MyAnimeList Scraper and Sentiment Analysis

Kaiwen Xue

Kxue2

5/11/2018

I . Abstract

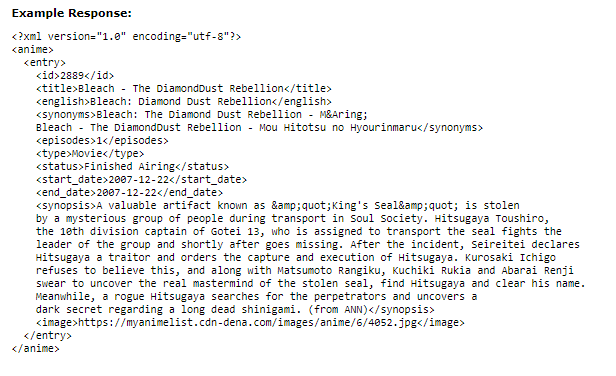
Movie reviews are essential in a reader’s process of evaluating any media. These allow readers to get a glimpse into what the reviewer was thinking when they provided their final score, this is an important line of communication. Otherwise, we would have a platform where scores hold very little meaning with there being no weight or reasoning behind the numerical values. For many reviewers, this is also a chance for them to explain any bias behind their views, or lack thereof. Essential, reviews benefit both the reader and the writer and it provides far more insight than simply providing a score. Media such as movies, tv shows, and animated shows require this textual component even further due to the nature of their evaluated works, these works can often be interpreted in multiple different ways and a true aggregate score should be thoughtful of this process. For the purposes of this paper, I shall refer to MyAnimeList as MAL.

II. Introduction

MAL provides an essential service for rating and reviewing Japanese animated shows. The website allows any user to score a show from 1 to 10, and optionally provide a written review for others to read. Like most other websites that provide a similar service (eg. IMDB, rottentomatoes), MAL takes these scores and creates a weighted aggregate to score and list these shows in a list that ranks the top-rated show to least rated show. The purpose of this project is to implement a web scraper to take a comprehensive amount of reviews for a show and do some text analysis, specifically in the form of topic modelling and sentiment analysis. The topic modelling should let us get a closer look at how reviews are typically broken down, the logical next step in this project should be to create a summarized review that takes these topics and gathers sentences that best match these topics in forming a general review that matches the mined text as closely as possible. However, we will later see why that was a failure. Consequently, instead of a summarizer I performed sentiment analysis on these reviews to reveal the nature of these reviews and how they really reflect a given score.

III. Methods/Functions

The first function that we needed to implement is a web scraper to extract the necessary data to perform the analysis and modelling on. MAL provides a very non-comprehensive API that was a good starting point, but ultimately fails to provide the most crucial data: the text for reviews themselves. The API returns a lot of unessential metadata relating to the animated show you search for, but I cannot use this data meaningfully.



The figure provided above is an example of what the API returns in XML form. Because of this, we needed to gather the data using a more rudimentary web scrape. Using the BeautifulSoup Python Package to extract the text from reviews. I scraped approximately 10000 sentences worth of reviews from two shows to test my implementation. These reviews varied massively in brevity, quality and relatedness to the topic at hand. This would be hard to control due to the openness of the website and lack of moderation in the platform. Due to this, I predicted that there would be some failure down the line.

Topic Modeling is the next step to this project. The metapy package was used to generate an index, learn and then perform topic modelling using Latent Dirichlet Allocation. I proposed that the reviews hold 6 distinct topics. This was a prediction I made since reviews scores are averaged using six different categories (Overall, Story, Animation, Sound, Character, Enjoyment). Therefore, I envisioned that reviewers would explain their numerical evaluations for each of these categories.



The results were unsatisfactory, these words did not convey a sense of six distinct topics. This is largely caused by the lack of a structure format across all the reviews, as well as the possibility that certain lingo related to the animated show were unable to be modelled.

Due to the nature of the data I decided to steer away from an aggregated text summary and perform some basic sentiment analysis instead. I used the Natural Language Toolkit (NLTK) to tokenize these reviews into individual sentences that would be easier to analysis. The NLTK Vader Sentiment Intensity Analyzer would return an independent result for each sentence with four given quantities, compound, positive, negative and neutral. These values tell us whether the sentence was positive, negative and by what magnitude, or if it was neutral. The compound score is derived from a combination of all three.

------------Sample data from running the reviews for Cowbop Bebop through the script-------------

It is a very well-made series that raises many interesting questions and concepts; and while some episodes frustrated me, others fascinated me.

compound: 0.34, neg: 0.125, neu: 0.662, pos: 0.213,

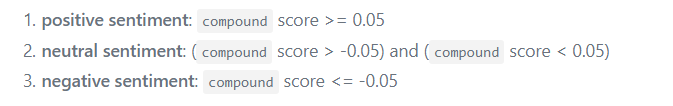
There's nothing wrong with the series, and I know many people like it's meandering nature and enjoy the lack of an over-arching plot (which, by the way, was rather tidily summed up in two eps at the end, nicely done).

compound: 0.834, neg: 0.048, neu: 0.716, pos: 0.235,

While my objective opinion is that it is a well-made series, well worth a watch, my subjective view wishes it was different, that it had been a story about Spike's past, about Vicious, about their friendship, about Julia's love and Jet and Faye's stories as well.

compound: 0.8834, neg: 0.047, neu: 0.675, pos: 0.278,

The NLTK package is quite capable in analyzing the sentiments behind a sentence, and the results are quite effective in giving a quantitative measure to an otherwise arbitrary system. The vaderSentiment package gives the following threshold values in assessing whether a statement is positive or negative.



I performed this analysis on two shows that I pre-scraped. “Cowboy Bebop” and “Mayoiga”, these two shows represent opposite spectrums in terms of total score on the website, “Cowboy Bebop” is one of the most world renowned animes of all time, with countless critics praising it and having an incredibly high score to boot (ranked 27th best anime of all time). On the other hand, “Mayoiga” is considerably less well received and has received some backlash for being controversial and poor executed, having a low score that gives it a rank of 8084th best anime of all time.

IV. Results and Conclusion

The compound, positive, negative and neutral scores averaged over all reviews for the two aforementioned shows are listed below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | MAL Score | Compound | Positive | Negative | Neutral |
| Cowboy Bebop | 8.81 | 0.2113 | 0.1419 | 0.0540 | 0.8025 |
| Mayoiga | 5.66 | 0.0338 | 0.1221 | 0.1053 | 0.7725 |

The results of performing sentiment analysis on a small sampling of reviews for these two shows are promising, but not resounding. Specifically, “Mayoiga” is perhaps one of the least liked show on the website, however the compound score calculated over all these reviews demonstrates that the overall sentiment for ‘Mayoiga’ is merely neutral. I believe that these results are largely due to the fact that reviewers try very hard to be unbiased and fair in their judgement, and as a result fail to represent the population’s opinion as a whole. Perhaps reviewers are ere on the side of caution when it comes to speaking out about these shows, otherwise it stands to show that either the Sentiment Analysis is weak or reviewers are an unfitting representation of the scores they themselves give.