Team member: Yuchen Zeng , yuchenz8

Captain: Yuchen Zeng , yuchenz8

Topic: Movie review classifier

Description: Following the algorithm described in Learning Word Vectors for Sentiment Analysis(https://aclanthology.org/P11-1015/), I plan to build a movie review classifier web application. This movie review web application will let the user to input a paragraph of review and the application will try to judge whether this paragraph is a positive review or a negative review. This task is interesting and can be used in several applications. It can be used by movie review websites like Rotten Tomatoes to automatically recognize and classify users' reviews.

Methods: I plan to train my model using the n-gram model. I will compare between unigram, bigram, and 3-gram model to find the best performance model. I will also try doing some performance to increase my accuracy to predict the tags, like using the smoothing method and log. I will be using the dataset provided by the paper, it is a training dataset with reviews and its property. I will expect the overall accuracy of classification to be higher than 85%. In order to evaluate this outcome, I will calculate my accuracy based on a testing dataset. I may create the testing dataset myself.

Programming language: Python

Workload:
Read the paper and configure the dataset --------2hr
Build unigram, bigram, and 3-gram model on the dataset ---6hr
Build testing dataset to test my model, search methods to improve performance---3hr
Build smoothing method ---2hr
Build the backend for the web application -----3hr
Build the frontend for the web application ----4hr