CS410
Tony Jiao
Nov 15th

Progress Report

Progress made:

I have made an overall structure of my application. I finished files such as loading in the training and testing data, a testing program that outputs the accuracy after running the test data set through, an initial iteration of the Naïve Bayes bigram algorithm, and a main driver for the whole application.

Remaining tasks:

The rest of the tasks are mainly focused on testing and tweaking the algorithm to have more accurate results. Currently, the training dataset I made only consists of around 200 reviews. This is why the test data results are not as accurate as I hope it'll be. As I input more training data in, some of the constants of the bigram algorithm might also need to be changed.

Challenges/issues faced:

The main challenge so far is the large amount of time spent on making the training data. I'll have to go into a game review website and manually copy and paste each review, read through it, and creating a label to determine if it's a positive or negative review. This is extremely time consuming as I plan on having thousands of reviews in my training data. I am considering making another scrapping app very quickly to automatically scrap all the reviews from a certain website from the HTML code. This is very similar to one of the MPs in this course so hopefully it won't take too long to make. This also raises another issue since I'm getting those

reviews from different websites, so I'll have to change the scrapping program for each website as well.