

Sentiment analysis of book reviews scraped from various websites

- What is the problem you want to solve?

Detecting sentiment from a piece of text is a key part of Natural Language Processing (NLP). It is a foundational part of recommender systems which track a user's affinities and aversions with the aim of matching the user with a product/service that's helpful for them. Much work on sentiment analysis has been done on short-form text data such as those derived from micro-blogging social media sites. Because of the succinct nature of these social media posts, it is perhaps easier to glean points of view than from longer pieces of text.

I would therefore like to perform sentiment parsing of long-form text in the form of book reviews. The aim is to see if NLP models are just as capable in situations where the user leaves a more verbose review.

- Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?

Book cataloging websites might benefit from insights gained from this project in order to improve their recommendation system. If longer reviews do result in less reliable recommendations the website might decide to implement a word limit on reviews.

- What data are you using? How will you acquire the data?

I am using book reviews scraped from various websites including Goodreads stored on Kaggle as CSV files.

- Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.

I will use a recurrent neural network to classify whether a review is positive or negative and contrast its performance with both a Naive Bayes classifier and a Long Short Term Memory neural network.

- What are your deliverables? Typically, this includes code, a paper, or a slide deck

I will present code, a paper and a slide deck presentation.