Visualizing Storytelling with Script Sentiment Analysis

JohnHenry Ward

Motivation

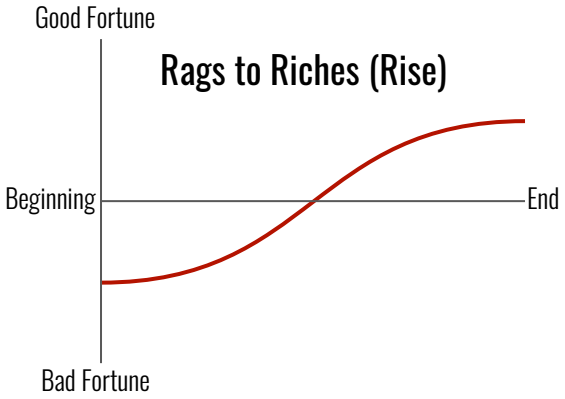
American author Kurt Vonnegut came up with the idea that graphs can show the plot lines of stories (seen on next slide) in his rejected master's thesis

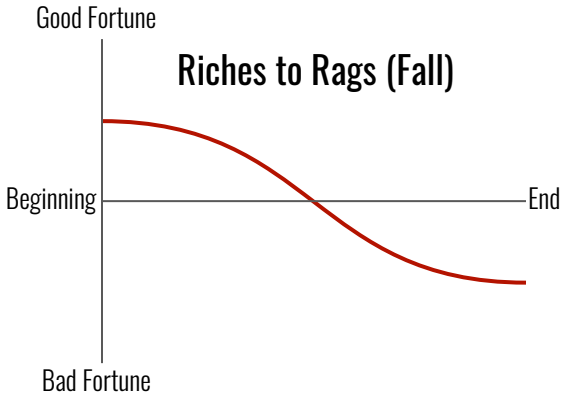
"There is no reason why the simple shapes of stories can't be fed into computers, they are beautiful shapes." -Kurt Vonnegut

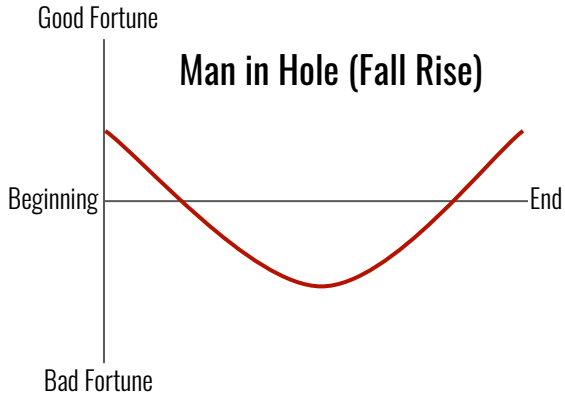
Kurt Vonnegut on the Shapes of Stories (Link)

Is it possible to visualize a movie as one of the six main arcs of storytelling using sentiment analysis?

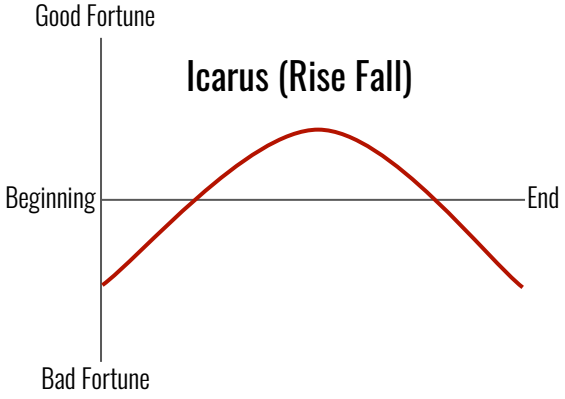
I chose movies because I love movies, but this could work with any piece of text

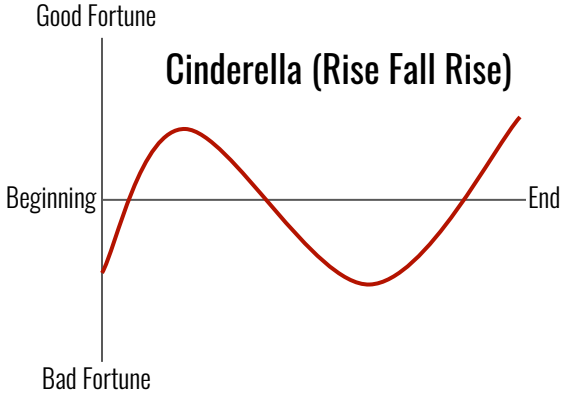


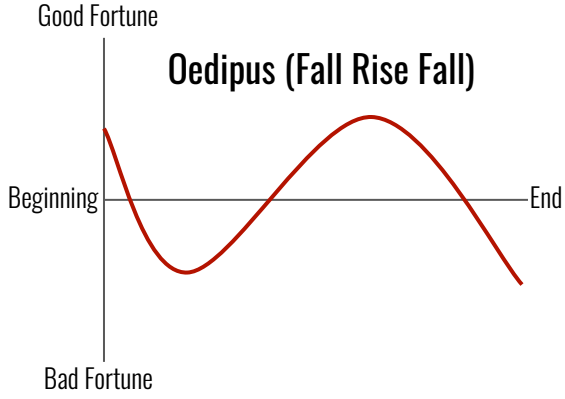




The Six Main Arcs of Storytelling







Related Work

Visualizing the Emotional Arcs of Movie Scripts Using Rule-Based Sentiment Analysis (Link)

-Project that compares and displays similar movies in term of the emotional trajectory

-Uses NRC Valence, Arousal, and Dominance Lexicon for sentiment analysis

The emotional arcs of stories are dominated by six basic shapes (Link)

-Research paper that takes things a few steps further using books

-Finds which story arcs are most popular

The Data

labMT: Language Assessment by Mechanical Turk (Link)

-A dataset that contains over 10,000 words, each with a happiness rank, score, etc.

-For the purpose of this project, only the happiness score will be used

word	rank	score
laughter	1	8.50
happiness	2	8.44
love	3	8.42
happy	4	8.30
laughed	5	8.26
...	...	...
terror	10206	1.76
die	10207	1.74
killing	10208	1.70
arrested	10209	1.64
deaths	10210	1.64

Various movie scripts/transcripts found online

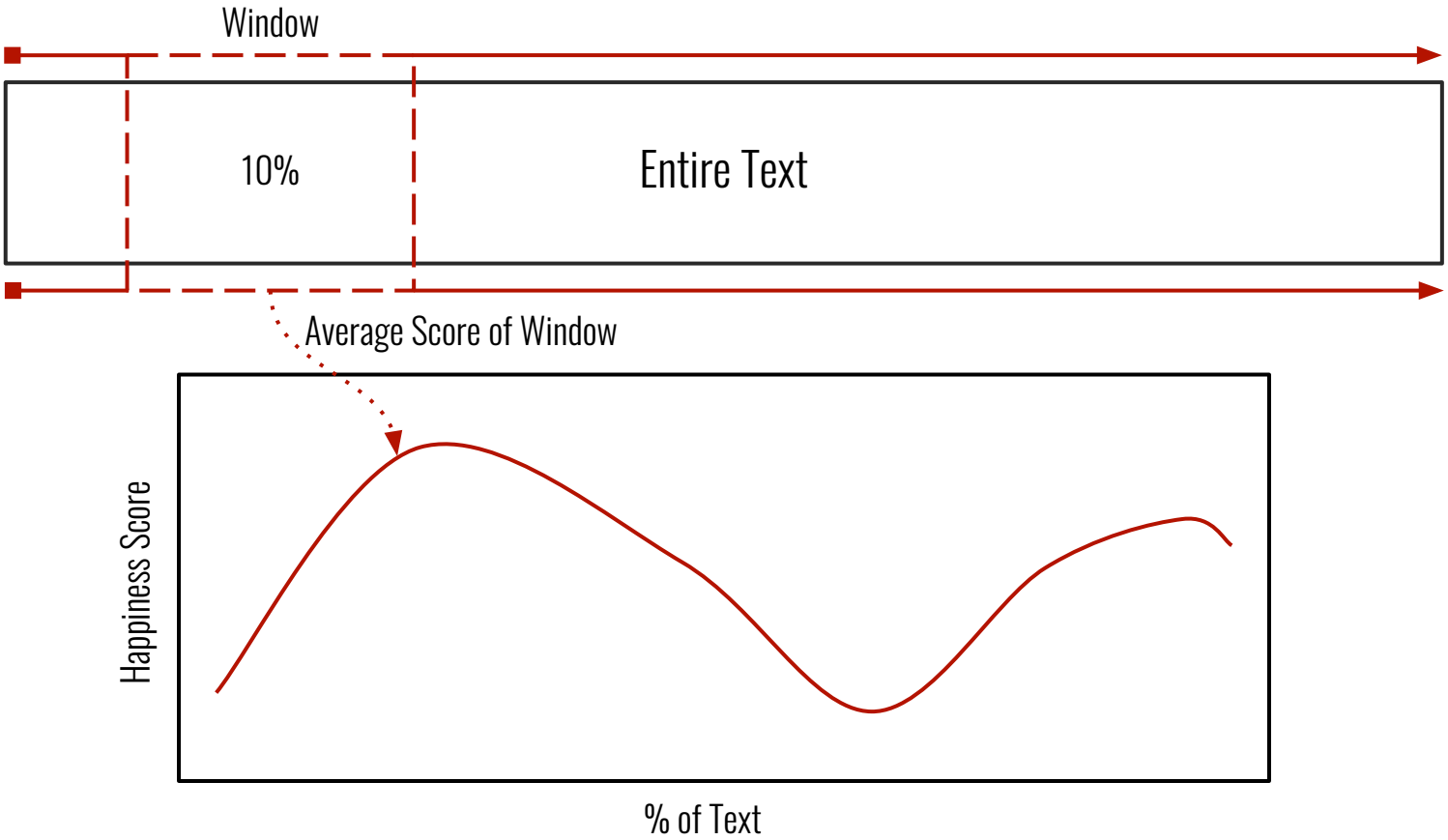
-The Internet Movie Script Database (Link)

-Some movie scripts are the true script, while others are only the spoken dialog with minimal action descriptions

My Approach

1) Read in the text file that contains the words with the happiness score and store that in a dictionary for easy access
2) Read in a given script specified by the command line, remove non-alphanumeric characters, tokenize the text, and store it in a list where each index represents one word
3) Using a sliding window that is 10% of the given text, calculate the average happiness score of that window and plot it (see next slide)
4) Slide the window through the text until we reach the end
5) Display the graph

Sliding Window

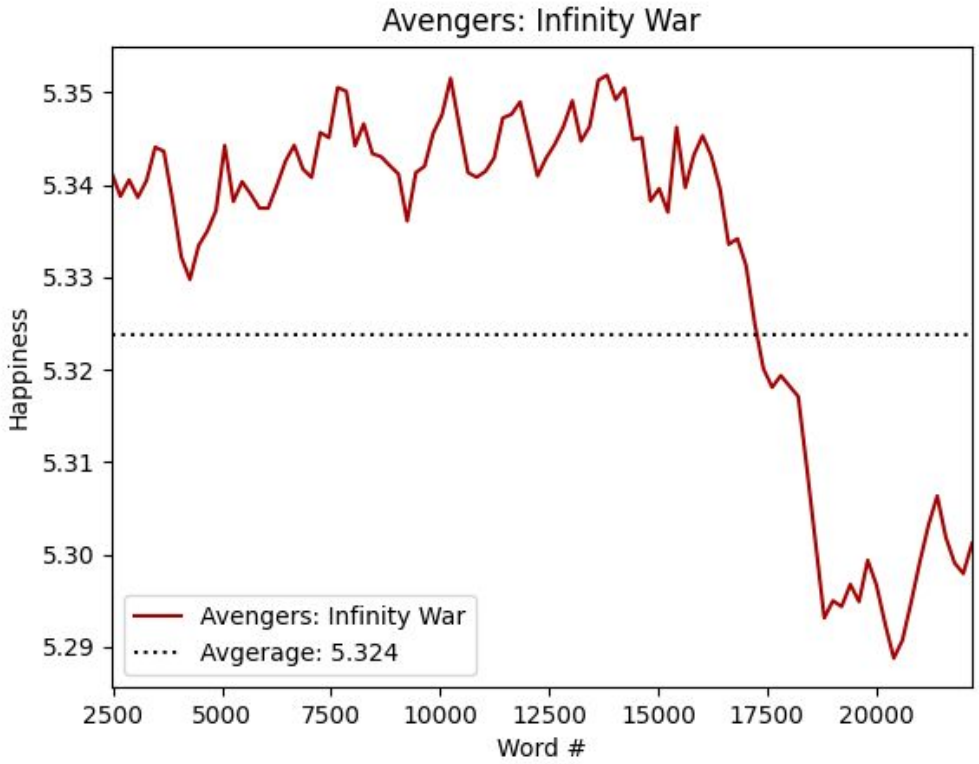


Results

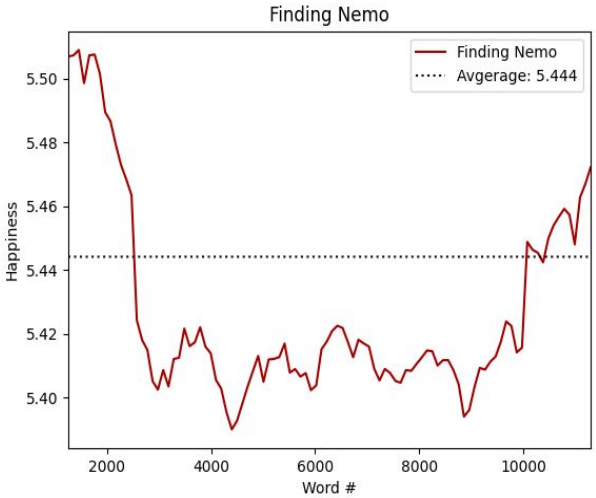
-Wanted to initially see if this would work at all with a movie that I thought should be cleary a Riches to Rags overall story arc

-From this initial test, my approach seems to work

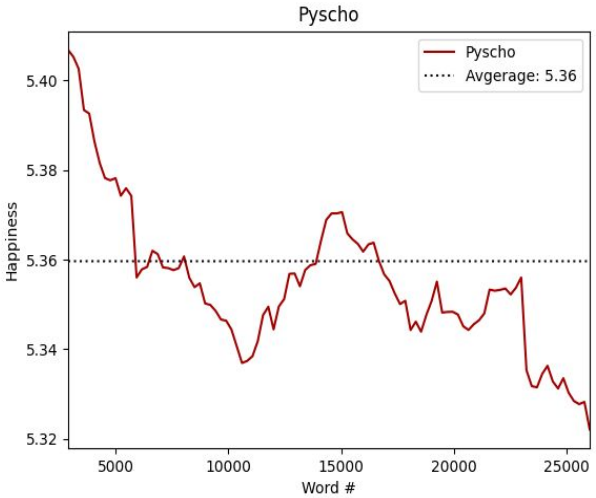
-Avengers: Infinity War = Riches to Rags (Fall)



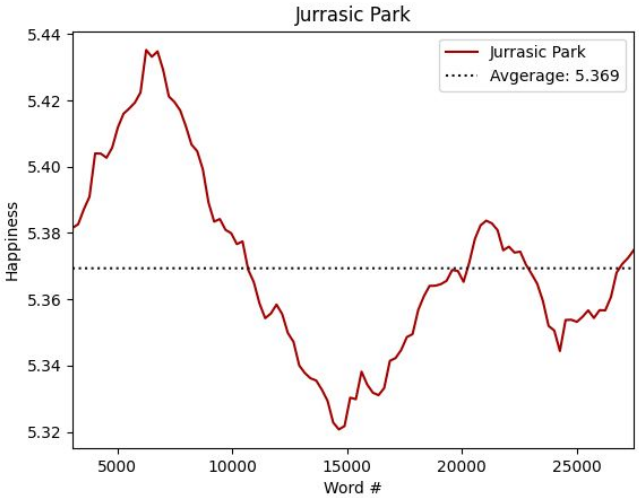
Results



Finding Nemo = Man in Hole (Fall Rise)



Psycho = Oedipus (Fall Rise Fall)



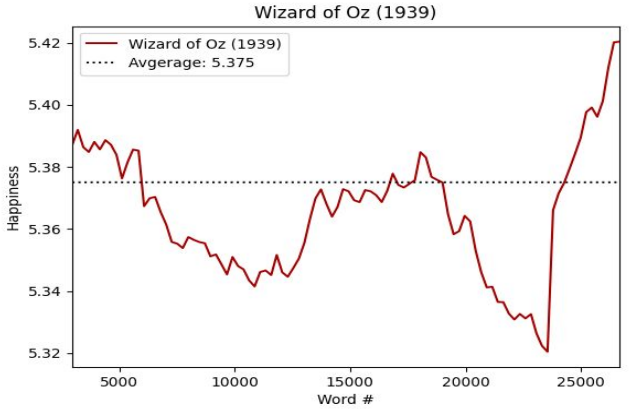
Jurassic Park = Icarus (Rise Fall)

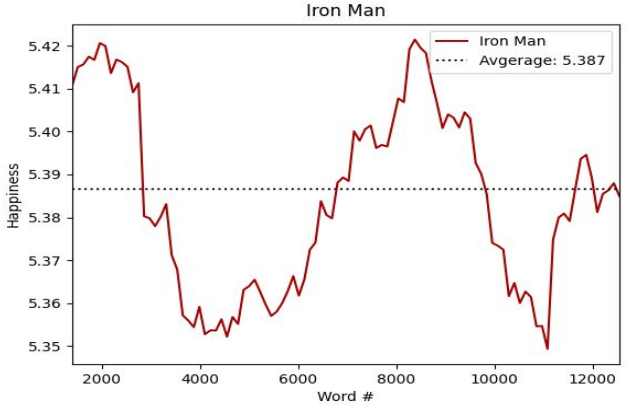
Results

-As expected, some movies don't follow any of the 6 arcs exactly

-Iron Man and The Wizard of Oz (1939) have fairly similar graphs just with different intensities

-Could be categorized as double Man in Hole arcs





Results

-Seems to work well with getting overall sentiment trend

-Most movies are a bit more nuanced than one single plot arc

-The arc of a subplot is sometimes more visible than the entire movie arc

-Works better with longer texts that are more descriptive

-Movie script are difficult in that most of it is dialogue where a book is much more descriptive

Future Work

See if retellings of classic stories have similar graphs to the original

-Hamlet vs. The Lion King
-The Taming of the Shrew vs. 10 things I hate about you

Analyze multiple movies to see which are most similar using K-means clustering

Give percentage of how similar a script is to one the six arcs based on the graph

Questions?