

The Emotional Impact of Highly Rated Movies on their Audiences

Victor Verma
Boston University
vpverm@bu.edu

Abstract

1 Introduction

I plan to look at user reviews for a subset of the top 250 highest rated movies on Letterboxd and determine which emotions were the most prevalent in the reviews. Specifically, I will look at the 65 movies in the list that I have personally seen to avoid spoilers. To collect the source data, I will write a python script that scrapes the reviews (several thousand for each movie) from Letterboxd, collecting the text as well as additional metadata such as the date and number of likes. My initial emotion analysis will only use the review text data, but if I have extra time I could work on an extension incorporating the other metadata. After scraping the reviews, I will remove special characters (and likely emojis as well) to make them easier to work with. For the sake of simplicity, I will only keep english language reviews. In total, I expect to aggregate over one-hundred thousand total reviews for analysis. I can also extend to a larger set of movies if more reviews are needed.

Once I have prepared the data, I will use named-entity recognition and sentiment analysis to classify the emotions that people express in their reviews. As someone who loves movies, I am curious to learn if there is a common set of emotions that all highly rated movies invoke upon its audience. If the project is successful, this question will be at least partially answered.

For my final report, I intend to present statistics regarding the emotions expressed in the reviews, such as those that appear the most and least frequently. If time permits, I will stratify this data across other movie features. Since movies are often a response to the world around them,

it appears likely that highly-rated movies from different decades invoke different emotions in their reviews, and it would be fascinating to confirm this suspicion with empirical evidence. It might also be interesting to figure out if there are traditionally unrelated genres that generate similar emotional reactions. In addition to this paper, all of my results would be presented on my movie recommendation website, which generates recommendations using machine learning based on a user's Letterboxd profile. Over 1600 people have used the website, and I am sure they would be excited to learn about any findings from this large-scale analysis of Letterboxd reviews.