Adapted Arcs: Sentiment Analysis and *The Sorcerer's Stone*

Erin Shaheen
Senior Seminar, Integrated Program in Humane Studies, Kenyon College

Introduction

For my project, I've chosen to apply a variety of different sentiment analysis tools to the screenplay of *Harry Potter and the Sorcerer's Stone*. The term sentiment analysis refers to a type of AI-based program that reads a text and makes judgments on the positive or negative sentiment contained within.

The models that I chose to work with read the text, break it down into sections, then score the sections based on positive or negative sentiment. They position the sections in chronological order, then draw a line connecting them. This serves as a visualization of the emotional arc of the story.

I chose to work with sentiment analysis because it focuses on the emotional arc instead of the story's plot. I am an avid reader of popular literature, who grew up watching as many of my favorite books were adapted for the screen. As a writer who has worked with both narrative prose and screenwriting, I was curious to see if sentiment analysis could offer new insight into the choices that screenwriters and directors make while adapting novels for the screen.

Methodology

Before implementing the algorithms, I worked with the other human readers to create our own projection of an emotional arc, based on prior viewings of *The Sorcerer's Stone* and my own reading of the screenplay.

I began the technical portion of my project with six different models: SyuzhetR, NRC, SentiWord, Jockers-Rinker, VADER, and a Multinomial Naive Bayes algorithm that I will refer to here as Textblob. SyuzhetR is made up of four different lexicons – collections of words that have already been labelled with their "directional sentiment" (Chun 3-4). NRC, SentiWord, Jockers-Rinker, and VADER are all heuristic models, which combine lexicons with sets of rules that are meant to detect specific contextual patterns that might trip a lexical model up (Chun 4). The last model that I worked with, Textblob, is a probabilistic model that uses traditional machine learning to predict the likelihood of a certain section of the text being tagged as positive or negative (Shriram).

I was able to remove both SentiWord and NRC from the lineup, as their arcs were outliers with significant divergence from the opinions of the human readers. For instance, NRC marked the scene where Harry meets the Weasley family on the train platform as the most negative point in the story, surpassing both of his encounters with Voldemort, as well as Ron's near-death experience with the living chess set.

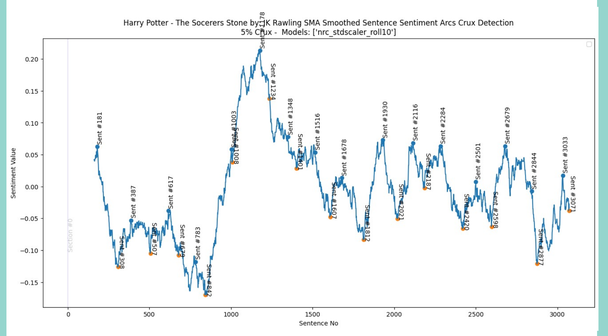

FIGURE 1: NRC's interpretation of the screenplay's emotional arc. The extremely "negative" early points include Harry's first meeting with the Weasley family, around Sentence 842.

In the end I selected VADER as the "best fit" model as its results represented a happy medium of the other three successes, and its peaks and valleys were corroborated by the human readers.

Methodology

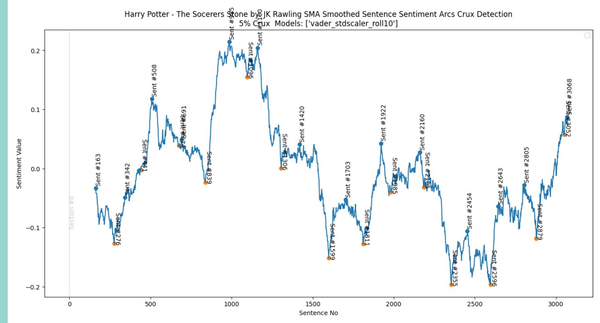

FIGURE 2: VADER's interpretation of the screenplay's emotional arc. Note the more realistic placement of Harry's meeting with the Weasley family, around Sentence 839.

Results

In a recent study done by Professors Elkins and Chun through the Digital Humanities Lab at Kenyon College, sentiment analysis of the novel *Harry Potter and the Sorcerer's Stone* created an arc that, while similar to that of the screenplay, contained some key differences.

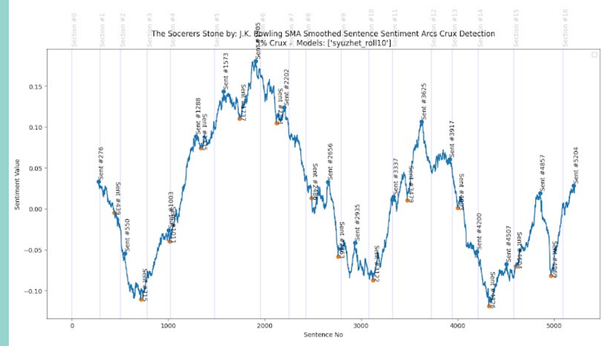

FIGURE 3: SyuzhetR's interpretation of the emotional arc of J.K. Rowling's novel, *Harry Potter and the Sorcerer's Stone.* Note the more regular rise and fall of the first 2,000 sentences, especially in comparison to VADER's graph of the screenplay.

Some of the most important differences between the book and the film appear during the beginning of the emotional arc. For instance, VADER's analysis of the screenplay drops down when Harry meets the Weasley family, which SyuzhetR marks as one of the most positive points in the book. This is likely due to the creative choices made by screenwriter Steve Kloves, who uses this scene to contrast the orphaned Harry's loneliness with the bustling Weasley family.

As a whole, VADER tends to mark moments in the screenplay that deal with loneliness and isolation as negative, and moments where Harry was able to expand his social network by engaging with the rest of the wizarding community as positive. The negative moments include but are not limited to: the destruction of the letters, Harry's nerves before his first Quidditch match, his encounter with Voldemort in the Forbidden Forest, the moment the students realize that Dumbledore has left the school, and the game of living chess. Notably, there is one major low point that does not have to do with loneliness– when Hermione overhears Ron making fun of her, and runs to the bathroom crying.

Results

Some of the most significant positive moments include: the trip to Diagon Alley, Harry and Ron's conversation on board the Hogwarts Express, the Welcome Feast, winning the Quidditch game, and Harry's final conversation with Dumbledore. Though the graph is cut off before the of the film, it is extremely likely that the emotional arc would continue on its upward trajectory, giving the subject matter of the final scenes.

Interestingly enough, two scenes that display a mixture of positive and negative emotions were marked near neutral by VADER. When Harry encounters the Weasley family in the train station, the warmth of their goodbye is juxtaposed with Harry not having anyone to say goodbye to. In a later scene, Harry's attempt to hide from his teachers leads to him discovering the Mirror of Erised, a magical object that shows its user their heart's desire.

After narrowly escaping being caught by Snape, Harry ducks into a seemingly empty room in the castle and pulls off his invisibility cloak. Curious, he begins to explore, and finds a tall, gilt-framed mirror tucked between two of the columns. The camera lingers on a series of strange, illegible words carved into the mirror's frame.

As Harry walks up to it, he sees a man and woman reflected behind him. It takes him a moment to recognize them as James and Lily Potter, the parents he has never known. A wide shot shows him looking over his shoulder, shocked, only to find the room empty. The reflections of his parents smile at him, but when Harry reaches out to touch them, his fingers brush the glass. The music turns slow and wistful as Lily puts her hand on Harry's shoulder. He attempts to cover it with his own, but the shot changes, and the audience sees that his shoulder is bare. He looks at the reflection in the mirror and smiles. This is the closest he has been to his parents in ten years.

Though this scene is brief and contains minimal dialogue, it is also one of the most emotionally complex scenes in the film. It invokes a bittersweet feeling in the human viewer, and leaves a lasting impression behind. The complex mixture of positive and negative emotions during Harry's discovery of the mirror and his encounter with the Weasleys at the train station causes VADER to mark them as near neutral. (It is worth noting that, when tracing the emotional arc of Rowling's novel, SyuzhetR marked both of these scenes as significantly more positive than they appeared in the screenplay).

There was some question of whether the screenplay's arc would match up with a reader's emotional analysis of the film itself since, by virtue of genre, a screenplay lacks the full breadth of audio and visual elements present in a finished movie. During my rewatch of the film, I was glad to find that various elements that were not present in the screenplay– including music, lighting, and the actors' facial expressions – served to enhance the emotions already present in the screenplay.

This was particularly evident in the scene where Harry, Ron, and Hermione confront the living chessmen, which VADER marked as one of the lowest points in the film.

Ron takes the lead, directing Harry and Hermione across the gigantic chessboard that the screenplay compares to "a kind of battlefield… studded with living soldiers." The viewer learns along with the characters that losing the game has serious consequences – just like in "real wizard's chess," any piece lost to the other team gets destroyed, including the human players. In the "thunderous collision" of a pawn being destroyed is accompanied by a loud crash, as well as a cloud of white smoke. The music picks up in volume and tempo immediately after the first piece is broken, and continues through a series of rapid cuts showing the destruction of several different pieces. The montage ends when the white queen shatters the black rook and turns towards Ron, who has climbed on top of a stone horse to better play the part of a knight.

The queen moves slowly as the music shifts into something quiet and creeping, mimicking the sense of "cruel indifference" described in the text of the screenplay. After arguing with Harry and Hermione, a frightened Ron moves his own piece forward, putting himself within range of the queen so Harry can win the game and continue on to fight Voldemort. The music swells again, reaching a climax and stopping just before the queen drives her sword through Ron's horse. Ron's scream overpowers the crashing sound that the audience has grown familiar with, and he falls to the ground amidst the pieces of the broken chessmen.

Conclusion

The four successful algorithms were able to create coherent emotional arcs, and even made thematic connections between scenes that the human readers had initially missed. The majority of low or lowest points marked by VADER were scenes where Harry wrestles with loneliness and isolation, while the highest points showed him expanding his social network by making personal connections with other characters.

This collection of scenes includes places where Harry's struggle with loneliness was more heavily emphasized in the screenplay than it was in the book. The scene with the living chessmen is more than just a violent conflict; it is also the moment where Harry realizes that he will have to face Voldemort alone, without the help of his friends. Sentiment analysis also showed that the emotional arc of the film is more complex than the classic Cinderella story, combining quick switches between positive and negative sentiment with complex, bittersweet moments that hover near the middle of the graph.

The agreement between different models reinforces the argument that sentiment analysis can function as a valuable tool for screenwriters. Though I needed to remove two of the heuristic models, there were substantial results from the other four models, showing that lexical (SyuzhetR), heuristic (VADER and Jockers-Rinker), and Traditional Machine Learning (Textblob) models all have the potential to provide a writer with fresh perspective. It may be easier to accept commentary from an AI than it would be from a competitive peer, or a mentor that you may want to impress.

Sentiment analysis has the ability to offer a screenwriter with a fresh perspective, piecing together connections and highlighting certain emotionally vibrant scenes before they bring their work to directors or focus groups. When working with adaptation, it's often more difficult to invoke a specific emotional response in the audience than it is to simply replicate events that appeared in the original text. *Harry Potter and the Sorcerer's Stone* was hailed as a successful adaptation by critics and fans alike, and praised for the ways that it stayed close to the beloved children's story. Whether the screenwriter in question wants to keep the emotional valence of certain scenes as close to the book as possible or shift the story in new and exciting ways, the use of sentiment analysis can help them figure out what they need to change.

Sentiment analysis offers a wide range of possibilities when used on screenplays. Whether it's through the comparison of a negatively received adaptation with a successful book, or the combined arcs of a single text split into multiple parts, there's plenty of room to experiment and explore. In future research, I would be particularly interested in comparing the emotional arc of the combined screenplays of *Harry Potter and the Deathly Hallows Parts One and Two* with the text of the novel they were adapted from. *Part One* was a particularly divisive adaptation, as many viewers found it boring compared to the other action-packed additions to the later part of the series.

Works Cited

Chun, Jon. "SentimentArcs: A Novel Method for Self-Supervised Sentiment Analysis of Time Series Shows SOTA Transformers Can Struggle Finding Narrative Arcs." *Cornell University*, 18 Oct. 2021, https://arxiv.org/abs/2110.09454.

Ebert, Roger. *Harry Potter and the Sorcerer's Stone* . 16 Nov. 2001, https://www.rogerebert.com/reviews/harry-potter-and-the-sorcerers-stone-2001.

Rowling, J. K. *Harry Potter and the Sorcerer's Stone.* Scholastic Inc., 1997.

Shriram. "Multinomial Naive Bayes Explained: Function, Advantages & Disadvantages, Applications in 2021." *UpGrad,* UpGrad Education Private Limited, 3 Jan. 2021, https://www.upgrad.com/blog/multinomial-naive bayes-explained/.

Acknowledgements

Many thanks to Professor Chun for his help with coding, Professor Elkins for all of her advice and expertise, and to my classmates, Reanna Phillips and Cameron Catana.