

# Hitchhiker's Guide to Sentiment Analysis

## A comparison between movie and film

Ella Gans

IPHS 200 Programming Humanity (Fall 2022) Prof Elkins and Chun, Kenyon College

### Abstract

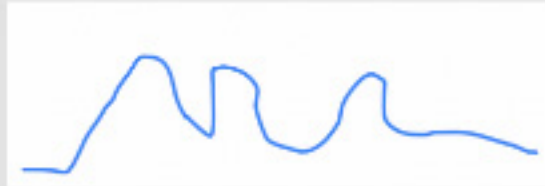
Though the 2005 film *The Hitchhiker's Guide to the Galaxy* is an adaptation of the original Douglas Adams novel of the same name, there are a multitude of differences. The film has plot points not at all present in the book, as well as a handful of characters. I wanted to see if Sentiment Analysis could pick up on the differences between the two works, and if so, what it marked those differences as. Would the plot structure look the same? Would the sentiments be changed?

I of course acknowledge that a film is far more than just a screenplay. Music, lighting, line delivery, all of these are lost in text. Thus, I must put these as caveats, that could affect the Sentiment Analysis. To this, I apologize, as I am unsure how I could've solved this problem.

Another quick confounding factor is that, by using multiple models for the analysis, that could make this a bit flawed. However, in my eyes, Sentiment Analysis is Sentiment Analysis. I simply chose what I felt best represented the shape of the story for both screenplay and novel.

### Introduction

I began this project by making two rough sketches of what I felt the book and film respectively are shaped like, in the form of arcs.



The one on the top is for the book, and the bottom is the movie. Though the two are both in waves, there are more bumps in the movie. This is since the movie has more plot points that were not featured in the novel, so some parts are more "exciting." I would say that though the film features many things not present in the original, the tone and script overall feels true to the source material.

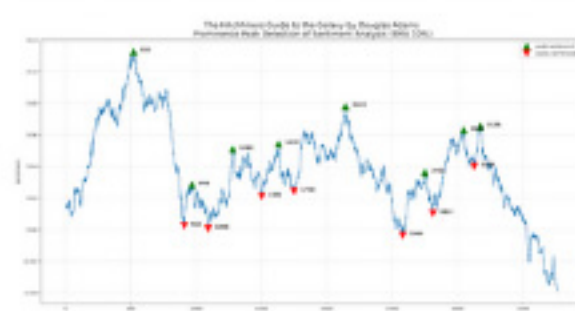
I had a worry that, considering the humorous nature of these books, the sentiment analysis may get confused, as a computer may not be able to read irony or sarcasm.

### Materials

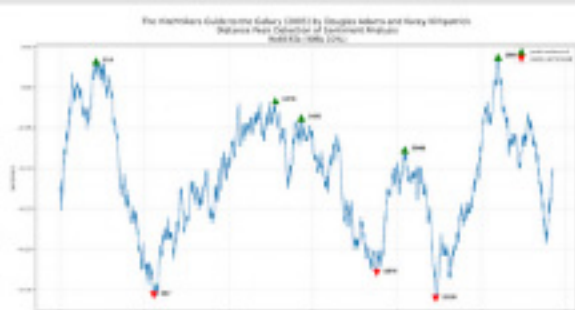
The materials included plain text versions of both the original novel and the screenplay for the film. For the Sentiment Analysis, I utilised the notebook given to us for this class.

### Methodology

I firstly input the plain text of the novel into the Sentiment Analysis notebook, and I chose to use VADER, since it seemed to get the shape down well. The result was as follows:



I tried putting the plain text of the film into VADER, Textblob, and DistilBERT, but none of those captured the shape too well. So I used RoBERTa, and got this:



### References/Acknowledgements

*The Hitchhiker's Guide to the Galaxy* by Douglas Adams  
*The Hitchhiker's Guide to the Galaxy* (2005) written by Karey Kirkpatrick and Douglas Adams  
*The Shapes of Stories* by Katherine Elkins  
Sentiment Analysis Notebook:  
<https://colab.research.google.com/drive/12-cUAvkN0qNctYYX7n3RwIRahaxytz3h?usp=sharing>



### Results

Reading through the crux points for the novel, I believe it is a mixed bag. The peaks were mostly what I would consider emotional peaks, whether it be due to suspense, excitement, or anger, with a handful of ones that were more in line with sarcasm or dry wit. It, rightfully I would say, put the first peak at main character Arthur Dent screaming "WHAT THE HELL IS THAT?!" at a spaceship. Pretty emotionally charged statement. The valleys were often found in longer passages that lead up to a joke, which are usually delivered in the form of an entry in the eponymous guide. These are delivered more passively, so it would make sense. I was pleased to find that, superficially, the arc I drew was similar to the result from VADER.

The screenplay, on the other hand, was harder. I used all three of the original models in the notebook, yet I didn't feel that any of the three captured the shape of the plot too well. RoBERTa fortunately made a lot more sense. The peaks, again, were at points I would consider emotionally charged, and the valleys were at more neutral points of the story. The shape of the plot that RoBERTa presented for the film was comparable to the one that VADER presented for the novel, which made sense to me, seeing as they are in essence the same story.

A shared peak point between both the film and novel was the arrival at the planet Magrathea, which in the context of the story is a planet that many believe not to exist. It is home to a race of people who make custom-made planets for the very wealthy. By the time the main cast arrives, it has been closed down due to economic recession, and the security system activates to try and kill them. Though the two models picked up on different lines, they picked up on the same plot point. And seeing as the main cast is at risk of death, this could easily be seen as an emotional peak.

Multiple other peaks in both analyses related to an important part of the story: The Ultimate Question and Answer. The story goes that a race of people created the most powerful computer in the universe, Deep Thought, to answer the Ultimate Question of Life, The Universe, and Everything. After a very, very long time, Deep Thought provided the answer 42. It then provided designs for an even more powerful computer than itself, called Earth, where even its biological components would be part of its computations. It would provide the Ultimate Question to the Ultimate Answer. Just before Earth completed its calculation, though, it was blown up, with its only survivors being two of our main characters and two mice. What happens next varies between the two adaptations. The point being, the search for the Ultimate Question is a major aspect of the story, which leads directly to the climax in both versions (though the climax itself differs). Due to its importance, I believe it makes perfect sense for the Sentiment Analysis to mark scenes involving it as a high point.

A consistent low point was the reading of Vogan Poetry. In-universe, Vogan Poetry is so awful that many would prefer to die than hear it. Though the response to it involves an emotional reaction (mostly agony), the poetry itself is so bad because of how dull it is. I thought it was funny that even the Sentiment Analysis picked up on that.

My biggest surprise was that none of these models picked up on Marvin, whose entire character is charged with one emotion: sadness.

### Conclusion

As I suspected, Sentiment Analysis could in fact pick up on some of the differences between the original novel and the 2005 film adaptation. It also picked up on similarities that, although I noticed they were present, I didn't think about how they were used to the same emotional effect in both works.

My fear that the humor of the writing would negatively affect the Sentiment Analysis wasn't completely unfounded. Some of the valleys were actually quite funny, but their phrasing made them appear to be neutral, or unemotional. It *did*, though, pick up on the joke that Vogan Poetry was supposed to be dull and awful, marking it as such. That's a step in the right direction.

What I took from this is that the movie intensified the emotional arcs from the book, at least the ones that they shared. Movies need to be more engaging, they have to keep the audience watching. Books of course need to be engaging, but if you put down a book, you can get back to it. If you leave the movie theater, you're missing the rest of the movie.

I also found that RoBERTa worked best with the sparser format of the screenplay, which is harder to analyze. This could be useful information for future projects and analysis.

### Future and Ethics Statement

I still think that we are far from the point where Sentiment Analysis will be more reliable than human interpretation. Though I think these models are fascinating, and I believe they can certainly *enhance* one's understanding of a piece of media, they are not yet sound enough to give a flawless interpretation. If I were to choose between asking a human or a machine for a summary of emotional high and low points, I would choose a human every time.

Looking toward the future, even as this technology gets smarter, I doubt it will ever phase out human interpretation. Sentiment Analysis can tell you a lot about a story's emotional beats, and maybe even emotional arcs, but one of the most important parts of the discussion of art is difference in opinions. Everything can be interpreted in a different way by a different person. In theory, if these Sentiment Analysis models are perfected, they'll give the same result every time. There is no right or wrong answer to what one takes away from a piece of media, whether it be a book or a film, so I don't think Sentiment Analysis can give us one. It may, though, prove to be a useful tool in analyzing media.

