



Five Books, Same Story

–Understanding Percy Jackson through Sentiment Analysis–

Huijun Mao, Professor Katherine Elkins, Professor Jon Chun
IPHS 200, Fall 2021

Integrated Program in Humane Studies, Kenyon College, Gambier, Ohio, 43022

Background

Sentiment Analysis, also known as Opinion Mining and Affective AI, is a subfield of Natural Language Processing which devotes to understanding human languages through computer science. The study of Sentiment Analysis has gained increasing attention in society over the last decade (Chun, 2021). Wide applications of Sentiment Analysis include but not restricted to directing customer services, developing intelligent chatbots and tracking drivers in self-driving vehicles (Elkins and Chun, 2019).

In literary field where emphasis is placed on the interrogation of underlying patterns and structures in literature, the advancement in Sentiment Analysis models opens a new door for scholars to closely examine the emotional shifts of story plots. Patrick Colm Hogan defines the analysis of this specific feature of narratives as “affective narratology” and hypothesizes that every story should have an underlying narrative arc of emotions (Hogan, 2011). The concept of emotional arc, however, was proposed much earlier by Kurt Vonnegut shortly after WWII (Chun, 2021). “Stories have shapes which can be drawn on graph paper”, wrote Vonnegut in his rejected master thesis as a graduate student at the University of Chicago (Vonnegut, 2005). It was only in the recent years that, with the introduction of computational techniques, study on emotional arc became popular.

There are a variety of models deployed on Sentiment Analysis of texts, ranging from simple lexicon models to more sophisticated multimodal Sentiment Analysis including Deep Neural Network and machine learning. With the massive computational power to process a large number of texts in a relatively short time, researchers are able to gain profound insights that could never have been gained through traditional literary analysis. By studying the emotional arcs of 1,327 books from Project Gutenberg, A.J. Reagan et al. discovered six basic shapes of stories (Reagan, 2016). Implementing simple models on the work of Virginia Wolf, Prof. Katherine Elkins and Prof. Jon Chun demonstrated an effective way of understanding modernist novels by detecting strong emotional ups and downs from plotless stories(Elkins and Chun, 2019).

There are, of course, shortcoming of this computational approach on literary analysis. Elements such as irony, slangs and idioms often pose difficulties for accurate emotional valence measurement (Elkins, 2019). However, new and more powerful models are constantly being updated to accommodate the inadequacy and Sentiment Analysis has never been rapidly progressing as it is right now.

Introduction

In this project, I employ two sentiment analysis models introduced in class on the Percy Jackson & Olympian series (PJO). It's a pentology of fantasy adventure novels written by Rick Riordan featuring the dangerous but exciting experience of a demigod called Percy Jackson in fighting monsters and helping the Olympian Gods defeat the Titans. The five books are *The Lightning Thief*, *The Sea of Monsters*, *The Titan's Curse*, *The Battle of the Labyrinth*, and *The Last Olympian*. Even though each novel tells a completely different story with varying characters and settings, the narrative structures of all five novels are very similar: the story always starts with an oracle that introduces a conflict; then three heroes will be sent on the quest; after numerous difficulties and a final twist, the story reaches its climax and the heroes complete the task successfully. Therefore, I'm inspired to look into the emotional arc of each novel to see if narrative structures have any influence on the sentimental ups and downs of the novels. I will also evaluate the effectiveness of the two sentiment analysis models by comparing their results.

Methodology

The two sentiment analysis models I implement in this project are VADER and Textblob. VADER is a lexical + heuristic model. Lexical model draws on a dictionary of words with a sentiment value assigned to each word. Without the need of training, lexical models can be executed quickly and are easily applicable under different conditions. However, the primary shortcoming of the lexical models is that they can't take context into account. The heuristic models are able to overcome this disadvantage by recognizing syntax patterns like negation (not happy) and amplification (very happy), thereby making the final analysis more accurate. Textblob is a traditional machine learning model. Compared to more complicated models like Deep Neural Network, the classical machine learning models are relatively easy to train with minimal overhead (Chun, 2021). It doesn't mean, however, that ML models will perform better than lexical models. Sometimes, simple models can yield more satisfying results.

Results

Emotional Arcs for five Percy Jackson Novels



Analysis

Book 1

The lightning Thief tells the story of three heroes adventuring to the underworld to retrieve the lightning bolt for Zeus. The blue line is the Vader model and the orange is the Textblob model. Seemingly, the biggest disagreement between the two models happens around sentence No. 2000, when Vader is at the peak while Textblob is at the valley. However, if we zoom out and take the overall story structure into account, we will realize that the difference is not that significant and the big pictures actually resemble each other. The story plot of *The lightning Thief* can be roughly divided into three parts depending on the change of settings. Between the first two peaks, Both arcs display a V-shape; this part refer to the experience that happens in Percy Jackson's school life, or more generally, the ordinary world. After the second peak, the setting changes to the half-blood camp, also understood as the monster world, and it's from here the quest starts. The following series of ups and downs that crowd together reflect the numerous difficulties the heroes encounter on their way to destination. This second part of emotional fluctuations is the journey of the quest. The climax of the story comes at the large fall-and-rise at the end of the plot, including reversal and resolution. Both models clearly display this striking twist at the end.

Book 2

The Sea of Monster tells the story of three heroes' quest to find the Golden Fleece on the Polyphemus' island to save the camp. Both models yield similar results for this book as the emotional arcs recognize peaks and valleys at similar places. The plot can be interpreted in the same way as in the first book by dividing the story into three parts. The big V-shape of the first part of the plot covers Percy's experience in the ordinary world. His arrival to the half-blood camp normally serves as the turning point as it's from here that the quest starts; in this story, it's around sentence no. 2492. The following series of ups and downs represent the difficulties during the journey of quest, while the final fall-and-rise marks the climax of the story. We are starting to see consistent patterns that emerge in each book.

Book 3

The Titan's Curse tells the story of a group of heroes' adventure to Mount Othrys to rescue the God Artemis and their friend Annabeth. The two models also work relatively well for this book as similar peaks and valleys are recognized. The first part of the novel is a bit messier than the previous two books but a V-shape can still be recognized between sentence 1500 and sentence 3000, marking events that happen under the ordinary-world setting. Compared to the first two books, the third book does start with more complications as more characters and fighting are involved before the heroes arrive at the half-blood camp; thus, it makes sense for the first part of the plot to appear slightly more complicated than others. Regardless of the shapes though, the second part of the story—the journey of the quest—always starts at around the same time, around sentence 3000. The climax can be recognized rather easily as the most striking pattern at the end.

Book 4

The Battle of the Labyrinth tells the story of the heroes' adventure in the labyrinth to find Daedalus' workshop. There seems to be a big disagreement around sentence 7000 when Vader recognizes a peak while Textblob recognizes a valley. The actual event, itself, however, is not that important because it's about a series of predictive dreams, which is a consistent element in all five books. I don't feel there is a special emotional fluctuation at this part because the dreams are not even centered upon the protagonists; their purpose is merely to supplement information that readers don't know so that story can be pushed forward. The other unique thing about the plot is that the climax is not as smooth as the previous three. I think that's because in this book, the climax is more complex than the others because it contains two parts. The story doesn't stop at where the heroes successfully reach Daedalus' workshop in the climax, but follows by a huge battle that takes place in the half-blood camp. With two different settings and more events involved, the emotional arc at the climax turns out to contain more ups and downs. Yet, a sharp fall-and-rise shape can still be detected, which is typical of climax at the end of every book. Besides the climax, the former two parts of the novel correspond well with the patterns we found previously

Book 5

The Last Olympians describes the battle between the Olympus and the Titans, focusing mostly the demigods' defence of New York against the monster army. Even though in this book, there is not necessarily a quest which heroes have to walk for a physically long distance to complete, the emotional arcs turn out to have similar patterns. The V-shape describes the preparatory stage of the war, as usual with a change of setting between the first valley (around no.1600) and the second peak (around no.3000); the war that takes place in New York is covered in the second part with consistent ups and downs; the final fight between protagonist and antagonist, which is also the climax, takes place on Mount Olympic and is denoted by the final big fall and big rise. Both models agree on most peaks and valleys.

Conclusion

Even though nuances and disagreements are detected between individual arcs and models, there are certain consistent patterns that can be extracted in general, which reflect the overall narrative structure that underlies all five novels of the Percy Jackson & the Olympians Series (PJO). Just like most story can be interpreted with the three-act structure, the story plot of PJO can also be divided into three parts—the ordinary world, the quest journey and the climax. In a more academic way, they are inciting actions, rising actions, and climax with reversal and resolution. I put climax, reversal and resolution together into one part is because they tend to show up in one unified pattern in the emotional arc of the novels—the final glider.

Settings tend to play an important role in the novels, since changes of settings usually indicate a new stage of narrative. The ordinary world part usually takes place in the common people's world, where the conflict is introduced. This is the hook of the novel, also the foundation of the success of a book. Based on the large V-shape at the start of each arc, Rick Riordan clearly shows his deep understanding of how to start a story that best capture readers' attention. The quest journey, consisted of micro-level conflicts and resolutions, usually takes place in the Olympian world where monsters and Gods are present. The climax takes place at the destination of the quest and it is the most recognizable part of the arc with the sharpest falling and rising.

Interestingly, the climax of book 2 is considerably smaller and less striking than the others. By comparing individual book ratings on Goodreads and Douban, two authoritative book review websites, I found that book 2 has the lowest book ratings of all five, though the difference is not very big. The final book receives the highest ratings. It's not definite whether there is any relationship between readers' response of the book and the amplitude of the climax, but it's certainly a point that worth further investigation.

By comparing with the seven basic plotlines introduced in “The Bestseller Code”, PJO fit best with Plotline 5 the Voyage and Return plot (Archer and Jockers, 2016). Stories that are assigned to this group are usually about a protagonist journey into an unknown world; after some twists and turns, usually a dark turn at the end, the story ends with a happy return to the normal world. One example is *Alice in Wonderland*. Based on our analysis previously, PJO, featuring the demigods' adventures in different settings, no doubt belong to this group with a highly similar story structure.

Future

For future improvement, there are two possible directions. First, we can implement more complicated models such as Deep Neural Network and Transformers for sentiment analysis. By comparing the results, we can see if other models also yield similar emotional patterns and whether complicated models will work better than simple models. It will also be insightful to look at how other models interpret the parts that the current two models disagree upon.

Second, now that we've recognized the similarities between the five novels, we can shift the focus towards the nuances between each emotional arc. The comparison of the amplitude of climax and its potential relation to book ratings is a good starting point. We can go from there to further investigate if certain features in the emotional arc will influence the popularity of the book; hopefully, new insights on what constitute a best seller can be derived.

References

Archer, Jodie. “The Sensations, or, How to Form Some Perfect Curves.” *Bestseller Code: Why Some Books Sell a Million Copies*. St Martin'S Press, 2016. pp. 73–111.

Chun, Jon. "SentimentArcs: A Novel Method for Self-Supervised Sentiment Analysis of Time Series Shows SOTA Transformers Can Struggle Finding Narrative Arcs." *arXiv preprint arXiv:2110.09454* (2021).

Elkins, Katherine, and Jon Chun. “Can Sentiment Analysis Reveal Structure in a ‘Plotless’ Novel?” 31 Aug. 2019, <https://doi.org/10.01441>.

Hogan, Patrick Colm. *Affective Narratology: The Emotional Structure of Stories*. Bison, 2011.

K. Vonnegut. At the blackboard, 2005. URL<https://www.laphamsquarterly.org/arts-letters/blackboard>.

Reagan, Andrew J, et al. “The Emotional Arcs of Stories Are Dominated by Six Basic Shapes.” *EPJ Data Science*, vol. 5, no. 1, 2016, <https://doi.org/10.1140/epjds/s13688-016-0093-1>.