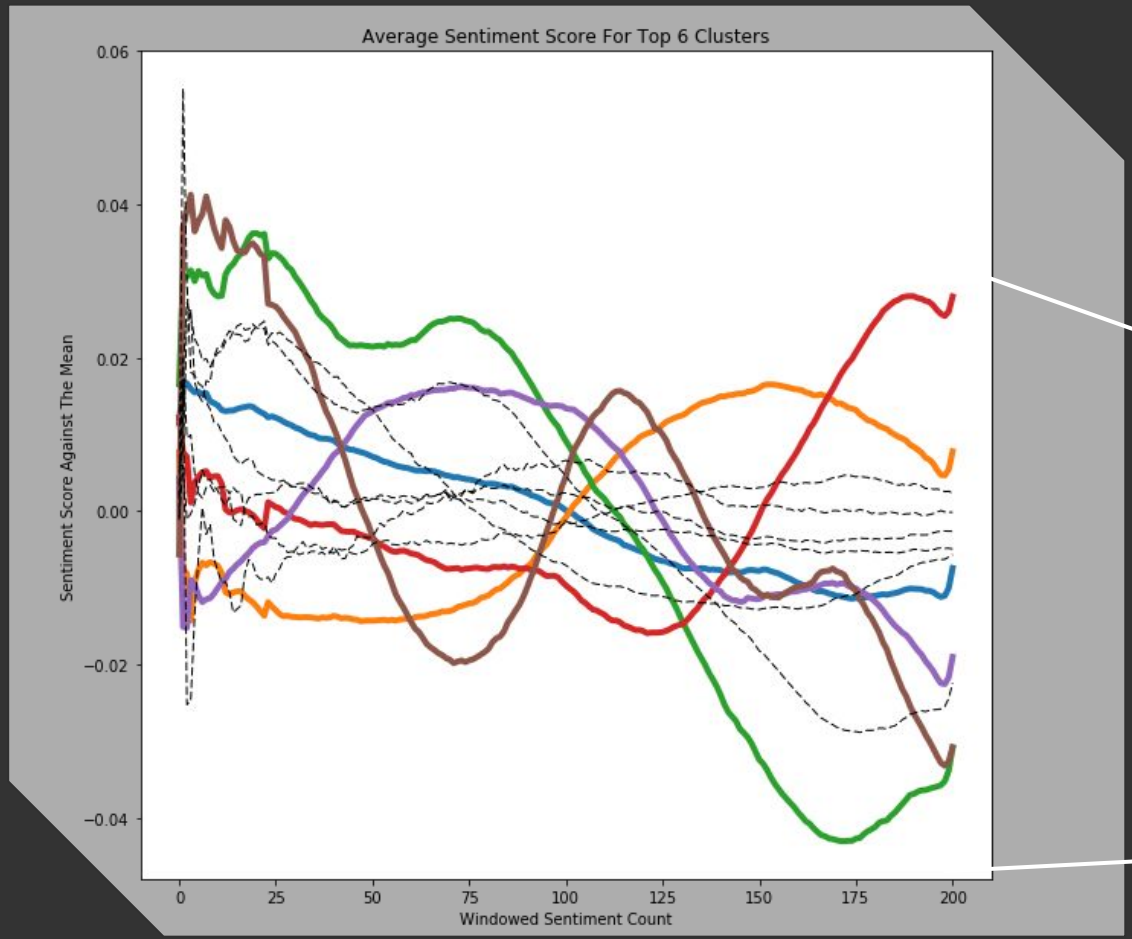
Explained Variance Ratio Comparison



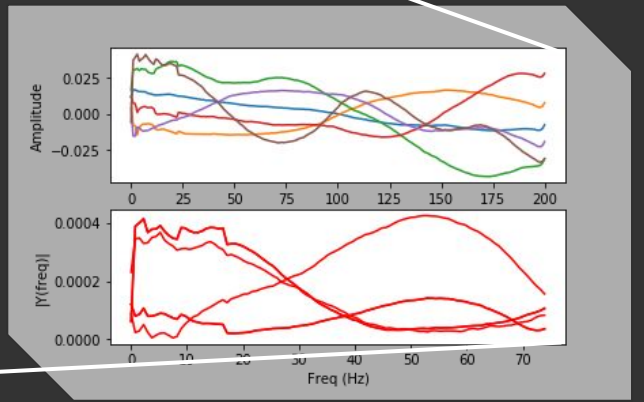


The percentage of variance explained by each of our chosen dimensions.

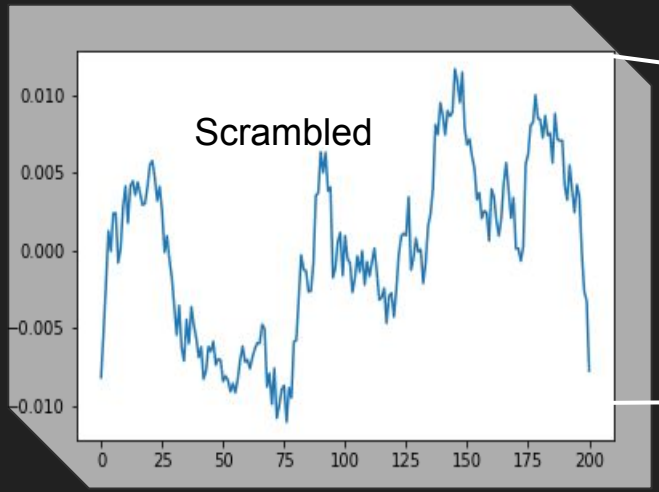
Emotional Arcs

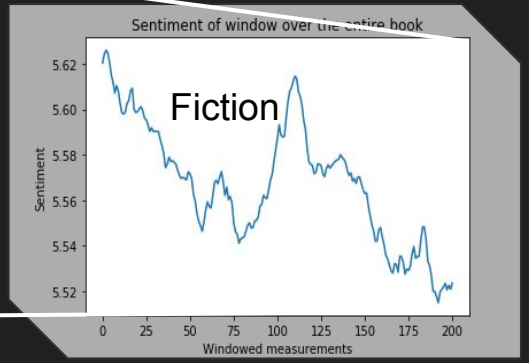


Fast Fourier Transform Of Emotional Arcs

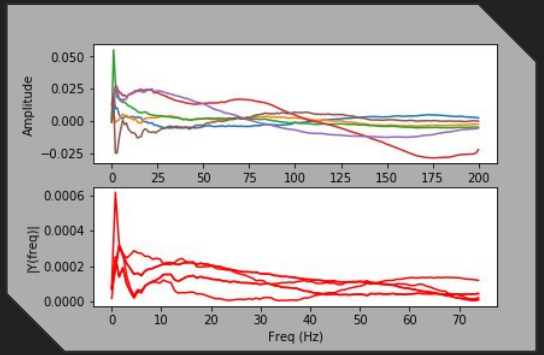


Example of scrambled sentiment for a “book”!





FFT of scrambled books



Recommender

Combine vectorized and SVD'd text from all books in collection and combine with SVD'd emotional arcs and NLP derived features.

Results

- Recommender system favors certain books more than others from selection.
- Not a strong distinction between the null books and fiction.

INPUT:
Flatland: A Romance of Many Dimensions

OUTPUT:

1.) 'Notes from the Underground'
2.) "Abraham Lincoln's First Inaugural Address"
3.) 'Steep Trails: California, Utah, Nevada, Washington, Oregon, the Grand Canyon'
4.) 'The Life and Adventures of Robinson Crusoe'
5.) 'Notes on Life & Letters'
6.) 'The History and Practice of the Art of Photography'
7.) 'The Autobiography of Benjamin Franklin'
8.) 'Collected Articles of Frederick Douglass'
9.) 'Spoon River Anthology'
10.) 'McTeague: A Story of San Francisco'

^ Approximately Discovery Topics